Communicating Short Messages with and without feedback

Richard Wesel
2026 IEEE North American School of Information Theory
Provo, Utah
June 22, 2026

These slides are provided in the original PowerPoint for review and further study by attendees at the 2026 IEEE North American School of Information Theory.

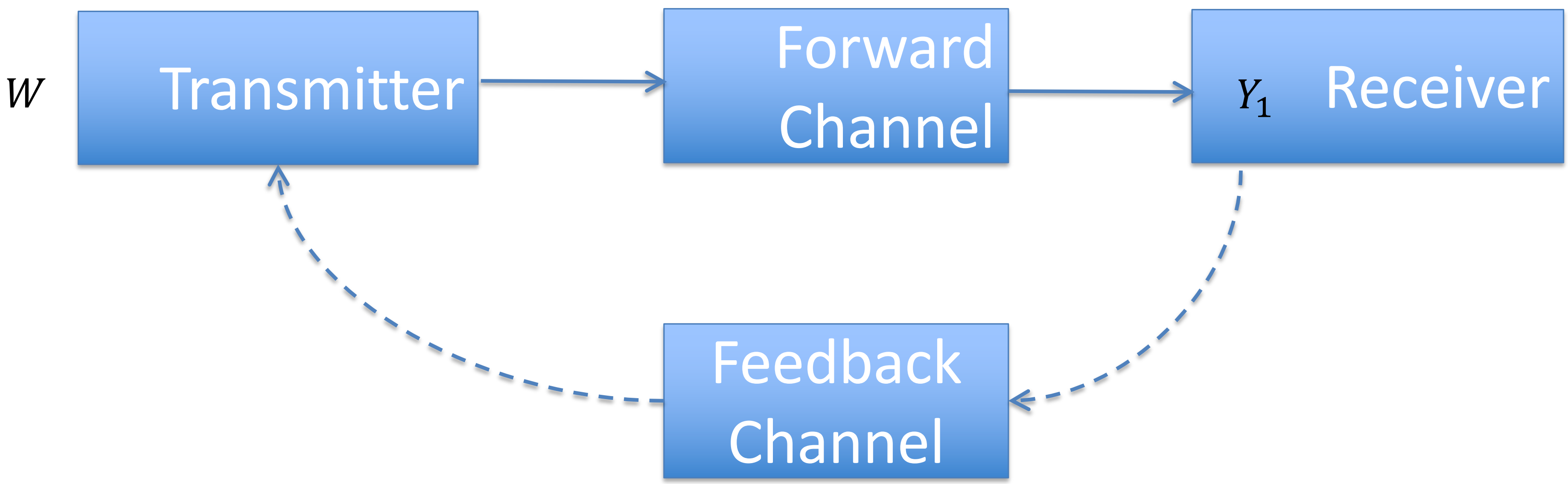
Engineer Change.

Communicating Short Messages with Feedback

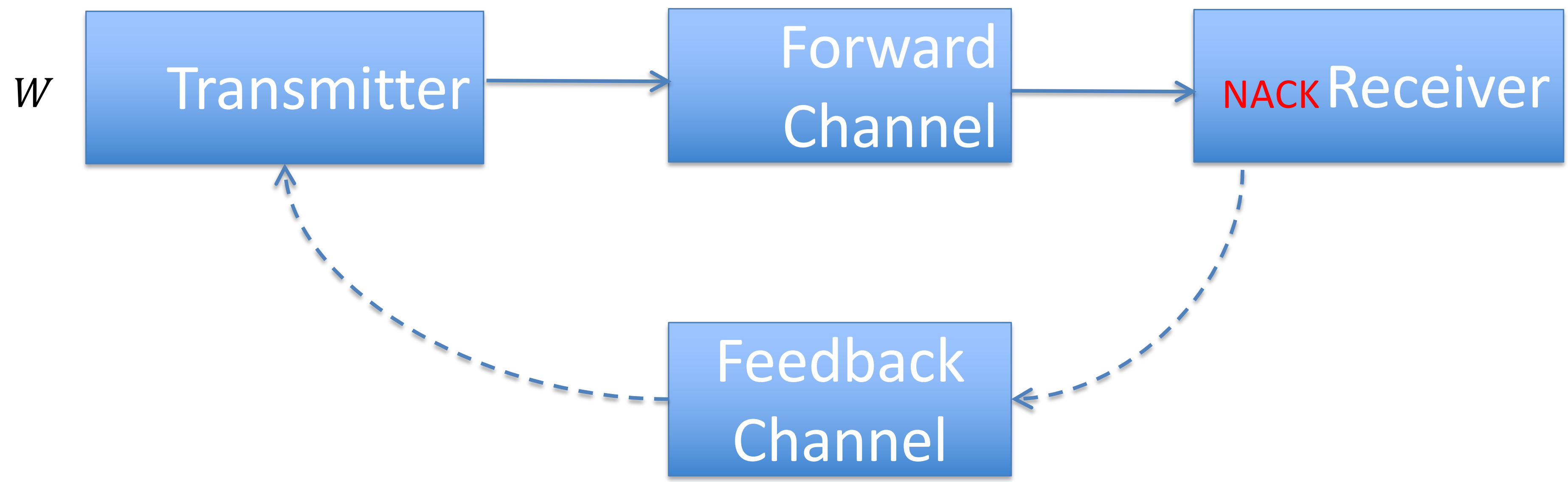
- Feedback does not increase capacity, *but*
- For short messages, it *can* increase the rate
- Two main ways to use feedback
  1. Stop Feedback - that only decides when to stop transmission
  2. Full Feedback - that helps determine each transmitted symbol

UCLA Samueli School of Engineering

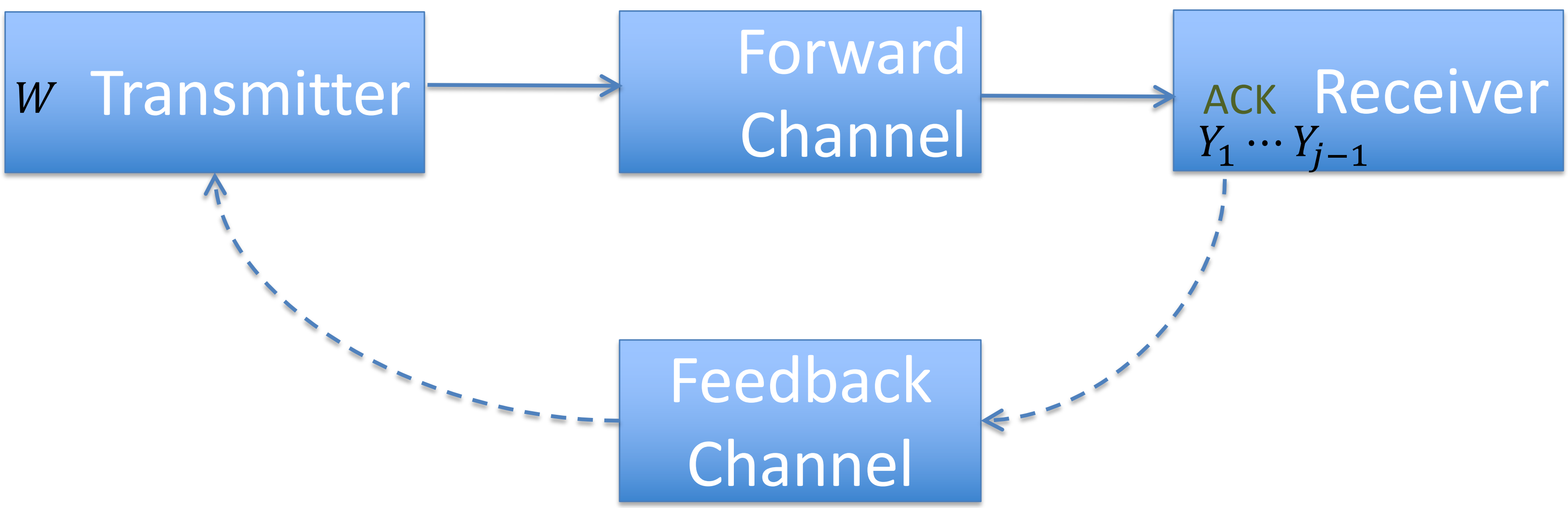
Full Feedback System Model



Stop Feedback System Model



Stop Feedback System Model



Feedback Timeline

Shannon 1956

Even Full Feedback does not increase capacity

Horstein 1963

Brilliant algorithm using full feedback on the binary symmetric channel

Burnashev 1976

Analytical proof that feedback improves error exponent

Shayevitz & Feder 2011

Posterior Matching – generalizes Horstein for any DMC

Polyanskiy, Poor, Verdu 2011

Achievability bounds for stop-Feedback for finite block-lengths

Vakilinia, Williamson, Ranganathan, Wang, Yang, Wesel

Stop feedback for practical communication

Yang, Antonini, Gimelshein, Wesel 2020-2024

Improved bounds and low complexity posterior matching for full feedback on the BSC

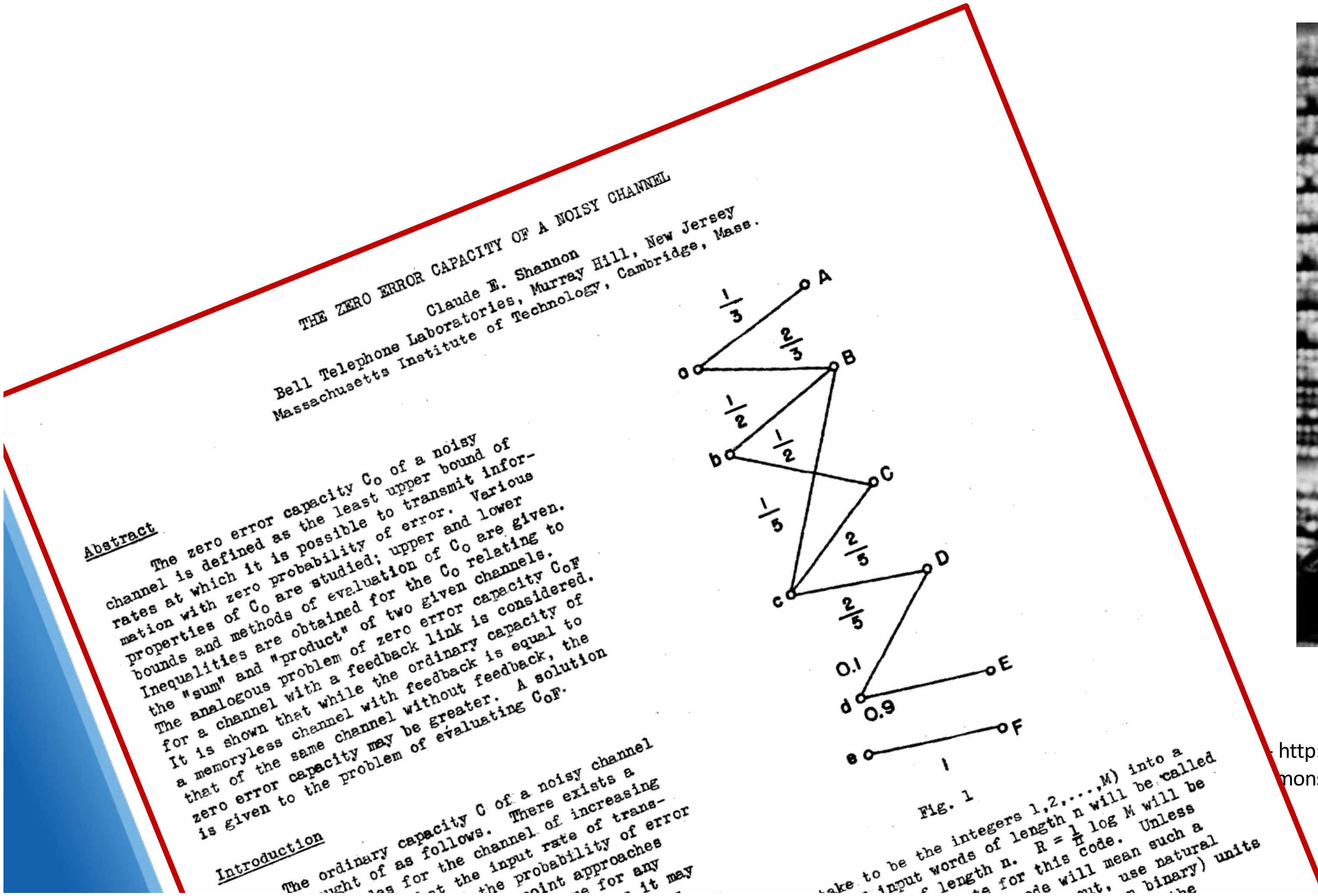
Communicating Short Messages

A Play in Three Acts

1. Feedback does not increase capacity
2. Stop feedback increases the rate of short messages that achieve a target reliability
3. Full feedback further increases the rate of short messages that achieve a target reliability

Act 1:
Feedback does not increase capacity.

Even Full Feedback does not increase capacity



THE ZERO ERROR CAPACITY OF A NOISY CHANNEL

Claude E. Shannon

Bell Telephone Laboratories, Murray Hill, New Jersey
Massachusetts Institute of Technology, Cambridge, Mass.

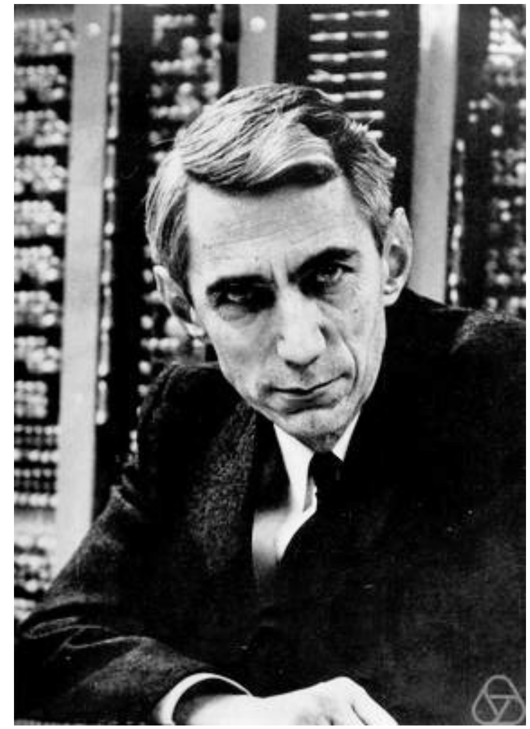
Abstract

The zero error capacity C_o of a noisy channel is defined as the least upper bound of rates at which it is possible to transmit information with zero probability of error. Various properties of C_o are studied; upper and lower bounds and methods of evaluation of C_o are given. Inequalities are obtained for the C_o relating to the "sum" and "product" of two given channels. The analogous problem of zero error capacity C_{oF} for a channel with a feedback link is considered. It is shown that while the ordinary capacity of a memoryless channel with feedback is equal to that of the same channel without feedback, the zero error capacity may be greater. A solution is given to the problem of evaluating C_{oF}.

Introduction

The ordinary capacity C of a noisy channel ...

Fig. 1



Shannon

Consider the gain in mutual information ΔI from transmitting $X_j = f(W, Y_1 \cdots Y_{j-1})$.

feedBack: $B = Y_1 \cdots Y_{j-1}$

Mutual information between message W and receiver's knowledge

Before we send $X_j = f(W, Y_1 \cdots Y_{j-1})$: $I(W; B)$

After we send $X_j = f(W, Y_1 \cdots Y_{j-1})$: $I(W; B, Y_j)$

$$\Delta I = I(W; B, Y_j) - I(W; B)$$

It turns out that ΔI cannot exceed C

$$\Delta I = \sum_{b} P_B(b) \Delta I(b)$$

For each possible feedback b, $\Delta I(b) \leq C$.

Act 2:
Stop feedback increases the rate of short messages that achieve a target reliability

“Stop” Feedback for Short Messages


IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 57, NO. 8, AUGUST 2011 4903

Feedback in the Non-Asymptotic Regime

Yury Polyanskiy, *Member, IEEE*, H. Vincent Poor, *Fellow, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—Without feedback, the backoff from capacity due to non-asymptotic blocklength can be quite substantial for blocklengths and error probabilities of interest in many practical applications. In this paper, novel achievability bounds are used to demonstrate that in the non-asymptotic regime, the maximal achievable rate improves dramatically thanks to variable-length coding and feedback. For example, for the binary symmetric channel with capacity 1/2 the blocklength required to achieve 90% of the capacity is smaller than 200, compared to at least 3100 for the best fixed-blocklength code (even with noiseless feedback). Virtually all the advantages of noiseless feedback are shown to be achievable, even if the feedback link is used only to send a single signal informing the encoder to terminate the transmission (stop-feedback). It is demonstrated that the non-asymptotic behavior of the fundamental limit depends crucially on the particular model chosen for the “end-of-packet” control signal. Fixed-blocklength codes and related questions concerning communicating with a guaranteed delay are discussed, in which situation feedback is demonstrated to be almost useless even non-asymptotically.**

***Index Terms*—Achievability bounds, channel capacity, converse bounds, feedback, memoryless channels, non-asymptotic analysis, Shannon theory, stop-feedback.**

I. INTRODUCTION

IN the context of fixed blocklength communication, Shannon [1] showed that noiseless feedback does not increase the capacity of memoryless channels but can increase the zero-error capacity. For a class of symmetric discrete memoryless channels (DMCs), Dobrushin [2] demonstrated that the ... holds even in the presence of noiseless ...

ditional output distributions. Moreover, zero-error capacity may improve from zero to the Shannon capacity (as in the case of the binary erasure channel (BEC)) if variable length is allowed. Furthermore, since existing communication systems with feedback (such as ARQ) have variable length, in the analysis of fundamental limits for channels with feedback, it is much more relevant and interesting to allow codes whose length is allowed to depend on the channel behavior.

We mention a few extensions of Burnashev’s work [3], [4] relevant to this paper. Yamamoto and Itoh proposed a simple and conceptually important two-phase coding scheme, attaining the optimal error exponent [5]. Using the notion of Goppa’s empirical mutual information (EMI) several authors have constructed universal coding schemes attaining rates arbitrarily close to capacity with small probability of error [6], [7], exponentially decaying probability of error [8] and even attaining the optimal Burnashev exponent [9], [10] simultaneously for a collection of channels. An extension to arbitrary varying channels with full state information available at the decoder has been recently proposed as well [11].

In contrast to the error exponent analysis of variable-length coding with feedback, which focuses on the regime of asymptotically long average blocklength at fixed rate, in this paper, following [12] we focus on the regime of fixed probability of error and finite average blocklength. Another aspect that was not previously addressed in the literature is the following. In practice, control information (such as initiation and termination) is not under the purview of the physical layer. However, the information theory literature typically assumes that all the feed-forward control information is carried through the same noisy channel as the information payload. This is most notably illustrated by Burnashev’s model in which the error exponent is, in fact, dictated by the reliability with which the termination information is conveyed to the receiver through the DMC while at the same time assuming that the feedback link has infinite reliability to carry the termination symbol but the whole sequence of channel ...


Polyanskiy


Poor


Verdu

By Royal Society uploader - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=34583438

An information density analysis that bit on their earlier work.

UCLA Samueli School of Engineering

Information Density i

- Information density i is The stochastic symbol-wise random variable whose expectation is the mutual information.

- Computed *per possible codeword* using a received symbol y_j and the corresponding symbol $\hat{x}_j$from a potential codeword.

$$i(y_j, \hat{x}_j) = \log_2 \frac{f_{Y|X}(y_j|\hat{x}_j)}{f_Y(y_j)} \tag{21}$$

Information Density

- Computed using a received symbol y_j and the corresponding symbol $\hat{x}_j$ from a potential codeword.

$$i(y_j, \hat{x}_j) = \log_2 \frac{f_{Y|X}(y_j|\hat{x}_j)}{f_Y(y_j)} \tag{21}$$

- Accumulated information density (AID) adds up these symbol-wise terms.

$$i(y^n, \hat{x}^n) = \sum_{j=1}^{n} \log_2 \frac{f_{Y|X}(y_j|\hat{x}_j)}{f_Y(y_j)} \tag{22}$$

Accumulated Information Density and $I(X;Y)$

$$i(y^n, \hat{x}^n) = \sum_{j=1}^{n} \log_2 \frac{f_{Y|X}(y_j|\hat{x}_j)}{f_Y(y_j)} \tag{22}$$

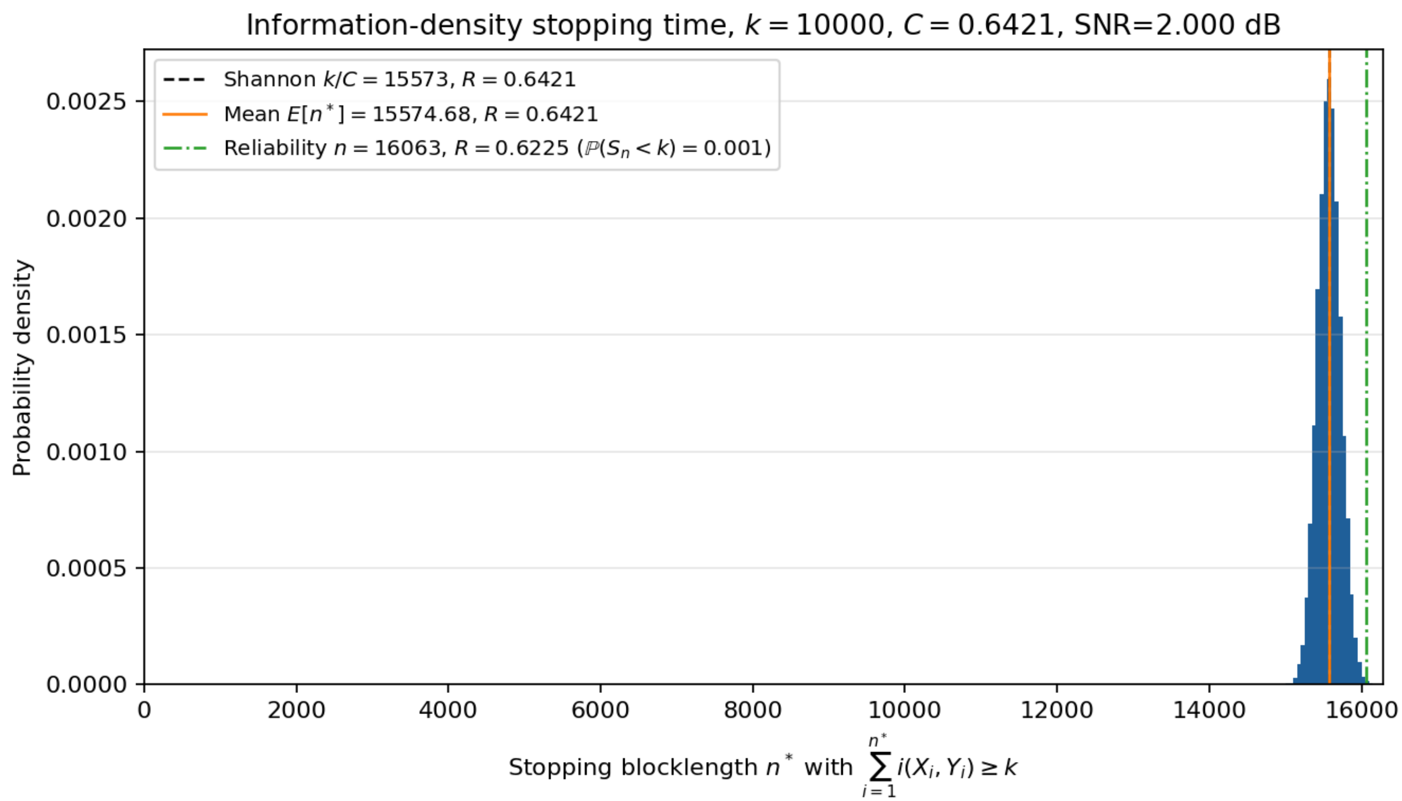
The Information Density Stopping Rule

- The receiver decides to terminate decoding when the accumulated information density for one of the possible codewords is above γ.
- Reasonable values of γ are on the order of the number of bits k in the message.
- Increasing γ lowers the probability of error but also increases the stopping time.

$$\tau = \inf\{n \geq 0 : \imath(X^n; Y^n) \geq \gamma\} \quad (25)$$

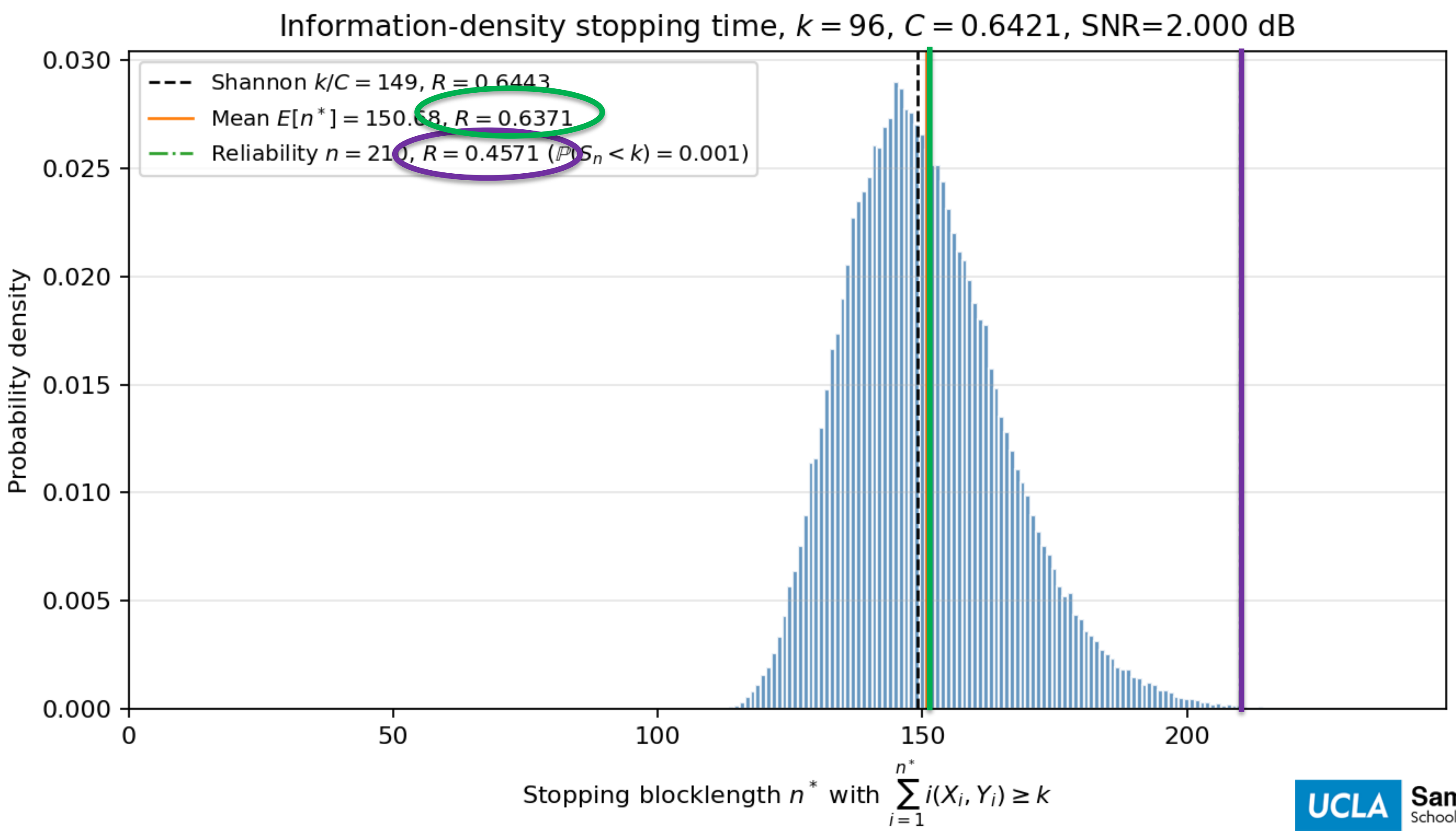
$$\bar{\tau} = \inf\{n \geq 0 : \imath(\bar{X}^n; Y^n) \geq \gamma\}. \quad (26)$$

