Variable-length Feedback Codes with Several Decoding Times for the Gaussian Channel

Recep Can Yavas

California Institute of Technology

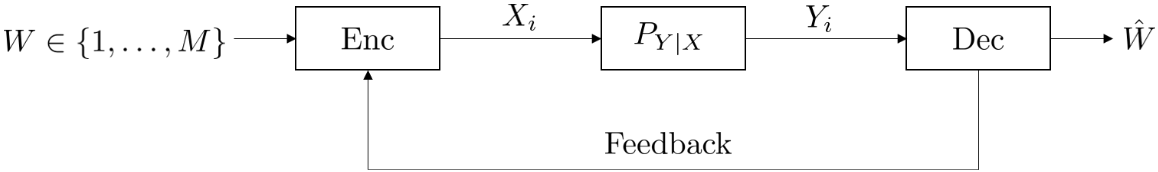
July 12-20, 2021

Joint work with Victoria Kostina and Michelle Effros

ISIT 2021

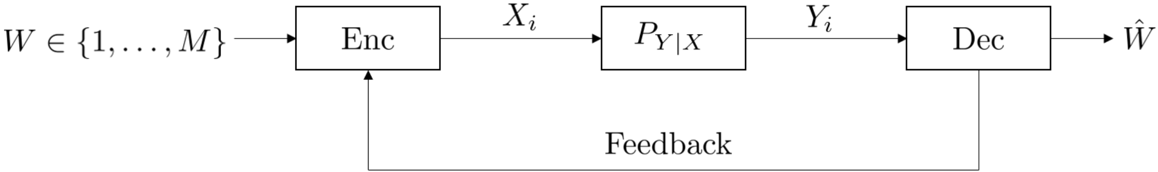
This work was supported in part by the National Science Foundation (NSF) under grant CCF-1817241 and CCF-1956386.

Variable-length stop-feedback (VLSF) codes



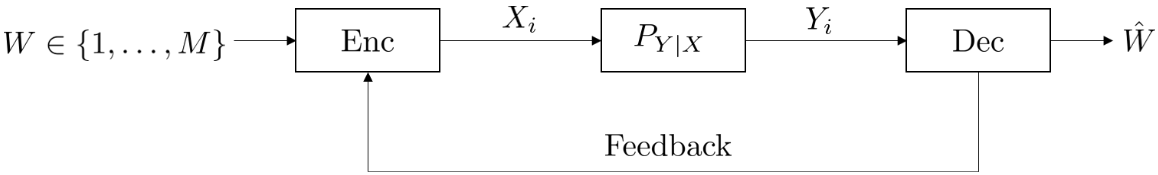
- Variable-length: Decoding at arbitrary times $\tau \in \{0, 1, 2, \ldots\}$

Variable-length stop-feedback (VLSF) codes



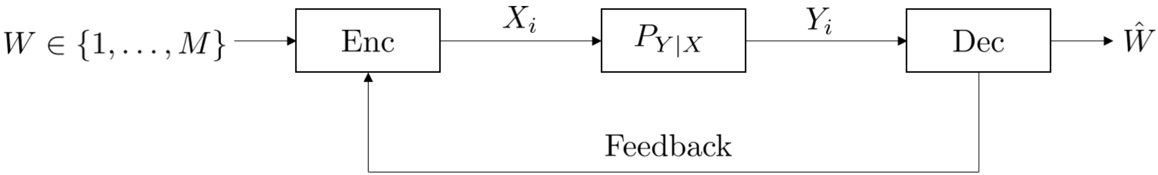
- Variable-length: Decoding at arbitrary times $\tau \in \{0, 1, 2, \ldots\}$
- Stop-feedback $\in \{0 = \text{continue}, 1 = \text{stop}\}$

Variable-length stop-feedback (VLSF) codes



- Variable-length: Decoding at arbitrary times $\tau \in \{0, 1, 2, \ldots\}$
- Stop-feedback $\in \{0 = \text{continue}, 1 = \text{stop}\}$
- Earlier decoding when the noise is low, later decoding when the noise is high $\implies$ High reliability

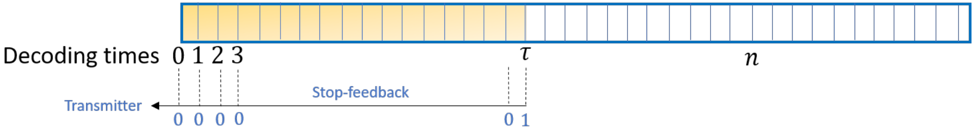
Variable-length stop-feedback (VLSF) codes



- Variable-length: Decoding at arbitrary times $\tau \in \{0, 1, 2, \ldots\}$
- Stop-feedback $\in \{0 = \text{continue}, 1 = \text{stop}\}$
- Earlier decoding when the noise is low, later decoding when the noise is high
  $\Longrightarrow$ High reliability
- Feedback at each time instant is impractical!

