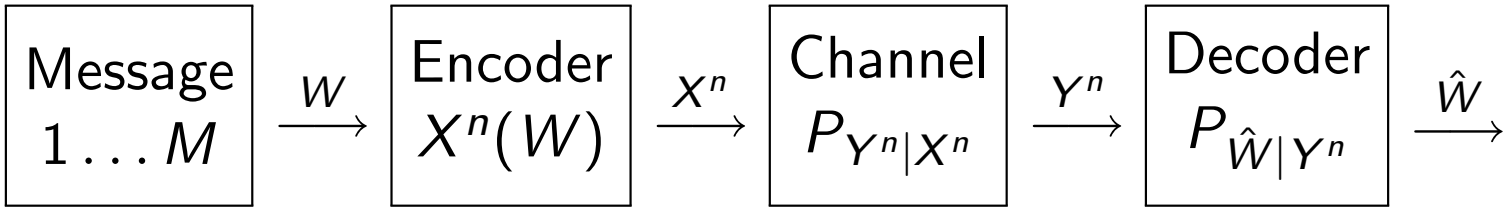
Channel Coding Rate in the Finite Blocklength Regime

H. Vincent Poor
(poor@princeton.edu)

Joint work with Y. Polyanskiy and S. Verdú

Princeton University

A Fundamental Model in Communications



- M: the number of values the message can take ($\log_2 M = \#$ of bits).
- n: the blocklength of the channel code.
- $P[\hat{W} \neq W]$: the probability of error.

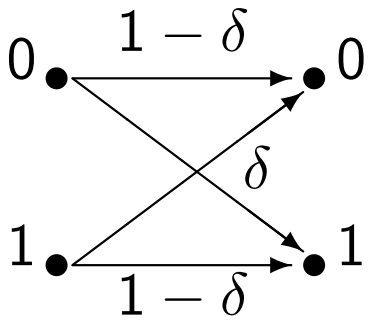
Important Examples of Channels

- Additive White Gaussian Noise (AWGN) channel:

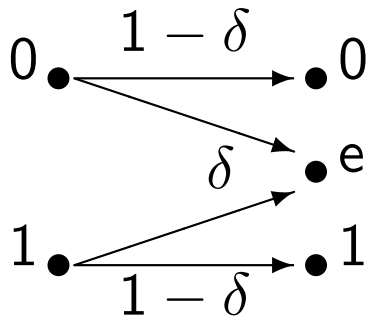
$$\begin{array}{ccccc} & & Z \sim \mathcal{N}(0,1) & & \\ & & \downarrow & & \\ X & \longrightarrow & \bigoplus & \longrightarrow & Y \end{array}$$

$$\mathbb{E}\left[|X|^2\right] \leq P$$

- Binary Symmetric (BSC) and Binary Erasure (BEC) channels:

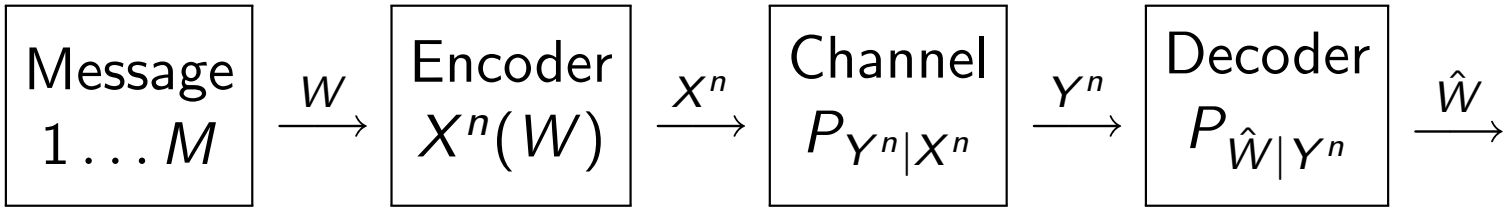


BSC



BEC

A Fundamental Problem in Communications



- (n, M, ϵ)-code:

$$\mathbb{P}[\hat{W} \neq W] \leq \epsilon$$

- Fundamental limit:

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

- Shannon: As $n \to \infty, \epsilon \to 0$

$$log \frac{M^*(n, \epsilon)}{n} \to C \quad \text{(capacity)}\,.$$

- This work: analyze $\frac{\log M^*(n,\epsilon)}{n}$ for a fixed ϵ and fixed n.