Block-length n^* where $i \geq k$: k=10,000 bits



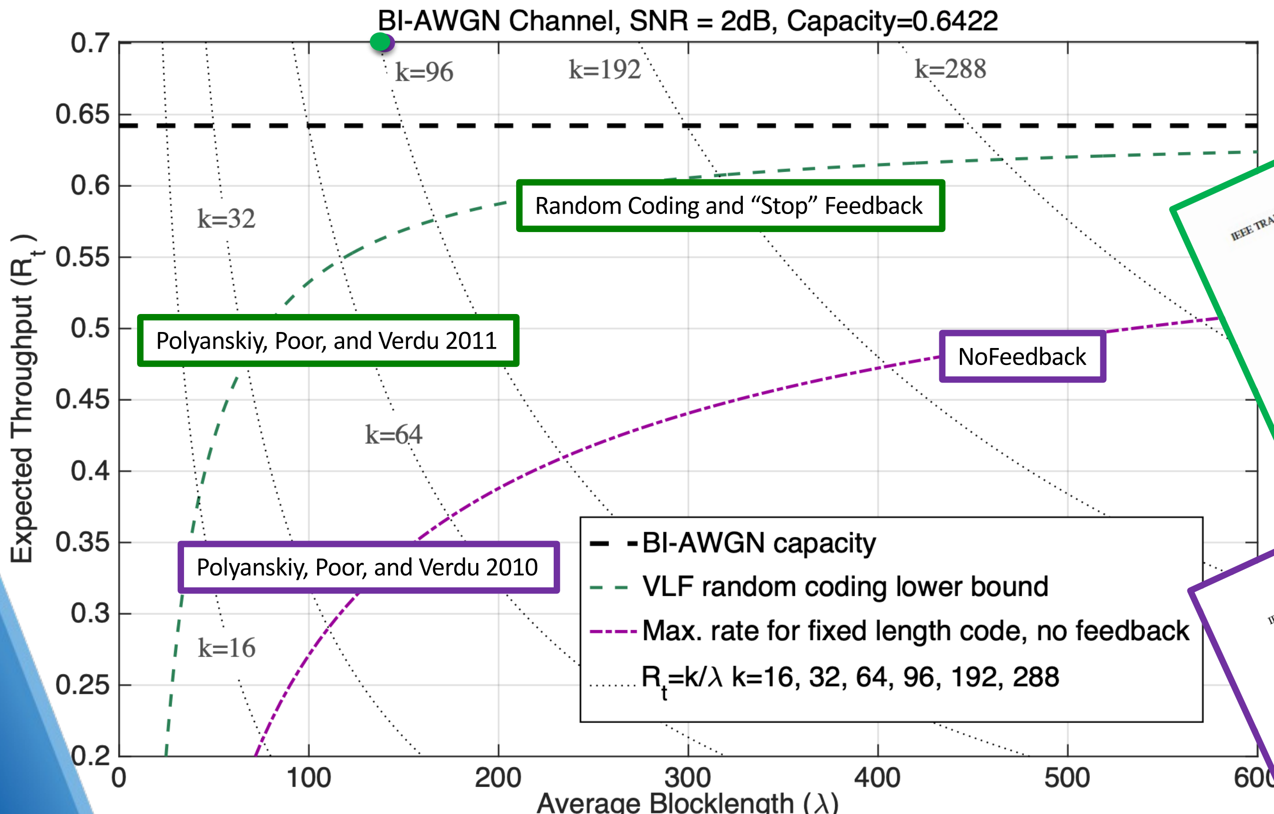
UCLA Samueli School of Engineering

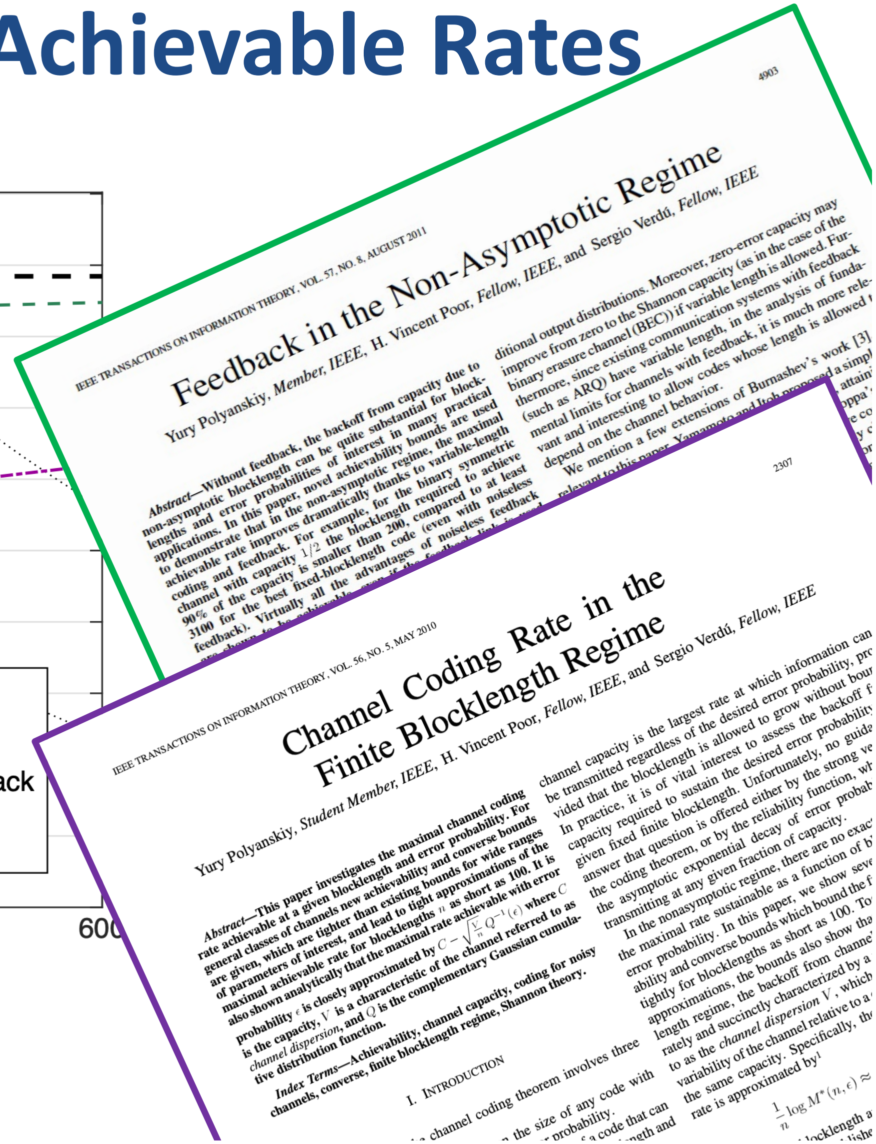
Block-length n^* where $i \geq k$: k=96 bits



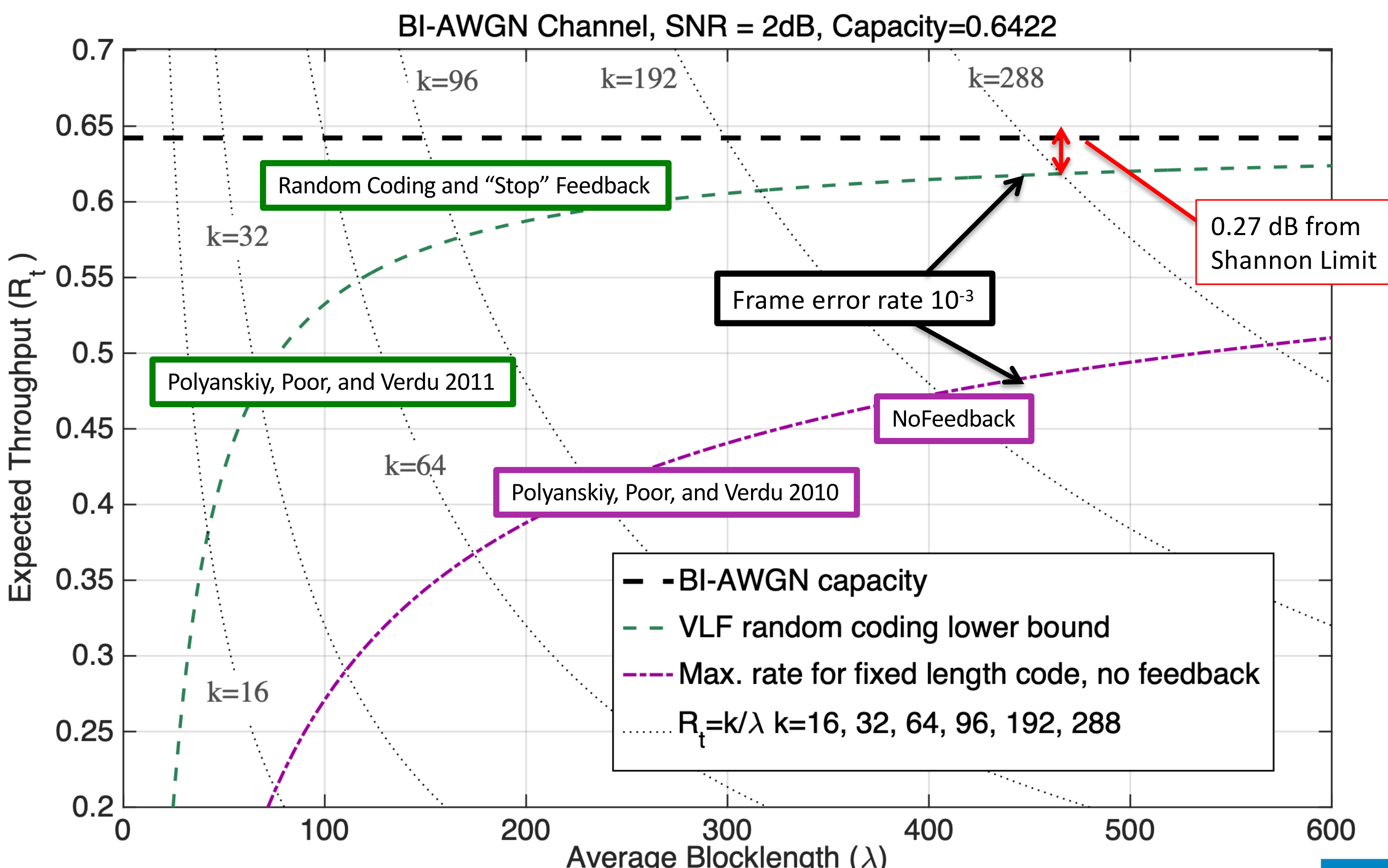
UCLA Samueli School of Engineering

Polyanskiy,Poor, Verdu Achievable Rates



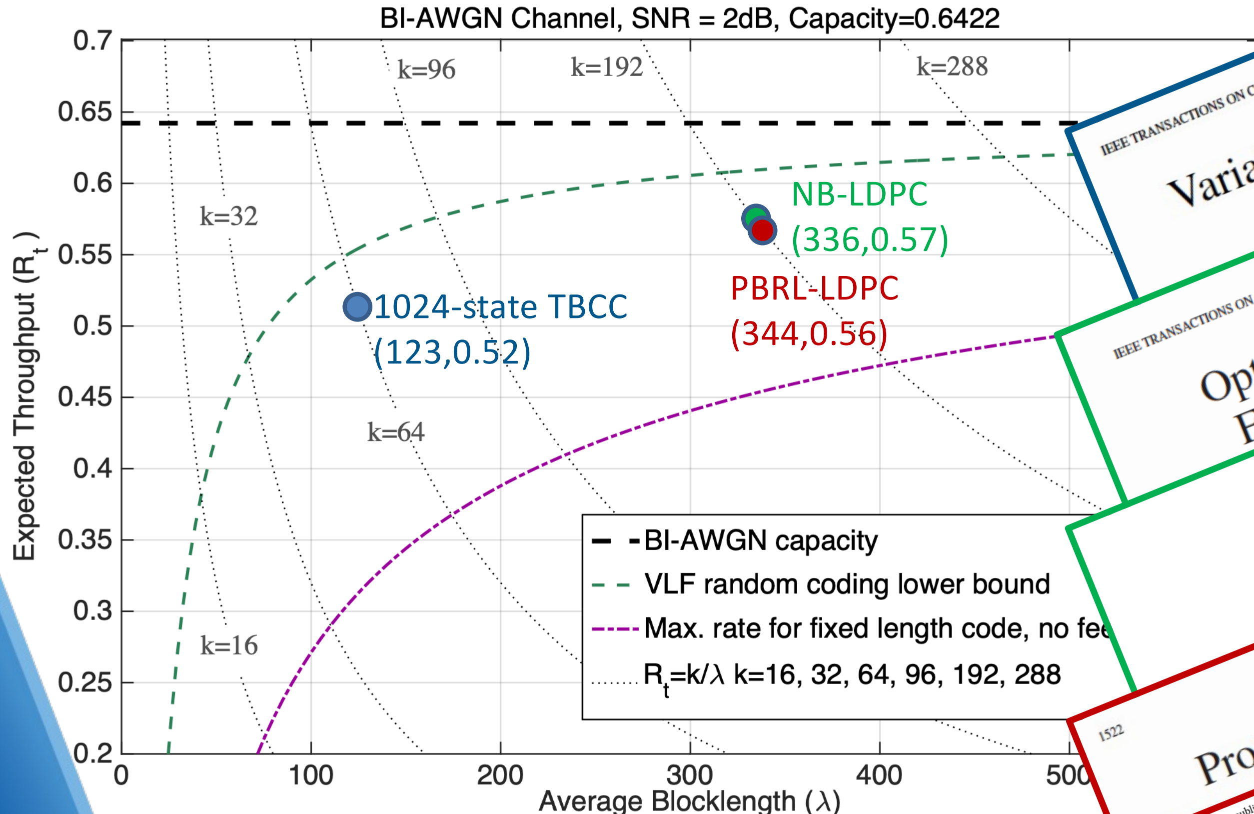


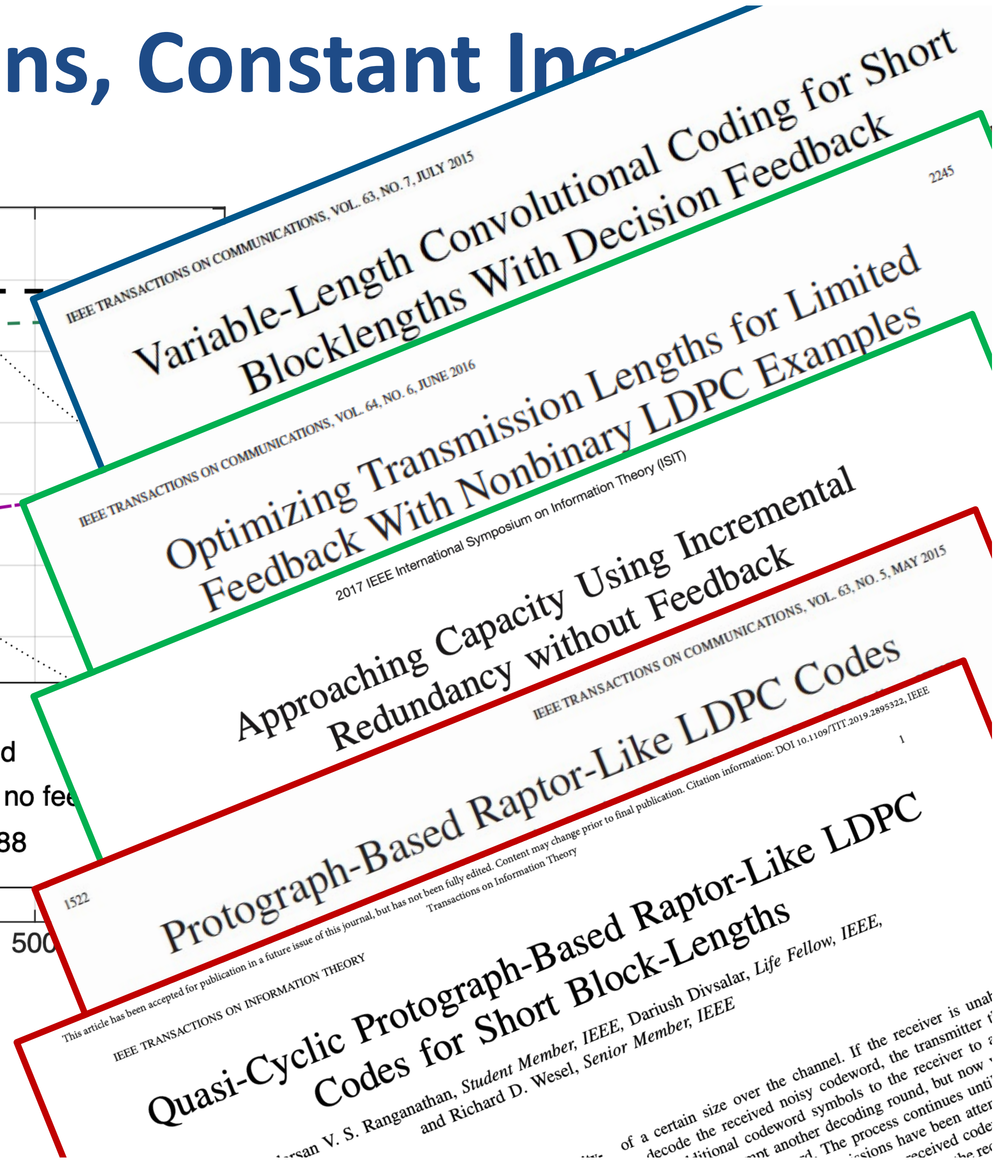
A dramatic increase in rate at short blocklength



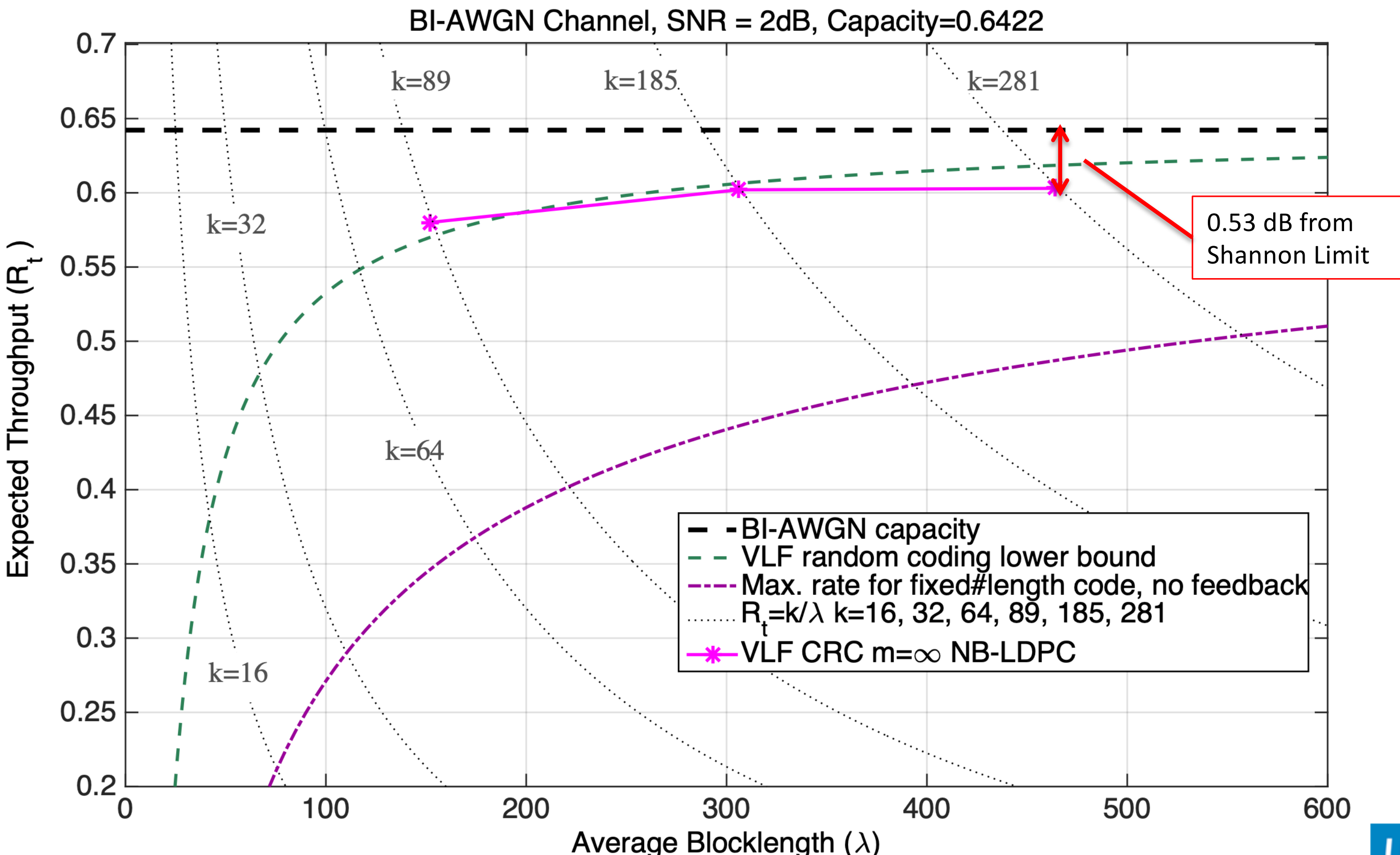
UCLA Samueli School of Engineering

Real codes, 5 Transmissions, Constant Inc





Limite Feedback, but more than Stop Feedback



UCLA Samueli School of Engineering

Rate-Compatible Coding with Limited Feedback

IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. 64, NO. 6, JUNE 2016 2245

Optimizing Transmission Lengths for Limited Feedback With Nonbinary LDPC Examples

Kasra Vakilinia, *Student Member, IEEE*, Sudarsan V. S. Ranganathan, *Member, IEEE*, Dariush Divsalar, *Life Fellow, IEEE*, and Richard D. Wesel, *Senior Member, IEEE*

***Abstract*—This paper presents a general approach for optimizing the number of symbols in increments (packets of incremental redundancy) in a feedback communication system with a limited number of increments. This approach is based on a tight normal approximation on the rate for successful decoding. Applying this approach to a variety of feedback systems using nonbinary (NB) low-density parity-check (LDPC) codes shows that greater than 90% of capacity can be achieved with blocklengths fewer than 500 transmitted bits. One result**

paper is not exclusively for NB-LDPC codes, but NB-LDPC codes are used for demonstration because they perform well in the short-blocklength regime (150 to 600 bits) that is of interest.

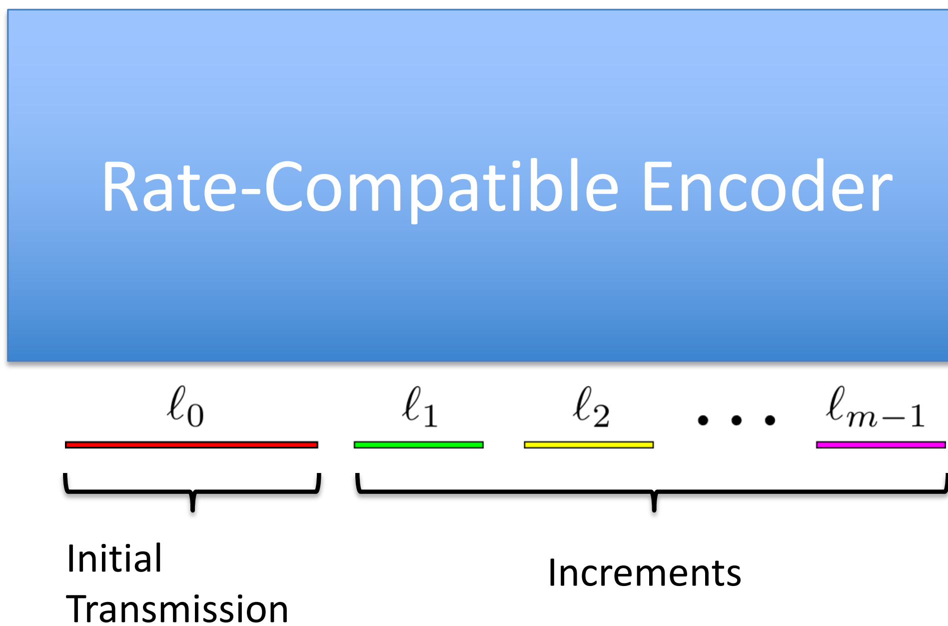
In VLFT analysis of [2], the receiver provides full noiseless feedback to the transmitter. The transmitter sends additional incremental bits until it knows the receiver has decoded the message correctly, resulting in zero probability of error. The "T" in VLFT stands for termination and corresponds to a noiseless transmitter confirmation (NTC) bit that the transmitter uses to terminate the transmission. The NTC is transmitted through a channel different from the main communication channel. In contrast, VLF (without the "T") does not have the advantage of an NTC. All VLF forward transmissions go over the same noisy channel. Thus, there is a nonzero probability of undetected error