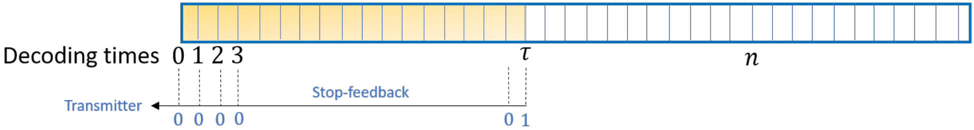
VLSF codes with $K = \infty$ vs. $K < \infty$

- VLSF code ($K = \infty$): Impractical!



VLSF codes with $K = \infty$ vs. $K < \infty$

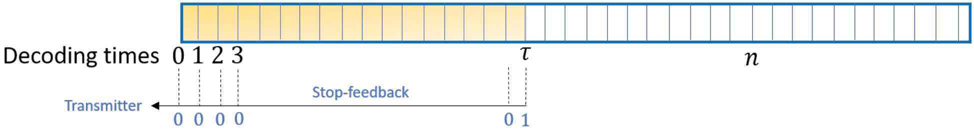
- VLSF code ($K = \infty$): Impractical!



  - Transmitter constantly listens to the feedback signal

VLSF codes with $K = \infty$ vs. $K < \infty$

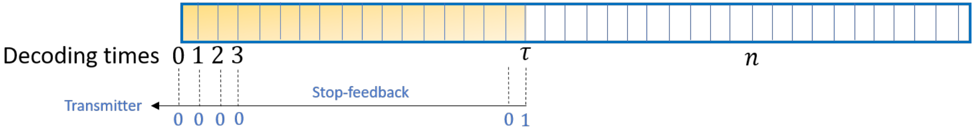
- VLSF code ($K = \infty$): Impractical!



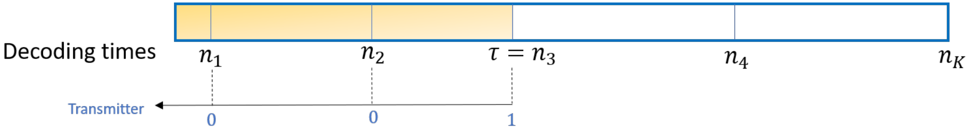
  - Transmitter constantly listens to the feedback signal
  - High feedback rate: 1 bit/channel use

VLSF codes with $K = \infty$ vs. $K < \infty$

- VLSF code ($K = \infty$): Impractical!

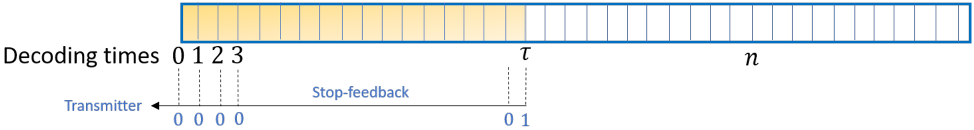


  - Transmitter constantly listens to the feedback signal
  - High feedback rate: 1 bit/channel use
- VLSF code with $K < \infty$ decoding times



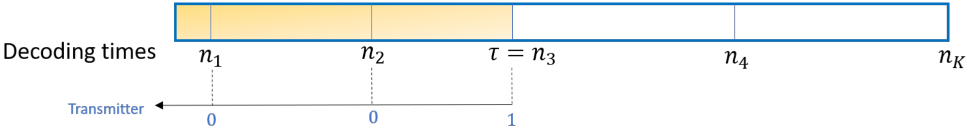
VLSF codes with $K = \infty$ vs. $K < \infty$

- VLSF code ($K = \infty$): Impractical!



  - Transmitter constantly listens to the feedback signal
  - High feedback rate: 1 bit/channel use

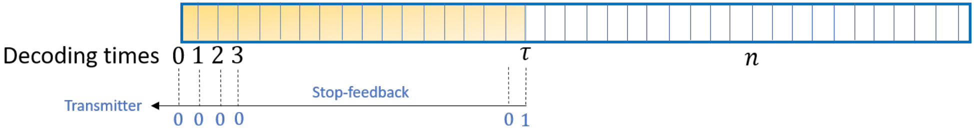
- VLSF code with $K < \infty$ decoding times



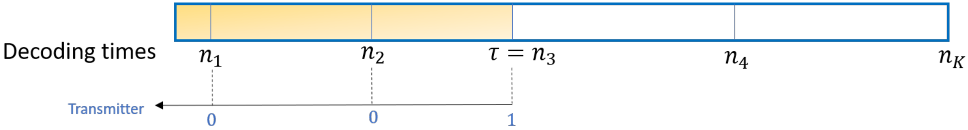
  - Sporadic feedback

VLSF codes with $K = \infty$ vs. $K < \infty$

- VLSF code ($K = \infty$): Impractical!



