What Hockey Teams and Foraging Animals Can Teach Us About Feedback Communication

Part II: Timid/Bold Coding

Aaron Wagner
Cornell University

In collaboration with
Nirmal Shende and Yücel Altuğ
On Feedback
On Feedback

- How can one use feedback to improve block coding performance in point-to-point channels?
On Feedback

- How can one use feedback to improve block coding performance in point-to-point channels?
  - If the channel has memory, we can predict the future.
On Feedback

- How can one use feedback to improve block coding performance in point-to-point channels?
  - If the channel has memory, we can predict the future.
  - If the channel is unknown, we can learn its law.
On Feedback

‣ How can one use feedback to improve block coding performance in point-to-point channels?
  ▶ If the channel has memory, we can predict the future.
  ▶ If the channel is unknown, we can learn its law.
  ▶ If the decoding time is not fixed, we can decode early or late opportunistically.
On Feedback

- How can one use feedback to improve block coding performance in point-to-point channels?
  - If the channel has memory, we can predict the future.
  - If the channel is unknown, we can learn its law.
  - If the decoding time is not fixed, we can decode early or late opportunistically.
  - If there is an average cost (e.g., power) constraint, we can use resources opportunistically.
On Feedback

- How can one use feedback to improve block coding performance in point-to-point channels?
  - If the channel has memory, we can predict the future.
  - If the channel is unknown, we can learn its law.
  - If the decoding time is not fixed, we can decode early or late opportunistically.
  - If there is an average cost (e.g., power) constraint, we can use resources opportunistically.
  - If the rate is low, we can increase the effective minimum distance of the code.
On Feedback

- How can one use feedback to improve block coding performance in point-to-point channels?
  - If the channel has memory, we can predict the future.
  - If the channel is unknown, we can learn its law.
  - If the decoding time is not fixed, we can decode early or late opportunistically.
  - If there is an average cost (e.g., power) constraint, we can use resources opportunistically.
  - If the rate is low, we can increase the effective minimum distance of the code.
- [Other contexts: networks, control over noisy channels, streaming codes, complexity-constrained coding....]
On Feedback

- How can one use feedback to improve block coding performance in point-to-point channels?
  - If the channel has memory, we can predict the future.
  - If the channel is unknown, we can learn its law.
  - If the decoding time is not fixed, we can decode early or late opportunistically.
  - If there is an average cost (e.g., power) constraint, we can use resources opportunistically.
  - If the rate is low, we can increase the effective minimum distance of the code.
On Feedback

- How can one use feedback to improve block coding performance in point-to-point channels?
  - If the channel has memory, we can predict the future.
  - If the channel is unknown, we can learn its law.
  - If the decoding time is not fixed, we can decode early or late opportunistically.
  - If there is an average cost (e.g., power) constraint, we can use resources opportunistically.
  - If the rate is low, we can increase the effective minimum distance of the code.

Is this it?
Discrete Memoryless Channels without Feedback

- Given:
  - input alphabet: $\mathcal{X}$ (finite)
  - output alphabet: $\mathcal{Y}$ (finite)
  - channel matrix: $W(y|x)$ (indep. over time)
Given:
- input alphabet: \( \mathcal{X} \) (finite)
- output alphabet: \( \mathcal{Y} \) (finite)
- channel matrix: \( W(y|x) \) (indep. over time)
Given:
- input alphabet: $\mathcal{X}$ (finite)
- output alphabet: $\mathcal{Y}$ (finite)
- channel matrix: $W(y|x)$ (indep. over time)
Discrete Memoryless Channels without Feedback

- input alphabet: \( \mathcal{X} \) (finite)
- output alphabet: \( \mathcal{Y} \) (finite)
- channel matrix: \( W(y|x) \) (indep. over time)
Discrete Memoryless Channels without Feedback

- Given:
  - input alphabet: $\mathcal{X}$ (finite)
  - output alphabet: $\mathcal{Y}$ (finite)
  - channel matrix: $W(y|x)$ (indep. over time)
Discrete Memoryless Channels without Feedback

Given:
- input alphabet: \( \mathcal{X} \) (finite)
- output alphabet: \( \mathcal{Y} \) (finite)
- channel matrix: \( W(y|x) \) (indep. over time)
DMCs with Feedback

- Given:
  - input alphabet: $\mathcal{X}$ (finite)
  - output alphabet: $\mathcal{Y}$ (finite)
  - channel matrix: $W(y|x)$ (indep. over time)
DMCs with Feedback

- Given:
  - input alphabet: \( \mathcal{X} \) (finite)
  - output alphabet: \( \mathcal{Y} \) (finite)
  - channel matrix: \( W(y|x) \) (indep. over time)
DMCs with Feedback

- Given:
  - input alphabet: \( \mathcal{X} \) (finite)
  - output alphabet: \( \mathcal{Y} \) (finite)
  - channel matrix: \( W(y|x) \) (indep. over time)
DMCs with Feedback

Given:
- input alphabet: $\mathcal{X}$ (finite)
- output alphabet: $\mathcal{Y}$ (finite)
- channel matrix: $W(y|x)$ (indep. over time)
DMCs with Feedback

- Given:
  - input alphabet: $\mathcal{X}$ (finite)
  - output alphabet: $\mathcal{Y}$ (finite)
  - channel matrix: $W(y|x)$ (indep. over time)
DMCs with Feedback

Given:
- input alphabet: $\mathcal{X}$ (finite)
- output alphabet: $\mathcal{Y}$ (finite)
- channel matrix: $W(y|x)$ (indep. over time)
DMCs with Feedback

- Given:
  - input alphabet: $\mathcal{X}$ (finite)
  - output alphabet: $\mathcal{Y}$ (finite)
  - channel matrix: $W(y|x)$ (indep. over time)
DMCs with Feedback

- Given:
  - input alphabet: $\mathcal{X}$ (finite)
  - output alphabet: $\mathcal{Y}$ (finite)
  - channel matrix: $W(y|x)$ (indep. over time)
DMCs with Feedback

- Given:
  - input alphabet: $\mathcal{X}$ (finite)
  - output alphabet: $\mathcal{Y}$ (finite)
  - channel matrix: $W(y|x)$ (indep. over time)
DMCs with Feedback

- Given:
  - input alphabet: \( \mathcal{X} \) (finite)
  - output alphabet: \( \mathcal{Y} \) (finite)
  - channel matrix: \( W(y|x) \) (indep. over time)

[Real feedback is never so ideal ....]
Figures of Merit

\[ U_1, \ldots, U_k \rightarrow \text{Encoder} \rightarrow X_1, \ldots, X_n \rightarrow W(\cdot | \cdot) \rightarrow Y_1, \ldots, Y_n \rightarrow \hat{U}_1, \ldots, \hat{U}_k \]
Figures of Merit