Vakilinia

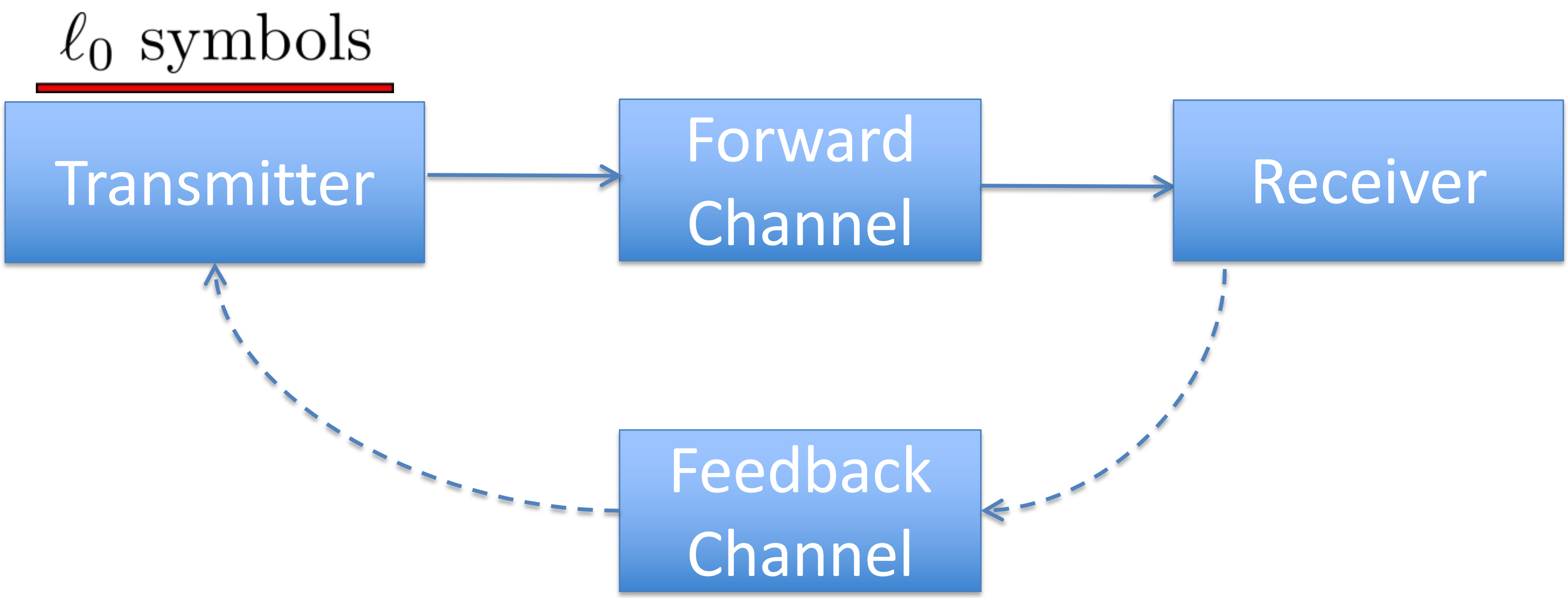
UCLA Samueli School of Engineering

A Rate-compatible Encoder

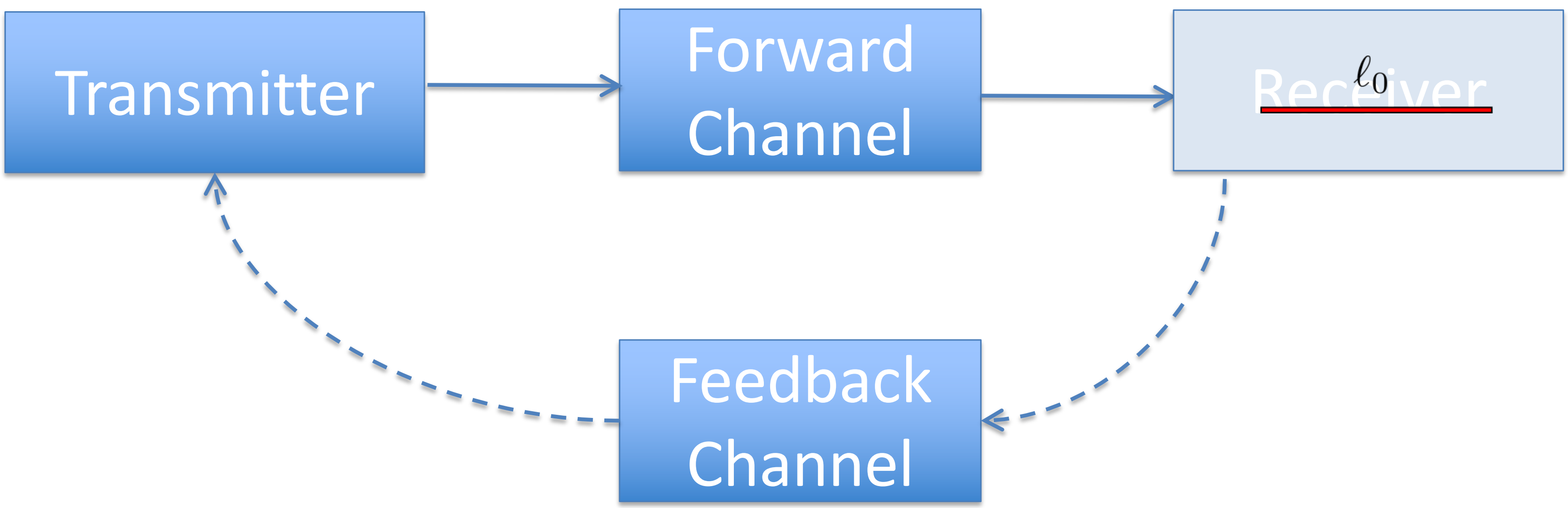
k bits of user information



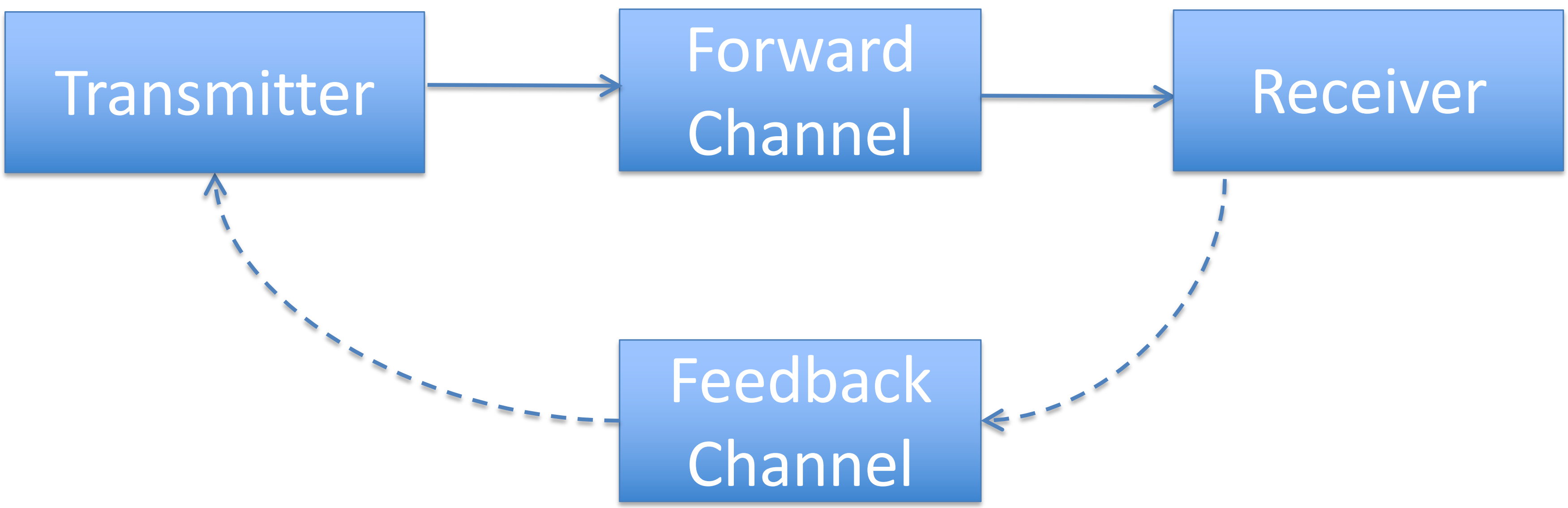
Incremental-Redundancy Hybrid ARQ



Incremental-Redundancy Hybrid ARQ



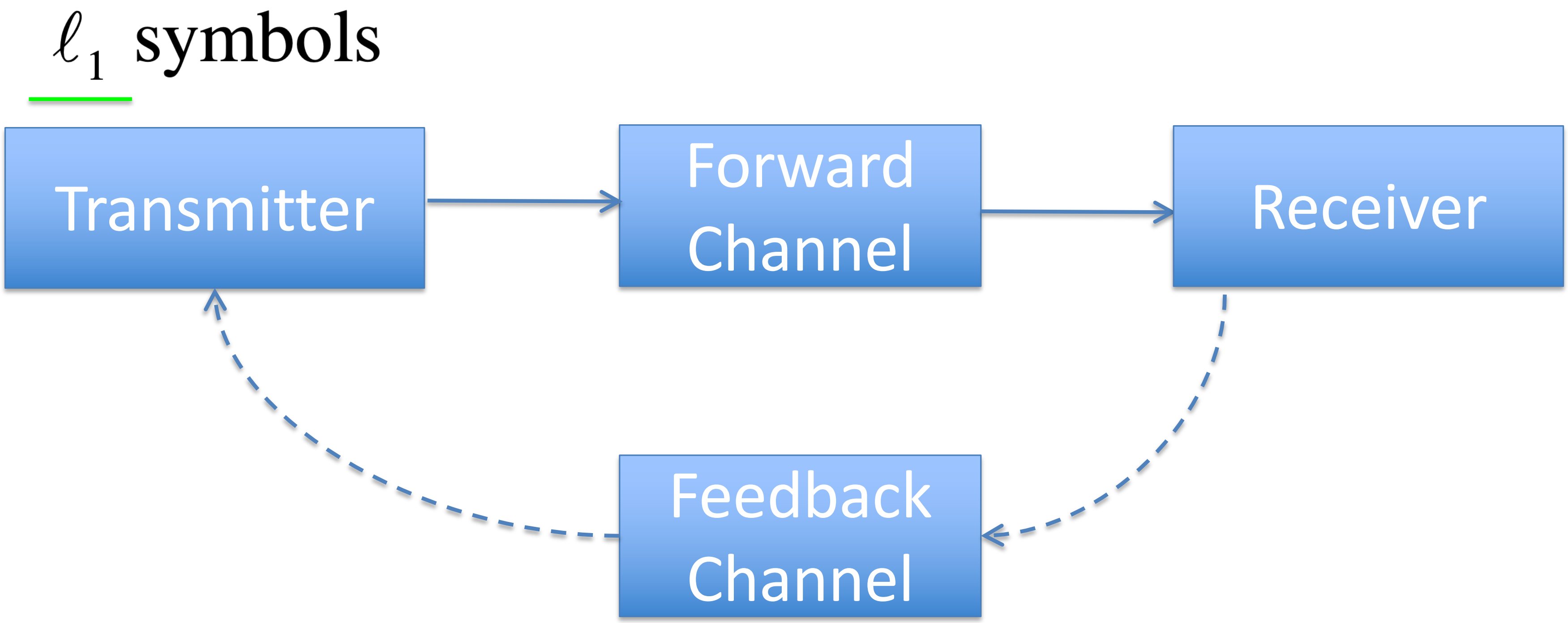
Incremental-Redundancy Hybrid ARQ



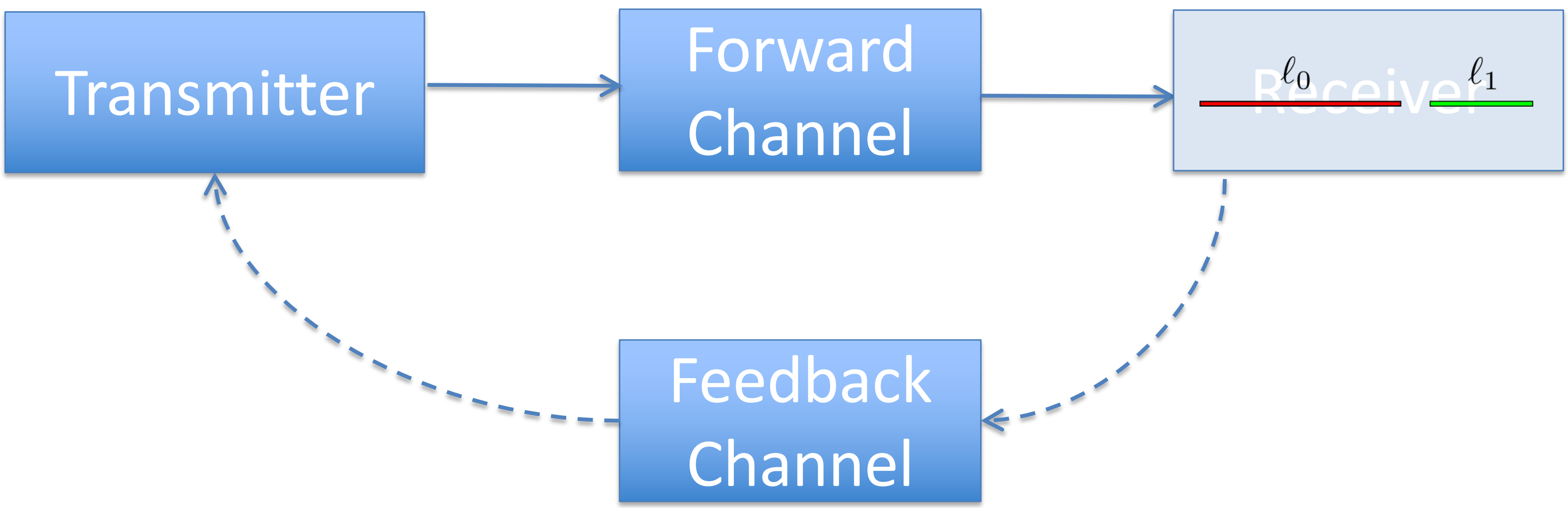
The probability that the decoder declares success is $P_{\mathrm{ACK}}^{(\ell_1)}$.

The probability that the decoder declares failure is $P_{NACK}^{(\ell_1)}$.

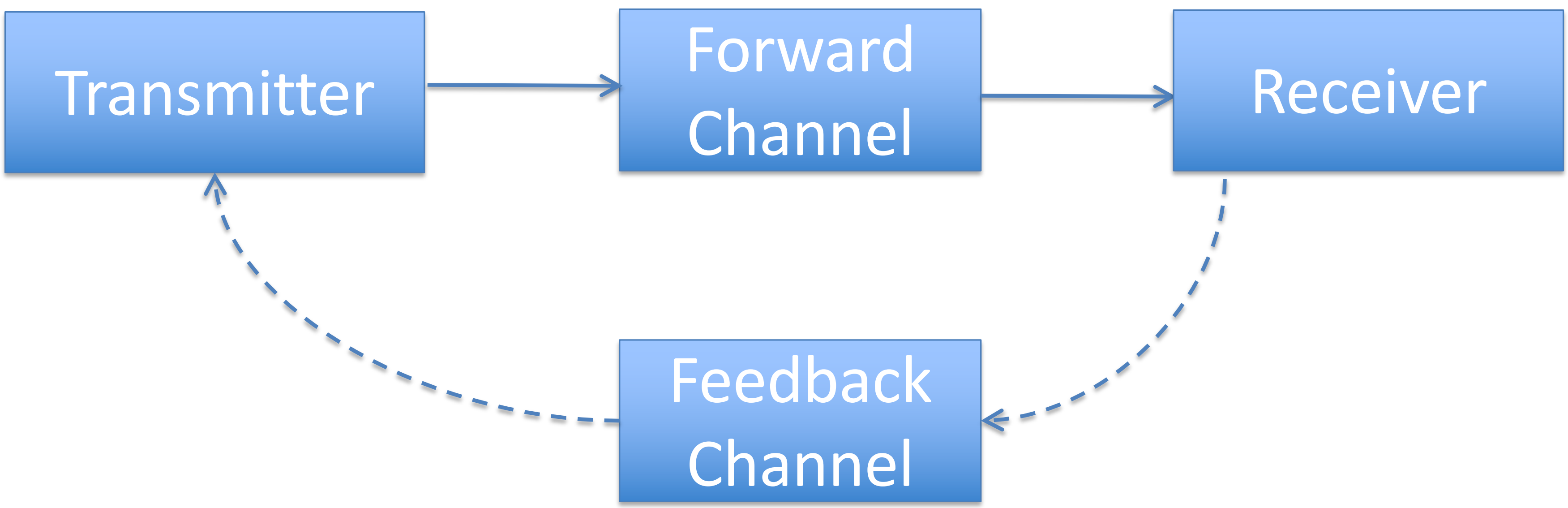
Incremental-Redundancy Hybrid ARQ



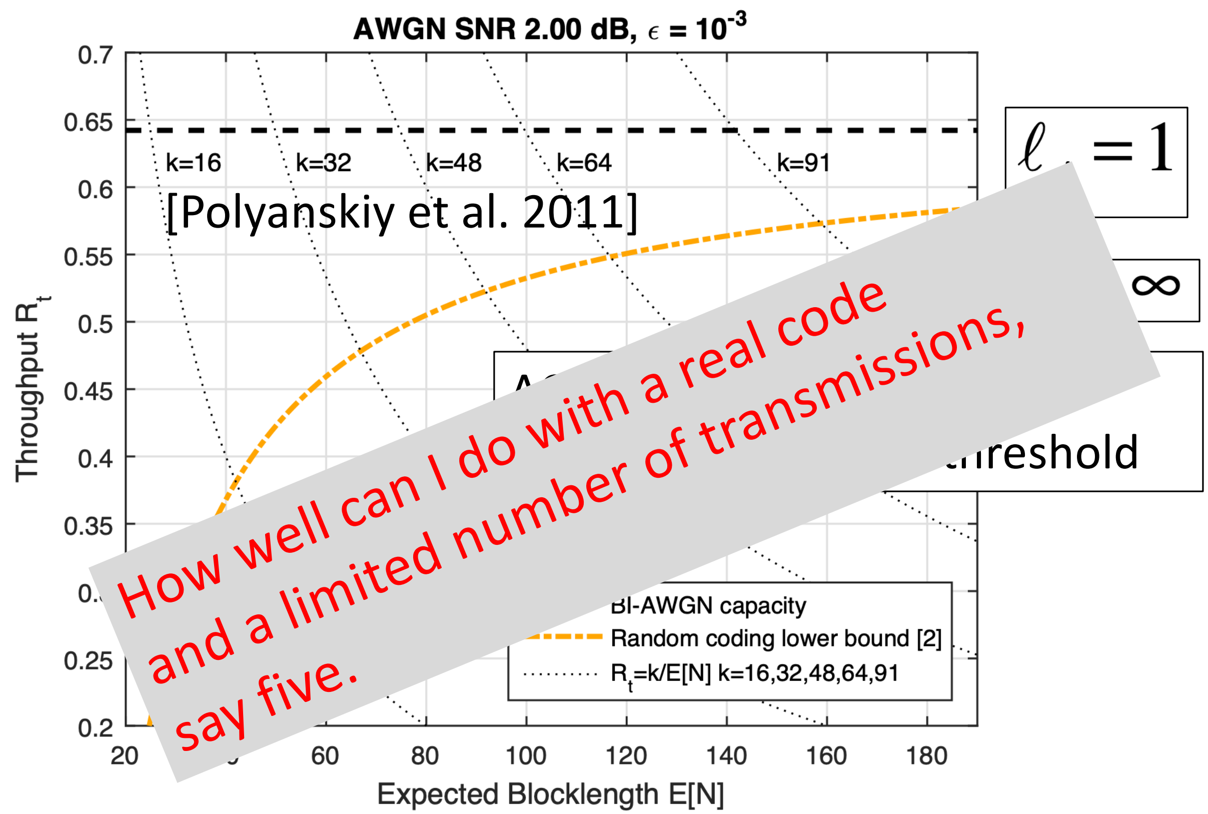
Incremental-Redundancy Hybrid ARQ



Incremental-Redundancy Hybrid ARQ

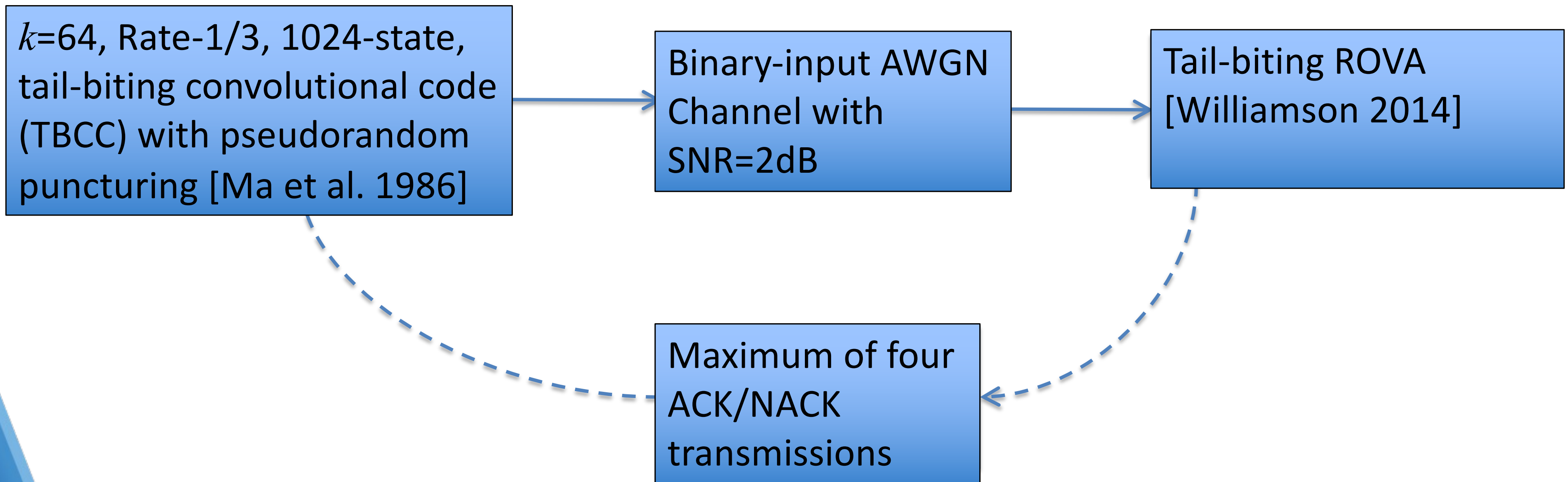


What if I am limited to five transmissions?

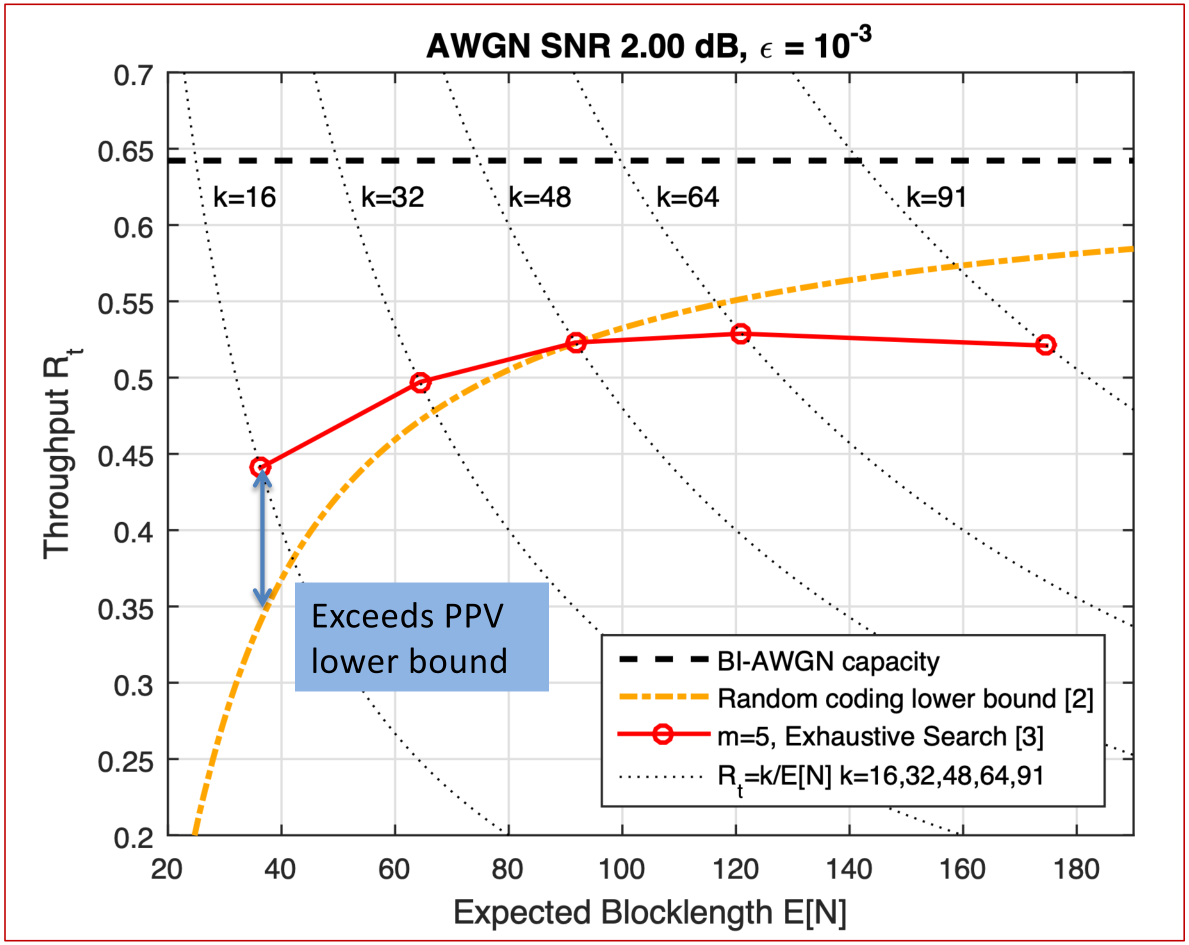


How well can I do with a real code and a limited number of transmissions, say five.

A real system [Williamson 2015]



A real system [Williamson 2015]



Increments and cumulative blockengths

$$\underbrace{\ell_1}_{} \quad \ell_2 \quad \ell_3 \quad \ell_4 \quad \cdots \quad \ell_m$$

Increments and cumulative blockengths

$$\underbrace{\ell_1 \quad \ell_2}\quad \ell_3 \quad \ell_4 \quad \cdots \quad \ell_m$$

Increments and cumulative blockengths

$$\underbrace{\ell_1 \quad \ell_2 \quad \ell_3 \quad \ell_4 \quad \cdots \quad \ell_m}$$

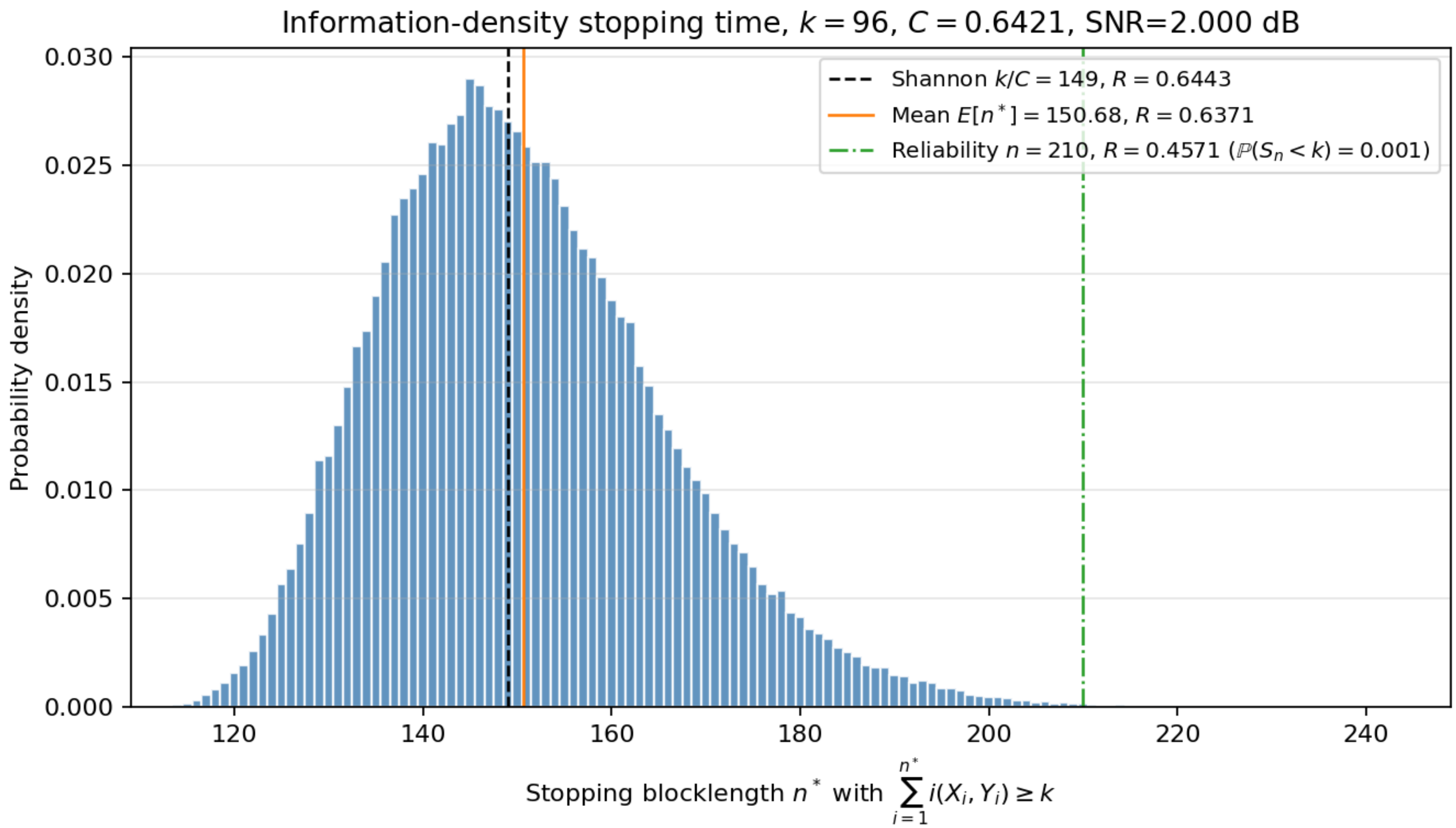
Stopping after m transmissions leads to little loss in many cases [Heindlmaier and Soljanin].