  - Transmitter constantly listens to the feedback signal
  - High feedback rate: 1 bit/channel use
- VLSF code with $K < \infty$ decoding times



  - Sporadic feedback
  - Practical codes: Incremental redundancy hybrid automatic repeat request codes

Prior work: $K = \infty$

- [Burnashev (1976)]: error exponent $\frac{-\ln P_e}{\mathbb{E}[\tau]}$ for DMCs
- [Polyanskiy et al. (2011)]: VLSF codes for DMCs under non-vanishing error value ϵ

$$\ln M^*(N, 1, \epsilon) = NC - \sqrt{NV} Q^{-1}(\epsilon) + O(\ln N)$$
$$\frac{NC}{1-\epsilon} - \ln N + O(1) \leq \ln M^*(N, \infty, \epsilon) \leq \frac{NC}{1-\epsilon} + O(1)$$

where C = capacity, V = dispersion, $M^*(N, K, \epsilon)$ = maximum achievable message size compatible with average decoding time N, average error probability ϵ and K decoding times

Prior work: $K < \infty$

- [Vakilinia et al. (2016)]: VLSF codes with K decoding times for LDPC codes over the binary-input Gaussian channel

Prior work: $K < \infty$

- [Vakilinia et al. (2016)]: VLSF codes with K decoding times for LDPC codes over the binary-input Gaussian channel
- To minimize N, they estimate mean and variance of τ through simulation and use Gaussian approximation for τ

$$\mathbb{P}\left[\tau \leq n\right] \approx Q\left(\frac{\mathbb{E}\left[\tau\right] - n}{\sqrt{\mathrm{Var}\left[\tau\right]}}\right)$$

Prior work: $K < \infty$

- [Vakilinia et al. (2016)]: VLSF codes with K decoding times for LDPC codes over the binary-input Gaussian channel
- To minimize N, they estimate mean and variance of τ through simulation and use Gaussian approximation for τ

$$\mathbb{P}\left[\tau \leq n\right] \approx Q\left(\frac{\mathbb{E}\left[\tau\right] - n}{\sqrt{\mathrm{Var}\left[\tau\right]}}\right)$$

- Given a fixed n_1, the decoding times $n_2, \ldots, n_K$ are computed using a recursive formula, which results from differential equations $\frac{\partial N}{\partial n_i} = 0$

Prior work: $K < \infty$

- [Vakilinia et al. (2016)]: VLSF codes with K decoding times for LDPC codes over the binary-input Gaussian channel
- To minimize N, they estimate mean and variance of τ through simulation and use Gaussian approximation for τ

$$\mathbb{P}[\tau \le n] \approx Q\left(\frac{\mathbb{E}[\tau] - n}{\sqrt{\mathrm{Var}[\tau]}}\right)$$

- Given a fixed n_1, the decoding times $n_2, \ldots, n_K$ are computed using a recursive formula, which results from differential equations $\frac{\partial N}{\partial n_i} = 0$
- Then they search for the optimal $n_1 \implies$ sequential differential optimization

Prior work: $K < \infty$

- [Vakilinia et al. (2016)]: VLSF codes with K decoding times for LDPC codes over the binary-input Gaussian channel
- To minimize N, they estimate mean and variance of τ through simulation and use Gaussian approximation for τ

$$\mathbb{P}\left[\tau \leq n\right] \approx Q\left(\frac{\mathbb{E}\left[\tau\right] - n}{\sqrt{\mathrm{Var}\left[\tau\right]}}\right)$$

- Given a fixed n_1, the decoding times $n_2, \ldots, n_K$ are computed using a recursive formula, which results from differential equations $\frac{\partial N}{\partial n_i} = 0$
- Then they search for the optimal $n_1 \implies$ sequential differential optimization
- They do not solve the problem analytically $\implies$ no second-order analysis

Channel model

- Memoryless Gaussian channel: the channel output at time i is

$$
\begin{aligned}
Y_i &= X_i + Z_i \\
Z_i &\sim \mathcal{N}(0,1),
\end{aligned}
$$

where Z_i's are i.i.d. and X_i and Z_i are independent.