- Number of bits sent: $k$
Figures of Merit

- Number of bits sent: $k$
- Transmission time: $n$
Figures of Merit

- Number of bits sent: $k$
- Transmission time: $n$
- Rate $R = k/n$
Figures of Merit

- Number of bits sent: \( k \)
- Transmission time: \( n \)
- Rate \( R = \frac{k}{n} \)
- Error probability: \( P_e = P(U^k \neq \hat{U}^k) \)
Second-Order Coding Rate

Def:

\[ R(n, \epsilon) = \max \left\{ \frac{k}{n} : \exists \text{ an } (n, k, P_e) \text{ code with } P_e \leq \epsilon \right\} \]

\[ R^{(fb)}(n, \epsilon) = \max \left\{ \frac{k}{n} : \exists \text{ an } (n, k, P_e) \text{ feedback code with } P_e \leq \epsilon \right\} \]
Second-Order Coding Rate

Def:

\[ R(n, \epsilon) = \max \left\{ \frac{k}{n} : \text{exists an } (n, k, P_e) \text{ code with } P_e \leq \epsilon \right\} \]

\[ R^{(fb)}(n, \epsilon) = \max \left\{ \frac{k}{n} : \text{exists an } (n, k, P_e) \text{ feedback code with } P_e \leq \epsilon \right\} \]

Theorem (cf. Shannon ’56):

\[ \lim_{n \to \infty} R(n, \epsilon) = \lim_{n \to \infty} R^{(fb)}(n, \epsilon) = C \quad \text{if } 0 < \epsilon < 1 \]

where \( C \) is the capacity:

\[ C = \max_P I(P; W) = \max_P E_{P \circ W} \left[ \log \frac{W(Y|X)}{PW(Y)} \right] \]
Second-Order Coding Rate

Def:

\[ R(n, \epsilon) = \max \left\{ \frac{k}{n} : \exists \text{ an } (n, k, P_e) \text{ code with } P_e \leq \epsilon \right\} \]

\[ R^{(fb)}(n, \epsilon) = \max \left\{ \frac{k}{n} : \exists \text{ an } (n, k, P_e) \text{ feedback code with } P_e \leq \epsilon \right\} \]
Def:

\[ R(n, \epsilon) = \max \left\{ \frac{k}{n} : \exists \text{ an } (n, k, P_e) \text{ code with } P_e \leq \epsilon \right\} \]

\[ R^{(fb)}(n, \epsilon) = \max \left\{ \frac{k}{n} : \exists \text{ an } (n, k, P_e) \text{ feedback code with } P_e \leq \epsilon \right\} \]

Think: \( R(n, \epsilon) \approx C + \frac{\beta(\epsilon)}{\sqrt{n}} + \cdots \)
Def: Second-Order Coding Rate (SOCR):

\[ R(n, \epsilon) = \max \left\{ \frac{k}{n} : \exists \text{ an } (n, k, P_e) \text{ code with } P_e \leq \epsilon \right\} \]

\[ R^{(fb)}(n, \epsilon) = \max \left\{ \frac{k}{n} : \exists \text{ an } (n, k, P_e) \text{ feedback code with } P_e \leq \epsilon \right\} \]

\[
\begin{bmatrix}
\text{Think: } R(n, \epsilon) \approx C + \frac{\beta(\epsilon)}{\sqrt{n}} + \cdots
\end{bmatrix}
\]

Def: Second-Order Coding Rate (SOCR):

\[ \beta(\epsilon) = \lim_{n \to \infty} (R(n, \epsilon) - C) \sqrt{n} \]

\[ \beta^{(fb)}(\epsilon) = \lim_{n \to \infty} (R^{(fb)}(n, \epsilon) - C) \sqrt{n} \]
Does Feedback Help?

- **Def:** Second-Order Coding Rate (SOCR):

  \[ \beta(\epsilon) = \lim_{n \to \infty} (R(n, \epsilon) - C)\sqrt{n} \]
  \[ \beta^{(fb)}(\epsilon) = \lim_{n \to \infty} (R^{(fb)}(n, \epsilon) - C)\sqrt{n} \]
Does Feedback Help?

- **Def:** Second-Order Coding Rate (SOCR):

\[
\beta(\varepsilon) = \lim_{n \to \infty} (R(n, \varepsilon) - C)\sqrt{n}
\]

\[
\beta^{(fb)}(\varepsilon) = \lim_{n \to \infty} (R^{(fb)}(n, \varepsilon) - C)\sqrt{n}
\]

- **Theorem** [Polyanskiy, Poor, Verdú (‘11)]: For *symmetric* channels, the SOCR is not improved with feedback:

\[
\beta(\varepsilon) = \beta^{(fb)}(\varepsilon)
\]

(\(\beta(\varepsilon)\) earlier characterized by Strassen ’62).
Def: A channel (stochastic matrix) $W$ is *symmetric* if its columns (outputs) can be partitioned so that, within each partition, the columns are permutations of each other, as are the rows.
Def: A channel (stochastic matrix) \( W \) is *symmetric* if its columns (outputs) can be partitioned so that, within each partition, the columns are permutations of each other, as are the rows.

\[
\begin{bmatrix}
1-p & 0 & \mid & p \\
0 & 1-p & \mid & p
\end{bmatrix}
\]

Symmetric
Symmetric Channels

**Def:** A channel (stochastic matrix) $W$ is symmetric if its columns (outputs) can be partitioned so that, within each partition, the columns are permutations of each other, as are the rows.

$$\begin{bmatrix} 1-p & 0 & p \\ 0 & 1-p & p \end{bmatrix}$$

Symmetric

$$\begin{bmatrix} 3/4 & 1/4 \\ 1/3 & 2/3 \end{bmatrix}$$

Not symmetric
Does Feedback Help?

- **Def:** Second-Order Coding Rate (SOCR):
  \[ \beta(\epsilon) = \lim_{n \to \infty} (R(n, \epsilon) - C) \sqrt{n} \]
  \[ \beta^{(fb)}(\epsilon) = \lim_{n \to \infty} (R^{(fb)}(n, \epsilon) - C) \sqrt{n} \]

- **Theorem** [Polyanskiy, Poor, Verdú ('11)]: For symmetric channels, the SOCR is not improved with feedback:
  \[ \beta(\epsilon) = \beta^{(fb)}(\epsilon) \]
  \( (\beta(\epsilon) \) earlier characterized by Strassen ’62).
For symmetric channels, feedback does not improve:
In Fact...

- For symmetric channels, feedback does not improve:
  - the error exponent (for large rates) [Haroutunian ’77; Dobrushin ’62]
In Fact...

- For symmetric channels, feedback does not improve:
  - the error exponent (for large rates) [Haroutunian ’77; Dobrushin ’62]
  - the order of the polynomial pre-factor in the error exponent (for large rates) [Altuğ-Wagner ’21]
In Fact...