But How To Optimize Transmission Lengths?

$$N_f = \sum_{j=1}^{f} \ell_j. \tag{1}$$

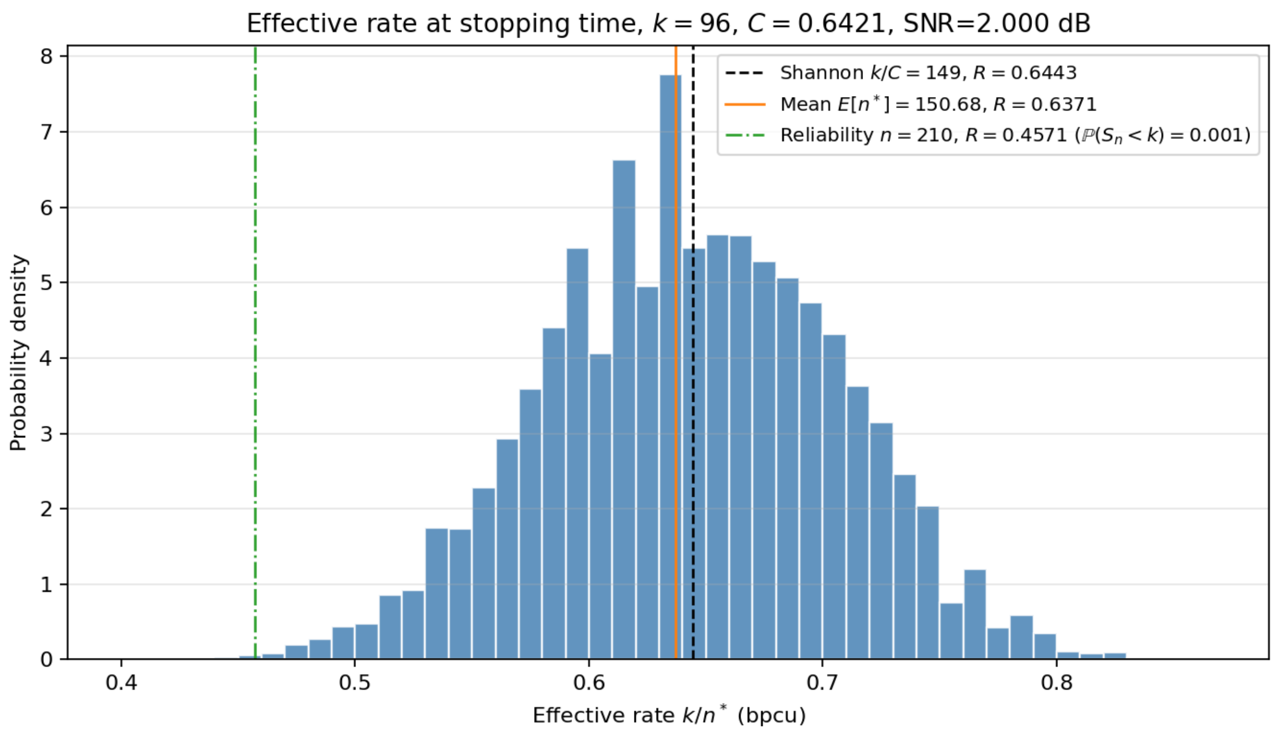
$$R_T = \frac{E[I]}{E[N]} = \frac{k\left(1 - P_{\text{NACK}}^{(m)} - P_{\text{UE}}\right)}{E[N]}, \tag{2}$$

Block Length n^* where $i \geq k$: k=96 bits

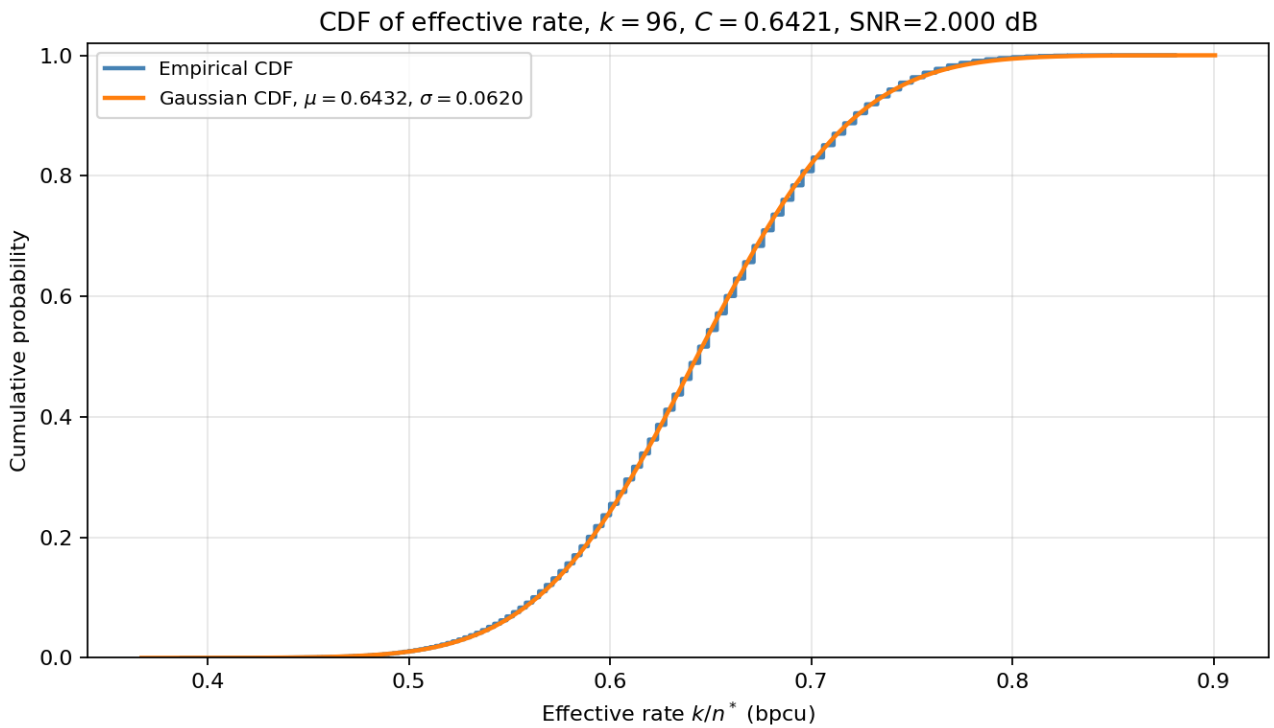


UCLA Samueli School of Engineering

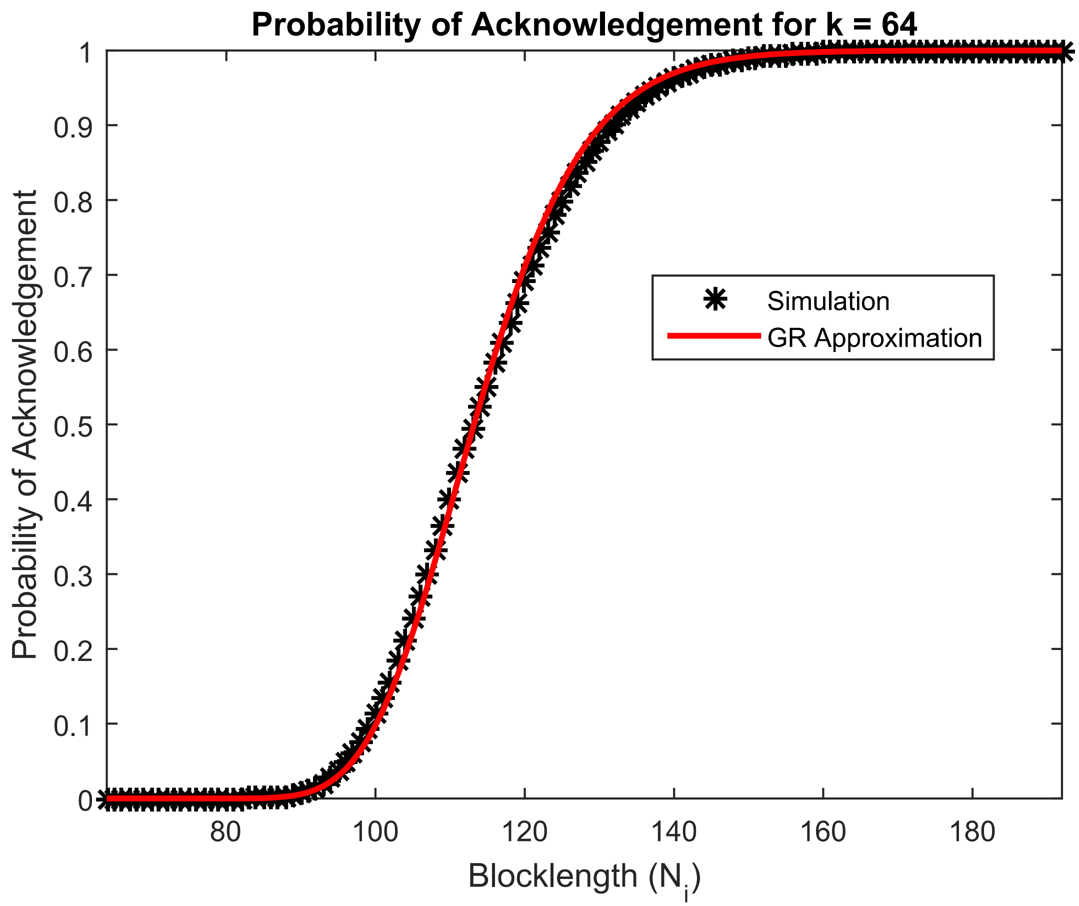
Effective rate k/n^* where $i \geq k$: k=96 bits



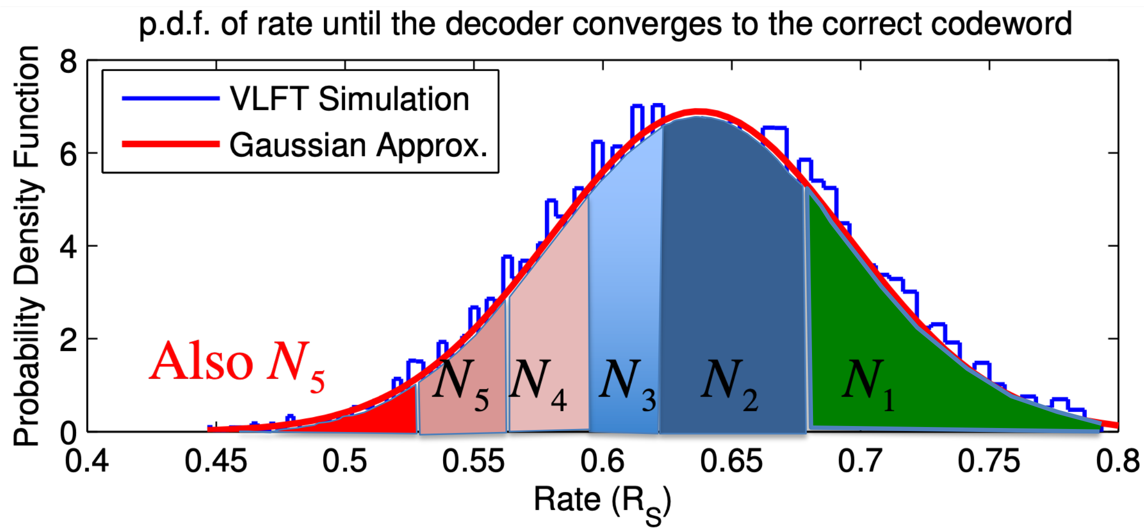
CDF of Effective rate k/n^* where $i \geq k$: k=96



P_{ACK} for actual code also matches Gaussian



A Closer Look at E[N]



$$E[N] = N_1 Q\left(\frac{\frac{k}{N_1} - \mu_S}{\sigma_S}\right) + \sum_{i=2}^{m} N_i \left[Q\left(\frac{\frac{k}{N_i} - \mu_S}{\sigma_S}\right) - Q\left(\frac{\frac{k}{N_{i-1}} - \mu_S}{\sigma_S}\right)\right] + N_m \left[1 - Q\left(\frac{\frac{k}{N_m} - \mu_S}{\sigma_S}\right)\right]$$

But How To Optimize Transmission Lengths?

$$\frac{\partial E[N]}{\partial N_1} = P_{\text{ACK}}^{(N_1)} + (N_1 - N_2)P_{\text{ACK}}'^{(N_1)} = 0 \qquad (4)$$

UCLA Samueli School of Engineering

But How To Optimize Transmission Lengths?

For $j > 2$, $\frac{\partial E[N]}{\partial N_{j-1}} = 0$ depends only on $\{N_{j-2}, N_{j-1}, N_j\}$ as follows:

$$\frac{\partial E[N]}{\partial N_{j-1}} = P_{\text{ACK}}^{(N_{j-1})} + (N_{j-1} - N_f) P_{\text{ACK}}^{\prime (N_{j-1})} - P_{\text{ACK}}^{(N_{j-2})} . \quad (6)$$

Sequential Differential Optimization

$$N_2 = N_1 + \frac{P_{\text{ACK}}^{(N_1)}}{P_{\text{ACK}}'^{(N_1)}} . \qquad (5)$$

$$N_j = N_{j-1} + \frac{P_{\text{ACK}}^{(N_{j-1})} - P_{\text{ACK}}^{(N_{j-2})}}{P_{\text{ACK}}'^{(N_{j-1})}} . \qquad (7)$$

Note that SDO requires a differentiable PDF for P_{ACK}.

UCLA Samueli School of Engineering

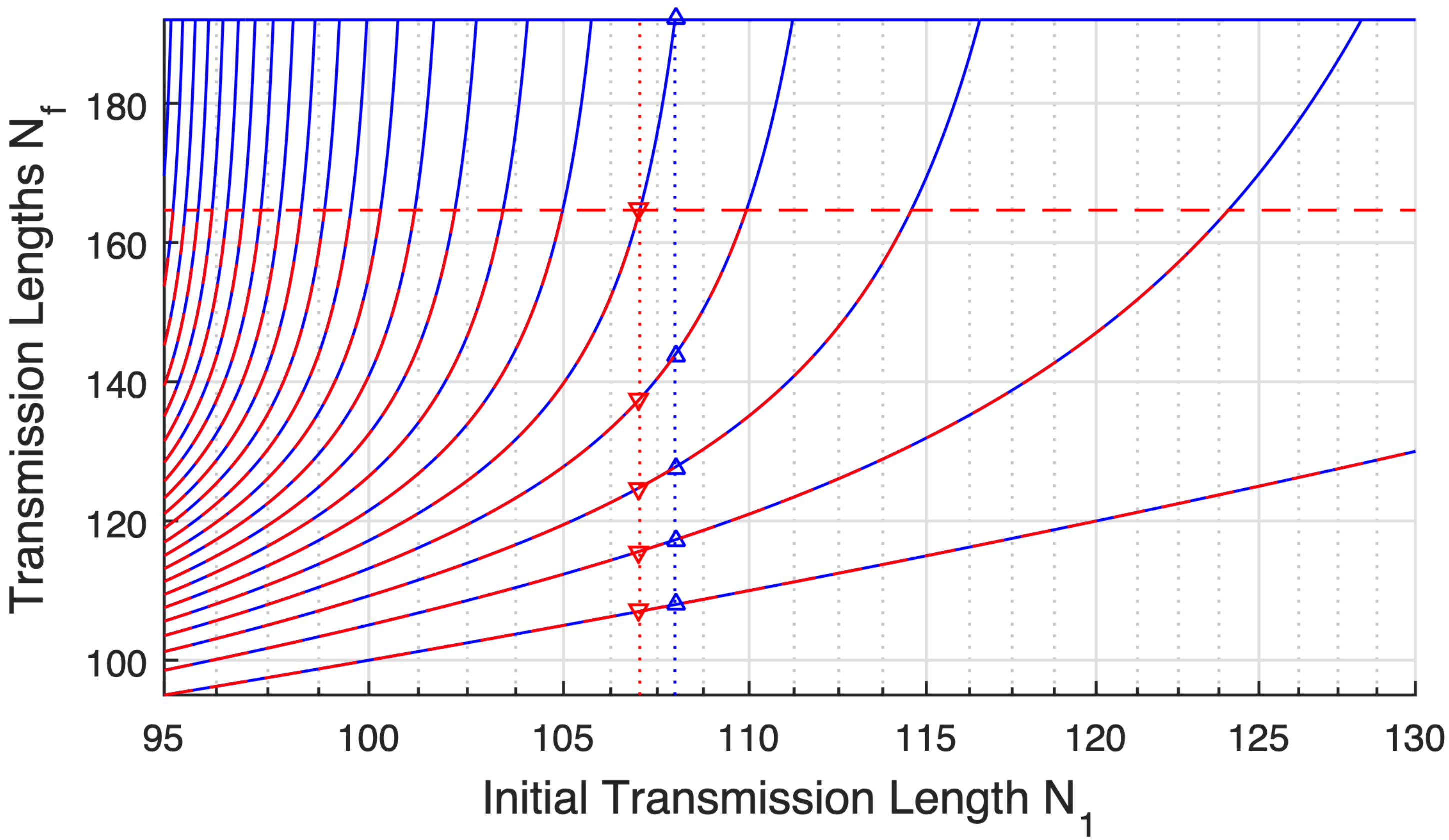
Sequential Differential Optimization

$$N_2 = N_1 + \frac{P_{\text{ACK}}^{(N_1)}}{P_{\text{ACK}}^{\prime(N_1)}}\,. \tag{5}$$

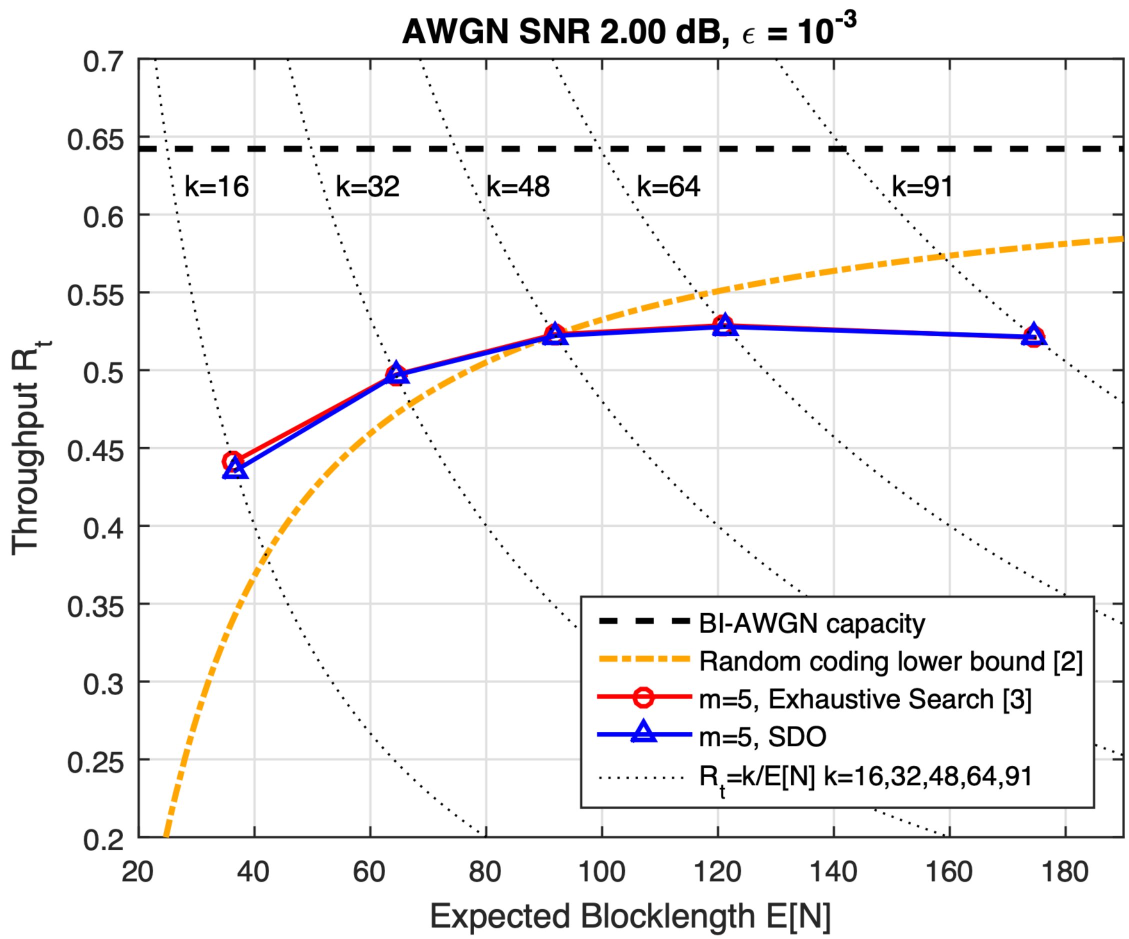
$$N_j = N_{j-1} + \frac{P_{\text{ACK}}^{(N_{j-1})} - P_{\text{ACK}}^{(N_{j-2})}}{P_{\text{ACK}}^{\prime(N_{j-1})}}\,. \tag{7}$$

Note that SDO gives a solution for each possible value of N_1

The evolution of transmission lengths

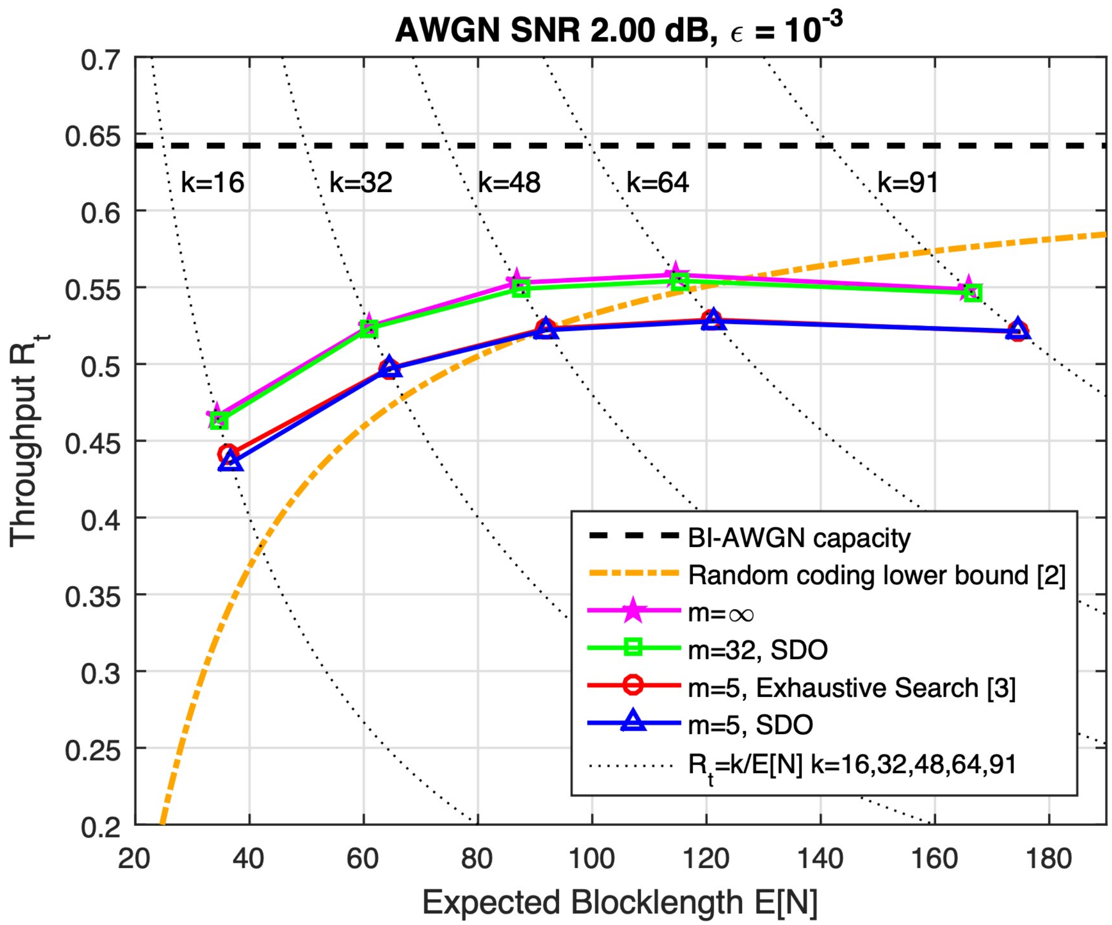


SDO matches Exhaustive Search

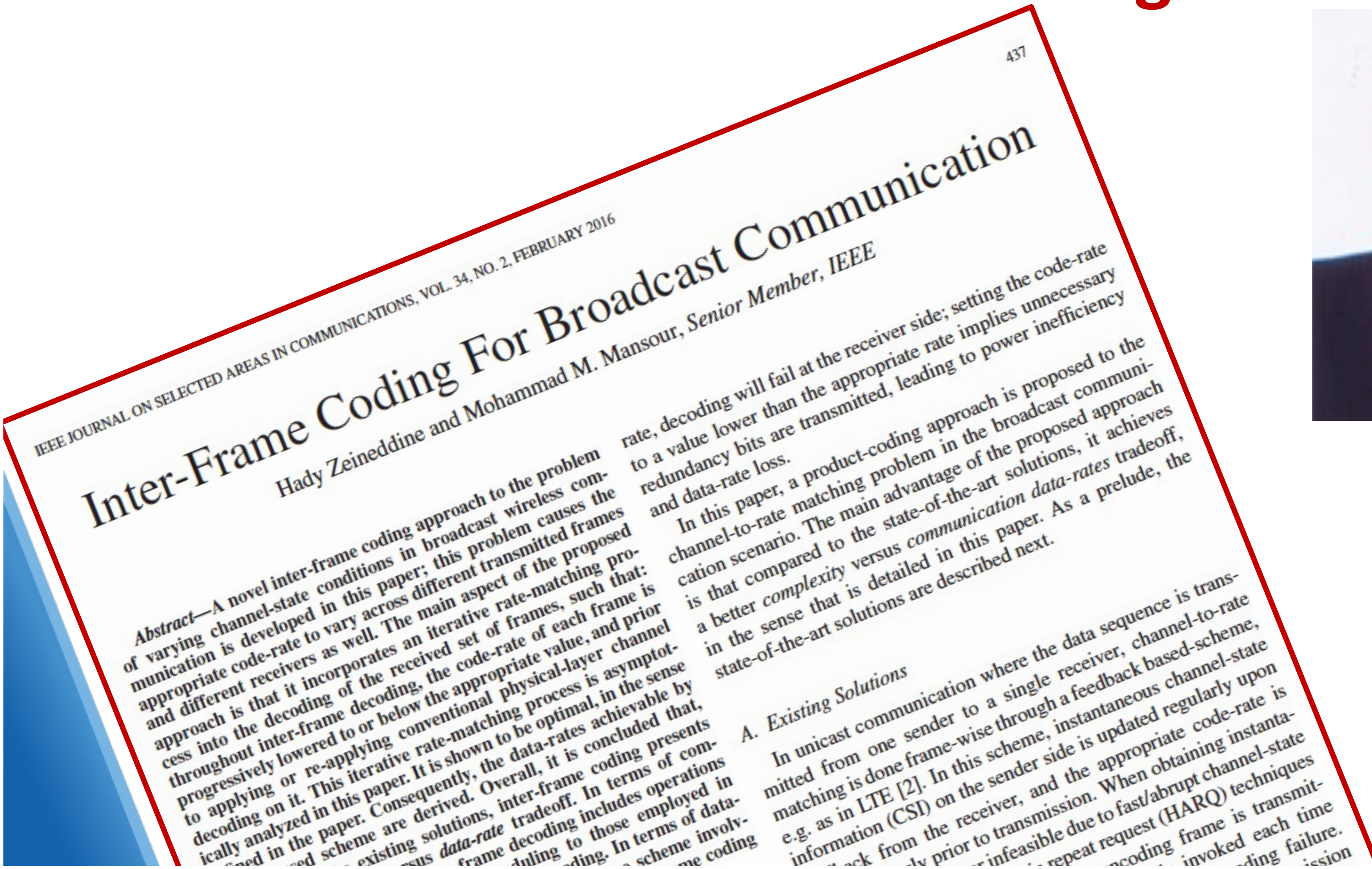


UCLA Samueli School of Engineering

SDO can optimize any number of lengths



Inter-Frame Coding



IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 34, NO. 2, FEBRUARY 2016 437

Inter-Frame Coding For Broadcast Communication

Hady Zeineddine and Mohammad M. Mansour, *Senior Member, IEEE*

Abstract—A novel inter-frame coding approach to the problem of varying channel-state conditions in broadcast wireless communication is developed in this paper; this problem causes the appropriate code-rate to vary across different transmitted frames and different receivers as well. The main aspect of the proposed approach is that it incorporates an iterative rate-matching process into the decoding of the received set of frames, such that: throughout inter-frame decoding, the code-rate of each frame is progressively lowered to or below the appropriate value, and prior to applying or re-applying conventional physical-layer channel decoding on it. This iterative rate-matching process is asymptotically analyzed in this paper. It is shown to be optimal, in the sense defined in the paper. Consequently, the data-rates achievable by the proposed scheme are derived. Overall, it is concluded that, compared to existing solutions, inter-frame coding presents a better complexity versus *data-rate* tradeoff. In terms of com-

rate, decoding will fail at the receiver side; setting the code-rate to a value lower than the appropriate rate implies unnecessary redundancy bits are transmitted, leading to power inefficiency and data-rate loss.