Code definition

Definition

An $(N, \{n_i\}_{i=1}^{K}, M, \epsilon, P)$ VLSF code comprises

1. encoding functions $\mathsf{f}_n \colon [M] \to \mathbb{R}$, $n = 1, \ldots, n_K$:

$$\mathsf{f}(m)^{n_K} \triangleq (\mathsf{f}_1(m), \ldots, \mathsf{f}_{n_K}(m))$$

Code definition

Definition

An $(N, \{n_i\}_{i=1}^{K}, M, \epsilon, P)$ VLSF code comprises

1. encoding functions $\mathsf{f}_n \colon [M] \to \mathbb{R}$, $n = 1, \ldots, n_K$:

$$\mathsf{f}(m)^{n_K} \triangleq (\mathsf{f}_1(m), \ldots, \mathsf{f}_{n_K}(m))$$

2. a random stopping time $\tau \in \{n_1, \ldots, n_K\}$

Code definition

Definition

An $(N, \{n_i\}_{i=1}^{K}, M, \epsilon, P)$ VLSF code comprises

1. encoding functions $\mathsf{f}_n \colon [M] \to \mathbb{R}$, $n = 1, \ldots, n_K$:

$$\mathsf{f}(m)^{n_K} \triangleq (\mathsf{f}_1(m), \ldots, \mathsf{f}_{n_K}(m))$$

2. a random stopping time $\tau \in \{n_1, \ldots, n_K\}$
3. K decoding functions $\mathsf{g}_{n_k} : \mathbb{R}^{n_k} \to [M]$ for $k \in [K]$

Code definition

Definition

An $(N, \{n_i\}_{i=1}^K, M, \epsilon, P)$ VLSF code comprises

1. encoding functions $\mathsf{f}_n \colon [M] \to \mathbb{R}$, $n = 1, \ldots, n_K$:
$$\mathsf{f}(m)^{n_K} \triangleq (\mathsf{f}_1(m), \ldots, \mathsf{f}_{n_K}(m))$$
2. a random stopping time $\tau \in \{n_1, \ldots, n_K\}$
3. K decoding functions $\mathsf{g}_{n_k} : \mathbb{R}^{n_k} \to [M]$ for $k \in [K]$
4. a common randomness between the transmitter and the receiver

Code definition

Definition

An $(N, \{n_i\}_{i=1}^K, M, \epsilon, P)$ VLSF code comprises

1. encoding functions $\mathsf{f}_n \colon [M] \to \mathbb{R}$, $n = 1, \ldots, n_K$:
$$\mathsf{f}(m)^{n_K} \triangleq (\mathsf{f}_1(m), \ldots, \mathsf{f}_{n_K}(m))$$
2. a random stopping time $\tau \in \{n_1, \ldots, n_K\}$
3. K decoding functions $\mathsf{g}_{n_k} : \mathbb{R}^{n_k} \to [M]$ for $k \in [K]$
4. a common randomness between the transmitter and the receiver

such that

$$\begin{aligned} \text{Maximal power constraint:} \quad & \|\mathsf{f}(m)^{n_k}\|^2 \leq n_k P \quad \forall\, m \in [M], k \in [K] \\ \text{Average decoding time:} \quad & \mathbb{E}[\tau] \leq N \\ \text{Average error probability:} \quad & \mathbb{P}[\mathsf{g}_\tau(Y^\tau) \neq W] \leq \epsilon \end{aligned}$$

where the message W is uniformly distributed on the set $[M]$.

Main Result

Theorem (Achievability)

Fix $K \geq 2$, $P > 0$ and $\epsilon \in (0, 1)$. For the Gaussian channel

$$\ln M^*(N, K, \epsilon, P) \geq \frac{NC(P)}{1-\epsilon} - \sqrt{N \ln_{(K-1)}(N) \frac{V(P)}{1-\epsilon}} + o(\sqrt{N})$$

The decoding times satisfy $n_1 = 0$ and the equations

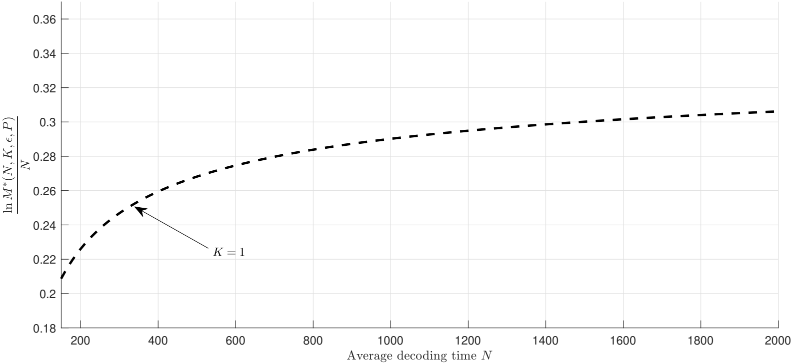
$$\ln M^*(N, K, \epsilon, P) = n_k C(P) - \sqrt{n_k \ln_{(K-k+1)}(n_k) V(P)} - \ln n_k + O(1)$$

for $k \in \{2, \ldots, K\}$.