- For symmetric channels, feedback does not improve:
  - the error exponent (for large rates) [Haroutunian ’77; Dobrushin ’62]
  - the order of the polynomial pre-factor in the error exponent (for large rates) [Altuğ-Wagner ’21]
  - the *third-order* coding rate [Polyanskiy et al. ’11, Altuğ-Wagner ’21]
For symmetric channels, feedback does not improve:

- the error exponent (for large rates) [Haroutunian ’77; Dobrushin ’62]
- the order of the polynomial pre-factor in the error exponent (for large rates) [Altuğ-Wagner ’21]
- the *third-order* coding rate [Polyanskiy et al. ’11, Altuğ-Wagner ’21]
- the moderate deviations performance [Altuğ-Poor-Verdú (’15)]
In Fact…

- For symmetric channels, feedback does not improve:
  - the error exponent (for large rates) [Haroutunian ’77; Dobrushin ’62]
  - the order of the polynomial pre-factor in the error exponent (for large rates) [Altuğ-Wagner ’21]
  - the third-order coding rate [Polyanskiy et al. ’11, Altuğ-Wagner ’21]
  - the moderate deviations performance [Altuğ-Poor-Verdú (’15)]

- For asymmetric channels,
In Fact...

- For symmetric channels, feedback does not improve:
  - the error exponent (for large rates) [Haroutunian ’77; Dobrushin ’62]
  - the order of the polynomial pre-factor in the error exponent (for large rates) [Altuğ-Wagner ’21]
  - the third-order coding rate [Polyanskiy et al. ’11, Altuğ-Wagner ’21]
  - the moderate deviations performance [Altuğ-Poor-Verdú (’15)]

- For asymmetric channels,
  - The high-rate error exponent is not improved by feedback [Nakiboğlu ’19, Augustin ’78]
In Fact...

- For symmetric channels, feedback does not improve:
  - the error exponent (for large rates) [Haroutunian ’77; Dobrushin ’62]
  - the order of the polynomial pre-factor in the error exponent (for large rates) [Altuğ-Wagner ’21]
  - the third-order coding rate [Polyanskiy et al. ’11, Altuğ-Wagner ’21]
  - the moderate deviations performance [Altuğ-Poor-Verdú (’15)]

- For asymmetric channels,
  - The high-rate error exponent is not improved by feedback [Nakiboğlu ’19, Augustin ’78]
  - We will show that the second-order coding rate can be improved by feedback via a novel mechanism.
A Puzzle

Two coins:

±$1

±$2
A Puzzle

Two coins:

- You begin with $0

±$1

±$2
A Puzzle

Two coins:

- You begin with $0
- You select which coin to flip at each of $n$ steps
A Puzzle

Two coins:

- You begin with $0
- You select which coin to flip at each of \( n \) steps
- How to minimize chance that your final wealth is < \( \alpha \)?
A Puzzle

Two coins:

- You begin with $0
- You select which coin to flip at each of $n$ steps
- How to minimize chance that your final wealth is < $\alpha$?
- Does “feedback” help?
A Puzzle
Strategies: No Feedback
Strategies: No Feedback

Wealth

Time

α

Play Timid

n
Strategies: No Feedback

Wealth vs Time graph with different curves representing different strategies.
Strategies: No Feedback

Play Bold
Strategies: Single Change-point
Strategies: Single Change-point
More Generally

Play Timid

Play Bold

\[ n \]

\( \alpha \)

0
More Generally

With feedback, we can do better.
Can You Flip Your Way To Victory?

Riddler Classic

From Abijith Krishnan comes a game of coin flipping madness:

You have two fair coins, labeled A and B. When you flip coin A, you get 1 point if it comes up heads, but you lose 1 point if it comes up tails. Coin B is worth twice as much — when you flip coin B, you get 2 points if it comes up heads, but you lose 2 points if it comes up tails.

To play the game, you make a total of 100 flips. For each flip, you can choose either coin, and you know the outcomes of all the previous flips. In order to win, you must finish with a positive total score. In your eyes, finishing with 2 points is just as good as finishing with 200 points — any positive score is a win. (By the same token, finishing with 0 or –2 points is just as bad as finishing with –200 points.)

If you optimize your strategy, what percentage of games will you win? (Remember, one game consists of 100 coin flips.)

Extra credit: What if coin A isn't fair (but coin B is still fair)? That is, if coin A comes up heads with probability $p$ and you optimize your strategy, what percentage of games will you win?
Strategies: Continuous-Time

- Let $B(\cdot)$ be a standard Brownian motion.
Strategies: Continuous-Time

- Let $B(\cdot)$ be a standard Brownian motion.

- Let $X(t) = \int_0^t \sigma(X(s), s) dB(s)$ be a controlled diffusion

where $\sigma(\cdot, \cdot) \in [\sigma_1, \sigma_2]; \sigma_1 > 0$
Strategies: Continuous-Time

- Let $B(\cdot)$ be a standard Brownian motion.

- Let $X(t) = \int_0^t \sigma(X(s), s) \, dB(s)$ be a controlled diffusion

  where $\sigma(\cdot, \cdot) \in [\sigma_1, \sigma_2]; \sigma_1 > 0$

- How to select $\sigma(\cdot, \cdot)$ to minimize $P(X(1) < \alpha)$?
Strategies: Continuous-Time

- Let $B(\cdot)$ be a standard Brownian motion.

- Let $X(t) = \int_0^t \sigma(X(s), s) \, dB(s)$ be a controlled diffusion
  
  where $\sigma(\cdot, \cdot) \in [\sigma_1, \sigma_2] ; \sigma_1 > 0$

- How to select $\sigma(\cdot, \cdot)$ to minimize $P(X(1) < \alpha)$?

