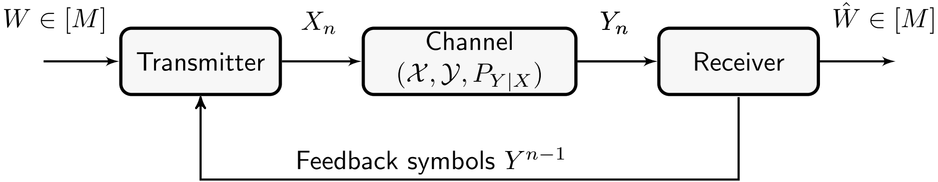
Variable-Length Feedback Codes over Known and Unknown Channels

Vincent Tan
Joint work with **Recep Can Yavas**

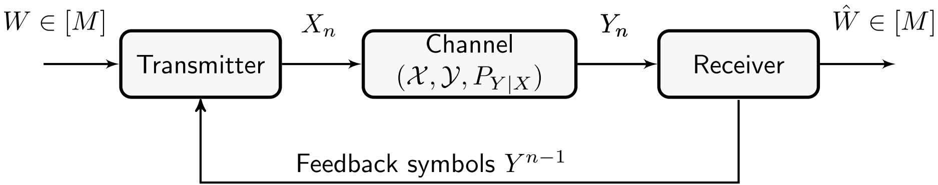
National University of Singapore

ITW 2024, Shenzhen, China

Variable-length feedback (VLF) codes



- Decoding occurs at a random stopping time $\tau \in \mathbb{N}$ (Burnashev '76)
  - Burnashev proved that the optimal error exponent is $\lim_{\epsilon \to 0} \frac{-\log \epsilon}{\mathbb{E}[\tau]} = C_1 \left(1 - \frac{R}{C}\right)$
- Polyanskiy et al. '11 showed an achievability bound for fixed $\epsilon \in (0, 1)$
  - Their scheme employs only stop-feedback



- An (N, M, ϵ)-VLF code consists of a sequence of encoders f_n, a stopping time τ, and a decoder g:

$$X_n = f_n(W, Y^{n-1}) \qquad W \sim \text{Unif}[M]$$
$$\hat{W} = g(Y^\tau) \qquad \tau: \text{ stopping time of filtration } \sigma\{Y^n\}$$
$$\mathbb{P}\left[W \neq \hat{W}\right] \leq \epsilon \qquad \text{average error probability}$$
$$\mathbb{E}[\tau] \leq N \qquad \text{average decoding time}$$

- $M^*(N, \epsilon) = \max\{M : \exists\ (N, M, \epsilon)\text{-VLF code}\}$

Preliminary definitions

- Information density:

$$\imath(x^n; y^n) = \sum_{i=1}^{n} \log \frac{P_{Y|X}(y_i|x_i)}{P_Y(y_i)}$$

- Mutual information:

$$I(P_X, P_{Y|X}) = \mathbb{E}\left[\imath(X;Y)\right] = \sum_{x,y} P_X(x) P_{Y|X}(y|x) \log \frac{P_{Y|X}(y|x)}{P_Y(y)}$$

- Capacity:

$$C = \max_{P_X} I(P_X, P_{Y|X})$$

- KL divergence between the two most distinguishable inputs:

$$C_1 = \max_{x_{\mathrm{A}}, x_{\mathrm{R}}} D(P_{Y|X=x_{\mathrm{A}}} \| P_{Y|X=x_{\mathrm{R}}})$$

Some Background on Second-order Terms in Channel Coding

- In fixed-length channel coding (decoding is done at a fixed time n):

1. Shannon (1949) shows

$$\log M^*(n, \epsilon) = nC + o(n).$$

2. Strassen (1961) and Polyanskiy et al. (2010) show

$$\log M^*(n, \epsilon) = nC - \sqrt{nV} Q^{-1}(\epsilon) + O(\log n)$$

where $V = \mathrm{Var}\left[\imath(X;Y)\right]$ and $Q^{-1}(\cdot)$ is complementary Gaussian CDF.