$C(P) = \frac{1}{2}\ln(1+P) = \text{capacity}$, $V(P) = \frac{P(P+2)}{2(1+P)^2} = \text{dispersion}$

$\ln_{(K)}(\cdot) \triangleq \underbrace{\ln(\ln(\ldots(\ln(\cdot))))}_{K \text{ times}}$

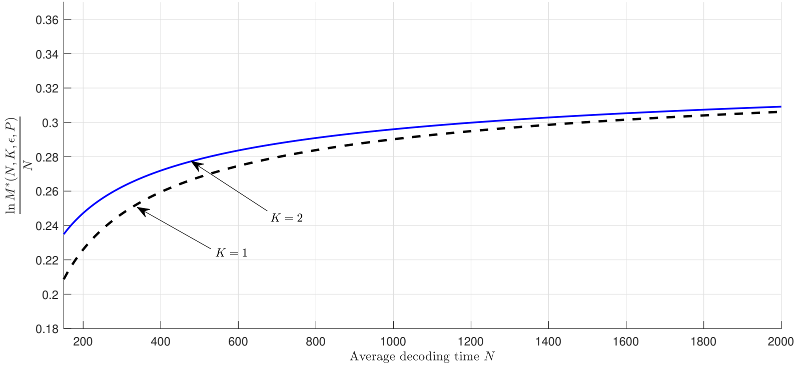
- **Bottom-line:** We derive an achievability bound and optimize the choices of the decoding times $n_1, \ldots, n_K$ to minimize average decoding time N for the given ϵ and M.

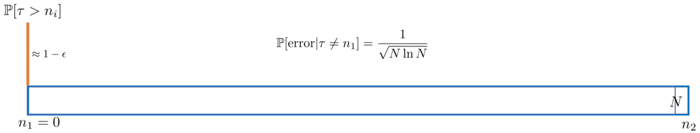


0.36
0.34
0.32
0.3
0.28
0.26
0.24
0.22
0.2
0.18
200
400
600
800
1000
1200
1400
1600
1800
2000
$\frac{\ln M^*(N, K, \epsilon, P)}{N}$
Average decoding time N
$K = 1$