Classical Bounds

- Notation: $X^n \to X$, $Y^n \to Y$
- Choose P_X, and define the **information density:**

$$i(X;Y) = \log \frac{P_{XY}(X,Y)}{P_X(X)P_Y(Y)} = \log \frac{P_{Y|X}(Y|X)}{P_Y(Y)}$$

- Feinstein bound (1955): exists a code with prob. of error ϵ and

$$M \geq \sup_{\beta \geq 0} \left\{ \beta(\epsilon - \mathbb{P}[i(X;Y) \leq \log \beta]) \right\}$$

- Shannon bound (1957): exists a code with M codewords and

$$\epsilon \leq \inf_{\beta \geq 0} \left\{ \mathbb{P}[i(X;Y) \leq \log \beta] + \frac{M-1}{\beta} \right\} .$$

- Gallager bound (1965): exists a code with M codewords and

$$\epsilon \leq \exp \left\{ -nE_r \left(\frac{\log M}{n} \right) \right\} .$$

RCU bound

Theorem (Random Coding Union Bound)

For any P_X there exists a code with M codewords and

$$\epsilon \leq \mathbb{E}\left[\min\left\{1, (M-1)\mathbb{P}[i(\bar{X}, Y) \geq i(X, Y) \,|\, X, Y]\right\}\right],$$

where $P_{XY\bar{X}}(a, b, c) = P_X(a)P_{Y|X}(b|a)P_X(c)$.

Proof:

- Key ingredients: random-coding + maximum likelihood (ML) decoding.
- Union bound on probability of having one of $M-1$ codewords closer to Y than the true codeword:

$$(M-1)\mathbb{P}[i(\bar{X}, Y) \geq i(X, Y) \,|\, X, Y]$$

RCU bound

Theorem (Random Coding Union Bound)

For any P_X there exists a code with M codewords and

$$\epsilon \leq \mathbb{E}\left[\min\left\{1, (M-1)\mathbb{P}[i(\bar{X}, Y) \geq i(X, Y) \,|\, X, Y]\right\}\right],$$

where $P_{XY\bar{X}}(a, b, c) = P_X(a)P_{Y|X}(b|a)P_X(c)$.

Highlights:

- Strictly stronger than Feinstein-Shannon and Gallager
- Hard to analyze asymptotics
- Hard to compute, complexity $O(n^{2(|X|-1)|Y|})$

DT bound

Theorem (Dependence Testing Bound)

For any P_X there exists a code with M codewords and

$$\epsilon \leq \mathbb{E}\left[\exp\left\{-\left|i(X,Y) - \log\frac{M-1}{2}\right|^+\right\}\right].$$

Highlights:

- Strictly stronger than Feinstein-Shannon
- . . . and no auxiliary constants
- Typically better than Gallager
- Easier to compute than the RCU bound
- Easier to analyze asymptotics: $\epsilon \leq \mathbb{E}\left[e^{-n\left|\frac{1}{n}i(X^n,Y^n)-R\right|^+}\right]$.