Some Background on Second-order Terms in Channel Coding

- In variable-length channel coding (decoding is done at a random time τ with $\mathbb{E}[\tau] \leq N$):

  1. Polyanskiy et al. (2011) show

$$\frac{NC}{1-\epsilon} - \log N + O(1) \leq \log M^*(N, \epsilon) \leq \frac{NC}{1-\epsilon} + O(1)$$

  - The first-order term is larger by a multiplicative factor of $\frac{1}{1-\epsilon}$
  - Faster convergence to the first-order term since the second-order term is $O(\log N)$ instead of $O(\sqrt{N})$

Our contributions

- We prove a novel achievability bound in the fixed-ϵ regime.
  Our result improves the second-order term from

$$-\log N \qquad \text{to} \qquad -\frac{C}{C_1}\log N$$

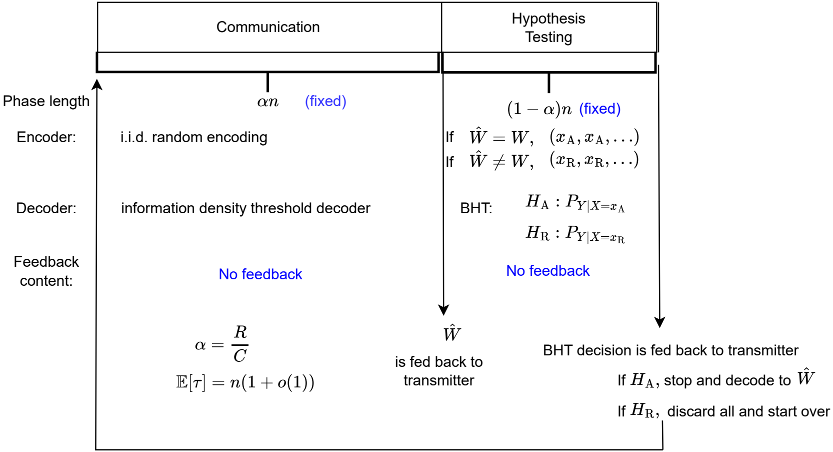
  ($C < C_1$ holds for all nontrivial DMCs)

- Our code is a refinement of the Yamamoto–Itoh scheme.

- We universalize our code for a scenario where the channel is unknown.

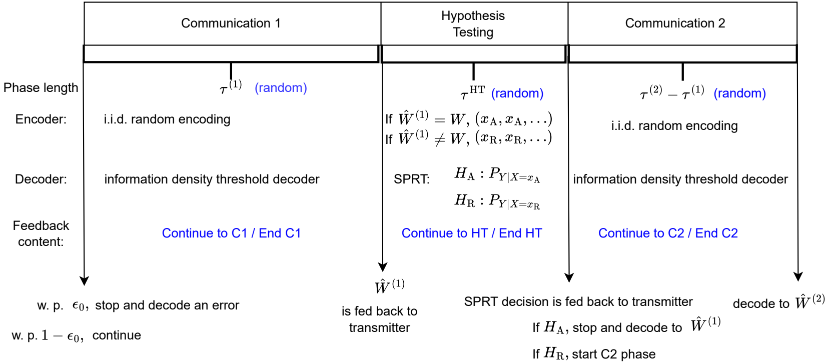
Original Yamamoto–Itoh Scheme

- Phases go indefinitely until decoding occurs.



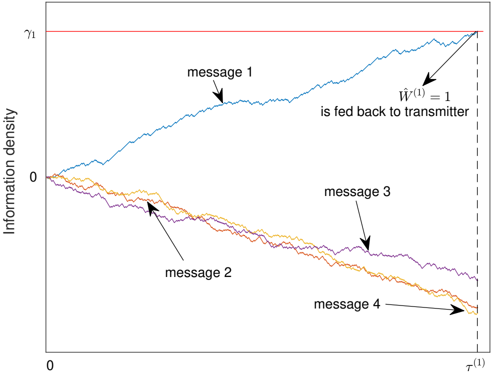
Our VLF code

- Consists of 3 phases: C1, HT, and C2.