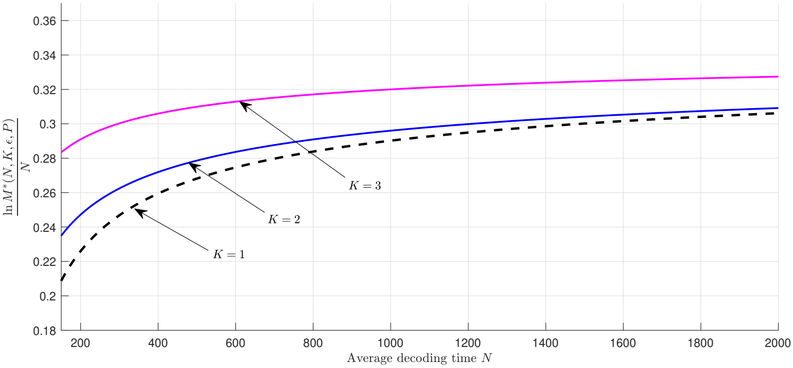
$n_1 = N$

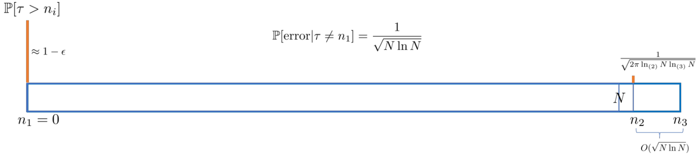
Example: $\epsilon = 10^{-3}$, $P = 1$



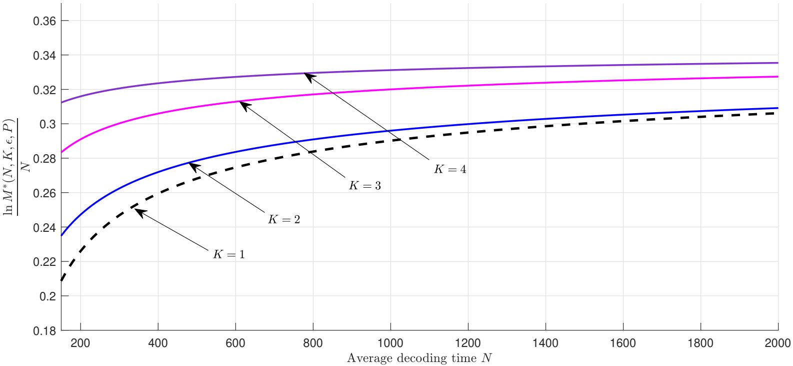


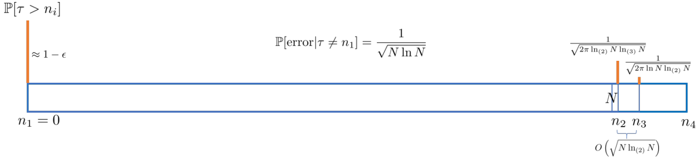
Example: $\epsilon = 10^{-3}$, $P = 1$



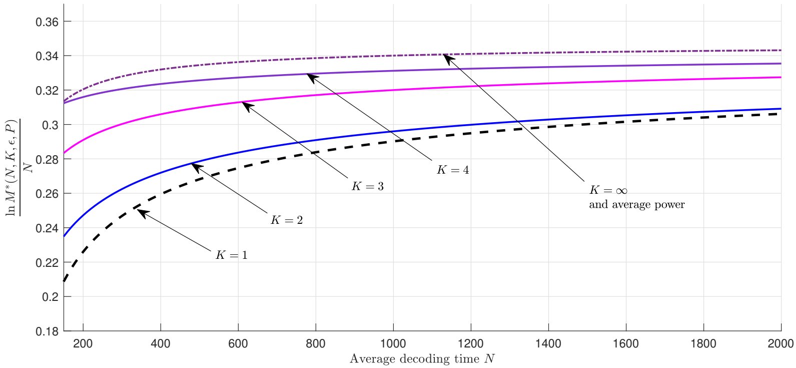


Example: $\epsilon = 10^{-3}$, $P = 1$



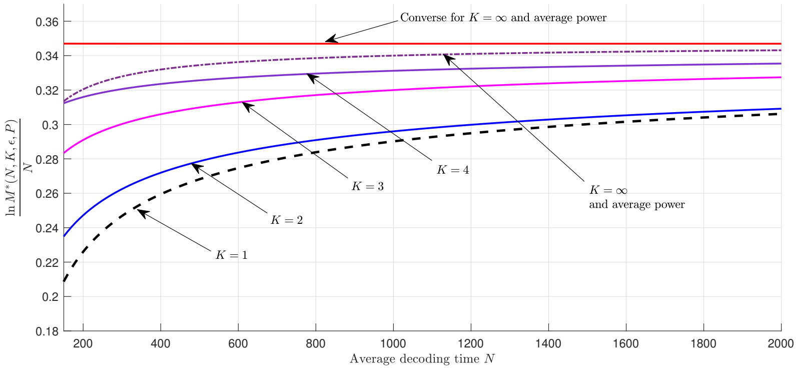


Example: $\epsilon = 10^{-3}$, $P = 1$



- Diminishing performance improvement as K increases!

Example: $\epsilon = 10^{-3}$, $P = 1$



- Diminishing performance improvement as K increases!

Comparison with prior work in extreme scenarios

- $2 \leq K < \infty$, maximal power:

$$\ln M^*(N, K, \epsilon, P) \geq \frac{NC(P)}{1-\epsilon} - \sqrt{N \ln_{(K-1)}(N) \frac{V(P)}{1-\epsilon}} + o(\sqrt{N})$$

Comparison with prior work in extreme scenarios

- $2 \leq K < \infty$, maximal power:

$$\ln M^* (N, K, \epsilon, P) \geq \frac{NC(P)}{1-\epsilon} - \sqrt{N \ln_{(K-1)}(N) \frac{V(P)}{1-\epsilon}} + o(\sqrt{N})$$

- $K = 1$, maximal power:
  [Polyanskiy et al. (2010) and Tan-Tomamichel (2015)]

$$\ln M^* (N, 1, \epsilon, P) = NC(P) - \sqrt{NV(P)} Q^{-1}(\epsilon) + \frac{1}{2} \ln N + O(1)$$

Comparison with prior work in extreme scenarios

- $2 \leq K < \infty$, maximal power:

$$\ln M^*(N, K, \epsilon, P) \geq \frac{NC(P)}{1-\epsilon} - \sqrt{N \ln_{(K-1)}(N) \frac{V(P)}{1-\epsilon}} + o(\sqrt{N})$$

- $K = 1$, maximal power:
  [Polyanskiy et al. (2010) and Tan-Tomamichel (2015)]

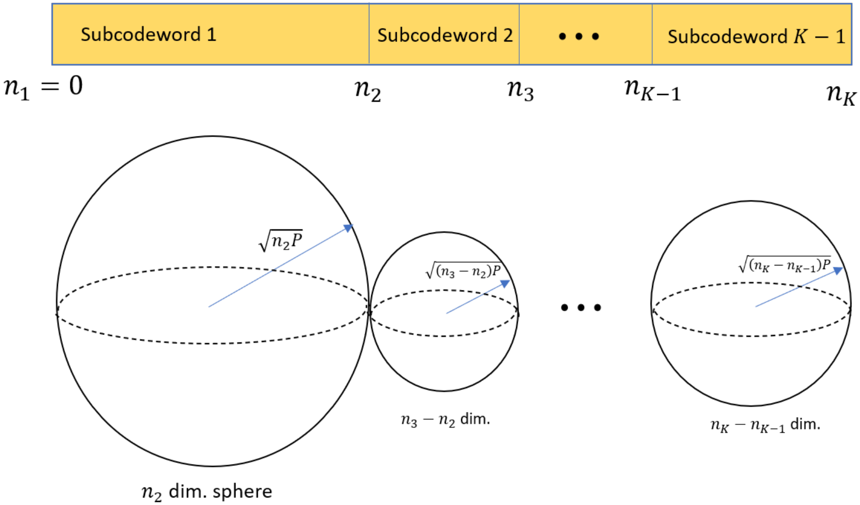
$$\ln M^*(N, 1, \epsilon, P) = NC(P) - \sqrt{NV(P)} Q^{-1}(\epsilon) + \frac{1}{2} \ln N + O(1)$$