DT bound: Proof

- Codebook: random selection of $C_1, \ldots C_M$, i.i.d. with P_X
- Feinstein decoder.
- j-th codeword's probability of error:

$$\mathbb{P}[\text{error} \mid W = j] \;\leq\; \mathbb{P}[i(X, Y) \leq \alpha] + (j-1)\mathbb{P}[i(\bar{X}, Y) > \alpha]$$

Average over W:

$$\mathbb{P}[\text{error}] \;\leq\; \mathbb{P}\left[i(X, Y) \leq \alpha\right] + \frac{M-1}{2}\mathbb{P}\left[i(\bar{X}, Y) > \alpha\right]$$

DT bound: Proof

- Recap: for every α there exists a code with

$$\epsilon \leq \mathbb{P}\left[i(X,Y) \leq \alpha\right] + \frac{M-1}{2}\mathbb{P}\left[i(\bar{X},Y) > \alpha\right] .$$

- Key step: closed form optimization of α.
  Let $Q_{XY} = P_X \times P_Y$:

$$\frac{M+1}{2}\left(\frac{2}{M+1}P_{XY}\left[\frac{dP_{XY}}{dQ_{XY}} \leq e^{\alpha}\right] + \frac{M-1}{M+1}Q_{XY}\left[\frac{dP_{XY}}{dQ_{XY}} > e^{\alpha}\right]\right)$$

  Bayesian dependence testing!

  Optimum threshold: Ratio of priors $\Longrightarrow$ $\boxed{\alpha^* = \log\frac{M-1}{2}}$

- Change of measure argument:

$$P\left[\frac{dP}{dQ} \leq \tau\right] + \tau Q\left[\frac{dP}{dQ} > \tau\right] = \mathbb{E}_P\left[\exp\left\{-\left|\log\frac{dP}{dQ} - \log\tau\right|^+\right\}\right] .$$

Extensions: maximal prob. of error and input-constraints

Theorem (κ-β bound)

For any Q_Y and ϵ there exists an **F***-constrained code of size*

$$M \geq \sup_{\tau \in [0,\epsilon]} \frac{\kappa_\tau(Q_Y)}{\sup_{x \in \mathbf{F}} \beta_{1-\epsilon+\tau}(x, Q_Y)}$$

$$\beta_\alpha(x, Q_Y) = \inf_{P_{Z|Y}: P[Z=1|X=x] \geq \alpha} Q_Y[Z=1]$$

a randomized test between $P_{Y|X=x}$ and Q_Y, which correctly detects $P_{Y|X=x}$ w.p. $\geq \alpha$

$$\kappa_\tau(Q_Y) = \inf_{P_{Z|Y}: \inf_{x \in \mathbf{F}} P[Z=1|X=x] \geq \tau} Q_Y[Z=1]$$

a randomized test between a collection $\{P_{Y|X=x}\}_{x \in \mathbf{F}}$ and Q_Y.

Converse: generalized sphere-packing

Theorem

Any (M, ϵ)-code over a constraint set $\mathbf{F}$ *satisfies*

$$M \leq \inf_{Q_Y} \sup_{x} \frac{1}{\beta_{1-\epsilon}(x, Q_Y)},$$

where

$$\beta_\alpha(x, Q_Y) = \inf_{P_{Z|Y}: P[Z=1|X=x] \geq \alpha} Q_Y[Z = 1]$$

a randomized test between $P_{Y|X=x}$ and Q_Y, which correctly detects $P_{Y|X=x}$ w.p. $\geq \alpha$.

Duality:

$$\frac{\kappa_\tau}{\beta_{1-\epsilon+\tau}} \leq M^*(n, \epsilon) \leq \frac{1}{\beta_{1-\epsilon}}$$

BEC: Conditional Channel Converse

Theorem (CC Converse)

For a BEC with erasure probability δ, the average error probability of a k-to-n code satisfies

$$\epsilon \geq \sum_{\ell=n-k+1}^{n} \binom{n}{\ell} \delta^{\ell}(1-\delta)^{n-\ell}\left(1-2^{n-\ell-k}\right)$$