- Message W is uniformly distributed on $[M] = \{1, \ldots, M\}$.
- **Encoder:** The transmitter transmits the codeword associated with W symbol by symbol.
- $\mathbf{c}(W) \sim P_X^{\infty}$
  infinite-length codeword corresponding to W
  $\mathbf{c}^n(W) =$ first n symbols of $\mathbf{c}(W)$

- **Decoder:**



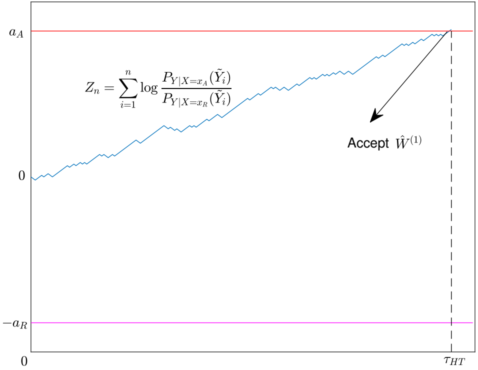
$$\tau^{(1)} = \inf\{n \geq 1 \colon \max_{m \in [M]} \imath(\mathbf{c}^n(m); Y^n) > \gamma_1\}$$

- Upon receiving $\hat{W}^{(1)}$, if $W = \hat{W}^{(1)}$ (correct estimate), the transmitter transmits $(x_{\mathrm{A}}, x_{\mathrm{A}}, \dots)$. If $W \neq \hat{W}^{(1)}$, it transmits $(x_{\mathrm{R}}, x_{\mathrm{R}}, \dots)$, where

$$(x_{\mathrm{A}}, x_{\mathrm{R}}) = \arg\max_{x_{\mathrm{A}}, x_{\mathrm{R}}} D(P_{Y|X=x_{\mathrm{A}}} \| P_{Y|X=x_{\mathrm{R}}})$$

The receiver receives $(\tilde{Y}_1, \tilde{Y}_2, \dots)$.

- **Decoder:** SPRT



$$\tau^{\mathrm{HT}} = \inf \left\{ n \geq 1 \colon \sum_{i=1}^{n} \log \frac{P_{Y|X=x_{\mathrm{A}}}(\tilde{Y}_i)}{P_{Y|X=x_{\mathrm{R}}}(\tilde{Y}_i)} \in [-a_{\mathrm{R}}, a_{\mathrm{A}}] \right\}$$

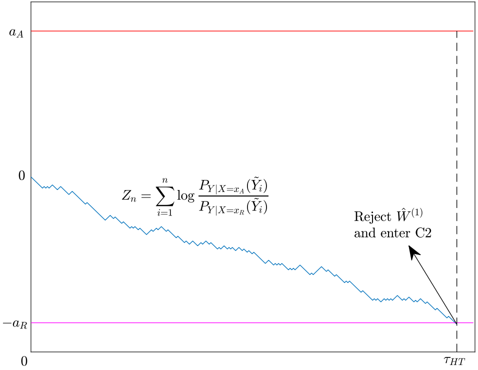
HT phase

- Upon receiving $\hat{W}^{(1)}$, if $W = \hat{W}^{(1)}$ (correct estimate), the transmitter transmits $(x_{\mathrm{A}}, x_{\mathrm{A}}, \ldots)$. If $W \neq \hat{W}^{(1)}$, it transmits $(x_{\mathrm{R}}, x_{\mathrm{R}}, \ldots)$, where

$$(x_{\mathrm{A}}, x_{\mathrm{R}}) = \arg\max_{x_{\mathrm{A}}, x_{\mathrm{R}}} D(P_{Y|X=x_{\mathrm{A}}} \| P_{Y|X=x_{\mathrm{R}}})$$

The receiver receives $(\tilde{Y}_1, \tilde{Y}_2, \ldots)$.

- **Decoder:** SPRT



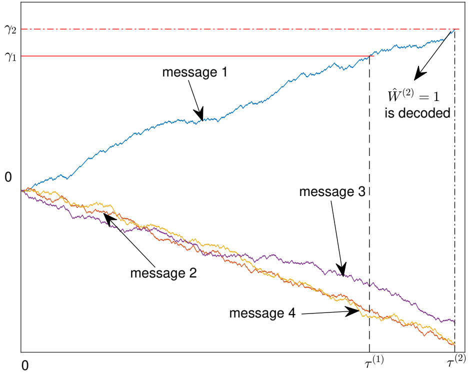
$$\tau^{\mathrm{HT}} = \inf\left\{n \geq 1 \colon \sum_{i=1}^{n} \log \frac{P_{Y|X=x_{\mathrm{A}}}(\tilde{Y}_i)}{P_{Y|X=x_{\mathrm{R}}}(\tilde{Y}_i)} \in [-a_{\mathrm{R}}, a_{\mathrm{A}}]\right\}$$

C2 phase

- **Encoder:** The transmitter continues to transmit the symbols from $\mathbf{c}(W)$.