- **Theorem** (McNamara '83): The bang-bang controller

  \[
  \sigma(x, s) = \begin{cases} 
  \sigma_1 & \text{if } x \geq \alpha \\
  \sigma_2 & \text{if } x < \alpha 
  \end{cases}
  \]

  is optimal.
In Simulation
In the Real World
In the Real World
In the Real World

- McNamara ’83: foraging animals
The McNamara Threshold

![Graph showing the McNamara Threshold]

The McNamara Threshold is a statistical measure used to evaluate the performance of classification systems. It compares the probability of detection (PD) against the probability of false alarm (PF) to determine the optimal threshold for classification.
Let $\Gamma^{(fb)}(\epsilon, \sigma_1, \sigma_2) = \max\{\alpha : \text{with feedback, } \Pr(\text{failure}) \leq \epsilon\}$. 
The McNamara Threshold

- Let $\Gamma^{(fb)}(\epsilon, \sigma_1, \sigma_2) = \max\{\alpha : \text{with feedback, Pr(failure)} \leq \epsilon\}$.
- Let $\Gamma(\epsilon, \sigma_1, \sigma_2) = \max\{\alpha : \text{without feedback, Pr(failure)} \leq \epsilon\}$.
Let $\Gamma^{(fb)}(\epsilon, \sigma_1, \sigma_2) = \max\{\alpha : \text{with feedback, } \Pr(\text{failure}) \leq \epsilon\}$. 
Let $\Gamma(\epsilon, \sigma_1, \sigma_2) = \max\{\alpha : \text{without feedback, } \Pr(\text{failure}) \leq \epsilon\}$. 
Both expressible in terms of inv. Gaussian CDF.
The McNamara Threshold

- Let $\Gamma^{(fb)}(\epsilon, \sigma_1, \sigma_2) = \max\{\alpha : \text{with feedback, Pr(failure)} \leq \epsilon\}$.
- Let $\Gamma(\epsilon, \sigma_1, \sigma_2) = \max\{\alpha : \text{without feedback, Pr(failure)} \leq \epsilon\}$.
- Both expressible in terms of inv. Gaussian CDF
- **Lemma**: $\Gamma(\epsilon, \sigma_1, \sigma_2) < \Gamma^{(fb)}(\epsilon, \sigma_1, \sigma_2)$ iff $\sigma_1 < \sigma_2$
A Key Lemma

- **Def**: a *controller* is a function

\[
f : (\mathcal{X} \times \mathcal{Y})^* \to \mathcal{P}(\mathcal{X})
\]

which along with the channel \( W \) defines a joint distribution

\[
(f \circ W)(x^n, y^n) = \prod_{i=1}^{n} (f(x^{i-1}, y^{i-1}))(x_i)W(y_i|x_i)
\]
A Key Lemma

- **Def:** a *controller* is a function
  \[ f : (\mathcal{X} \times \mathcal{Y})^* \to \mathcal{P}(\mathcal{X}) \]
  which along with the channel \( W \) defines a joint distribution
  \[ (f \circ W)(x^n, y^n) = \prod_{i=1}^{n} (f(x^{i-1}, y^{i-1}))(x_i)W(y_i|x_i) \]

- **Lemma** (cf. Shannon ’57, Fong-Tan ’17; Wang et al. ’09; Blahut ’74): The SOCR with feedback, \( \beta^{(fb)}(\epsilon) \), is the largest \( \alpha \) such that
  \[
  \lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} \leq nC + \alpha \sqrt{n} \right) < \epsilon
  \]
A Key Lemma

- **Def:** A controller is a function
  \[ f : (\mathcal{X} \times \mathcal{Y})^* \rightarrow \mathcal{P(\mathcal{X})} \]
  which along with the channel \( W \) defines a joint distribution
  \[ (f \circ W)(x^n, y^n) = \prod_{i=1}^{n} (f(x^{i-1}, y^{i-1})) \]

- **Lemma** (cf. Shannon '57, Fong-Tan '17; Wang et al. '09; Blahut '74): The SOCR with feedback, \( \beta^{(fb)}(\varepsilon) \), is the largest \( \alpha \) such that
  \[
  \lim_{n \to \infty} \inf \sum_{i=1}^{n} \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} \leq nC + \alpha \sqrt{n} < \varepsilon
  \]
A Key Lemma

- **Def:** a controller is a function $f : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Y}^n$, which along with the channel $W$ defines a joint distribution $(f \circ W)(x^n, y^n)$.

- **Lemma** (cf. Shannon ’57, Fong-T an ’17; Wang et al. ’09; Blahut ’74): The SOCR with feedback, is the largest $\alpha$ such that

$$\lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} \leq nC + \alpha \sqrt{n} \right) < \epsilon$$

...
A Key Lemma

**Def:** a *controller* is a function

\[ f : (\mathcal{X} \times \mathcal{Y})^* \rightarrow \mathcal{P} (\mathcal{X}) \]

which along with the channel \( W \) defines a joint distribution

\[
(f \circ W)(x^n, y^n) = \prod_{i=1}^{n} (f(x^{i-1}, y^{i-1}))(x_i)W(y_i|x_i)
\]

**Lemma** (cf. Shannon ’57, Fong-Tan ’17; Wang et al. ’09; Blahut ’74): The SOCR with feedback, \( \beta^{(fb)}(\epsilon) \), is the largest \( \alpha \) such that

\[
\lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} \leq nC + \alpha \sqrt{n} \right) < \epsilon
\]
A Key Lemma

- **Def**: a *controller* is a function

\[ f : (X \times Y)^* \rightarrow \mathcal{P}(X) \]

which along with the channel \( W \) defines a joint distribution

\[ (f \circ W)(x^n, y^n) = \prod_{i=1}^{n} (f(x^{i-1}, y^{i-1}))(x_i)W(y_i|x_i) \]

- **Lemma** (cf. Shannon ‘57, Fong-Tan ‘17; Wang *et al.* ’09; Blahut ‘74): The SOCR with feedback, \( \beta^{(fb)}(\epsilon) \), is the largest \( \alpha \) such that

\[
\lim_{n \to \infty} \inf \left( f \circ W \right) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} - C \right) \leq \alpha \sqrt{n} \right) < \epsilon
\]
A Key Lemma

- **Def:** a controller is a function

\[ f : (X \times Y)^* \to \mathcal{P}(X) \]

which along with the channel \( W \) defines a joint distribution

\[ (f \circ W)(x^n, y^n) = \prod_{i=1}^{n} (f(x^{i-1}, y^{i-1}))(x_i)W(y_i|x_i) \]

- **Lemma** (cf. Shannon ‘57, Fong-Tan ‘17; Wang et al. ’09; Blahut ‘74): The SOCR with feedback, \( \beta^{(fb)}(\epsilon) \), is the largest \( \alpha \) such that

\[ \lim_{n \to \infty} \inf_f (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} - C \right) \right) \leq \alpha \sqrt{n} < \epsilon \]

- Non-feedback version as well.
A Key Lemma

- **Def:** a *controller* is a function
  \[ f : (\mathcal{X} \times \mathcal{Y})^* \to \mathcal{P}(\mathcal{X}) \]
  which along with the channel \( W \) defines a joint distribution
  \[
  (f \circ W)(x^n, y^n) = \prod_{i=1}^{n} (f(x^{i-1}, y^{i-1}))(x_i)W(y_i|x_i)
  \]

- **Lemma** (cf. Shannon ’57, Fong-Tan ’17; Wang *et al.* ’09; Blahut ’74): The SOCR with feedback, \( \beta^{(fb)}(\epsilon) \), is the largest \( \alpha \) such that
  \[
  \lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} - C \right) \right) \leq \alpha \sqrt{n} < \epsilon
  \]

- Non-feedback version as well.
- $64K$ question: can we control the variance of the increments?
Fact: \[ PW = Q^* \text{ for all } P : I(P; W) = C \]
Fact: \( PW = Q^* \) for all \( P : I(P; W) = C \)

Def: 
\[
V_{\text{min}} = \min_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]
\[
V_{\text{max}} = \max_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]
Variance Definitions

Fact: \( PW = Q^* \) for all \( P \).