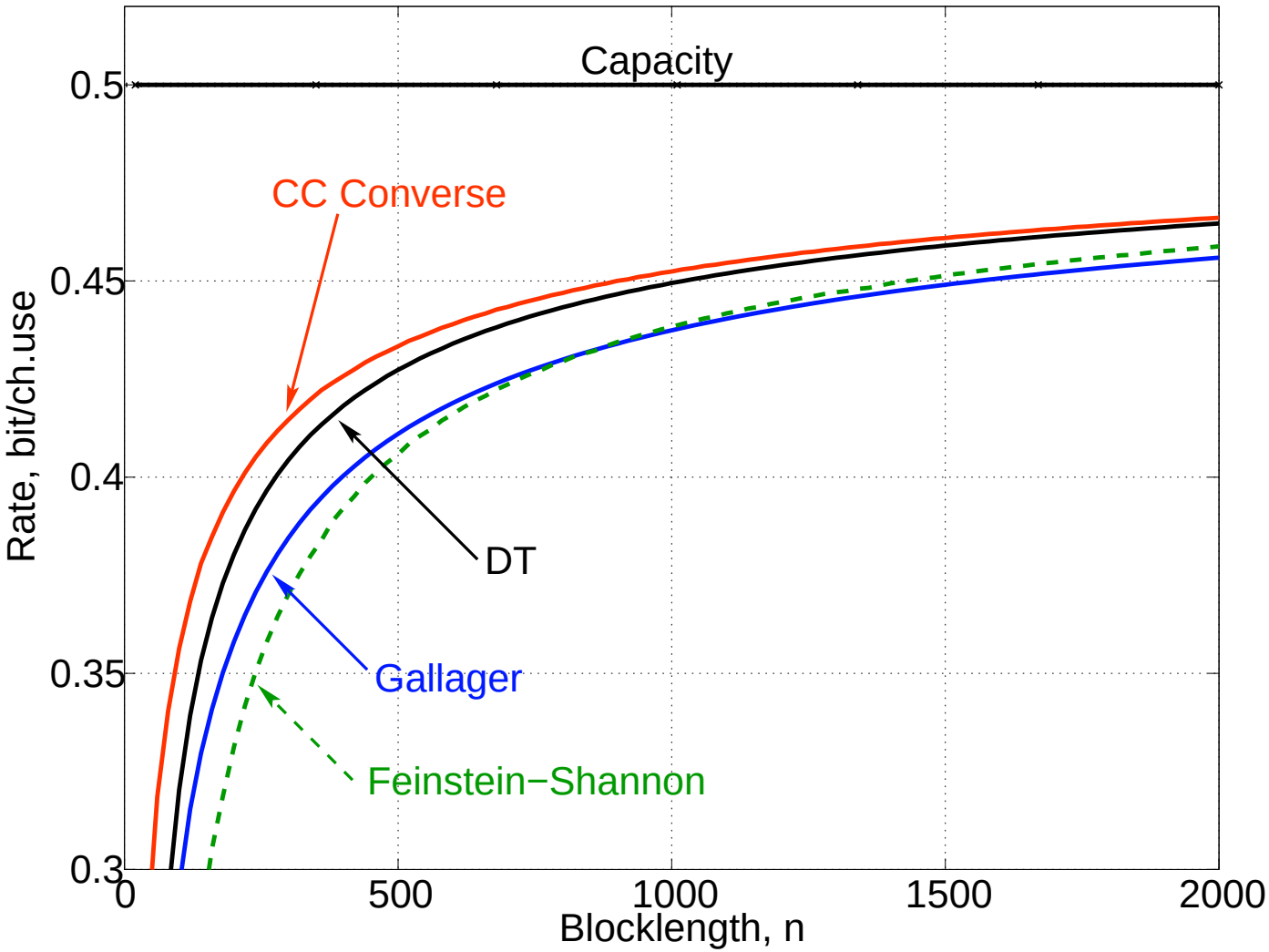
Proof:

- if ℓ erasures happened then

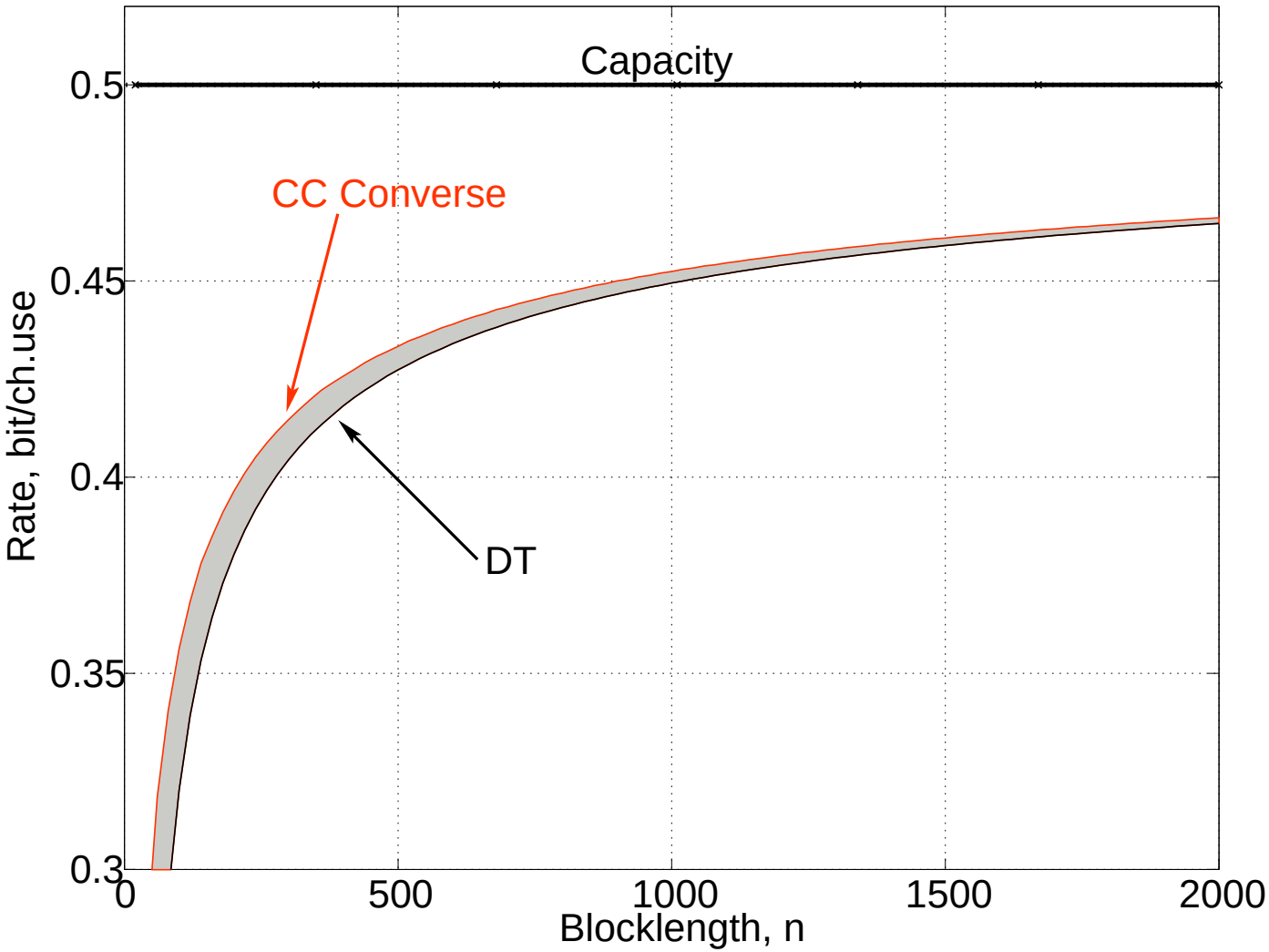
$$\mathbb{P}[\text{error}|\#\text{ erasures} = \ell] \geq \left(1-2^{n-\ell-k}\right)$$