- $\mathbf{c}_{\tau^{(1)}+1}(W), \mathbf{c}_{\tau^{(1)}+2}(W), \ldots$ are transmitted until the decoder makes a decision

- **Decoder:**



$$\tau^{(2)} = \inf\{n \geq 1 \colon \max_{m \in [M]} \imath(\mathbf{c}^n(m); Y^n) > \gamma_2\}$$

Stop-at-time-zero strategy

- To improve the performance, we employ Polyanskiy et al.'s *stop-at-time-zero* strategy:

$$\begin{aligned}&\text{With probability } \epsilon_0, \text{decode to an arbitrary message at time } \tau = 0\\&\text{With probability } 1-\epsilon_0, \text{ use the code described above}\end{aligned}$$

- Overall error probability

$$\epsilon \leq (1-\epsilon_0)\epsilon' + \epsilon_0$$

- Overall average decoding time

$$N = (1-\epsilon_0)N',$$

where N' and ϵ' are the average decoding time and average error probability of the Yamamoto–Itoh code.

Theorem 1

There exists an (N, M, ϵ)-VLF code with

$$N \leq (1-\epsilon_0)N' \quad \textit{and} \quad \epsilon \leq \epsilon_0 + (1-\epsilon_0)\epsilon'$$

where

$$\begin{aligned}
\epsilon' &= (M-1)\exp\{-(\gamma_1 + a_{\mathrm{A}})\} + (M-1)\exp\{-\gamma_2\} \\
N' &= \frac{\gamma_1 + b}{C} + ((M-1)\exp\{-\gamma_1\} + \exp\{-a_{\mathrm{R}}\})\frac{\gamma_2 - \gamma_1 + b}{C} + \frac{a_{\mathrm{A}} + b_{\mathrm{A}}}{D(P_{Y|X=x_{\mathrm{A}}} \| P_{Y|X=x_{\mathrm{R}}})} \\
&\quad + (M-1)\exp\{-\gamma_1\}\frac{a_{\mathrm{R}} + b_{\mathrm{R}}}{D(P_{Y|X=x_{\mathrm{R}}} \| P_{Y|X=x_{\mathrm{A}}})}
\end{aligned}$$

$b, b_{\mathrm{A}}, b_{\mathrm{R}}$ *are constants, and* $\epsilon_0, \gamma_1, \gamma_2, a_{\mathrm{A}}$, *and* a_{R} *are the parameters of our code design.*

Nonasymptotic bound

Theorem 1

There exists an (N, M, ϵ)*-VLF code with*

$$N \leq (1-\epsilon_0)N' \quad \textit{and} \quad \epsilon \leq \epsilon_0 + (1-\epsilon_0)\epsilon'$$

where

$$\epsilon' = (M-1)\exp\{-(\gamma_1 + a_{\mathrm{A}})\} + (M-1)\exp\{-\gamma_2\}$$

$$N' = \frac{\gamma_1 + b}{C} + ((M-1)\exp\{-\gamma_1\} + \exp\{-a_{\mathrm{R}}\})\frac{\gamma_2 - \gamma_1 + b}{C} + \frac{a_{\mathrm{A}} + b_{\mathrm{A}}}{D(P_{Y|X=x_{\mathrm{A}}} \| P_{Y|X=x_{\mathrm{R}}})}$$

$$+ (M-1)\exp\{-\gamma_1\}\frac{a_{\mathrm{R}} + b_{\mathrm{R}}}{D(P_{Y|X=x_{\mathrm{R}}} \| P_{Y|X=x_{\mathrm{A}}})}$$