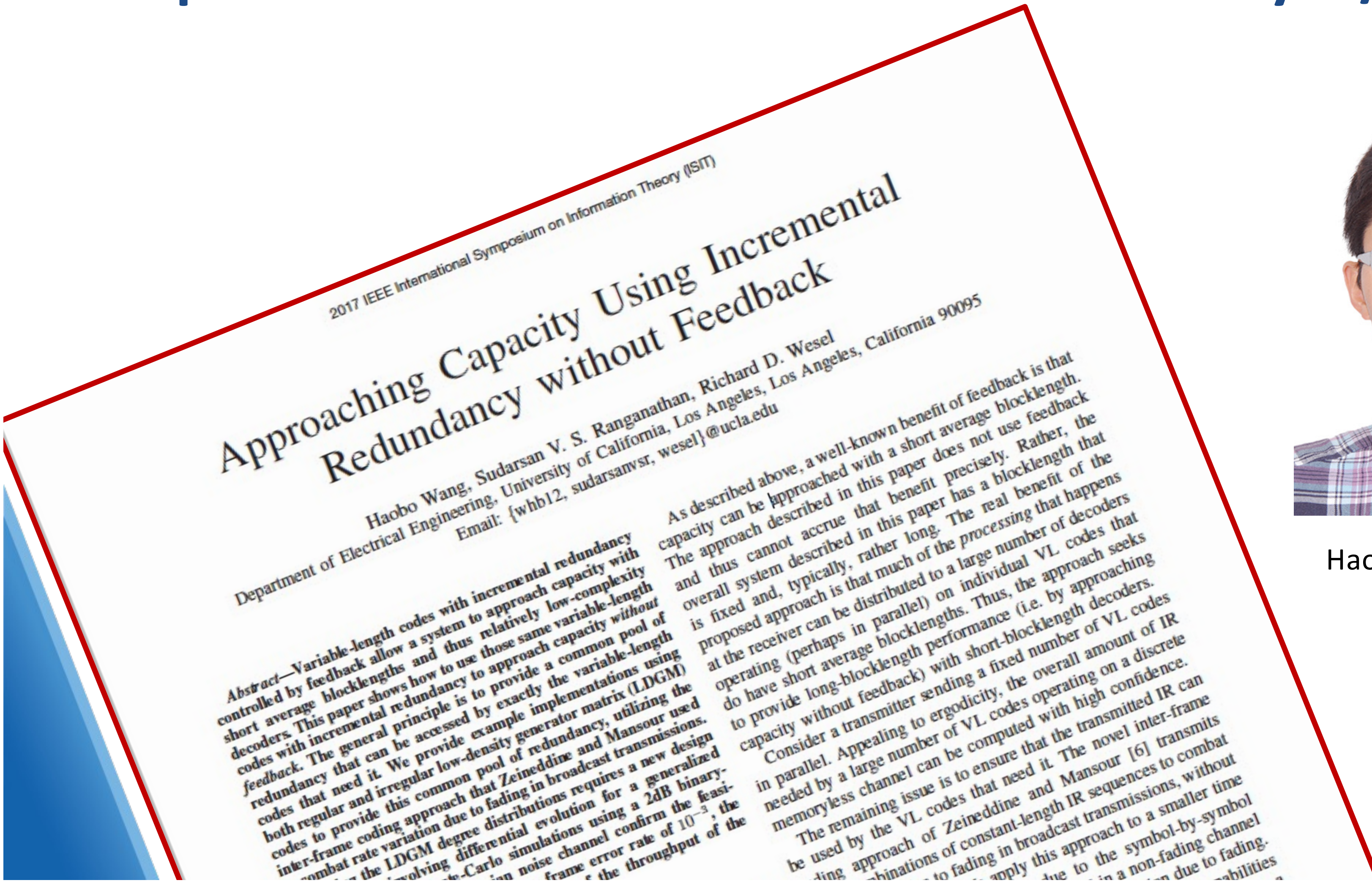
In this paper, a product-coding approach is proposed to the channel-to-rate matching problem in the broadcast communication scenario. The main advantage of the proposed approach is that compared to the state-of-the-art solutions, it achieves a better *complexity* versus *communication data-rates* tradeoff, in the sense that is detailed in this paper. As a prelude, the state-of-the-art solutions are described next.

A. Existing Solutions

In unicast communication where the data sequence is transmitted from one sender to a single receiver, channel-to-rate matching is done frame-wise through a feedback based-scheme, e.g. as in LTE [2]. In this scheme, instantaneous channel-state information (CSI) on the sender side is updated regularly upon feedback from the receiver, and the appropriate code-rate is ... prior to transmission. When obtaining instantaneous channel-state ... infeasible due to fast/abrupt channel-state ... repeat request (HARQ) techniques

Mansour

Implementation of Incremental Redundancy w/o Feedback



2017 IEEE International Symposium on Information Theory (ISIT)

Approaching Capacity Using Incremental Redundancy without Feedback

Haobo Wang, Sudarsan V. S. Ranganathan, Richard D. Wesel
Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, California 90095
Email: {whb12, sudarsanvsr, wesel}@ucla.edu

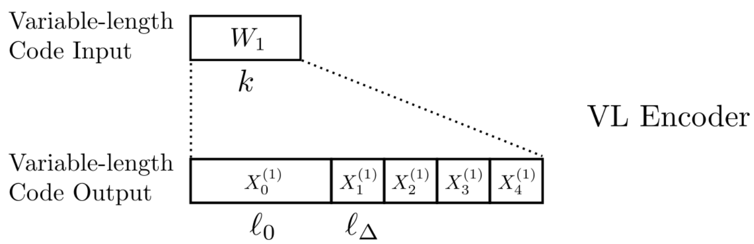
Abstract—Variable-length codes with incremental redundancy controlled by feedback allow a system to approach capacity with short average blocklengths and thus relatively low-complexity decoders. This paper shows how to use those same variable-length codes with incremental redundancy to approach capacity *without feedback*. The general principle is to provide a common pool of redundancy that can be accessed by exactly the variable-length codes that need it. We provide example implementations using both regular and irregular low-density generator matrix (LDGM) codes to provide this common pool of redundancy, utilizing the inter-frame coding approach that Zeineddine and Mansour used to combat rate variation due to fading in broadcast transmissions. ... the LDGM degree distributions requires a new design ... involving differential evolution for a generalized ... -Carlo simulations using a 2dB binary- ... noise channel confirm the feasi- ... frame error rate of 10^{-3}, the ... of the throughput of the

As described above, a well-known benefit of feedback is that capacity can be approached with a short average blocklength. The approach described in this paper does not use feedback and thus cannot accrue that benefit precisely. Rather, the overall system described in this paper has a blocklength that is fixed and, typically, rather long. The real benefit of the proposed approach is that much of the *processing* that happens at the receiver can be distributed to a large number of decoders operating (perhaps in parallel) on individual VL codes that do have short average blocklengths. Thus, the approach seeks to provide long-blocklength performance (i.e. by approaching capacity without feedback) with short-blocklength decoders.

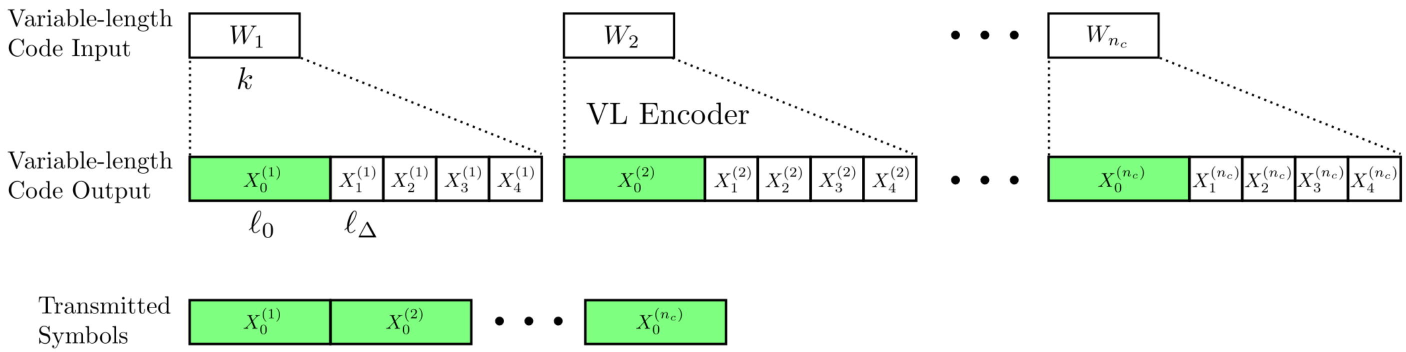
Consider a transmitter sending a fixed number of VL codes in parallel. Appealing to ergodicity, the overall amount of IR needed by a large number of VL codes operating on a discrete memoryless channel can be computed with high confidence. The remaining issue is to ensure that the transmitted IR can be used by the VL codes that need it. The novel inter-frame ... approach of Zeineddine and Mansour [6] transmits ... combinations of constant-length IR sequences to combat ... to fading in broadcast transmissions, without ... apply this approach to a smaller time ... due to the symbol-by-symbol ... in a non-fading channel ... due to fading.

Haobo Wang

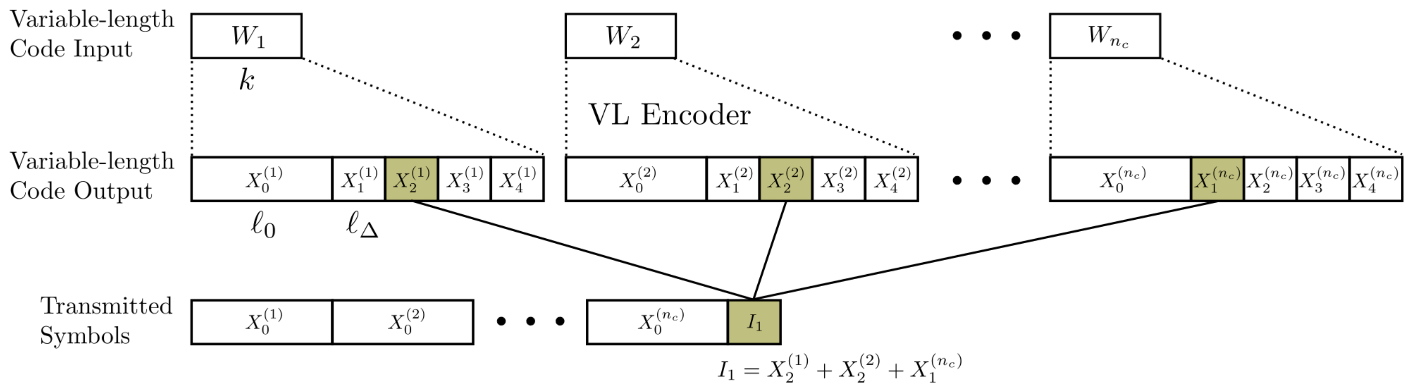
Inter-frame Coding



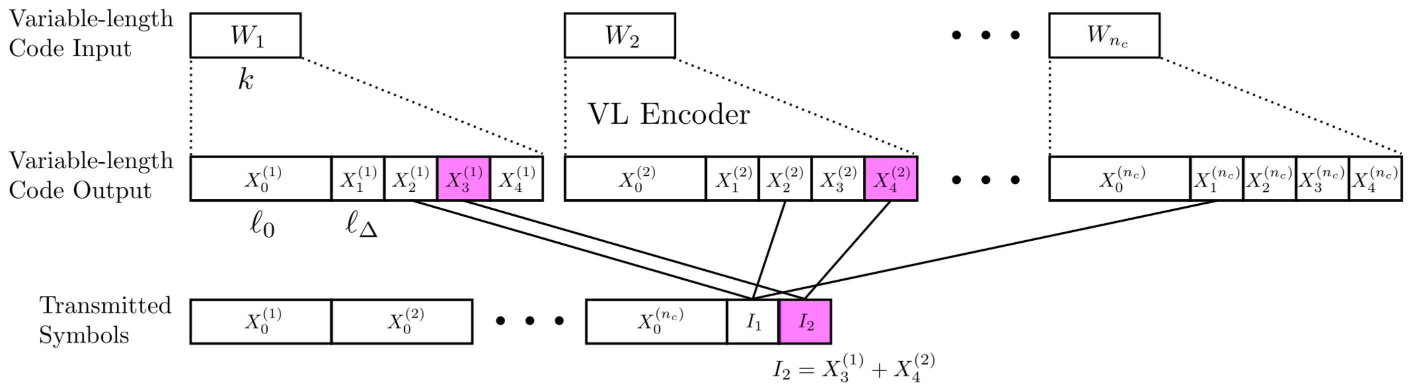
Inter-frame Coding



Inter-frame Coding

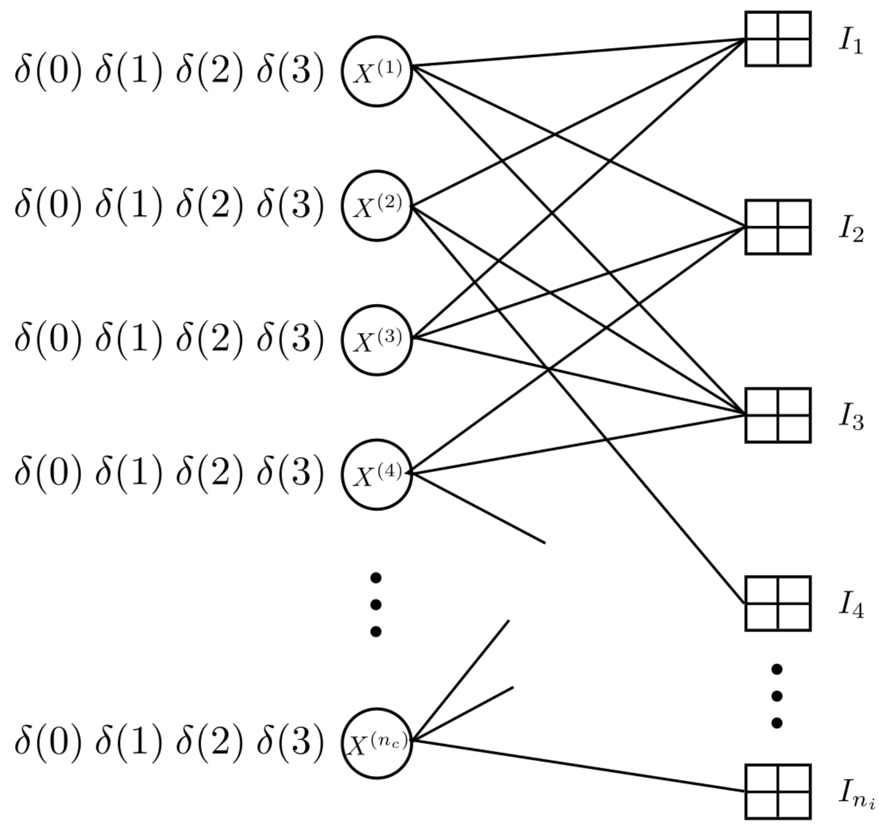


Inter-frame Coding



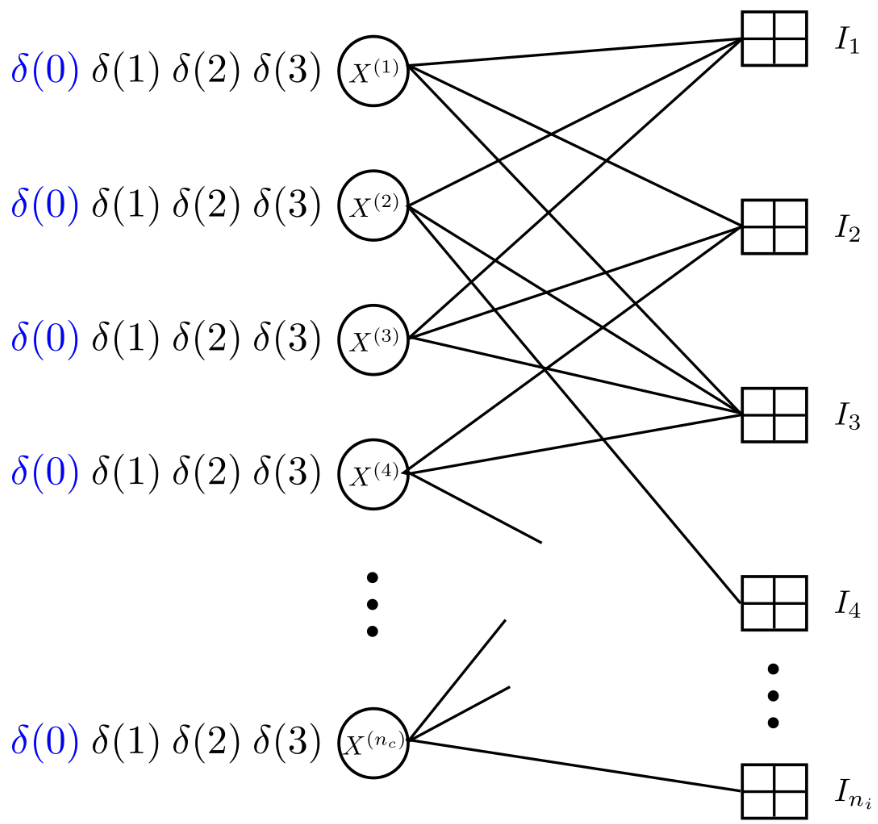
Inter-frame Coding: Peeling Decoder

$m = 3$



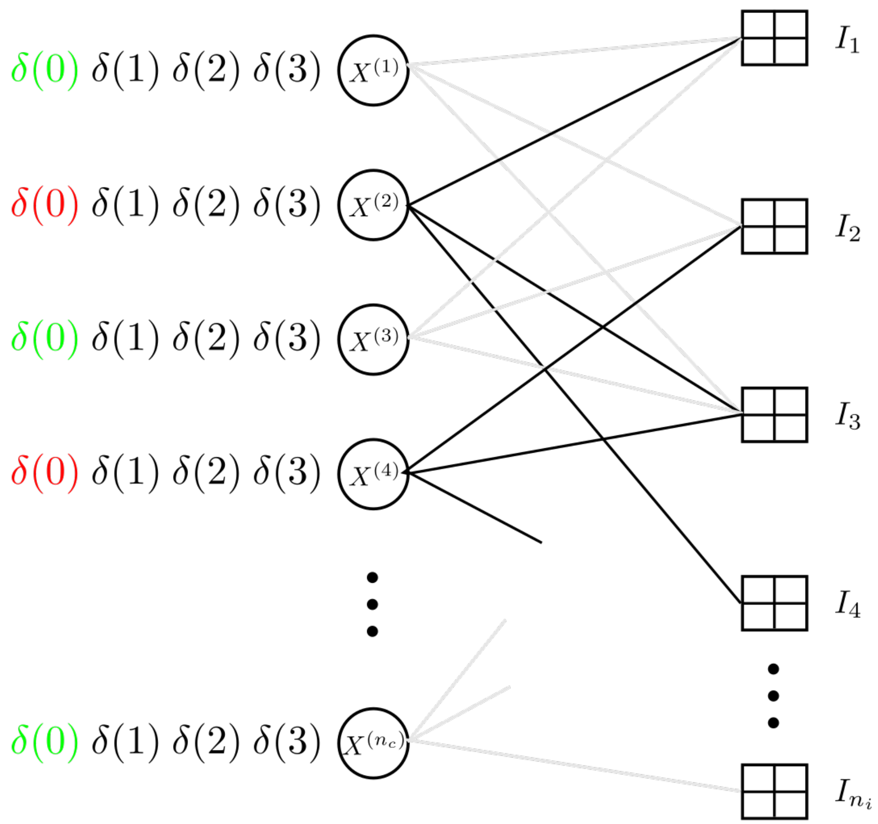
Inter-frame Coding: Peeling Decoder

Attempting decoding
Decoding Succeeded
Decoding Failed



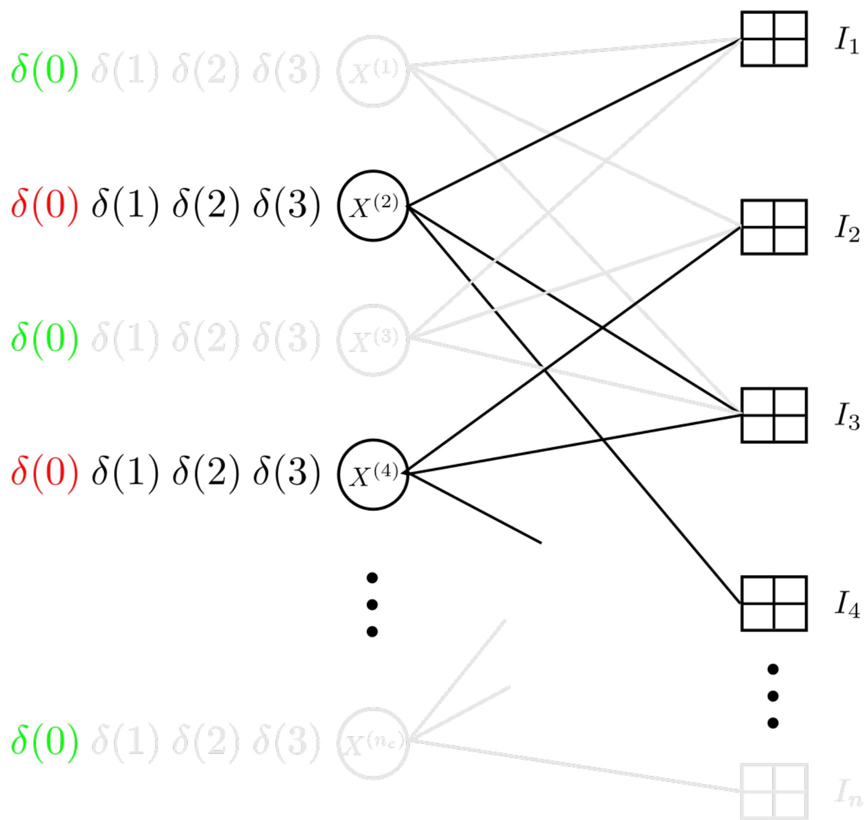
Inter-frame Coding: Peeling Decoder

Attempting decoding
Decoding Succeeded
Decoding Failed



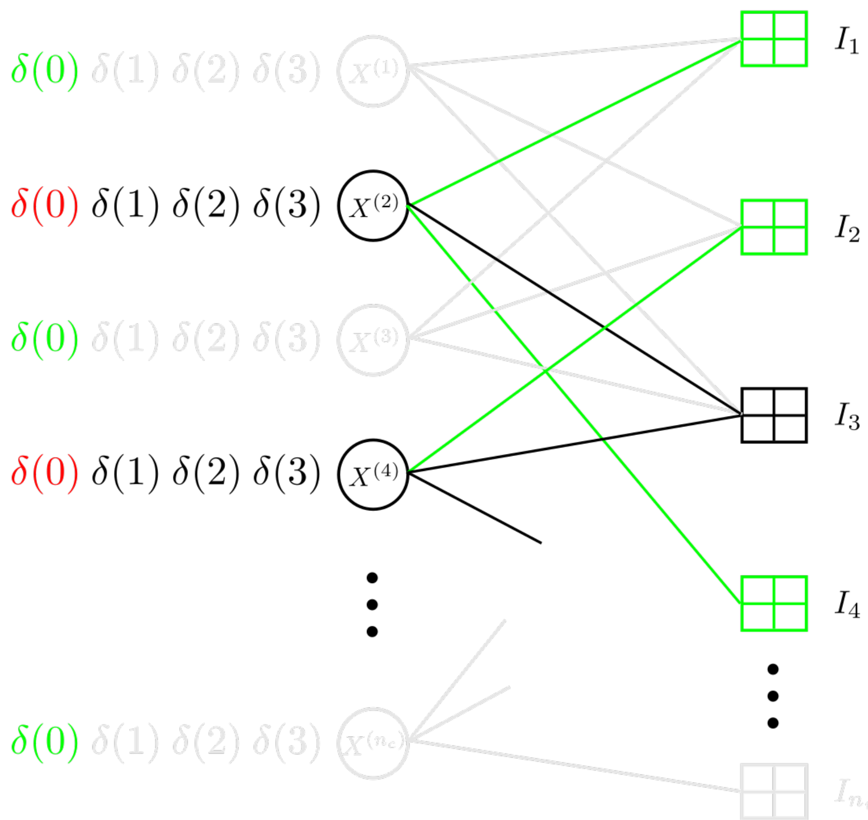
Inter-frame Coding: Peeling Decoder

Attempting decoding
Decoding Succeeded
Decoding Failed



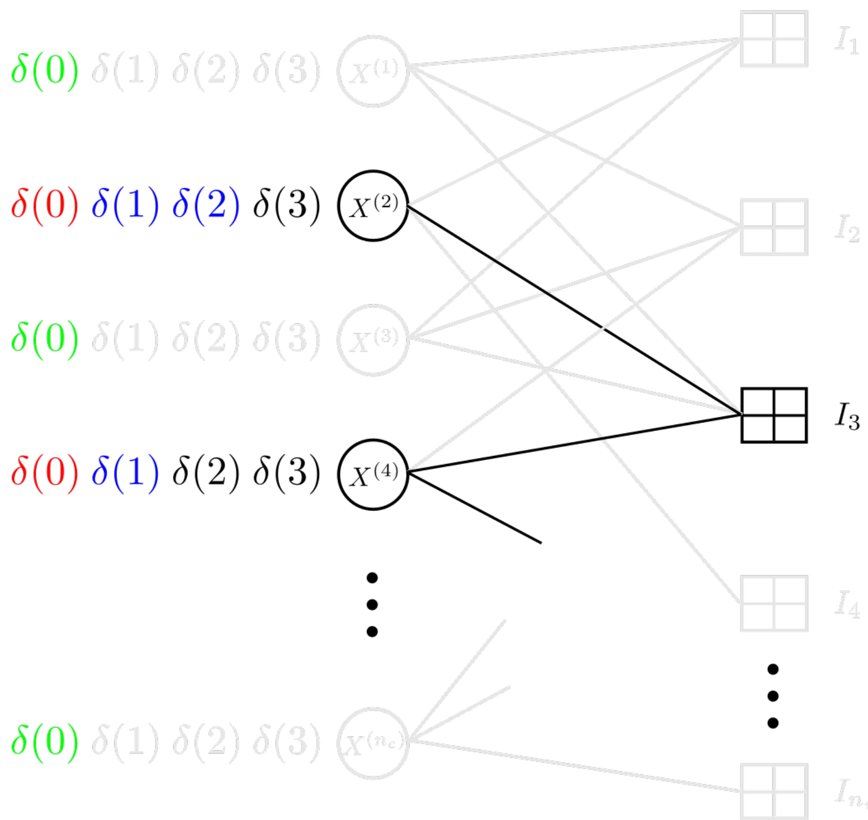
Inter-frame Coding: Peeling Decoder

Attempting decoding
Decoding Succeeded
Decoding Failed



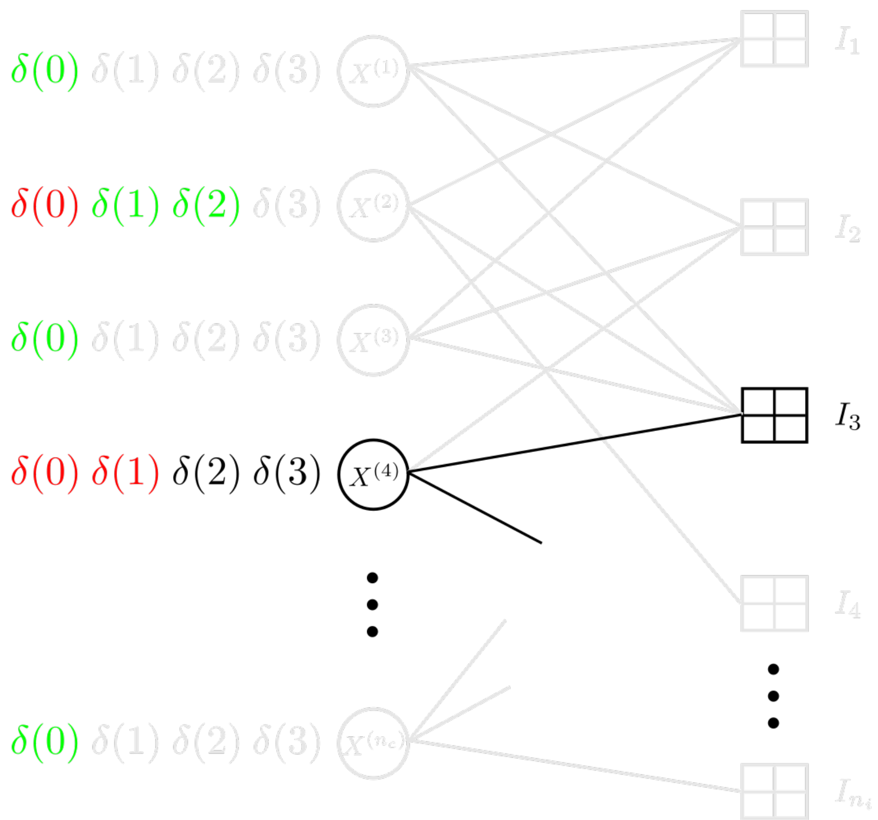
Inter-frame Coding: Peeling Decoder

Attempting decoding
Decoding Succeeded
Decoding Failed



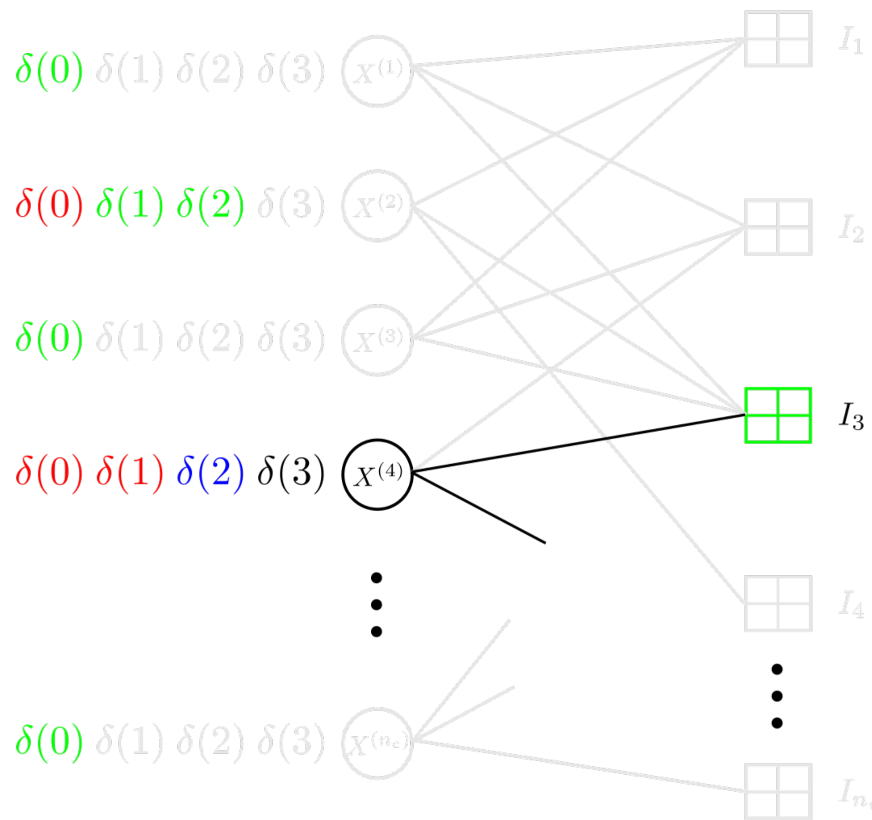
Inter-frame Coding: Peeling Decoder

Attempting decoding
Decoding Succeeded
Decoding Failed



Inter-frame Coding: Peeling Decoder

Attempting decoding
Decoding Succeeded
Decoding Failed



Stop Feedback Summary

- For short messages, stop feedback dramatically increases achievable rate.
- This is because it allows a variable-length code that stops when a reliable decision is available.
- The theoretical analysis assumes we check accumulated information density for every codeword after every symbol
- But real systems can achieve similar performance by checking decoder reliability at a few key times determined by sequential differential optimization.
- You can also use many variable-length short codes in parallel and use linear combinations of increments and a peeling decoder to avoid using feedback

UCLA Samueli School of Engineering

Act 3:
Full feedback further increases the rate of short messages that achieve a target reliability

Full Feedback

- Stop Feedback increases achievable rate by stopping transmission as soon as possible
- Full feedback also does this, but *additionally* increases achievable rate by avoiding telling the receiver things that it already knows.