Def: Let \( P_{\text{min}} \) be a minimizer.

\[
V_{\text{min}} = \min_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
V_{\text{max}} = \max_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]
Variance Definitions

**Fact:** \( PW = Q^* \) for all \( P : I(P; W) = C \)

**Def:**

\[
V_{\text{min}} = \min_{P:I(P;W)=C} \text{Var}_{P \circ W} \left[ \log \frac{Q^*(Y)}{Q^*(Y)} \right]
\]

\[
V_{\text{max}} = \max_{P:I(P;W)=C} \text{Var}_{P \circ W} \left[ \log \frac{Q^*(Y)}{Q^*(Y)} \right]
\]

Let \( P_{\text{max}} \) be a maximizer
Variance Definitions

Fact: \( PW = Q^* \) for all \( P : I(P; W) = C \)

Def: 

\[
V_{\text{min}} = \min_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
V_{\text{max}} = \max_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]
**Fact:** \( PW = Q^* \) for all \( P : I(P; W) = C \)

**Def:**

\[
V_{\text{min}} = \min_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
V_{\text{max}} = \max_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
\nu_{\text{min}} = \min_{X} \text{Var}_{W(\cdot|X)} \left[ \log \frac{W(Y|x)}{Q^*(Y)} \right]
\]

\[
\nu_{\text{max}} = \max_{X} \text{Var}_{W(\cdot|X)} \left[ \log \frac{W(Y|x)}{Q^*(Y)} \right]
\]
Variance Definitions

**Fact:** \( PW = Q^* \) for all \( P : I(P; W) = C \)

**Def:**

\[
V_{\text{min}} = \min_{P : I(P; W)=C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right] \\
V_{\text{max}} = \max_{P : I(P; W)=C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right] \\

\nu_{\text{min}} = \min_x \text{Var}_{W(\cdot|x)} \left[ \log \frac{W(Y|x)}{Q^*(Y)} \right] \\
\nu_{\text{max}} = \max_x \text{Var}_{W(\cdot|x)} \left[ \log \frac{W(Y|x)}{Q^*(Y)} \right]
\]

\( \nu_{\text{min}} \leq V_{\text{min}} \leq V_{\text{max}} \leq \nu_{\text{max}} \)
**Variance Definitions**

**Fact:** \( PW = Q^* \) for all \( P : I(P; W) = C \)

**Def:**

\[
V_{\min} = \min_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
V_{\max} = \max_{P : I(P; W) = C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
\nu_{\min} = \min_{X} \text{Var}_{W(\cdot|X)} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
\nu_{\max} = \max_{X} \text{Var}_{W(\cdot|X)} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[\nu_{\min} \leq V_{\min} \leq V_{\max} \leq \nu_{\max}\]
Variance Definitions

Fact: \( PW = Q^* \) for all \( P : I(P; W) = C \)

Def:

\[
V_{\text{min}} = \min_{P : I(P; W) = C} \text{Var}_{P \cdot W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
V_{\text{max}} = \max_{P : I(P; W) = C} \text{Var}_{P \cdot W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
\nu_{\text{min}} = \min_X \text{Var}_{W(.|X)} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right] \quad [\text{assumed } > 0 \ 	ext{throughout}]
\]

\[
\nu_{\text{max}} = \max_X \text{Var}_{W(.|X)} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

Compound dispersion if \( V_{\text{min}} < V_{\text{max}} \). Otherwise simple dispersion.
Variance Definitions

Fact: \( PW = Q^* \) for all \( P : I(P; W) = C \)

Def: 

\[
V_{\text{min}} = \min_{P : I(P; W)=C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
V_{\text{max}} = \max_{P : I(P; W)=C} \text{Var}_{P \circ W} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
\nu_{\text{min}} = \min_X \text{Var}_{W(\cdot | X)} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

\[
\nu_{\text{max}} = \max_X \text{Var}_{W(\cdot | X)} \left[ \log \frac{W(Y|X)}{Q^*(Y)} \right]
\]

[assumed \( > 0 \) throughout]

A channel with a unique capacity-achieving input distribution is necessarily simple dispersion.
A Compound Dispersion Example

\[
W(y|x) = \begin{bmatrix}
    p & 0.5(1-p) & 0.5(1-p) \\
    0.5(1-p) & p & 0.5(1-p) \\
    0.5(1-p) & 0.5(1-p) & p \\
    q & 1-q & 0 \\
    0 & q & 1-q \\
    1-q & 0 & q
\end{bmatrix}
\]

if \( p = 0.8 \)
and \( q \approx 0.337 \)
then \( V_{\text{min}} = .102 \)
\( V_{\text{max}} = .692 \)
A Compound Dispersion Example

\[ W(y|x) = \begin{bmatrix}
  p & 0.5(1-p) & 0.5(1-p) \\
  0.5(1-p) & p & 0.5(1-p) \\
  0.5(1-p) & 0.5(1-p) & p \\
  q & 1-q & 0 \\
  0 & q & 1-q \\
  1-q & 0 & q
\end{bmatrix} \begin{bmatrix}
1/3 \\
1/3 \\
1/3 \\
P_{\text{max}}
\end{bmatrix} \]

If \( p = 0.8 \)
and \( q \approx 0.337 \)
then \( V_{\text{min}} = 0.102 \)
\( V_{\text{max}} = 0.692 \)
A Compound Dispersion Example

\[ W(y|x) = \begin{bmatrix}
  p & 0.5(1-p) & 0.5(1-p) \\
  0.5(1-p) & p & 0.5(1-p) \\
  0.5(1-p) & 0.5(1-p) & p \\
  q & 1-q & 0 \\
  0 & q & 1-q \\
  1-q & 0 & q \\
\end{bmatrix} \begin{bmatrix}
  1/3 \\
  1/3 \\
  1/3 \\
\end{bmatrix} P_{\text{min}} \]

if \( p = 0.8 \)
and \( q \approx 0.337 \)
then \( V_{\text{min}} = 0.102 \)
\( V_{\text{max}} = 0.692 \)
A Compound Dispersion Example

\[ W(y|x) = \begin{bmatrix} 
    p & 0.5(1-p) & 0.5(1-p) \\
    0.5(1-p) & p & 0.5(1-p) \\
    0.5(1-p) & 0.5(1-p) & p \\
    q & 1-q & 0 \\
    0 & q & 1-q \\
    1-q & 0 & q 
\end{bmatrix} \]

if \( p = 0.8 \)
and \( q \approx 0.337 \)
then \( V_{\text{min}} = .102 \)
\( V_{\text{max}} = .692 \)