- Upper bound on $\mathbb{P}\,[\text{Error at the end of HT}]$

Theorem 1

There exists an (N, M, ϵ)-VLF code with

$$N \leq (1-\epsilon_0)N' \quad \textit{and} \quad \epsilon \leq \epsilon_0 + (1-\epsilon_0)\epsilon'$$

where

$$\epsilon' = (M-1)\exp\{-(\gamma_1 + a_{\mathrm{A}})\} + (M-1)\exp\{-\gamma_2\}$$

$$N' = \frac{\gamma_1 + b}{C} + ((M-1)\exp\{-\gamma_1\} + \exp\{-a_{\mathrm{R}}\})\frac{\gamma_2 - \gamma_1 + b}{C} + \frac{a_{\mathrm{A}} + b_{\mathrm{A}}}{D(P_{Y|X=x_{\mathrm{A}}} \| P_{Y|X=x_{\mathrm{R}}})}$$

$$+ (M-1)\exp\{-\gamma_1\}\frac{a_{\mathrm{R}} + b_{\mathrm{R}}}{D(P_{Y|X=x_{\mathrm{R}}} \| P_{Y|X=x_{\mathrm{A}}})}$$

- Upper bound on $\mathbb{P}$[Error at the end of C2]

Theorem 1

There exists an (N, M, ϵ)*-VLF code with*

$$N \leq (1-\epsilon_0)N' \quad \textit{and} \quad \epsilon \leq \epsilon_0 + (1-\epsilon_0)\epsilon'$$

where

$$\epsilon' = (M-1)\exp\{-(\gamma_1 + a_{\mathrm{A}})\} + (M-1)\exp\{-\gamma_2\}$$

$$N' = \boxed{\frac{\gamma_1 + b}{C}} + ((M-1)\exp\{-\gamma_1\} + \exp\{-a_{\mathrm{R}}\})\frac{\gamma_2 - \gamma_1 + b}{C} + \frac{a_{\mathrm{A}} + b_{\mathrm{A}}}{D(P_{Y|X=x_{\mathrm{A}}} \| P_{Y|X=x_{\mathrm{R}}})}$$
$$+ (M-1)\exp\{-\gamma_1\}\frac{a_{\mathrm{R}} + b_{\mathrm{R}}}{D(P_{Y|X=x_{\mathrm{R}}} \| P_{Y|X=x_{\mathrm{A}}})}$$

- Upper bound on $\mathbb{E}\left[\tau^{(1)}\right]$

Nonasymptotic bound

Theorem 1

There exists an (N, M, ϵ)*-VLF code with*

$$N \leq (1 - \epsilon_0) N' \quad \textit{and} \quad \epsilon \leq \epsilon_0 + (1 - \epsilon_0)\epsilon'$$

where

$$\epsilon' = (M-1)\exp\{-(\gamma_1 + a_{\mathrm{A}})\} + (M-1)\exp\{-\gamma_2\}$$

$$N' = \frac{\gamma_1 + b}{C} + ((M-1)\exp\{-\gamma_1\} + \exp\{-a_{\mathrm{R}}\})\frac{\gamma_2 - \gamma_1 + b}{C} + \frac{a_{\mathrm{A}} + b_{\mathrm{A}}}{D(P_{Y|X=x_{\mathrm{A}}} \| P_{Y|X=x_{\mathrm{R}}})}$$

$$+ (M-1)\exp\{-\gamma_1\}\frac{a_{\mathrm{R}} + b_{\mathrm{R}}}{D(P_{Y|X=x_{\mathrm{R}}} \| P_{Y|X=x_{\mathrm{A}}})}$$

- Upper bound on $\mathbb{E}\left[\tau^{(2)} - \tau^{(1)}\right]$

Nonasymptotic bound

Theorem 1

There exists an (N, M, ϵ)*-VLF code with*

$$N \le (1-\epsilon_0)N' \quad \textit{and} \quad \epsilon \le \epsilon_0 + (1-\epsilon_0)\epsilon'$$

where

$$\epsilon' = (M-1)\exp\{-(\gamma_1 + a_{\mathrm{A}})\} + (M-1)\exp\{-\gamma_2\}$$

$$N' = \frac{\gamma_1 + b}{C} + ((M-1)\exp\{-\gamma_1\} + \exp\{-a_{\mathrm{R}}\})\frac{\gamma_2 - \gamma_1 + b}{C} + \frac{a_{\mathrm{A}} + b_{\mathrm{A}}}{D(P_{Y|X=x_{\mathrm{A}}} \| P_{Y|X=x_{\mathrm{R}}})}$$

$$+ (M-1)\exp\{-\gamma_1\}\frac{a_{\mathrm{R}} + b_{\mathrm{R}}}{D(P_{Y|X=x_{\mathrm{R}}} \| P_{Y|X=x_{\mathrm{A}}})}$$