UCLA Samueli School of Engineering

Horstein's Full Feedback Algorithm for the BSC

136 IEEE TRANSACTIONS ON INFORMATION THEORY July

Sequential Transmission Using Noiseless Feedback*

MICHAEL HORSTEIN†, MEMBER, IEEE

Summary—The presence of a feedback channel makes possible a variety of sequential transmission procedures, each of which can be classified as either a block-transmission or a continuous-transmission scheme according to the way in which information is encoded for transmission over a noisy forward channel. A sequential continuous-transmission system employing a binary symmetric forward channel (but which is suitable for use with any discrete memoryless forward channel) and a noiseless feedback channel is described. Its error exponent is shown to be substantially greater than the optimum block-code error exponent at all transmission rates less than channel capacity. The average value and the first-order probability distribution of the effective constraint length, found by simulating the system on an IBM 709 computer, are also given.

INTRODUCTION

MOST OF THE existing coding schemes designed to improve the reliability of digital communication envision a channel which allows for the transmission of information in one direction only. The block length or constraint length of these codes is a fixed number. Many communication systems, on the other hand, have transmission facilities in both directions between a pair of terminals. This makes it possible to insert, in the …ward channel, redundancy based on information … results of previous transmissions in the … information is made available … *feedback* channel. A wide … procedures is thereby … a procedure, … symbols …

carding scheme" which is conceptually quite simple. When the feedback channel is error-free, the error probabilities decay to zero at an exponential rate; otherwise they converge to constant values. This scheme, which requires that one of the channel symbols be interpreted to mean, "Erase the last received symbol," can be used to transmit at capacity only under very special circumstances—namely, when there is no "residual information"[2] in an erased symbol. It cannot be used at all with a binary channel.

As a second example, Wozencraft[3] has proposed the use of his sequential decoding scheme in a two-way communication system. The efficiency of the procedure in each direction is increased by the periodic injection of symbols in the opposite direction which indicate whether the decoding is proceeding at a satisfactory rate. If it is not, the transmission is repeated from a point which guarantees the retransmission of the troublesome portion of the received sequence.

In two-way communication, as in one-way systems, the various methods of encoding information for transmission over a noisy channel may be divided into two general types. Suppose, for the sake of argument, that the information is in the form of a binary sequence of 0's and 1's, the two digits being equally likely in every position; that is, each digit of the sequence carries one bit of information. (If the information is not in this form, converting it to a binary sequence of independent digits is a useful intermediate step in matching the information source to the channel.) We may choose to divide the …tion sequence into blocks of fixed length and …ence of channel symbols corresponding to …mitted sequence may be of variable …f its symbols may depend on …mitter concerning the …ocedure will be …we may …

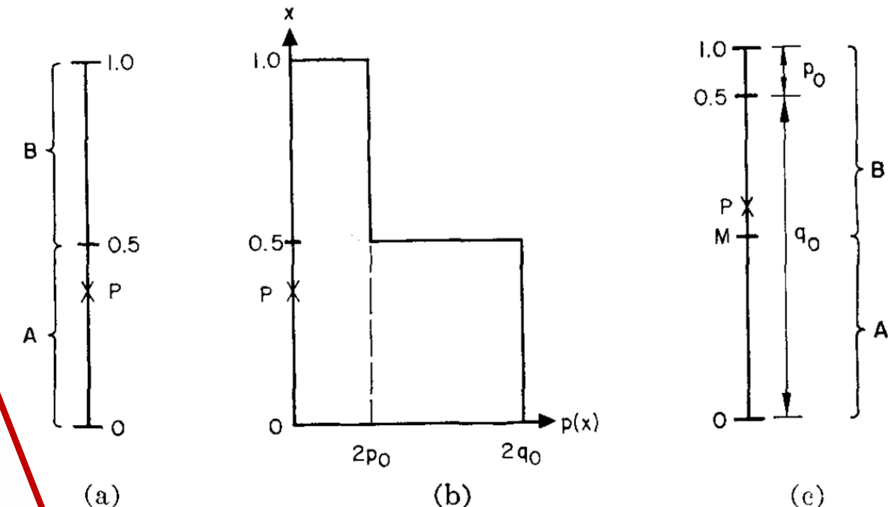


Fig. 2—Selection of the transmitted symbol.

From Horstein Trans. IT 1963

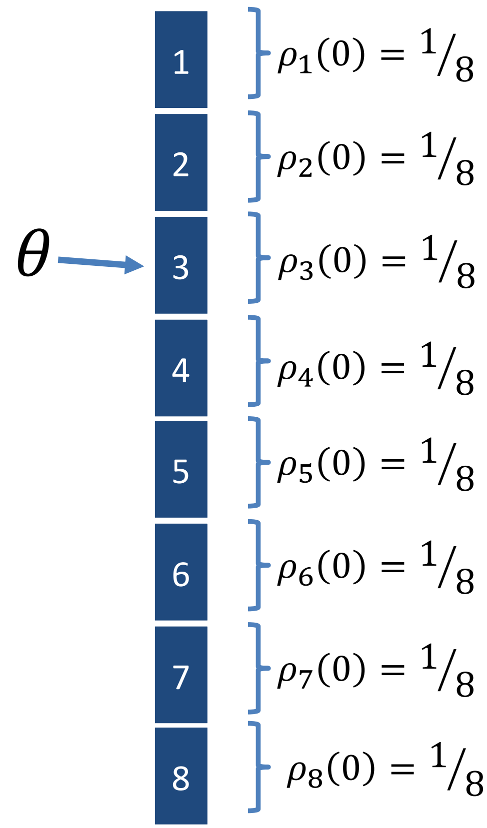
Make sure each transmitted bit is conveying a full bit of useful information in light of what the receiver already knows.

UCLA Samueli School of Engineering

Finite blocklength sequential transmission

AN EIGHT-MESSAGE EXAMPLE ON THE BSC

- There are $M = 2^k$ messages $\{1,2,\cdots,M\}$.
- The true message is θ.
- The original prior probabilities are all $\rho_i(0) = 2^{-k}$.



Finite blocklength sequential transmission

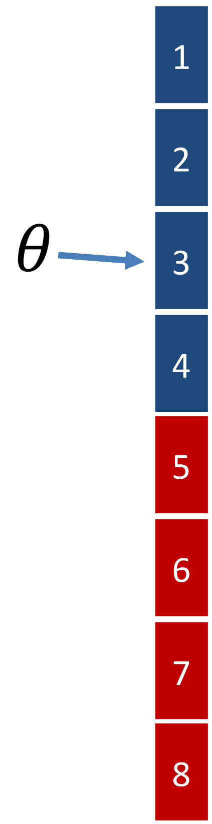
AN EIGHT-MESSAGE EXAMPLE ON THE BSC

- There are $M = 2^k$ messages $\{1, 2, \cdots, M\}$.
- The true message is θ.
- The original prior probabilities are all $\rho_i(0) = 2^{-k}$.

Finite blocklength sequential transmission

AN EIGHT-MESSAGE EXAMPLE ON THE BSC

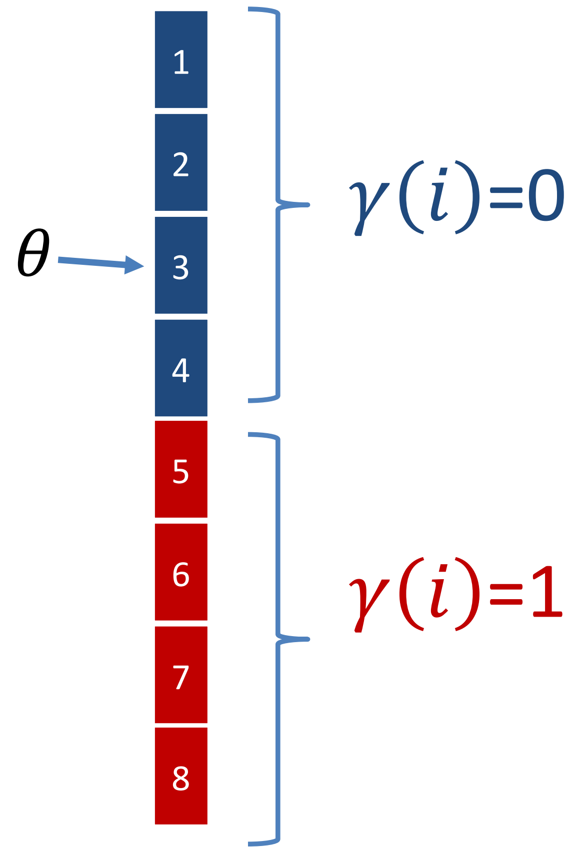
- There are $M = 2^k$ messages $\{1,2,\cdots,M\}$.
- The true message is θ.
- The original prior probabilities are all $\rho_i(0) = 2^{-k}$.
- Divide the interval into two sets, Set 0 and Set 1, Each having probability ½.



Finite blocklength sequential transmission

AN EIGHT-MESSAGE EXAMPLE ON THE BSC

- There are $M = 2^k$ messages $\{1,2,\cdots,M\}$.
- The true message is θ.
- The original prior probabilities are all $\rho_i(0) = 2^{-k}$.
- Divide the interval into two sets, Set 0 and Set 1, Each having probability ½.
- Transmit $\gamma(\theta)$, which is a 0 if θ is in S_0 and transmit a 1 if θ is in S_1.



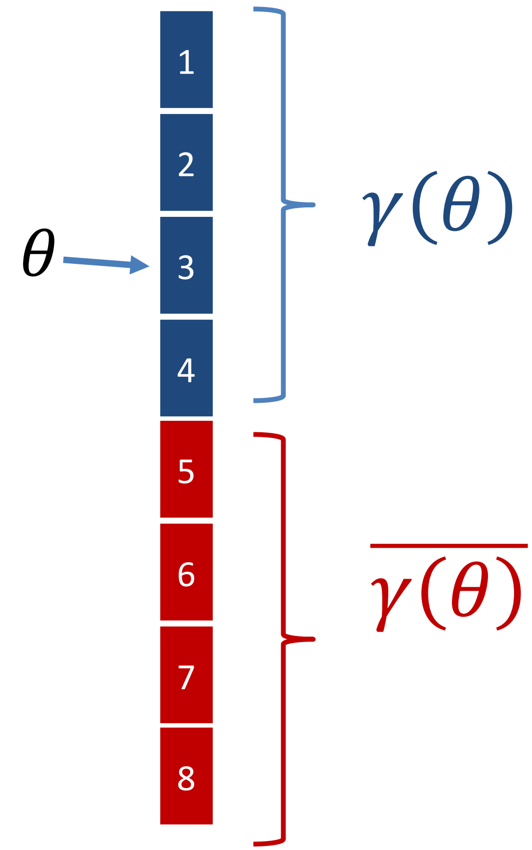
Computing the posterior $\rho_i(n+1)$ from $\rho_i(n)$

- It's a BSC with probability of error p so

$$Y = \begin{cases} \gamma(\theta) \text{ with probability } q = 1 - p \\ \overline{\gamma(\theta)} \text{ with probability } p. \end{cases}$$

- The posteriors are updated as follows:

$$\rho_i(n+1) = \begin{cases} \dfrac{\rho_i(n) P(Y = \gamma(i) | i)}{P(Y = \gamma(i))} & \text{when } Y = \gamma(i) \\[2ex] \dfrac{\rho_i(n) P\left(Y = \overline{\gamma(i)} | i\right)}{P(Y = \overline{\gamma(i)})} & \text{when } Y = \overline{\gamma(i)} \end{cases}$$



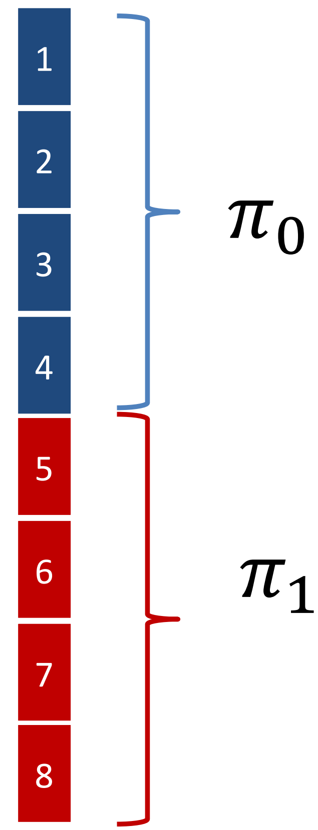
Computing marginal output probabilities $P(Y)$

MARGINAL OUTPUT PROBABILITIES FOR THE BSC

- Define the marginal probabilities of S_0 and S_1.
- $\pi_0 = P(S_0)$
- $\pi_1 = P(S_1)$

- It's a BSC with probability of error p so

$$P(Y) = \begin{cases} \pi_0 \mathrm{q} + \pi_1 \mathrm{p} & \text{for } Y = 0 \\ \pi_0 \mathrm{p} + \pi_1 \mathrm{q} & \text{for } Y = 1 \end{cases}$$

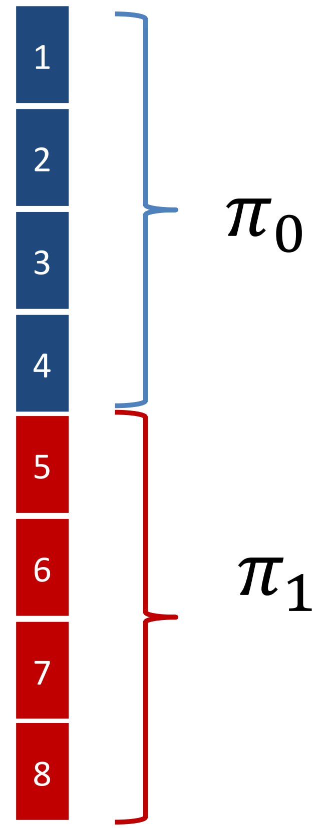


Computing the posterior $\rho_i(n+1)$ from $\rho_i(n)$

MARGINAL OUTPUT PROBABILITIES FOR THE BSC

- But if things are going well.
- $\pi_0 = {}^{1}/_{2}$
- $\pi_1 = {}^{1}/_{2}$

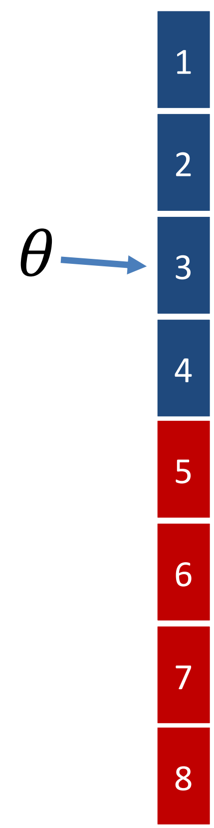
- $P(Y) = \begin{cases} {}^{1}/_{2} & \text{for } Y = 0 \\ {}^{1}/_{2} & \text{for } Y = 1 \end{cases}$



Computing the posterior $\rho_i(n+1)$ from $\rho_i(n)$

- The posteriors are updated as follows:

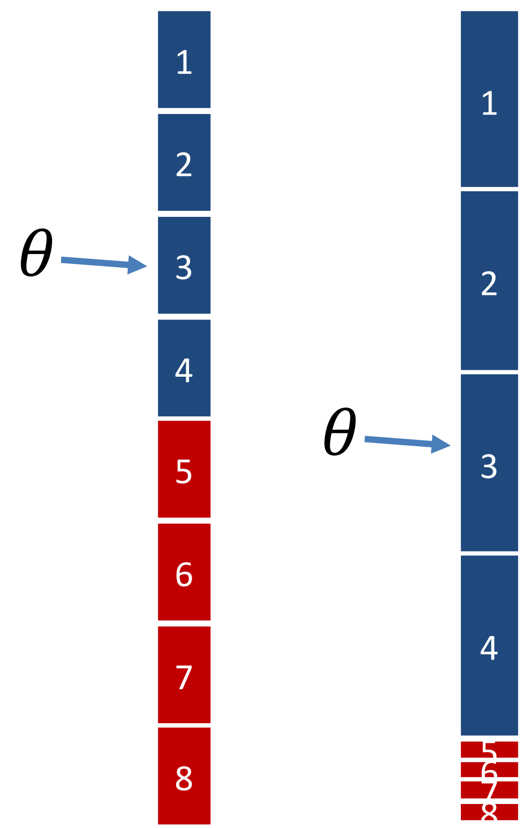
$$\rho_i(n+1) = \begin{cases} \dfrac{\rho_i(n)q}{P(Y = \gamma(i))} & \text{when } Y = \gamma(i) \\ \dfrac{\rho_i(n)p}{P(Y = \overline{\gamma(i)})} & \text{when } Y = \overline{\gamma(i)} \end{cases}$$



Computing the posterior $\rho_i(n+1)$ from $\rho_i(n)$

- The posteriors are updated as follows:

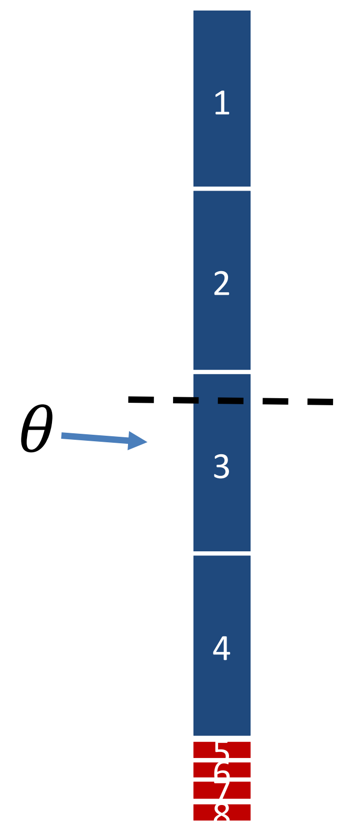
$$\rho_i(n+1) = \begin{cases} \rho_i(n)2q & \text{when } Y = \gamma(i) \\ \rho_i(n)2p & \text{when } Y = \overline{\gamma(i)} \end{cases}$$



Horstein's assignment* to S_0 and S_1

SOMETIMES THE MEDIAN LINE DIVIDES A MESSAGE

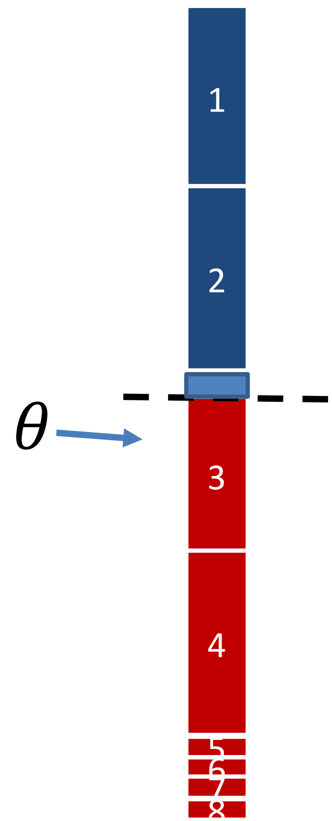
In this case Hornstein performs random encoding



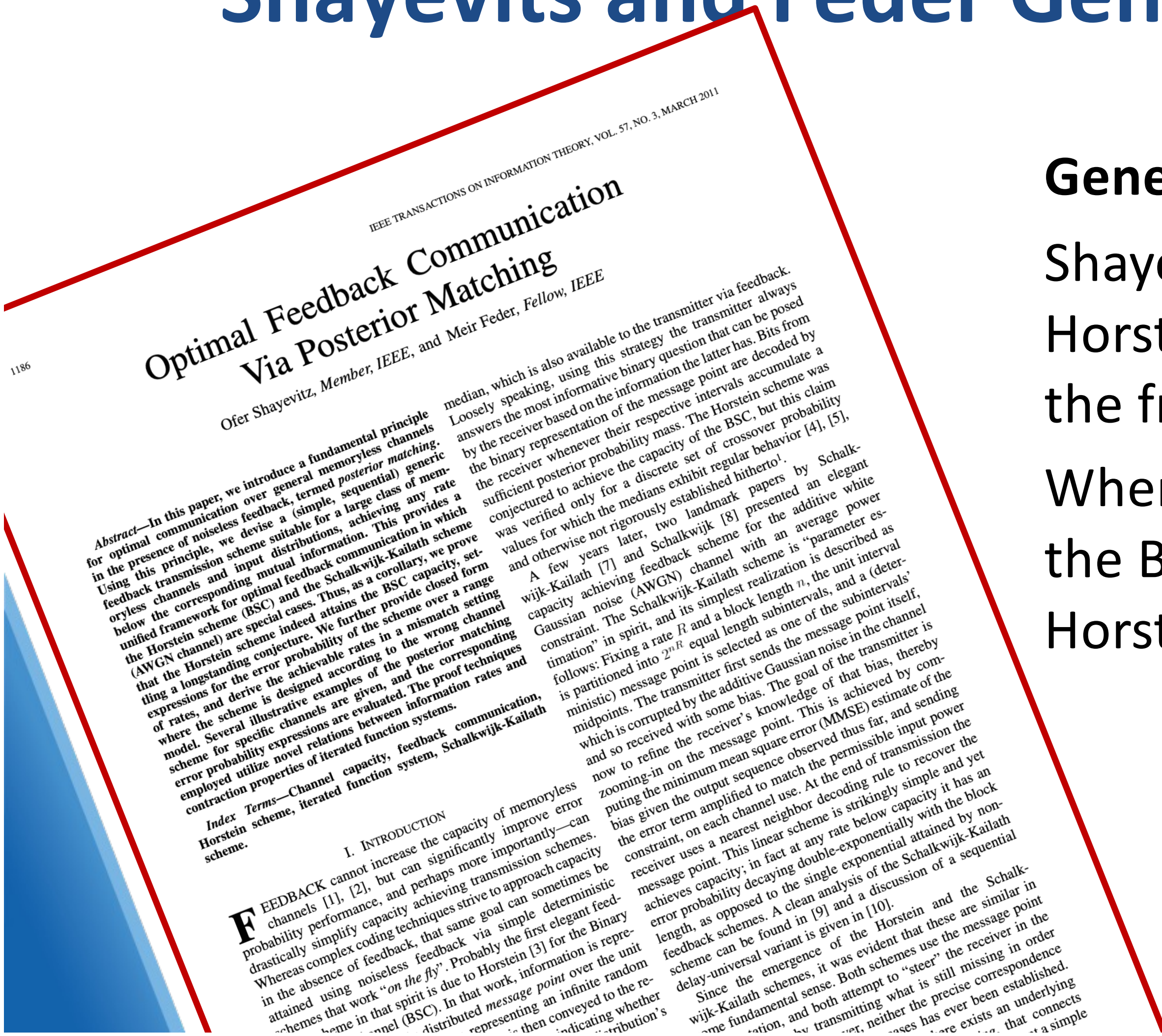
Horstein's assignment* to S_0 and S_1

SOMETIMES THE MEDIAN LINE DIVIDES A MESSAGE

In this case Hornstein performs random encoding



Shayevits and Feder Generalize Horstein



1186 IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 57, NO. 3, MARCH 2011

Optimal Feedback Communication Via Posterior Matching

Ofer Shayevitz, *Member, IEEE*, and Meir Feder, *Fellow, IEEE*

Abstract—In this paper, we introduce a fundamental principle for optimal communication over general memoryless channels in the presence of noiseless feedback, termed *posterior matching*. Using this principle, we devise a (simple, sequential) generic feedback transmission scheme suitable for a large class of memoryless channels and input distributions, achieving any rate below the corresponding mutual information. This provides a unified framework for optimal feedback communication in which the Horstein scheme (BSC) and the Schalkwijk-Kailath scheme (AWGN channel) are special cases. Thus, as a corollary, we prove that the Horstein scheme indeed attains the BSC capacity, settling a longstanding conjecture. We further provide closed form expressions for the error probability of the scheme over a range of rates, and derive the achievable rates in a mismatch setting where the scheme is designed according to the wrong channel model. Several illustrative examples of the posterior matching scheme for specific channels are given, and the corresponding error probability expressions are evaluated. The proof techniques employed utilize novel relations between information rates and contraction properties of iterated function systems.

Index Terms—Channel capacity, feedback communication, Horstein scheme, iterated function system, Schalkwijk-Kailath scheme.

I. Introduction

FEEDBACK cannot increase the capacity of memoryless channels [1], [2], but can significantly improve error probability performance, and perhaps more importantly—can drastically simplify capacity achieving transmission schemes. Whereas complex coding techniques strive to approach capacity in the absence of feedback, that same goal can sometimes be attained using noiseless feedback via simple deterministic schemes that work "*on the fly*". Probably the first elegant feedback scheme in that spirit is due to Horstein [3] for the Binary ...channel (BSC). In that work, information is repre... distributed *message point* over the unit ... representing an infinite random ... is then conveyed to the re... indicating whether ... distribution's

median, which is also available to the transmitter via feedback. Loosely speaking, using this strategy the transmitter always answers the most informative binary question that can be posed by the receiver based on the information the latter has. Bits from the binary representation of the message point are decoded by the receiver whenever their respective intervals accumulate a sufficient posterior probability mass. The Horstein scheme was conjectured to achieve the capacity of the BSC, but this claim was verified only for a discrete set of crossover probability values for which the medians exhibit regular behavior [4], [5], and otherwise not rigorously established hitherto[1].