- average over ℓ

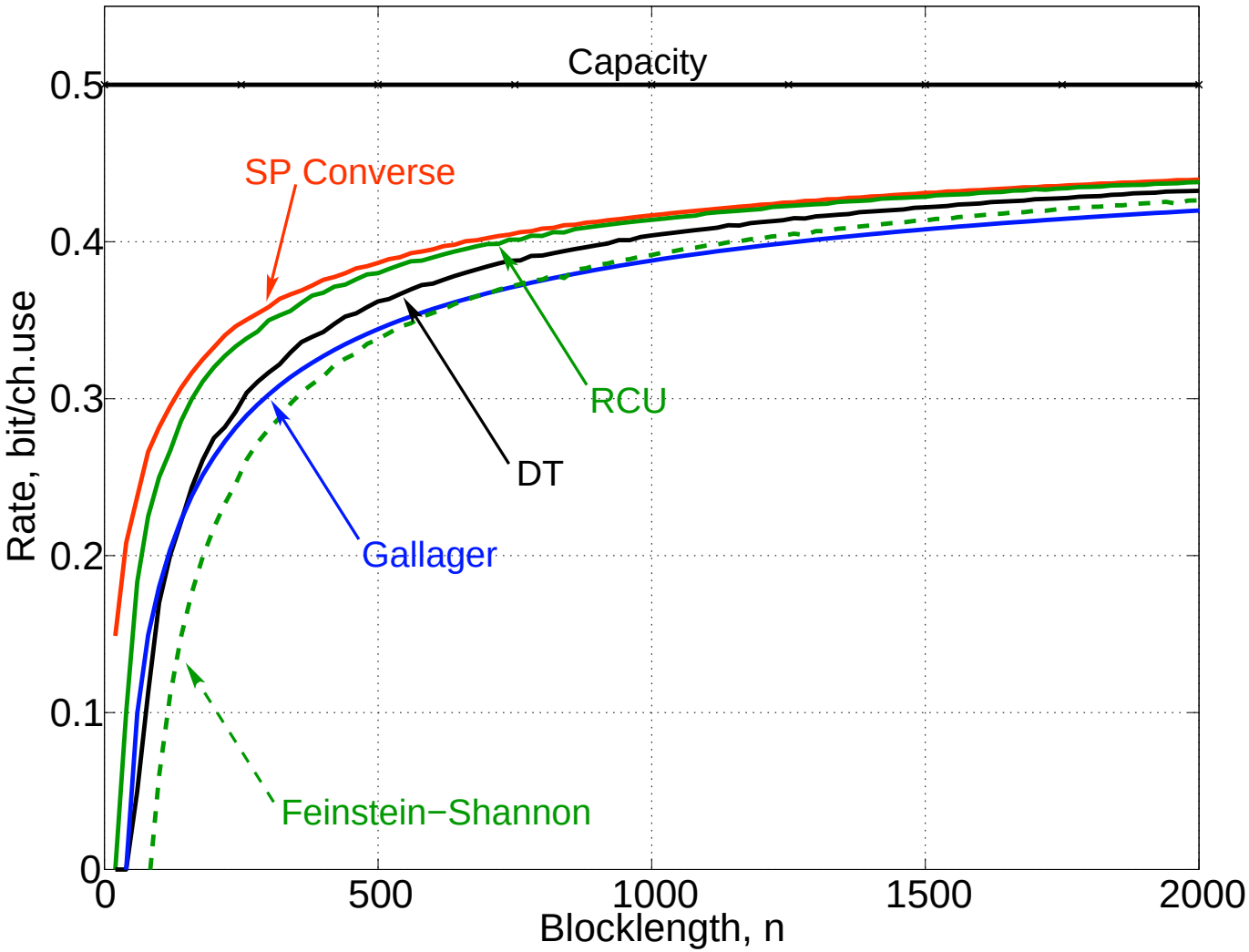
BEC with $\delta = 0.5$; $\epsilon = 10^{-3}$