[Many more examples when we consider cost constraints ...]
Theorem 0: (Strassen ’62) For any DMC, the SOCR satisfies:

$$\beta(\epsilon) = \Gamma(\epsilon, \sqrt{V_{\min}}, \sqrt{V_{\max}})$$
Theorem 0: (Strassen ’62) For any DMC, the SOCR satisfies:

\[ \beta(\epsilon) = \Gamma(\epsilon, \sqrt{V_{\min}}, \sqrt{V_{\max}}) \]

Intuition: By the key lemma, SOCR is the max \( \alpha \) such that

\[
\lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y_i^{i-1})} - C \right) \right) \leq \alpha \sqrt{n} < \epsilon
\]

where \( f \) is “open-loop.” Intuitively, optimal choice should be:

\[
\begin{cases} 
    P_{\min} & \text{if } \Gamma(\epsilon, \sqrt{V_{\min}}, \sqrt{V_{\max}}) < 0 \\
    P_{\max} & \text{if } \Gamma(\epsilon, \sqrt{V_{\min}}, \sqrt{V_{\max}}) > 0
\end{cases}
\]
Theorem 1 (Wagner-Shende-Altuğ ’20): For any DMC with feedback,

\[ \beta_{\text{fb}}(\epsilon) \geq \Gamma_{\text{fb}}(\epsilon, \sqrt{V_{\text{min}}}, \sqrt{V_{\text{max}}}) \]

> \Gamma(\epsilon, \sqrt{V_{\text{min}}}, \sqrt{V_{\text{max}}}) \text{ if } V_{\text{max}} > V_{\text{min}}

= \beta(\epsilon)
Theorem 1 (Wagner-Shende-Altuğ ’20): For any DMC with feedback,

\[
\beta^{(f_b)}(\epsilon) \geq \Gamma^{(f_b)} \left( \epsilon, \sqrt{V_{\min}}, \sqrt{V_{\max}} \right) \\
> \Gamma \left( \epsilon, \sqrt{V_{\min}}, \sqrt{V_{\max}} \right) \text{ if } V_{\max} > V_{\min} \\
= \beta(\epsilon)
\]

Corollary (Wagner-Shende-Altuğ ’20): Feedback improves the second-order coding rate for any compound-dispersion DMC.
Proof of Theorem 1
Proof of Theorem 1

- Key lemma:

\[
\lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_{i-1}|Y_{i-1})} - C \right) \right) \leq \alpha \sqrt{n} < \epsilon
\]
Proof of Theorem 1

- Key lemma:

\[
\lim_{n \to \infty} \inf_f (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y_i^{i-1})} - C \right) \right) \leq \alpha \sqrt{n} < \epsilon
\]

- Choose \( f(x^k,y^k) \) to be capacity-achieving for each \( (x^k,y^k) \):
  - Increment is zero mean
Proof of Theorem 1

- Key lemma:

\[
\lim_{n \to \infty} \inf_{f \circ W} \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} - C \right) \right) \leq \alpha \sqrt{n} < \epsilon
\]

- Choose \( f(x^k,y^k) \) to be capacity-achieving for each \( (x^k,y^k) \):
  - Increment is zero mean

- Select bang-bang \( f \):

\[
f(x^k,y^k) = \begin{cases} 
P_{\min} & \text{if running sum} > \alpha \sqrt{n} \\
P_{\max} & \text{if running sum} \leq \alpha \sqrt{n} \end{cases}
\]
Proof of Theorem 1

Key lemma:

\[ \lim_{n \to \infty} \inf_{f} \left( f \circ W \right) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y_{i-1})} - C \right) \right) \leq \alpha \sqrt{n} < \epsilon \]

Choose \( f(x^k, y^k) \) to be capacity-achieving for each \( (x^k, y^k) \):
- Increment is zero mean

Select bang-bang \( f \):

\[
  f(x^k, y^k) = \begin{cases} 
    P_{\text{min}} & \text{if running sum} > \alpha \sqrt{n} \\
    P_{\text{max}} & \text{if running sum} \leq \alpha \sqrt{n} 
  \end{cases}
\]

Show convergence to cont.-time controlled diffusion
- Not Lipschitz ...

Apply McNamara’s characterization of bang-bang controller
Proof of Theorem 1

- Key lemma:

\[
\lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y_{i-1})} - C \right) \right) \leq \alpha \sqrt{n} < \epsilon
\]

- Choose \( f(x^k,y^k) \) to be capacity-achieving for each \( (x^k,y^k) \):
  - Increment is zero mean

- Select bang-bang \( f \):

\[
f(x^k, y^k) = \begin{cases} 
\text{Timid/Bold Coding} & \text{if max running sum} \geq \alpha \sqrt{n} \\
\text{Timid/Bold Coding} & \text{if max running sum} \leq \alpha \sqrt{n}
\end{cases}
\]

- Show convergence to cont.-time controlled diffusion
  - Not Lipschitz ...

- Apply McNamara’s characterization of bang-bang controller
A Compound Dispersion Example

\[ W(y|x) = \begin{bmatrix}
p & 0.5(1-p) & 0.5(1-p) \\
0.5(1-p) & p & 0.5(1-p) \\
0.5(1-p) & 0.5(1-p) & p \\
q & 1-q & 0 \\
0 & q & 1-q \\
1-q & 0 & q \\
\end{bmatrix} \]

If \( p = 0.8 \)
and \( q \approx 0.337 \)
then \( V_{\text{min}} = .102 \)
\( V_{\text{max}} = .692 \)
Numerical Example
When Does Feedback Help?

Feedback Improves SOCR
- Compound-Dispersion Channels

Feedback Does Not Improve SOCR
- Symmetric Channels
When Does Feedback Help?