- $K = \infty$, average power:
  [Truong-Tan (2018)]

$$\ln M^*(N, \infty, \epsilon, P)_{\mathrm{ave}} \geq \frac{NC(P)}{1-\epsilon} - \ln N + O(1)$$
$$\ln M^*(N, \infty, \epsilon, P)_{\mathrm{ave}} \leq \frac{NC(P)}{1-\epsilon} + \frac{h_b(\epsilon)}{1-\epsilon}$$

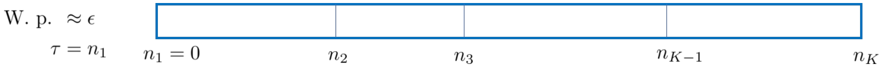
Random encoder design

- We generate M i.i.d. codewords of length-n_K so that maximal power constraint is satisfied with equality for each n_k, and the subcodewords are drawn independent of each other.



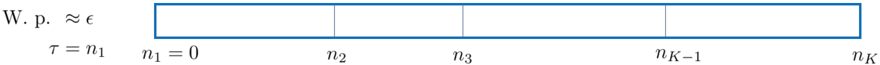
Decoder design: Polyanskiy et al. (2011)

- With probability $\approx \epsilon$, do not transmit any symbols and decode to an arbitrary message at time $n_1 = 0$.



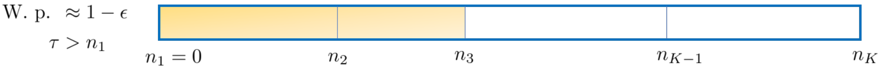
Decoder design: Polyanskiy et al. (2011)

- With probability $\approx \epsilon$, do not transmit any symbols and decode to an arbitrary message at time $n_1 = 0$.



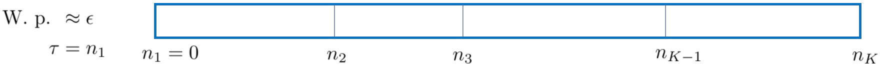
- With probability $\approx 1 - \epsilon$, transmit symbols and use
  Threshold decoder: Decode at the first time $n_k \in \{n_2, \ldots, n_K\}$ s.t. $\imath(\mathsf{f}(m)^{n_k}; Y^{n_k}) \geq \gamma$ for some m.

$$\overbrace{\imath(x^n; y^n)}^{\text{information density}} \triangleq \ln \frac{P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n}(y^n)}$$



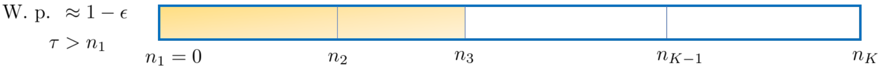
Decoder design: Polyanskiy et al. (2011)

- With probability $\approx \epsilon$, do not transmit any symbols and decode to an arbitrary message at time $n_1 = 0$.



- With probability $\approx 1 - \epsilon$, transmit symbols and use
  Threshold decoder: Decode at the first time $n_k \in \{n_2, \ldots, n_K\}$ s.t. $\imath(\mathsf{f}(m)^{n_k}; Y^{n_k}) \geq \gamma$ for some m.

$$\overbrace{\imath(x^n; y^n)}^{\text{information density}} \triangleq \ln \frac{P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n}(y^n)}$$



- Goal: optimize $n_2, \ldots, n_K$.

Optimizing decoding times $n_2, \ldots, n_K$ to minimize N

- (N', ϵ_N): average decoding time and error probability given $\tau > n_1$

$$
\begin{aligned}
\min \quad & N(n_2, \ldots, n_K, \gamma) = \frac{N'(1-\epsilon)}{1-\epsilon_N} \\
\text{s.t.} \quad & N' = n_2 + \sum_{i=2}^{K-1} (n_{i+1} - n_i) \mathbb{P}\left[\tau > n_i\right] \\
& \epsilon_N = \mathbb{P}\left[\imath(X^{n_K}; Y^{n_K}) < \gamma\right] + M \exp\{-\gamma\}
\end{aligned}
$$