A few years later, two landmark papers by Schalkwijk-Kailath [7] and Schalkwijk [8] presented an elegant capacity achieving feedback scheme for the additive white Gaussian noise (AWGN) channel with an average power constraint. The Schalkwijk-Kailath scheme is "parameter estimation" in spirit, and its simplest realization is described as follows: Fixing a rate R and a block length n, the unit interval is partitioned into 2^{nR} equal length subintervals, and a (deterministic) message point is selected as one of the subintervals' midpoints. The transmitter first sends the message point itself, which is corrupted by the additive Gaussian noise in the channel and so received with some bias. The goal of the transmitter is now to refine the receiver's knowledge of that bias, thereby zooming-in on the message point. This is achieved by computing the minimum mean square error (MMSE) estimate of the bias given the output sequence observed thus far, and sending the error term amplified to match the permissible input power constraint, on each channel use. At the end of transmission the receiver uses a nearest neighbor decoding rule to recover the message point. This linear scheme is strikingly simple and yet achieves capacity; in fact at any rate below capacity it has an error probability decaying double-exponentially with the block length, as opposed to the single exponential attained by non-feedback schemes. A clean analysis of the Schalkwijk-Kailath scheme can be found in [9] and a discussion of a sequential delay-universal variant is given in [10].

Since the emergence of the Horstein and the Schalkwijk-Kailath schemes, it was evident that these are similar in some fundamental sense. Both schemes use the message point ...tation, and both attempt to "steer" the receiver in the ... by transmitting what is still missing in order ... ver, neither the precise correspondence ... cases has ever been established. ... there exists an underlying ... that connects ... a simple

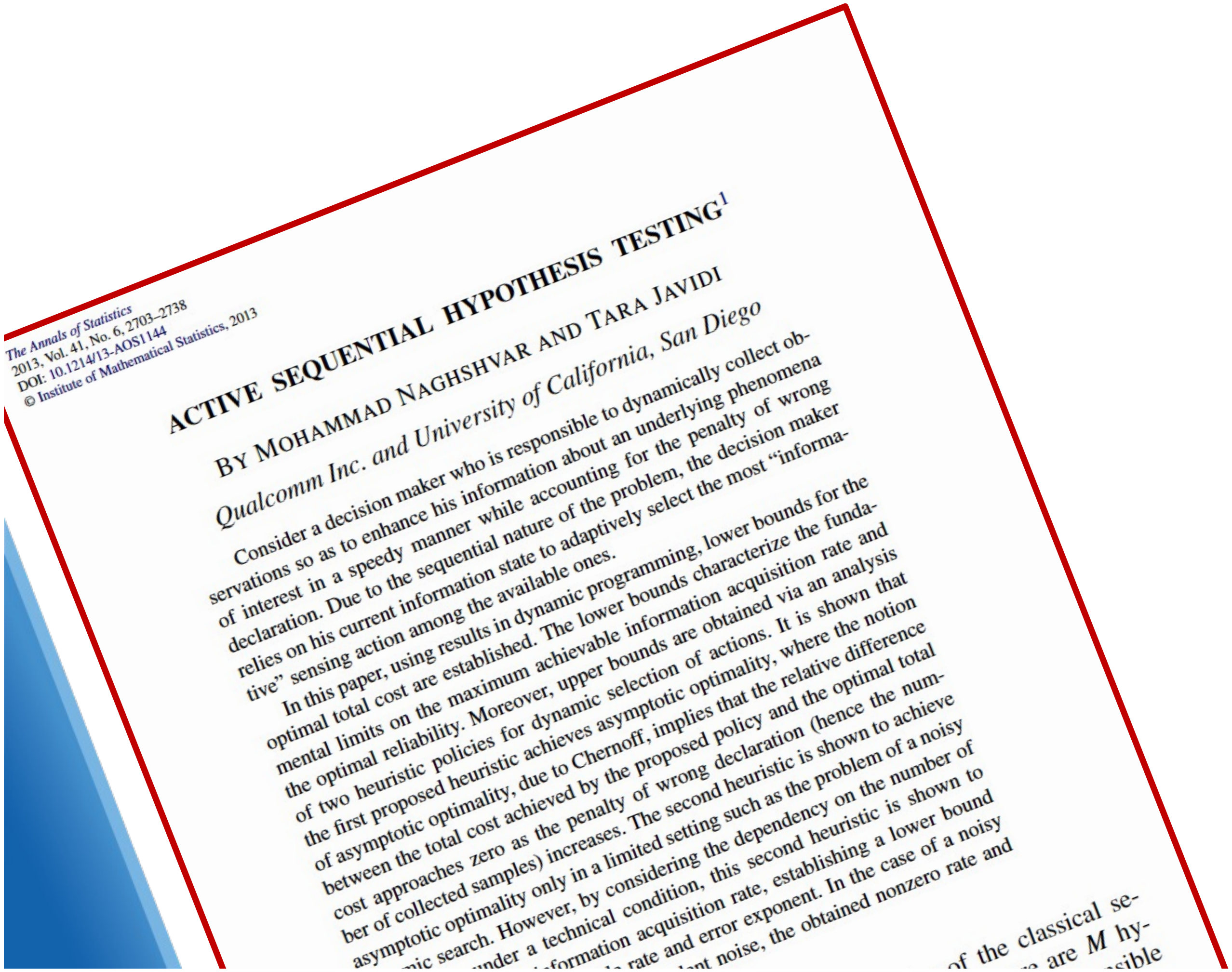
Posterior Matching

General and rigorous Treatment:

Shayevitz and Feder generalized Horstein's approach to any DMC using the framework of posterior matching.

When posterior matching is applied to the BSC, it naturally reduces to the Horstein scheme.

Active Sequential Hypothesis Testing (ASHT)


The Annals of Statistics
2013, Vol. 41, No. 6, 2703–2738
DOI: 10.1214/13-AOS1144

ACTIVE SEQUENTIAL HYPOTHESIS TESTING[1]

BY MOHAMMAD NAGHSHVAR AND TARA JAVIDI

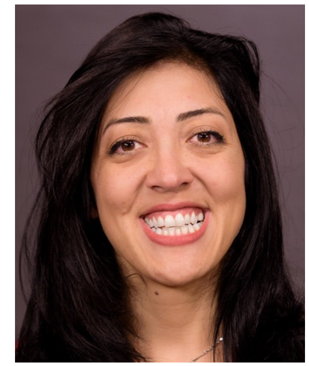
Qualcomm Inc. and University of California, San Diego

Consider a decision maker who is responsible to dynamically collect observations so as to enhance his information about an underlying phenomena of interest in a speedy manner while accounting for the penalty of wrong declaration. Due to the sequential nature of the problem, the decision maker relies on his current information state to adaptively select the most "informative" sensing action among the available ones.

In this paper, using results in dynamic programming, lower bounds for the optimal total cost are established. The lower bounds characterize the fundamental limits on the maximum achievable information acquisition rate and the optimal reliability. Moreover, upper bounds are obtained via an analysis of two heuristic policies for dynamic selection of actions. It is shown that the first proposed heuristic achieves asymptotic optimality, where the notion of asymptotic optimality, due to Chernoff, implies that the relative difference between the total cost achieved by the proposed policy and the optimal total cost approaches zero as the penalty of wrong declaration (hence the number of collected samples) increases. The second heuristic is shown to achieve asymptotic optimality only in a limited setting such as the problem of a noisy
mic search. However, by considering the dependency on the number of
under a technical condition, this second heuristic is shown to
information acquisition rate, establishing a lower bound
le rate and error exponent. In the case of a noisy
nt noise, the obtained nonzero rate and

of the classical se-
are M hy-
sible

Naghshvar

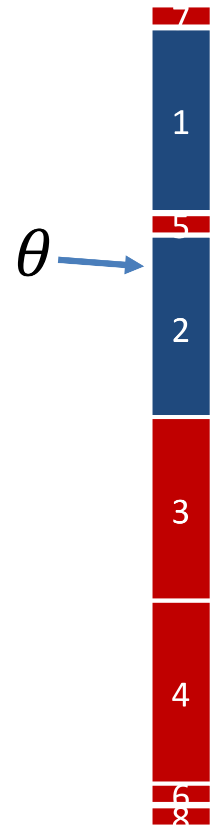


Javidi

Reordering (Naghshvar's MaxEJS)

BUT THERE ARE MANY SMALL MESSAGES

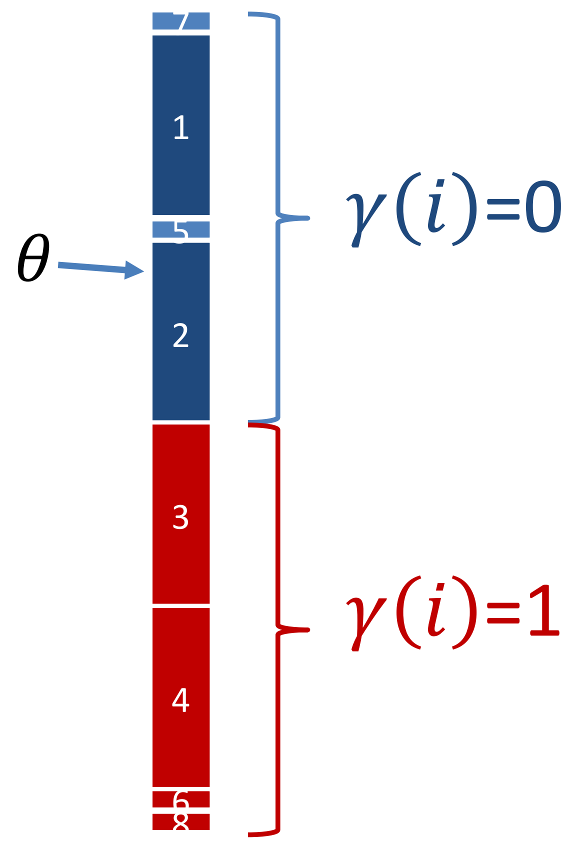
Naghshvar et al. simply move the messages around to allow a nice split right at (or very near to) ½.



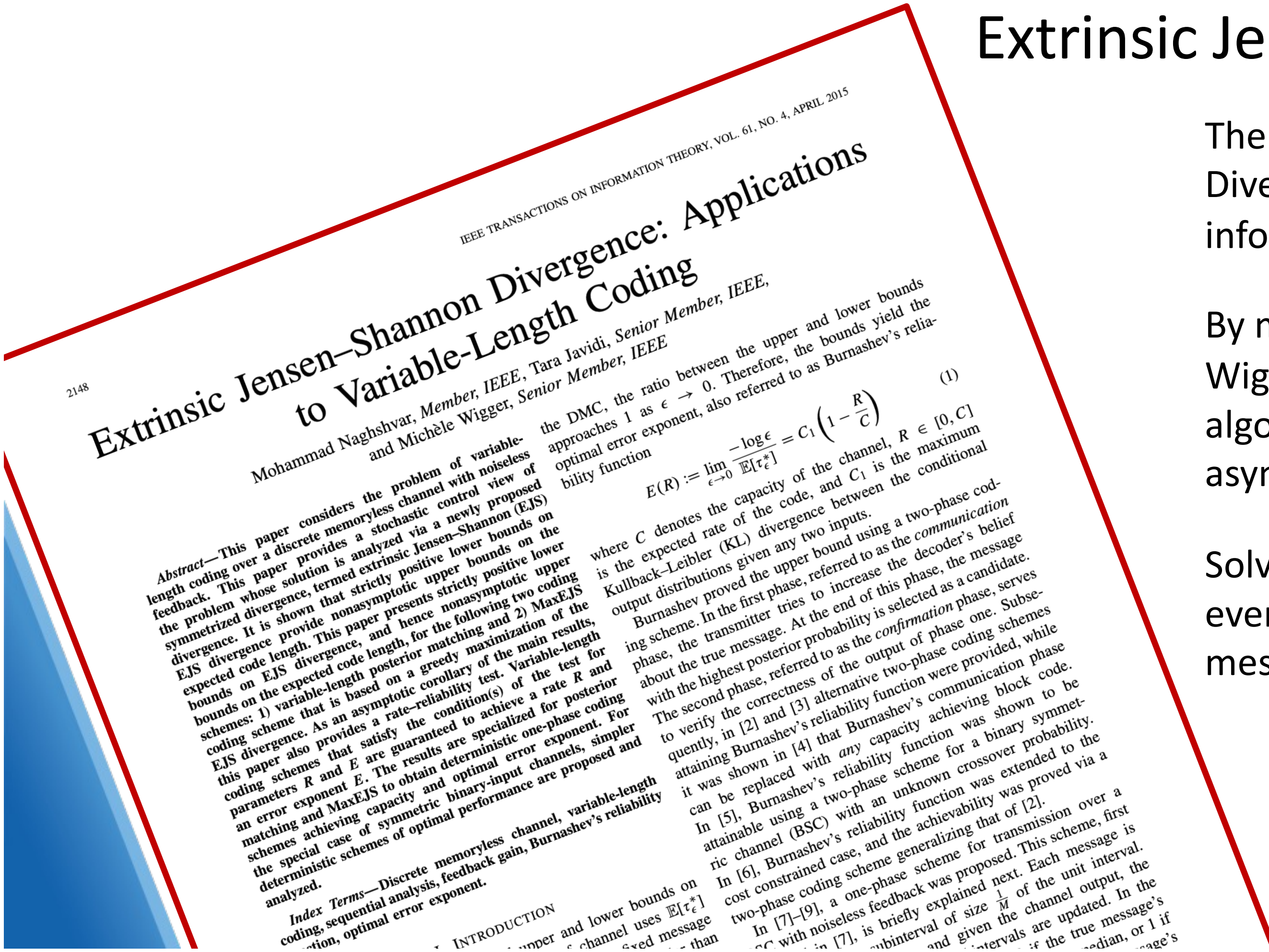
Reordering (Naghshvar's MaxEJS)

BUT THERE ARE MANY SMALL MESSAGES

Naghshvar et al. simply move the messages around to allow a nice split right at (or very near to) ½.



Finally, a non-asymptotic bound



Extrinsic Jensen-Shannon Divergence

The Extrinsic Jensen-Shannon (EJS) Divergence is a metric that indicates how informative a transmitted symbol will be.

By maximizing EJS, Naghshvar, Javidi, and Wigger provide a version of Horstein's algorithm that can be analyzed non-asymptotically

Solves the "coin-weighing" problem for every transmission by REORDERING the messages.

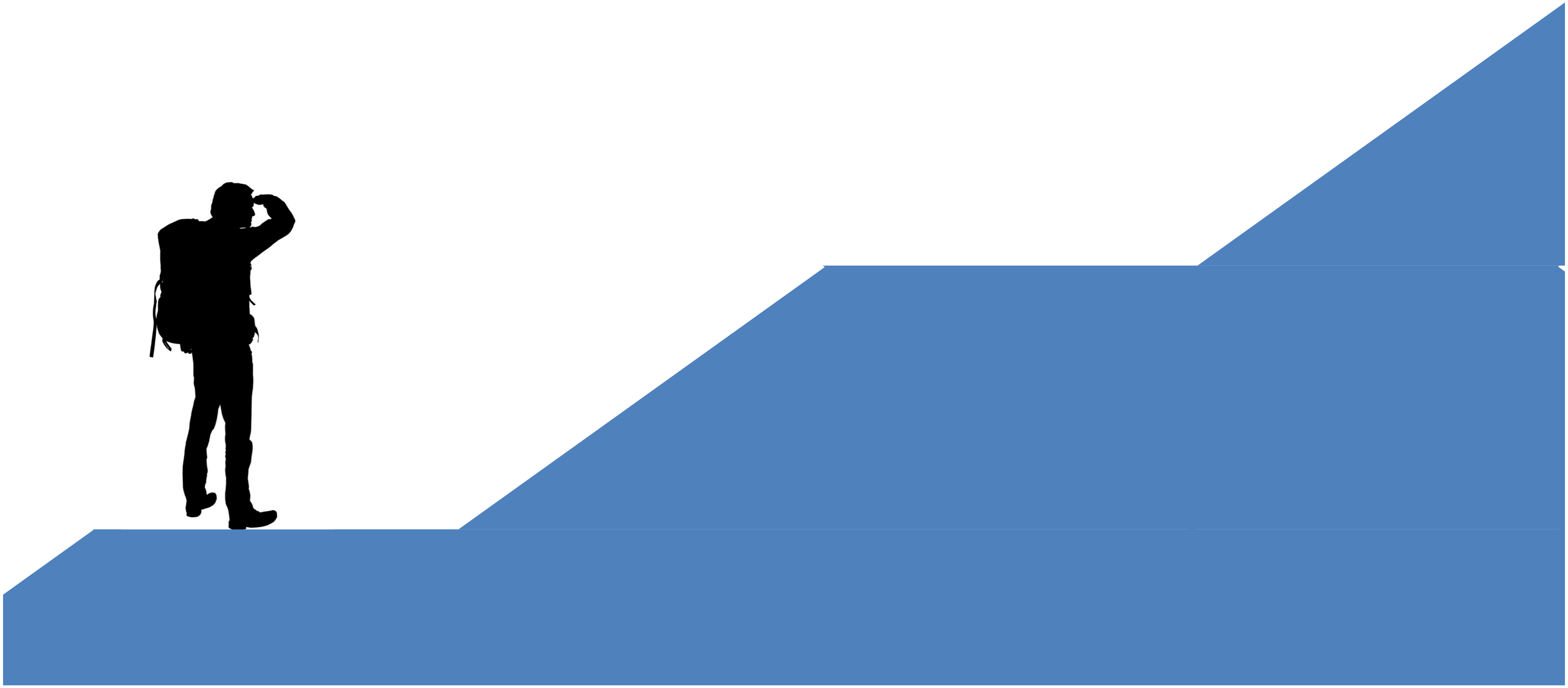
Finally, a non-asymptotic bound

Remark 4: By Theorem 1 and Proposition 2,

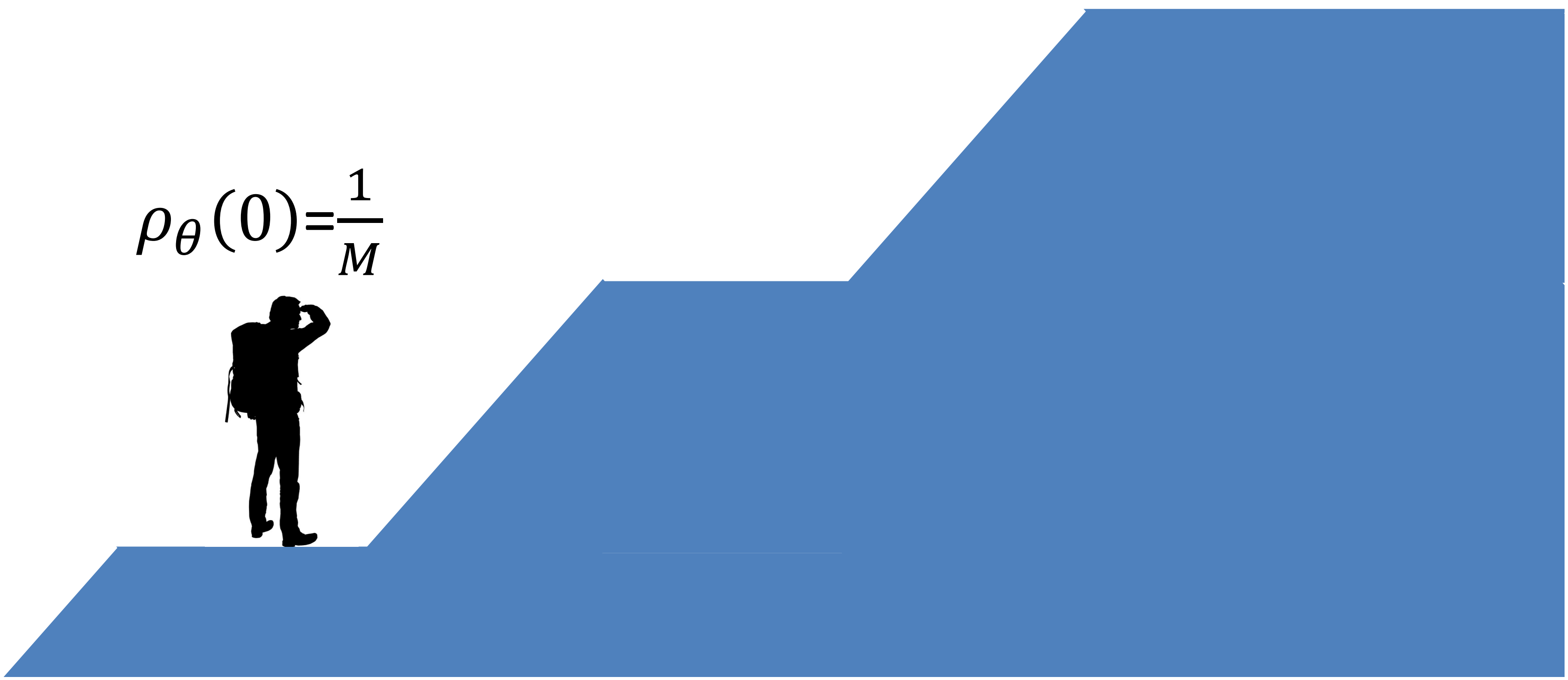
$$\mathbb{E}_{\Gamma^{\mathrm{MaxEJS}}}[\tilde{\tau}_\epsilon] \le \frac{\log M + \log\log\frac{M}{\epsilon}}{C} + \frac{\log\frac{1}{\epsilon}+1}{C_1} + \frac{6(4C_2)^2}{CC_1}, \qquad (40)$$

and thus MaxEJS encoding together with the decoding and stopping rules described in (26) and (28) achieves Burnashev's optimal asymptotic performance in (23), see Corollary 1.

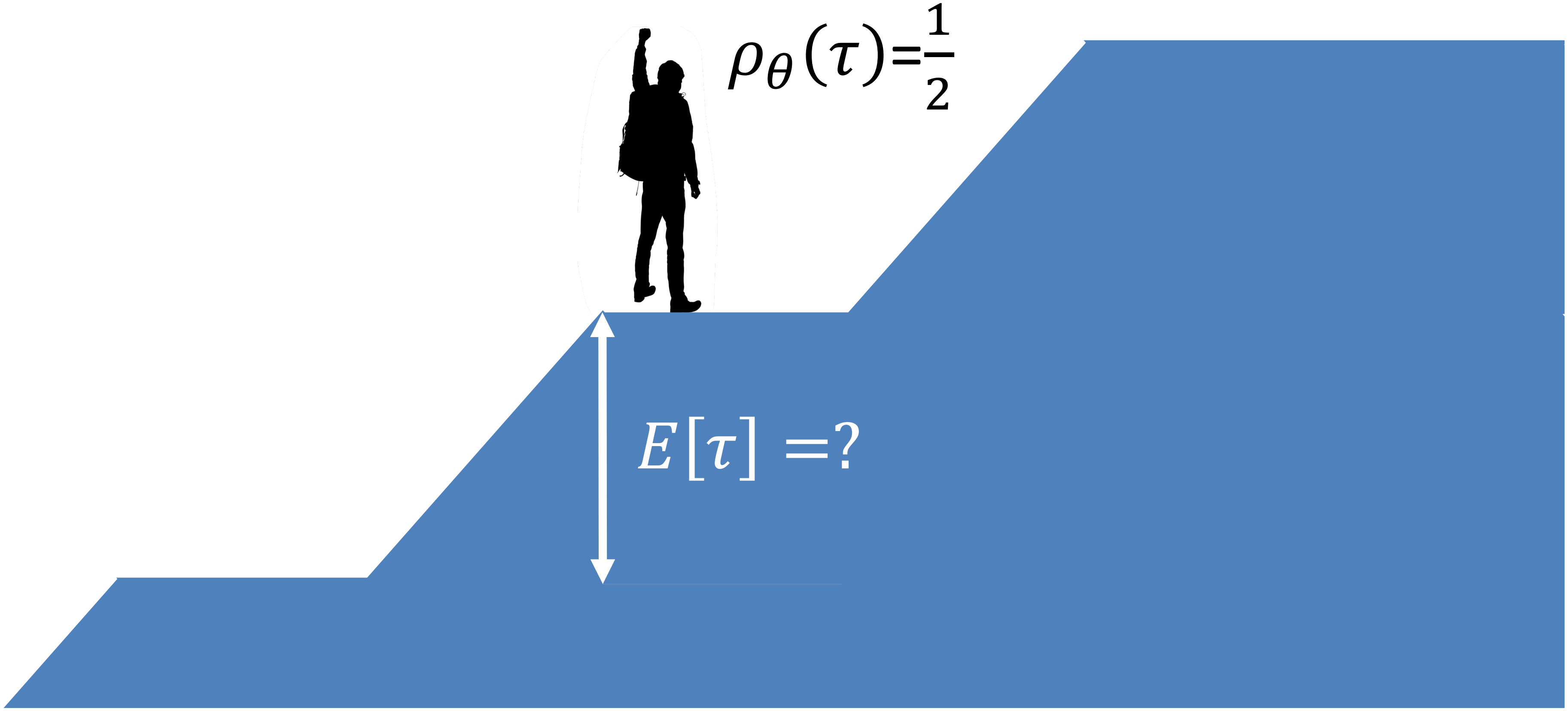
"Climbing Mount Posterior"



The Journey Begins…



How long to reach the Plateau of $\rho_\theta(\tau)=\frac{1}{2}$.