- Upper bound on $\mathbb{E}\left[\tau^{\mathrm{HT}}\right]$

Theorem 2

Assume that $C > 0$ and $C_1 < \infty$. Then,

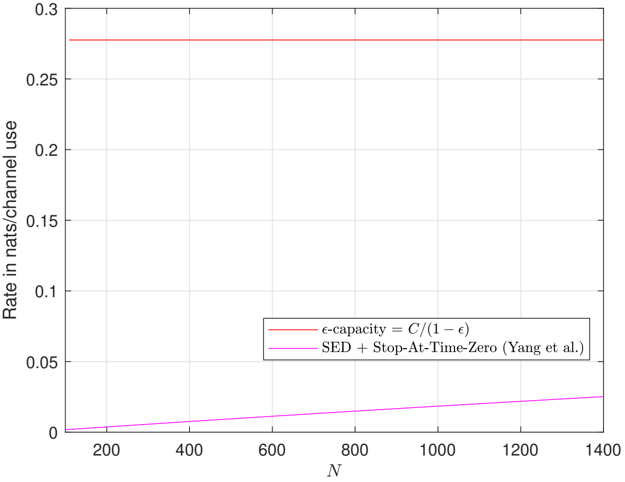
$$\log M^*(N,\epsilon) \geq \frac{NC}{1-\epsilon} - \frac{C}{C_1}\log N - \log\log N + O(1)$$

- It improves the best second-order term in the literature from $-\log N$ to $-\frac{C}{C_1}\log N$
- Upper bound:

$$\log M^*(N,\epsilon) \leq \frac{NC}{1-\epsilon} + \underbrace{\frac{-\epsilon\log(\epsilon) - (1-\epsilon)\log(1-\epsilon)}{1-\epsilon}}_{O(1)}$$

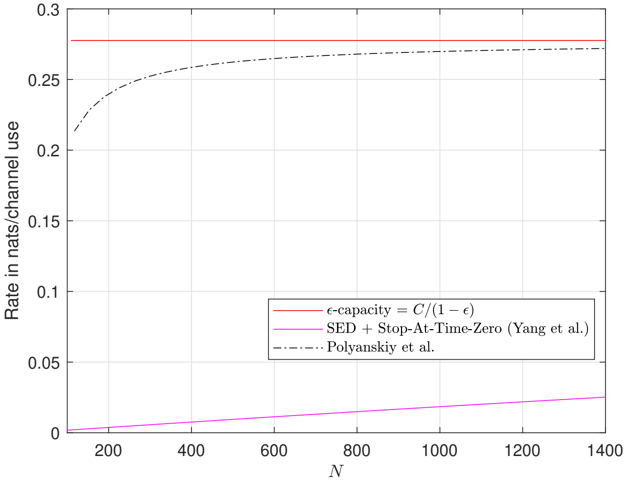
- The gap from the best upper bound is $O(\log N)$.

Channel: Cascade of BSC(0.11) and BEC(0.2)



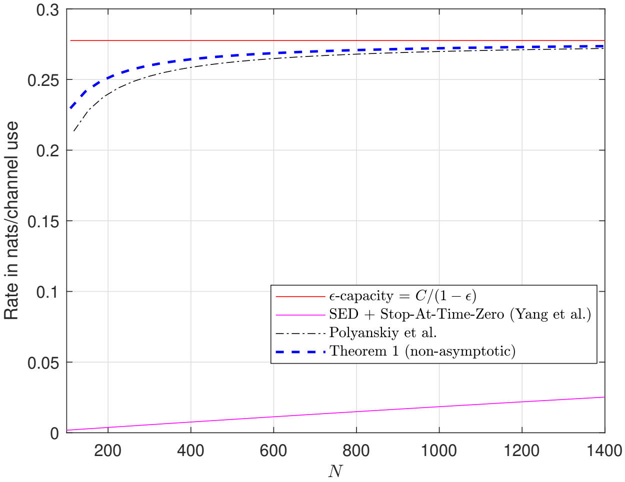
- The non-asymptotic bound for SED code is far from ϵ-capacity (it asymptotically achieves ϵ-capacity)