Optimizing decoding times $n_2, \ldots, n_K$ to minimize N

- (N', ϵ_N): average decoding time and error probability given $\tau > n_1$

$$\min \quad N(n_2, \ldots, n_K, \gamma) = \frac{N'(1-\epsilon)}{1-\epsilon_N}$$

$$\text{s.t.} \quad N' = n_2 + \sum_{i=2}^{K-1} (n_{i+1} - n_i) Q\left(\frac{n_i C(P) - \gamma}{\sqrt{n_i V(P)}}\right)(1 + o(1))$$

$$\epsilon_N = Q\left(\frac{n_K C(P) - \gamma}{\sqrt{n_K V(P)}}\right)(1 + o(1)) + M \exp\{-\gamma\}$$

Using moderate deviations theorem

Optimizing decoding times $n_2, \ldots, n_K$ to minimize N

- (N', ϵ_N): average decoding time and error probability given $\tau > n_1$

$$
\begin{aligned}
\min \quad & N(n_2, \ldots, n_K, \gamma) = \frac{N'(1-\epsilon)}{1-\epsilon_N} \\
\text{s.t.} \quad & N' = n_2 + \sum_{i=2}^{K-1} (n_{i+1} - n_i) Q\left(\frac{n_i C(P) - \gamma}{\sqrt{n_i V(P)}}\right)(1 + o(1)) \\
& \epsilon_N = Q\left(\frac{n_K C(P) - \gamma}{\sqrt{n_K V(P)}}\right)(1 + o(1)) + M \exp\{-\gamma\}
\end{aligned}
$$

- Find the optimal $(n_2^*, \ldots, n_K^*, \gamma^*)$ by solving $\nabla N = 0$.

Optimizing decoding times $n_2, \ldots, n_K$ to minimize N

- (N', ϵ_N): average decoding time and error probability given $\tau > n_1$

$$\begin{aligned} \min \quad & N(n_2, \ldots, n_K, \gamma) = \frac{N'(1-\epsilon)}{1-\epsilon_N} \\ \text{s.t.} \quad & N' = n_2 + \sum_{i=2}^{K-1} (n_{i+1} - n_i) Q\left(\frac{n_i C(P) - \gamma}{\sqrt{n_i V(P)}}\right)(1 + o(1)) \\ & \epsilon_N = Q\left(\frac{n_K C(P) - \gamma}{\sqrt{n_K V(P)}}\right)(1 + o(1)) + M \exp\{-\gamma\} \end{aligned}$$

- Find the optimal $(n_2^*, \ldots, n_K^*, \gamma^*)$ by solving $\nabla N = 0$.
- The optimal $\epsilon_N^* = \frac{1}{\sqrt{N \ln N}}$.

Conclusion

- We derived a second-order achievability bound for VLSF codes with K decoding times and optimized the values of K decoding times.

Conclusion

- We derived a second-order achievability bound for VLSF codes with K decoding times and optimized the values of K decoding times.
- Our theorem suggests that using $K > 4$ decoding times is not very beneficial.

Conclusion

- We derived a second-order achievability bound for VLSF codes with K decoding times and optimized the values of K decoding times.
- Our theorem suggests that using $K > 4$ decoding times is not very beneficial.
- We drew independent subcodewords, each drawn uniformly on a power sphere.

Future work

- Improve the converse result for $K < \infty$ and maximal power constraint.

Future work

- Improve the converse result for $K < \infty$ and maximal power constraint.
- Investigate maximal power constraint vs. average power constraint for VLSF codes with $K = \infty$.

$$\frac{NC(P)}{1-\epsilon} - \ln N + O(1) \leq \ln M^*_{\mathrm{ave}}(N, \infty, \epsilon, P) \leq \frac{NC(P)}{1-\epsilon} + O(1)$$

We show for the maximal power constraint:

$$\ln M^*(N, \infty, \epsilon, P) \geq \frac{NC(P)}{1-\epsilon} - O(\sqrt{N})$$

References

1. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," Problems of Information Transmission, vol. 12, no. 4, pp. 10–30, 1976.
2. L. V. Truong and V. Y. F. Tan, "On Gaussian macs with variable-length feedback and non-vanishing error probabilities," IEEE Trans. on Inf. Theory., vol. 64, no. 4, p. 2333–2346, Apr. 2018.
3. Y. Polyanskiy, H. V. Poor, and S. Verdu,"Feedback in the non-asymptotic regime," IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 4903–4925, Aug. 2011.
4. Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.
5. V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the AWGN channel," IEEE Transactions on Information Theory, vol. 61, no. 5, pp. 2430–2438, May 2015.
6. K. Vakilinia, S. V. S. Ranganathan, D. Divsalar, and R. D. Wesel, "Optimizing transmission lengths for limited feedback with non-binary LDPC examples", IEEE Transactions on Communications, vol. 64, no. 6,pp. 2245–2257, 2016.
7. R. C. Yavas, V. Kostina, and M. Effros, "Variable-length Feedback Codes with Several Decoding Times for the Gaussian Channel," ISIT, 2021. Available at: `https://arXiv:2103.09373`.