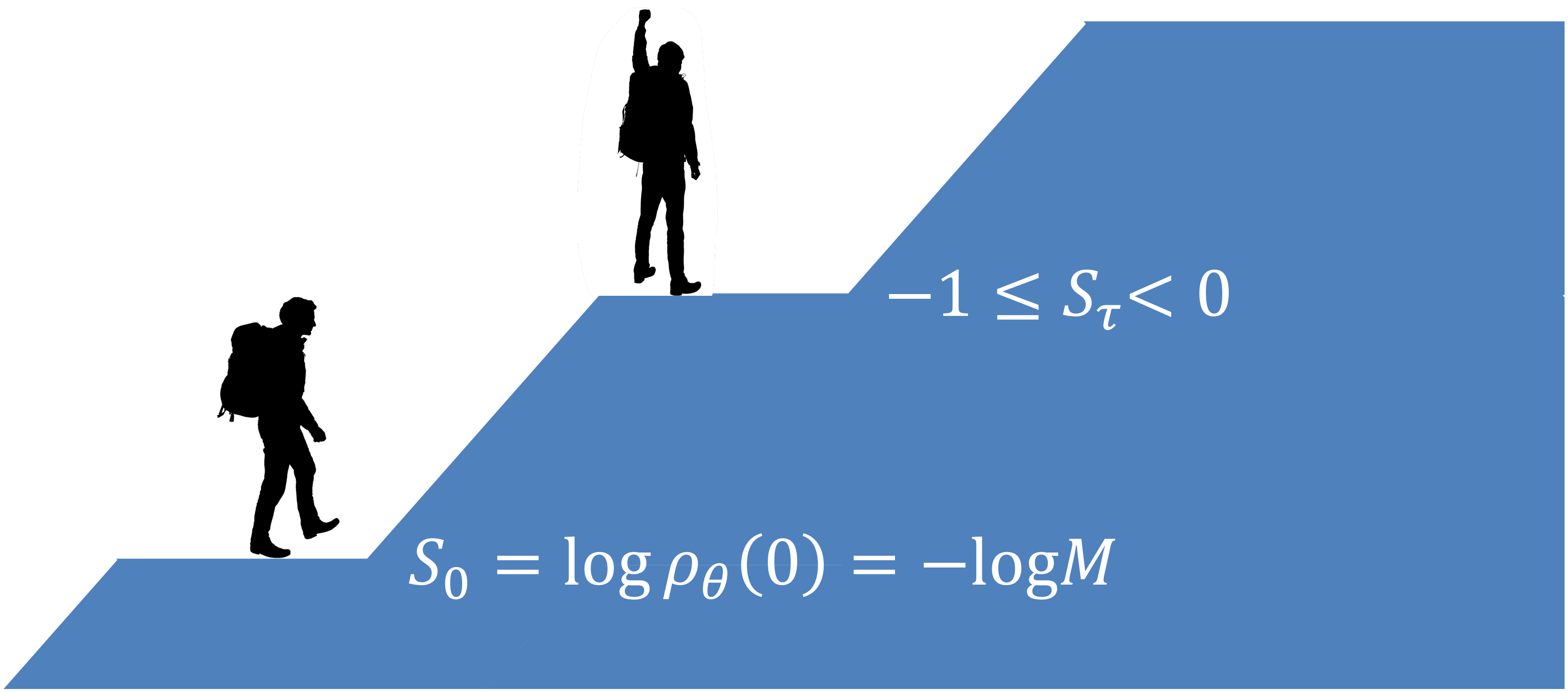


Random walk from $\rho_\theta(0)=\frac{1}{M}$ **to** $\rho_\theta(\tau) \geq \frac{1}{2}$

ASSUME (FOR NOW) THAT $\pi_0 = \pi_1 = {}^1/_2$

- Define a random walk based on the log of the posterior of the true message:
- $S_n = \log \rho_\theta(n)$
- $S_{n+1} = S_n + X$ where $X = \begin{cases} \log 2q & \text{with probability q} \\ \log 2p & \text{with probability } p \end{cases}$
- Consider the expected value of the step size:
- $EX = p \log 2p + q \log 2q$
- $EX = 1 - H(p)$
- $EX = C$

Random walk from $\rho_\theta(0)$=$\frac{1}{M}$ **to** $\rho_\theta(\tau) \geq \frac{1}{2}$



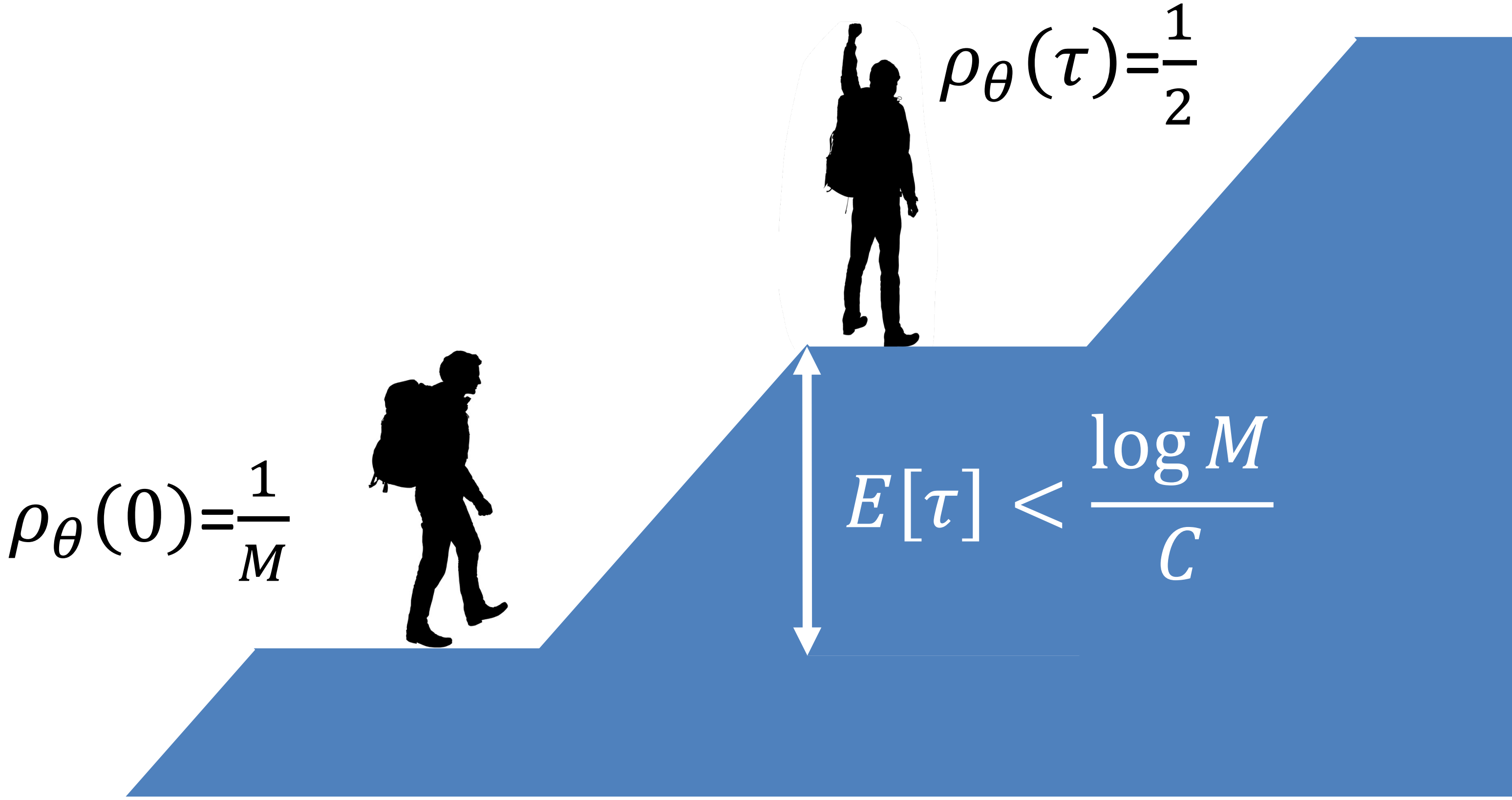
Wald's Equality

ABRAHAM WALD: $\boldsymbol{ES_\tau = S_0 + E\tau EX}$

- Using Wald's equality:
- $E\tau = \frac{ES_\tau - S_0}{EX}$
- $E\tau < \frac{0 + \log M}{C}$
- $E\tau < \frac{\log M}{C}$

Abraham Wald

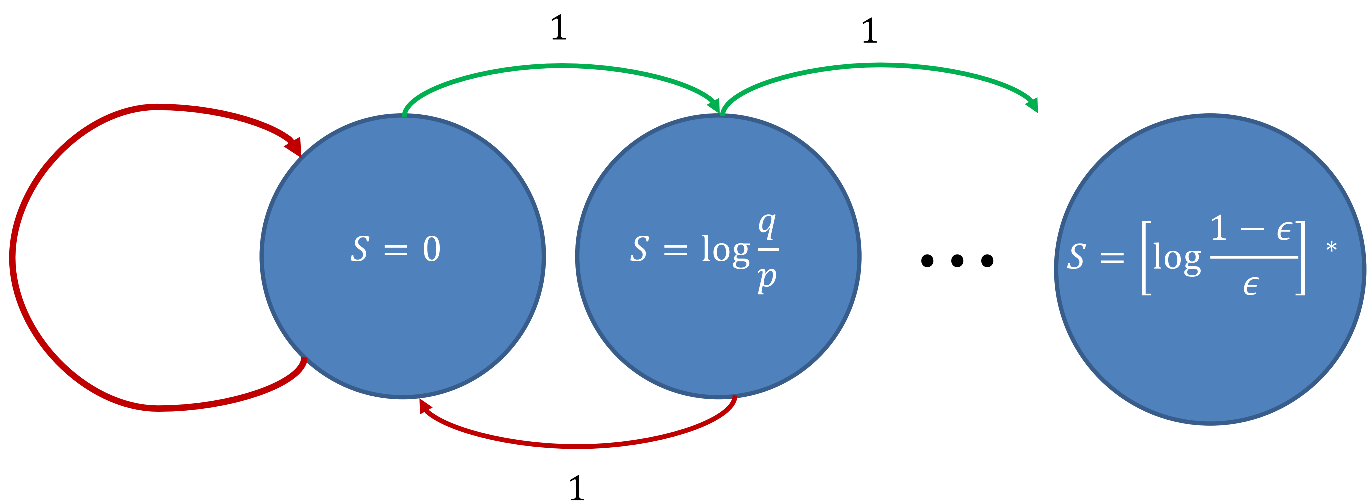
Random walk from $\rho_\theta(0)=\frac{1}{M}$ **to** $\rho_\theta(\tau) \geq \frac{1}{2}$



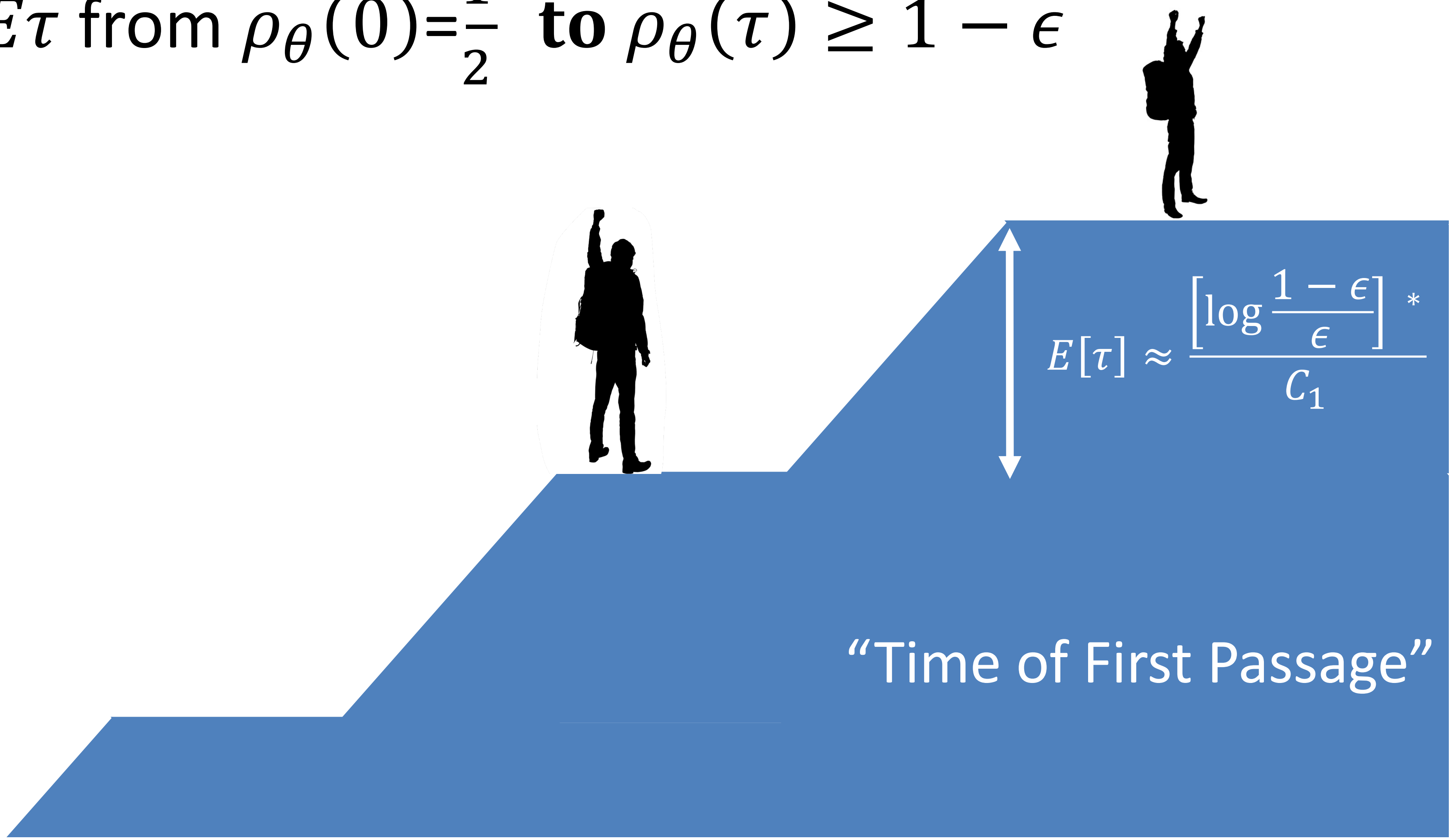
Abraham Wald

A Markov Chain

$$E\tau < 1 + \frac{\log \frac{1}{p}}{C}$$



$E\tau$ from $\rho_\theta(0) = \frac{1}{2}$ **to** $\rho_\theta(\tau) \geq 1 - \epsilon$



Amaael Antonini: Algorithm and Bounds for BSC

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 70, NO. 8, AUGUST 2024
5471

Achievable Rates for Low-Complexity Posterior Matching Over the Binary Symmetric Channel

Amaael Antonini, Student Member, IEEE, Rita Gimelshein, Student Member, IEEE, and Richard D. Wesel, Fellow, IEEE

Abstract—**Horstein, Burnashev, Shayevitz and Feder, Naghshvar et al. and others have studied sequential transmission of a k-bit message over the binary symmetric channel (BSC) with full, noiseless feedback using posterior matching. Yang et al. provide an improved lower bound on the achievable rate using martingale analysis that relies on the small-enough difference (SED) partitioning introduced by Naghshvar et al. SED requires a relatively complex encoder and decoder. To reduce complexity, this paper replaces SED with relaxed constraints that admit the small enough *absolute* difference (SEAD) partitioning rule. The main analytical results show that achievable-rate bounds higher than those found by Yang et al. (2021) are possible even under the new constraints, which are less restrictive than SED. The new analysis does not use martingale theory for the confirmation phase and applies a surrogate channel technique to tighten the results. An initial systematic transmission further increases the achievable rate bound. The simplified encoder associated with SEAD has a complexity below order $O(K^2)$ and allows simulations for message sizes of at least 1000 bits. For example, simulations achieve 99% of of the channel's 0.50-bit capacity with an average block size of 200 bits for a target codeword error rate of 10^{-3}.**

Index Terms—**Posterior matching, binary symmetric channel, noiseless feedback, random coding.**

I. INTRODUCTION

Source → θ → Encoder → X_t → BSC → Y_t → Decoder → $\hat{\theta}$; Y_{t-1}

Fig. 1. System diagram of a BSC with full, noiseless feedback.

The process terminates at stopping time $t = \tau$ when a reliability threshold is achieved, at which point the receiver computes an estimate $\hat{\theta} \in \Omega$ of θ from the received symbols $Y_1, Y_2, \ldots, Y_\tau$. The communication problem consists of obtaining a decoding estimate of θ at the smallest possible time index τ while keeping the error probability $\Pr(\hat{\theta} \neq \theta)$ bounded by a small threshold ϵ.

A. Background

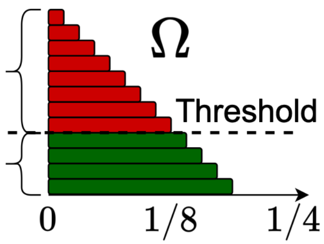
Shannon [3] showed that feedback cannot increase the capacity of discrete memoryless channels (DMC). However, when combined with variable-length coding, Burnashev [4] showed that feedback can help increase the frame error rate's (FER) decay rate as a function of blocklength. One such variable length coding method was pioneered by Horstein [5]. Horstein's sequential transmission scheme was presumed to achieve the capacity of the BSC, which was later proved by Shayevitz and Feder [6] showing that it satisfies the criteria of a *posterior matching scheme*. A *posterior matching (PM) scheme* was defined by Shayevitz and Feder as one that satisfies the two requirements of the *posterior matching principle*:

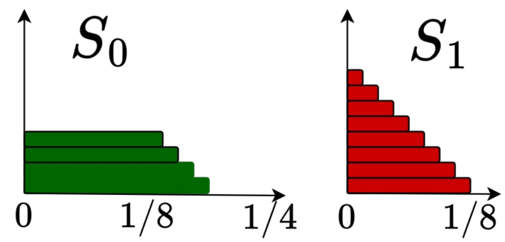
1) The input symbol at time $t+1$, X_{t+1}, is a fixed function of a random variable U, that is independent of the received $Y^t \triangleq \{Y_1, Y_2, \ldots, Y_t\}$; and

Amaael Antonini

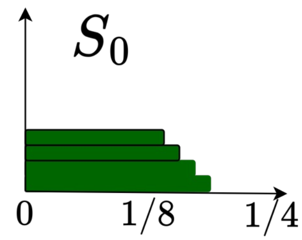
UCLA Samueli School of Engineering

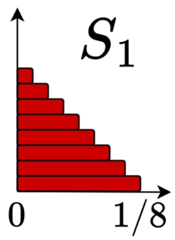
Amaael Antonini's Algorithm



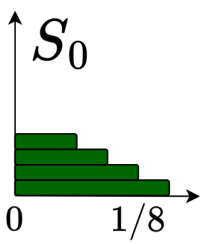


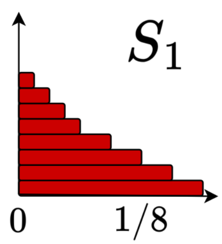
Amaael Antonini's Algorithm

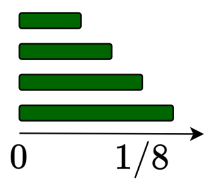


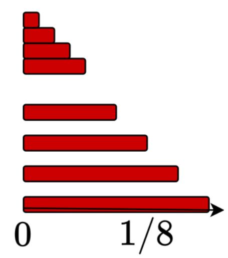


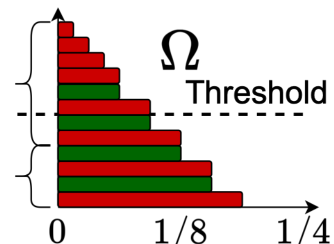
$Y = 1$



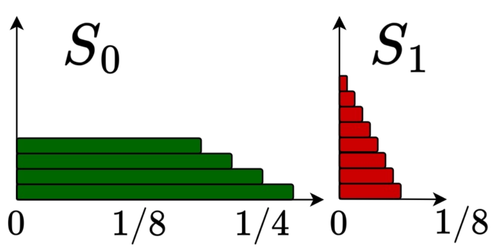


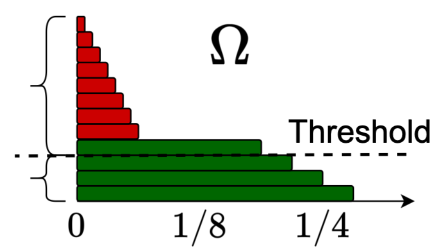




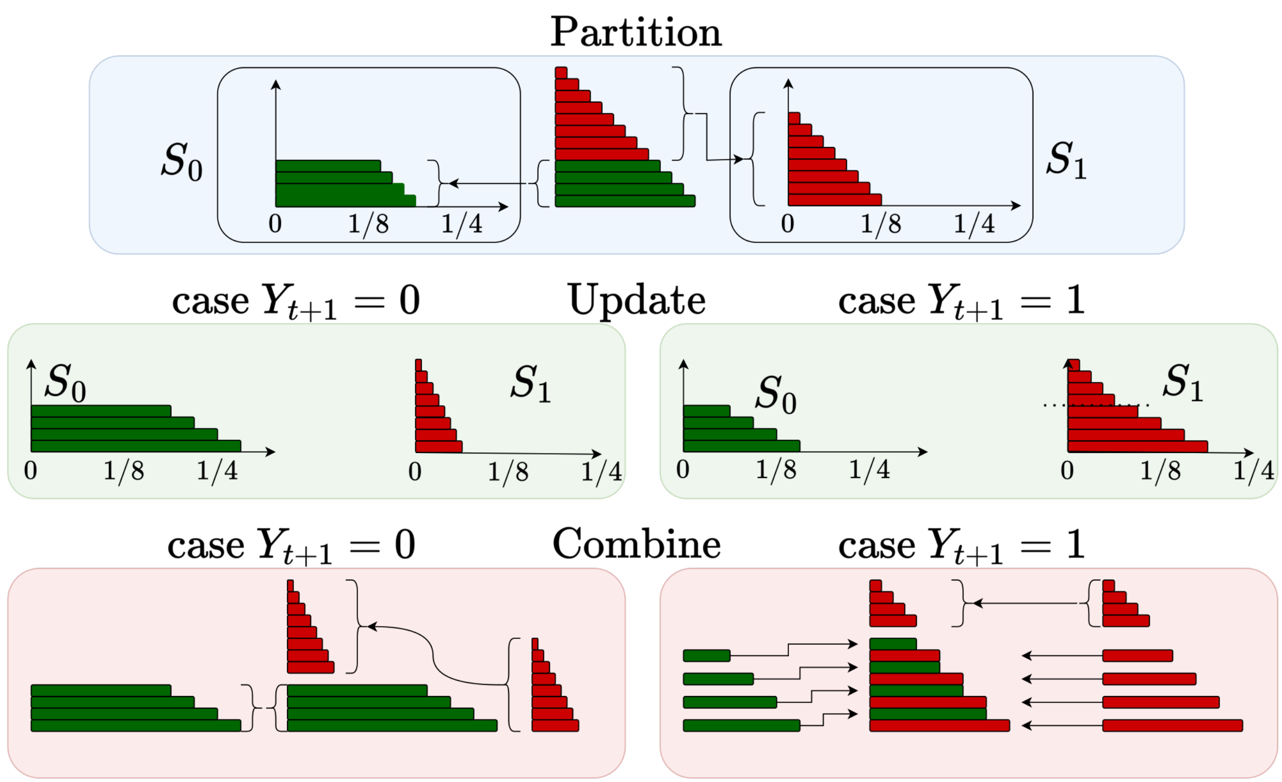


$Y = 0$

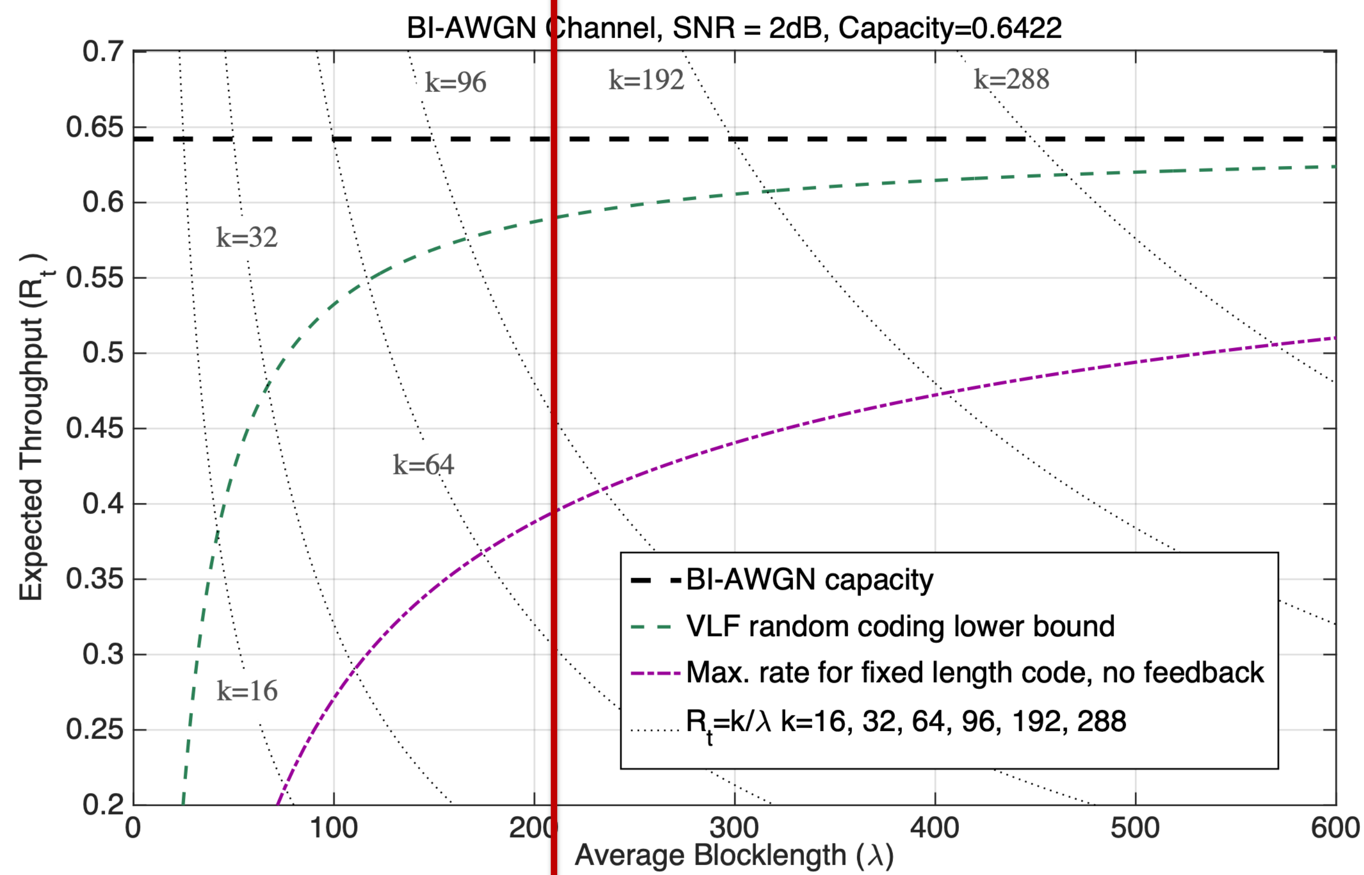




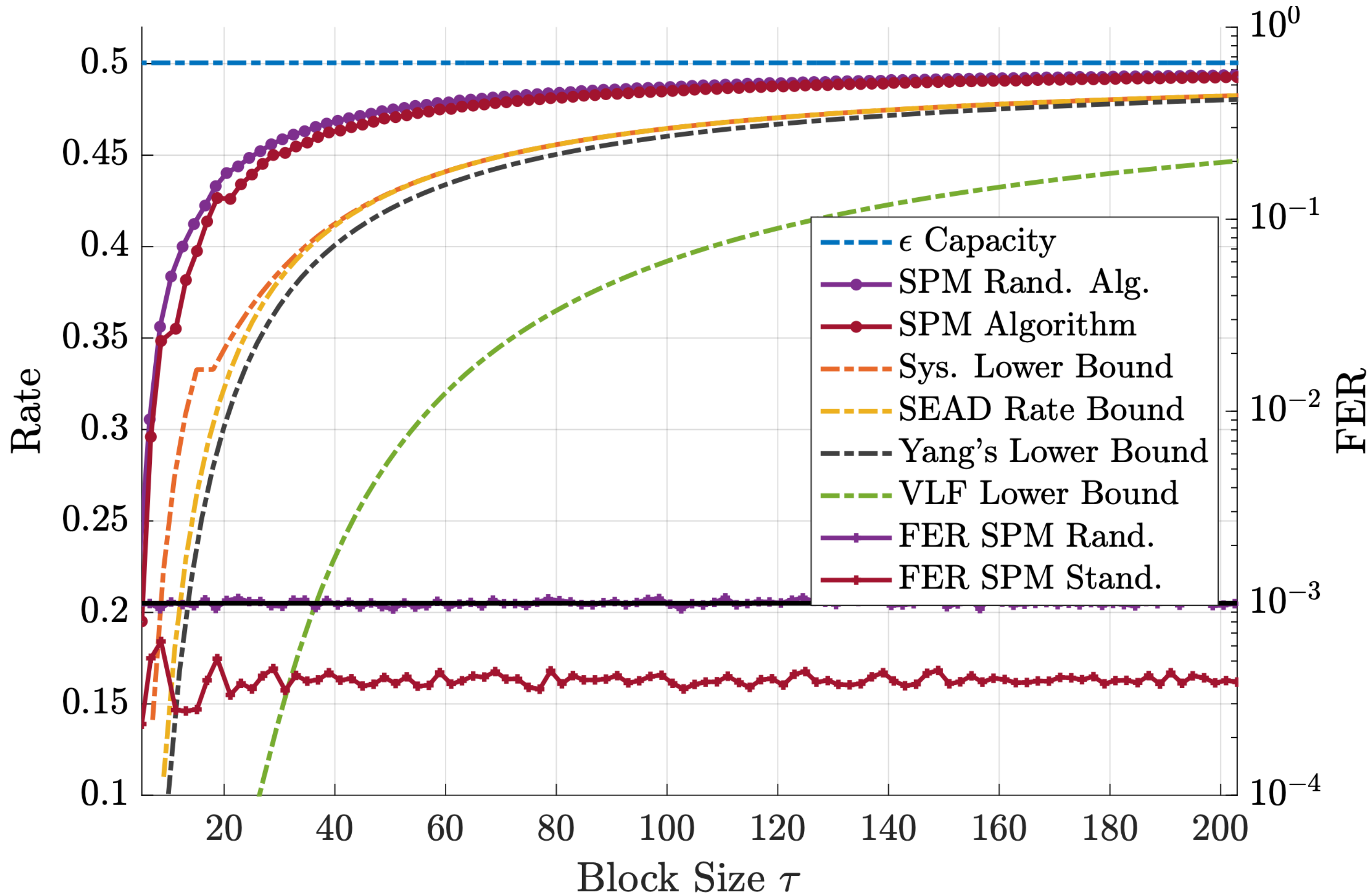
Amaael Antonini's Algorithm



Now we are looking at the first 200 symbols



Amaael Antonini's Bounds and Algorithm



CLA Samueli School of Engineering

Full Feedback Summary

- Full feedback increases achievable rate by avoiding telling the receiver things that it already knows.
- Capacity can be approached with even 100 channel uses.