- **Theorem 2** (Wagner-Shende-Altug): Feedback improves the second-order coding rate iff the channel is compound dispersion.
Proof of Theorem 2
Proof of Theorem 2

- By the key lemma, suffices to show that

\[
\lim_{n \to \infty} \inf_{f \circ W} \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} - C \right) \right) \leq \Gamma(\epsilon, \sqrt{V_{\text{min}}}, \sqrt{V_{\text{max}}}) \cdot \sqrt{n} > \epsilon
\]
Proof of Theorem 2

- By the key lemma, suffices to show that

$$\lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} - C \right) \right) \leq \Gamma(\varepsilon, \sqrt{V_{\text{min}}}, \sqrt{V_{\text{max}}}) \cdot \sqrt{n} > \varepsilon$$

- Can reduce to controller such that $X^n$ is empirically capacity achieving w.h.p. [Fong-Tan '17]
Proof of Theorem 2

- By the key lemma, suffices to show that

$$\lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} - C \right) \right) \leq \Gamma(\epsilon, \sqrt{V_{\min}}, \sqrt{V_{\max}}) \cdot \sqrt{n} > \epsilon$$

- Can reduce to controller such that $X^n$ is empirically capacity achieving w.h.p. [Fong-Tan ’17]

  - Simple dispersion $\to$ sum of conditional variances of the terms in the sum given the past is fixed.
Proof of Theorem 2

- By the key lemma, suffices to show that

\[ \lim_{n \to \infty} \inf_{f} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i | X_i)}{(fW)(Y_i | Y_{i-1})} - C \right) \right) \leq \Gamma(\epsilon, \sqrt{V_{\min}}, \sqrt{V_{\max}}) \cdot \sqrt{n} \] > \epsilon

- Can reduce to controller such that \( X^n \) is empirically capacity achieving w.h.p. [Fong-Tan ’17]
  
  - Simple dispersion \( \rightarrow \) sum of conditional variances of the terms in the sum given the past is fixed.

- Apply martingale CLT [Bolthausen ’82]
Theorem 1 (Wagner-Shende-Altuğ ’20): For any DMC with feedback,

\[ \beta^{(fb)}(\epsilon) \geq \Gamma^{(fb)}(\epsilon, \sqrt{V_{\min}}, \sqrt{V_{\max}}) \]
By How Much Does Feedback Help?

- **Theorem 1** (Wagner-Shende-Altuğ ’20): For any DMC with feedback,
  \[ \beta^{(fb)}(\epsilon) \geq \gamma^{(fb)}\left(\epsilon, \sqrt{V_{\text{min}}}, \sqrt{V_{\text{max}}}\right) \]

- **Theorem 3** (Wagner-Shende-Altuğ ’20): For any DMC with feedback,
  \[ \beta^{(fb)}(\epsilon) \leq \gamma^{(fb)}\left(\epsilon, \sqrt{V_{\text{min}}}, \sqrt{V_{\text{max}}}\right) \]
Proof of Theorem 3
Proof of Theorem 3

- By the key lemma, suffices to show that for any controller:

\[
\lim_{n \to \infty} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} - C \right) \right) \leq \Gamma^{(fb)}(\epsilon, \sqrt{\nu_{\min}}, \sqrt{\nu_{\max}}) \sqrt{n} > \epsilon
\]
Proof of Theorem 3

- By the key lemma, suffices to show that for any controller:

\[
\lim_{n \to \infty} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y_{i-1})} - C \right) \right) \leq \Gamma^{(fb)}(\varepsilon, \sqrt{\nu_{\text{min}}}, \sqrt{\nu_{\text{max}}}) \sqrt{n} > \varepsilon
\]

- Weaken by replacing \( fW \) with \( Q^* \):

\[
\lim_{n \to \infty} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{Q^*(Y_i)} - C \right) \right) \leq \Gamma^{(fb)}(\varepsilon, \sqrt{\nu_{\text{min}}}, \sqrt{\nu_{\text{max}}}) \sqrt{n} > \varepsilon
\]
Proof of Theorem 3

By the key lemma, suffices to show that for any controller:

\[
\lim_{n \to \infty} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y_{i-1})} - C \right) \right) \leq 
\Gamma^{(fb)}(\epsilon, \sqrt{\nu_{\min}}, \sqrt{\nu_{\max}}) \sqrt{n} > \epsilon
\]

Weaken by replacing \(fW\) with \(Q^*\):

\[
\lim_{n \to \infty} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{Q^*(Y_i)} - C \right) \right) \leq 
\Gamma^{(fb)}(\epsilon, \sqrt{\nu_{\min}}, \sqrt{\nu_{\max}}) \sqrt{n} > \epsilon
\]

DT martingale w.r.t. \(\sigma(X_{i-1}, Y_{i-1})\); cond. variance in \([\nu_{\min}, \nu_{\max}]\)
Proof of Theorem 3

- By the key lemma, suffices to show that for any controller:

\[
\lim_{n \to \infty} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{(fW)(Y_i|Y^{i-1})} - C \right) \right) \leq \Gamma^{(fb)}(\epsilon, \sqrt{\nu_{\min}}, \sqrt{\nu_{\max}}) \sqrt{n} > \epsilon
\]

- Weaken by replacing \(fW\) with \(Q^*\):

\[
\lim_{n \to \infty} (f \circ W) \left( \sum_{i=1}^{n} \left( \log \frac{W(Y_i|X_i)}{Q^*(Y_i)} - C \right) \right) \leq \Gamma^{(fb)}(\epsilon, \sqrt{\nu_{\min}}, \sqrt{\nu_{\max}}) \sqrt{n} > \epsilon
\]

[to apply McNamara, need to switch to cont.-time]

DT martingale w.r.t. \(\sigma(X^{i-1}, Y^{i-1})\); cond. variance in \(\nu_{\min}, \nu_{\max}\)
Proof of Theorem 3

- **Theorem (Strassen '67):** If \( \{S_n\} \) is a square-integrable martingale with \( S_0 = 0 \), then there exists a Brownian motion \( B(\cdot) \) and a sequence of stopping times \( 0 = T_0 \leq T_1 \leq \ldots \leq T_n \) such that

\[
(S_0, S_1, \ldots, S_n) \overset{d}{=} (B(T_0), B(T_1), \ldots, B(T_n))
\]

and

\[
E[T_k - T_{k-1}|S_1, \ldots, S_{k-1}, T_1, \ldots, T_{k-1}] = \text{Var}(S_k - S_{k-1}|S_1, \ldots, S_{k-1})
\]
Proof of Theorem 3

- So view $f$ as selecting stopping times:
Proof of Theorem 3

- So view $f$ as selecting stopping times:
Then view $f$ as speeding up the BM instead of waiting:
Proof of Theorem 3

- Then view \( f \) as speeding up the BM instead of waiting:
Proof of Theorem 3

- Then view $f$ as speeding up the BM instead of waiting:

- Such $f$ is *nearly* a feasible scheme in McNamara’s problem
Proof of Theorem 3

- Then view $f$ as speeding up the BM instead of waiting:

- Such $f$ is nearly a feasible scheme in McNamara’s problem
- Make feasible and apply McNamara’s optimality result
By How Much Does Feedback Help?