Channel: Cascade of BSC(0.11) and BEC(0.2)



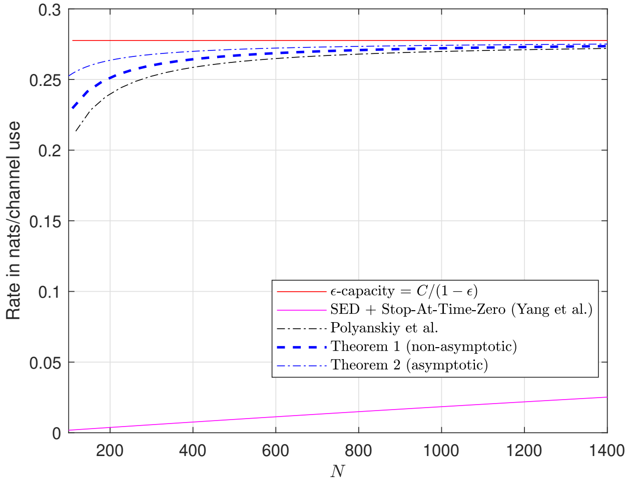
- Polyanskiy et al.'s bound outperforms SED code for this channel.

Channel: Cascade of BSC(0.11) and BEC(0.2)



- Our non-asymptotic bound outperforms Polyanskiy et al.'s and SED.

Channel: Cascade of BSC(0.11) and BEC(0.2)



- Our asymptotic bound is shown.

Universal VLF (UVLF) Codes

- We desire universality in code design because the channel statistics are not always available.
- The goal is to develop coding schemes whose performance is not so below from that of codes specific to a particular channel.
- An (N, M, ϵ)-UVLF code is such that

$$f_n \text{ and } g \text{ cannot depend on } P_{Y|X}$$

- $M^*_{\mathrm{U}}(N, \epsilon) = \{\max M : \exists\ (N, M, \epsilon)\text{-UVLF code}\}$

Our UVLF Code

- We universalize our previous code by employing a universal decoding metric.
- Empirical distribution:

$$\hat{P}_{x^n}(a) = \sum_{i=1}^{n} \frac{1}{n} 1\{x_i = a\}$$

- Instead of information density $\imath(X^n; Y^n)$, the decoder uses the **empirical mutual information** $nI(\hat{P}_{\mathbf{c}^n(m)}, \hat{P}_{Y^n|\mathbf{c}^n(m)})$
- Only a single communication phase is employed:

$$\tau^{(1)} = \inf\{n \geq 1 \colon \max_{m \in [M]} nI(\hat{P}_{\mathbf{c}^n(m)}, \hat{P}_{Y^n|\mathbf{c}^n(m)}) > \gamma_1\}$$

- This decoder keeps track of the empirical MI of each codeword and decodes as soon as one of them is larger than γ_1.
- We employ the stop-at-time-zero strategy similar to the known-channel case.

Main Result

Theorem 3

Assume that a capacity-achieving distribution for the channel is known. Assume that $C > 0$ and $C_1 < \infty$. Then,

$$\log M^*_{\mathrm{U}}(N,\epsilon) \geq \frac{NC}{1-\epsilon} - \log N - \min\left\{\frac{|\mathcal{X}||\mathcal{Y}|}{2}, \left(|\mathcal{X}| - \frac{3}{2}\right)\left(|\mathcal{Y}| - \frac{3}{2}\right) + \frac{3}{4}\right\}\log N + O(\log\log N).$$

If $P_{Y|X}$ is known to be a BSC with an unknown flip probability $p \in (0,1) \setminus \{\frac{1}{2}\}$,

$$\log M^*_{\mathrm{U}}(N,\epsilon) \geq \frac{NC}{1-\epsilon} - \frac{3}{2}\log N + o(\log\log N).$$

- When $P_{Y|X}$ is known, $-\log N$ is achieved by Polyanskiy et al.'s single-phase scheme.
- We characterize a penalty term of order $\log N$ due to not knowing the channel.