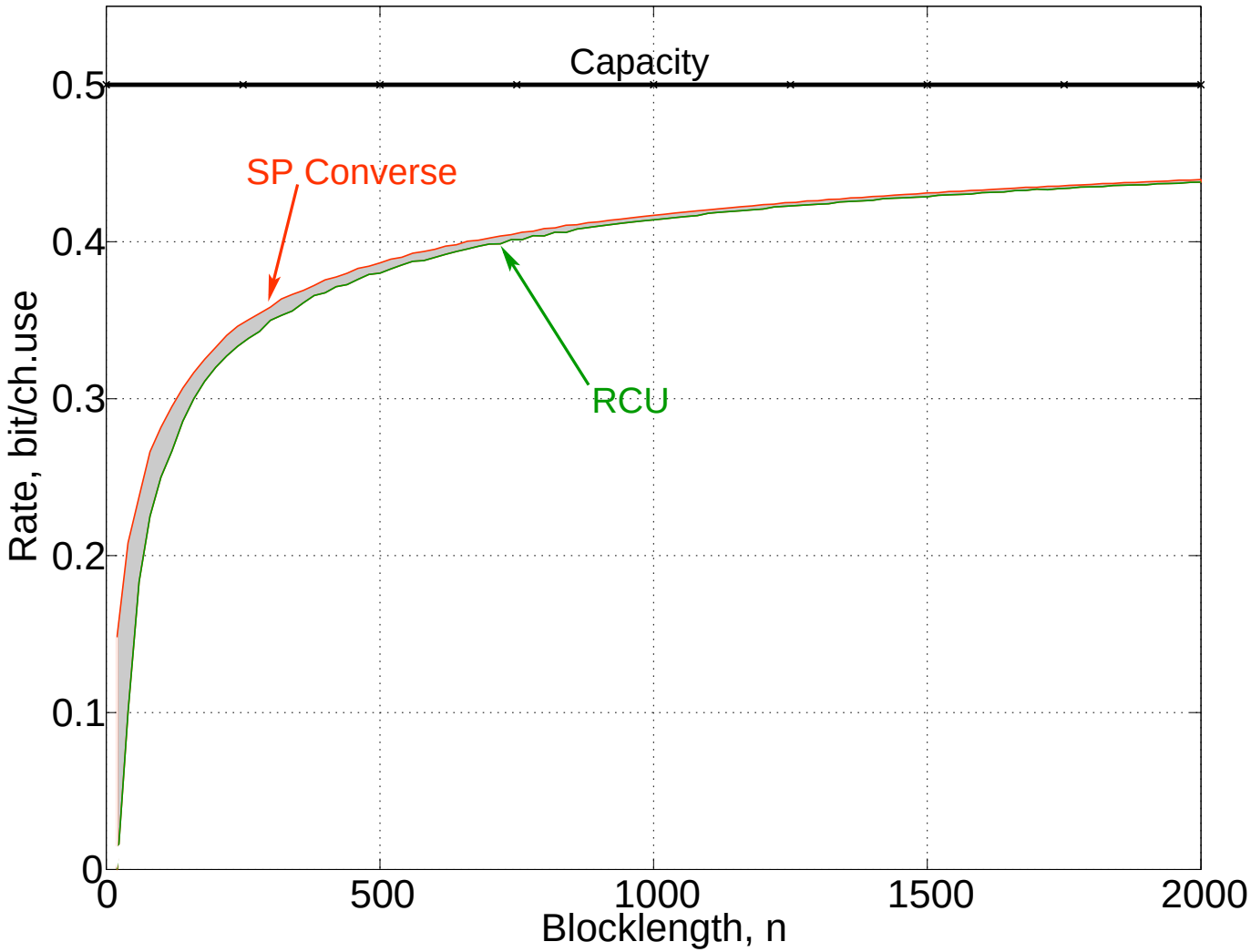
BEC with $\delta = 0.5$; $\epsilon = 10^{-3}$