- **Theorem 1** (Wagner-Shende-Altuğ ’20): For any DMC with feedback,
  \[ \beta^{(fb)}(\varepsilon) \geq \Gamma^{(fb)} \left( \varepsilon, \sqrt{V_{\text{min}}}, \sqrt{V_{\text{max}}} \right) \]

- **Theorem 3** (Wagner-Shende-Altuğ ’20): For any DMC with feedback,
  \[ \beta^{(fb)}(\varepsilon) \leq \Gamma^{(fb)} \left( \varepsilon, \sqrt{\mathcal{V}_{\text{min}}}, \sqrt{\mathcal{V}_{\text{max}}} \right) \]
A Compound Dispersion Example

\[
W(y|x) = \begin{bmatrix}
p & 0.5(1-p) & 0.5(1-p) \\
0.5(1-p) & p & 0.5(1-p) \\
0.5(1-p) & 0.5(1-p) & p \\
q & 1-q & 0 \\
0 & q & 1-q \\
1-q & 0 & q \\
\end{bmatrix}
\]

if \( p = 0.8 \)
and \( q \approx 0.337 \)
then \( V_{\min} = .102 \)
\( V_{\max} = .692 \)
A Compound Dispersion Example

\[
W(y|x) = \begin{bmatrix}
  p & 0.5(1-p) & 0.5(1-p) \\
  0.5(1-p) & p & 0.5(1-p) \\
  0.5(1-p) & 0.5(1-p) & p \\
  q & 1-q & 0 \\
  0 & q & 1-q \\
  1-q & 0 & q
\end{bmatrix}
\]

if \( p = 0.8 \)
and \( q \approx 0.337 \)
then \( V_{\text{min}} = .102 \)
\( V_{\text{max}} = .692 \)
\( \nu_{\text{min}} = V_{\text{min}} \)
\( \nu_{\text{max}} = V_{\text{max}} \)
A Compound Dispersion Example

\[ W(y|x) = \begin{bmatrix}
    p & 0.5(1-p) & 0.5(1-p) \\
    0.5(1-p) & p & 0.5(1-p) \\
    0.5(1-p) & 0.5(1-p) & p \\
    q & 1-q & 0 \\
    0 & 1-q & q \\
    1-q & 0 & 1-q \\
\end{bmatrix} \]

if \( p = 0.8 \)
and \( q \approx 0.337 \)
then \( V_{\min} = 0.102 \)
\( V_{\max} = 0.692 \)
\( \nu_{\min} = V_{\min} \)
\( \nu_{\max} = V_{\max} \)

... so the upper bound is tight in this case.
Numerical Example

Graph showing the second order coding rate as a function of error probability ($\varepsilon$) with and without feedback.
A Compound Dispersion Example

\[ W(y|x) = \begin{bmatrix}
  p & 0.5(1-p) & 0.5(1-p) \\
  0.5(1-p) & p & 0.5(1-p) \\
  0.5(1-p) & 0.5(1-p) & p \\
  q & 1-q & 0 \\
  0 & q & 1-q \\
  1-q & 0 & q
\end{bmatrix} \]

if \( p = 0.8 \)
and \( q \approx 0.337 \)
then \( V_{\text{min}} = 0.102 \)
\( V_{\text{max}} = 0.692 \)
\( \nu_{\text{min}} = V_{\text{min}} \)
\( \nu_{\text{max}} = V_{\text{max}} \)
The AWGN

\[ Z^n \sim \mathcal{N}(\bar{0}, I \cdot N) \]

\[ X^n \quad \oplus \quad Y^n \]

power constraint: \( P \)
The AWGN

\[ Z^n \sim \mathcal{N}(\mathbf{0}, I \cdot N) \]

\[ X^n \rightarrow Y^n \]

If \( X^n \) is drawn uniformly from the radius-\( \sqrt{nP} \) sphere:

\[
\frac{1}{n} \text{Var} \left[ \log \frac{W(Y^n|X^n)}{Q^*(Y^n)} \right] = \frac{P(P + 2N)}{2(P + N)^2}
\]

If \( X^n \) is drawn i.i.d. \( \mathcal{N}(0, P) \):

\[
\frac{1}{n} \text{Var} \left[ \log \frac{W(Y^n|X^n)}{Q^*(Y^n)} \right] = \frac{P}{P + N}
\]
The AWGN

\[ Z^n \sim \mathcal{N}(\mathbf{0}, I \cdot N) \]

\[ X^n \rightarrow Y^n \]

power constraint: \( P \)

If \( X^n \) is drawn uniformly from the radius-\( \sqrt{nP} \) sphere:

\[
\frac{1}{n} \text{Var} \left[ \log \frac{W(Y^n|X^n)}{Q^*(Y^n)} \right] = \frac{P(P + 2N)}{2(P + N)^2} \quad \text{[PSK]}
\]

If \( X^n \) is drawn i.i.d. \( \mathcal{N}(0, P) \):

\[
\frac{1}{n} \text{Var} \left[ \log \frac{W(Y^n|X^n)}{Q^*(Y^n)} \right] = \frac{P}{P + N} \quad \text{[QAM]}
\]
The AWGN

\[ Z^n \sim \mathcal{N}(\hat{0}, I \cdot N) \]

power constraint: \( P \)

\[ X^n \rightarrow Y^n \]

If \( X^n \) is drawn uniformly from the radius-\( \sqrt{nP} \) sphere:

\[
\frac{1}{n} \text{Var} \left[ \log \frac{W(Y^n|X^n)}{Q^*(Y^n)} \right] = \frac{P(P + 2N)}{2(P + N)^2} \tag{PSK}
\]

If \( X^n \) is drawn i.i.d. \( \mathcal{N}(0, P) \):

\[
\frac{1}{n} \text{Var} \left[ \log \frac{W(Y^n|X^n)}{Q^*(Y^n)} \right] = \frac{P}{P + N} \tag{QAM}
\]

[Similarly for any DMC with an active cost constraint]
How does one use feedback to improve block coding performance in point-to-point channels?
Conclusion

‣ How does one use feedback to improve block coding performance in point-to-point channels?
  ▸ Exploit channel memory to predict the future
How does one use feedback to improve block coding performance in point-to-point channels?

- Exploit channel memory to predict the future
- Learn the channel law
Conclusion

- How does one use feedback to improve block coding performance in point-to-point channels?
  - Exploit channel memory to predict the future
  - Learn the channel law
  - Opportunistically vary the decoding time
Conclusion

- How does one use feedback to improve block coding performance in point-to-point channels?
  - Exploit channel memory to predict the future
  - Learn the channel law
  - Opportunistically vary the decoding time
  - Opportunistically vary the power
Conclusion

- How does one use feedback to improve block coding performance in point-to-point channels?
  - Exploit channel memory to predict the future
  - Learn the channel law
  - Opportunistically vary the decoding time
  - Opportunistically vary the power
  - Increase the effective minimum distance
Conclusion

- How does one use feedback to improve block coding performance in point-to-point channels?
  - Exploit channel memory to predict the future
  - Learn the channel law
  - Opportunistically vary the decoding time
  - Opportunistically vary the power
  - Increase the effective minimum distance
  - *Use timid/bold coding*