Lemma 1 (Woodroofe Th. 4.5)

Let $g\colon \mathbb{R}^k \to \mathbb{R}$ be a twice differentiable continuous function. Let $Y_1, Y_2, \dots \in \mathbb{R}^k$ be i.i.d. random vectors. Let $\mu = g(\mathbb{E}[Y_1]) > 0$. Let $\gamma > 0$. Define

$$Z_n = ng\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right), \quad \tau = \inf\{n \geq 1\colon Z_n > \gamma\}.$$

As $\gamma \to \infty$

$$\mathbb{E}[\tau] = \frac{1}{\mu}\left(\gamma + \underbrace{\rho - \frac{1}{2}\mathrm{tr}(\mathrm{Cov}(Y_1)\nabla^2 g(\mathbb{E}[Y_1]))}_{O(1)}\right) + o(1)$$

where $\rho = \frac{\mathbb{E}\left[S_{\tau_+}^2\right]}{2\mathbb{E}\left[S_{\tau_+}\right]}$ with $S_n = n\mu + \sum_{i=1}^{n} \nabla g(\mathbb{E}[Y_1])^\top (Y_i - \mathbb{E}[Y_1])$, and $\tau^+ = \inf\{n \geq 1\colon Z_n > 0\}$.

Lemma 1 (Woodroofe Th. 4.5)

Let $g\colon \mathbb{R}^k \to \mathbb{R}$ *be a twice differentiable continuous function. Let* $Y_1, Y_2, \dots \in \mathbb{R}^k$ *be i.i.d. random vectors. Let* $\mu = g(\mathbb{E}[Y_1]) > 0$. *Let* $\gamma > 0$. *Define*

$$Z_n = ng\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right), \quad \tau = \inf\{n \geq 1\colon Z_n > \gamma\}.$$

As $\gamma \to \infty$

$$\mathbb{E}[\tau] = \frac{1}{\mu}\left(\gamma + \underbrace{\rho - \frac{1}{2}\mathrm{tr}(\mathrm{Cov}(Y_1)\nabla^2 g(\mathbb{E}[Y_1]))}_{O(1)}\right) + o(1)$$

- **Application:** We choose g to be the mutual information function. $Z_n = nI(\hat{P}_{X^n}, \hat{P}_{Y^n|X^n})$ and $S_n = \imath(X^n; Y^n)$ evaluated at P_{XY}. We get $\mathbb{E}[\tau] = \frac{\gamma}{C} + O(1)$ as in the known-channel case.

A Large Deviations Bound

Lemma 2

Let $(\bar{X}^n, Y^n) \sim P_X^n P_Y^n$ *and let* $\gamma > 0$*. Assume that* $P_X(x) > 0$ *and* $P_Y(y) > 0$ *for all* (x, y)*. Then,*

$$\mathbb{P}\left[nI(\hat{P}_{\bar{X}^n}, P_{Y^n|\bar{X}^n}) \geq \gamma\right] \leq K_1 (n+1)^d \exp\{-\gamma\}$$
$$d = \min\left\{\frac{|\mathcal{X}||\mathcal{Y}| - 2}{2}, \left(|\mathcal{X}| - \frac{3}{2}\right)\left(|\mathcal{Y}| - \frac{3}{2}\right) - \frac{1}{4}\right\}$$

where K_1 *is a positive constant depending only on* $|\mathcal{X}|$ *and* $|\mathcal{Y}|$*. By the union bound,*

$$\mathbb{P}\left[\exists\, n \leq n_0 \colon nI(\hat{P}_{\bar{X}^n}, P_{Y^n|\bar{X}^n}) \geq \gamma\right] \leq K_1 (n_0+1)^{d+1} \exp\{-\gamma\}$$

- For comparison: $\mathbb{P}\left[\exists\, n \in \mathbb{N} \colon \imath(\bar{X}^n; Y^n) \geq \gamma\right] \leq \exp\{-\gamma\}$
- Method of types: $\mathbb{P}\left[nI(\hat{P}_{\bar{X}^n}, P_{Y^n|\bar{X}^n}) \geq \gamma\right] \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|-1} \exp\{-\gamma\}$

Conclusion

- We develop a novel VLF code and analyze its performance, which improves the best achievability bound known.
- We universalize our code by employing empirical mutual information and characterize the back-off from the known-channel scenario.
- We extend our results to the Gaussian channel, where the universal metric becomes the mutual information assuming that the input-output pair is jointly Gaussian.

Thank you!