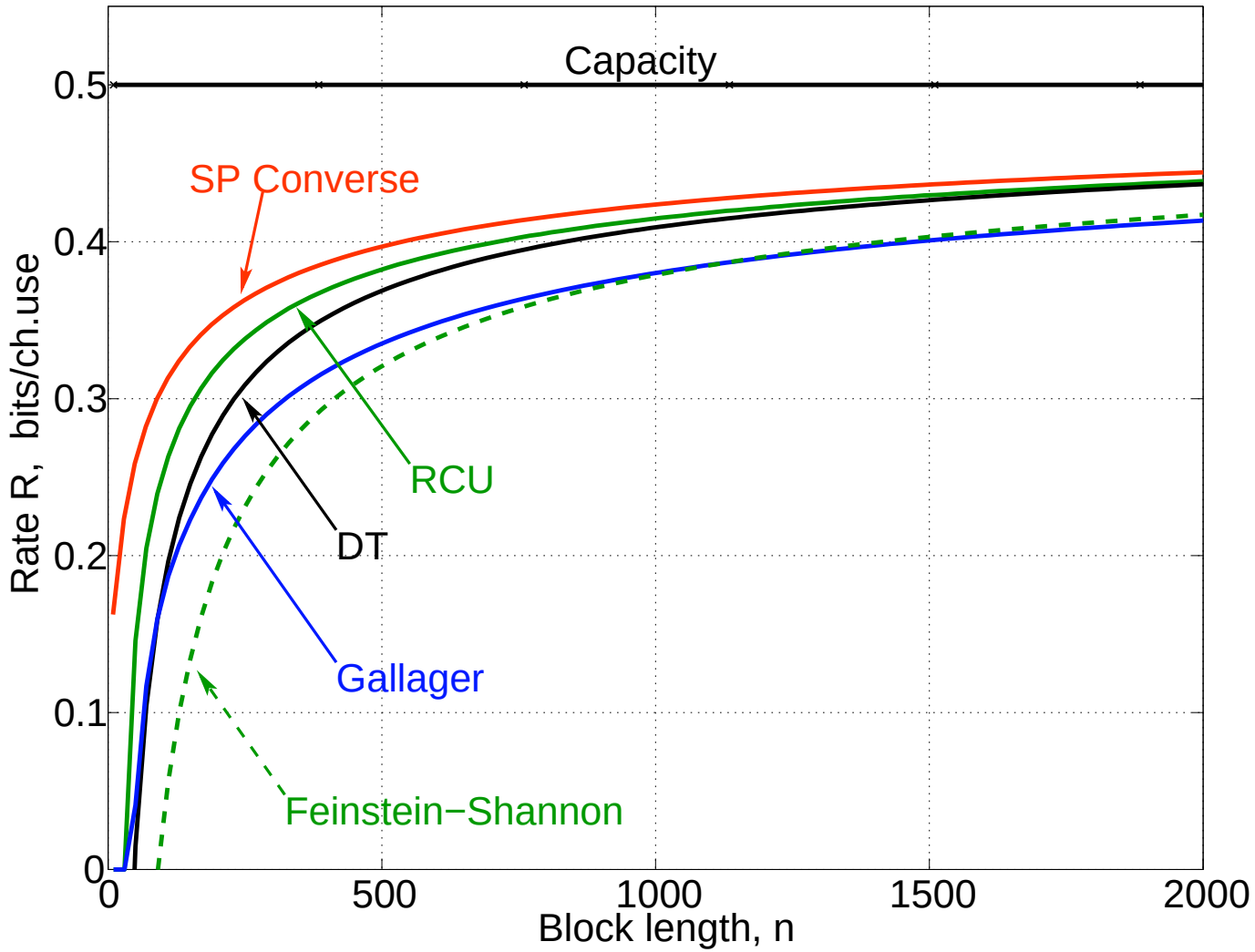
BSC with $\delta = 0.11$; $\epsilon = 10^{-3}$



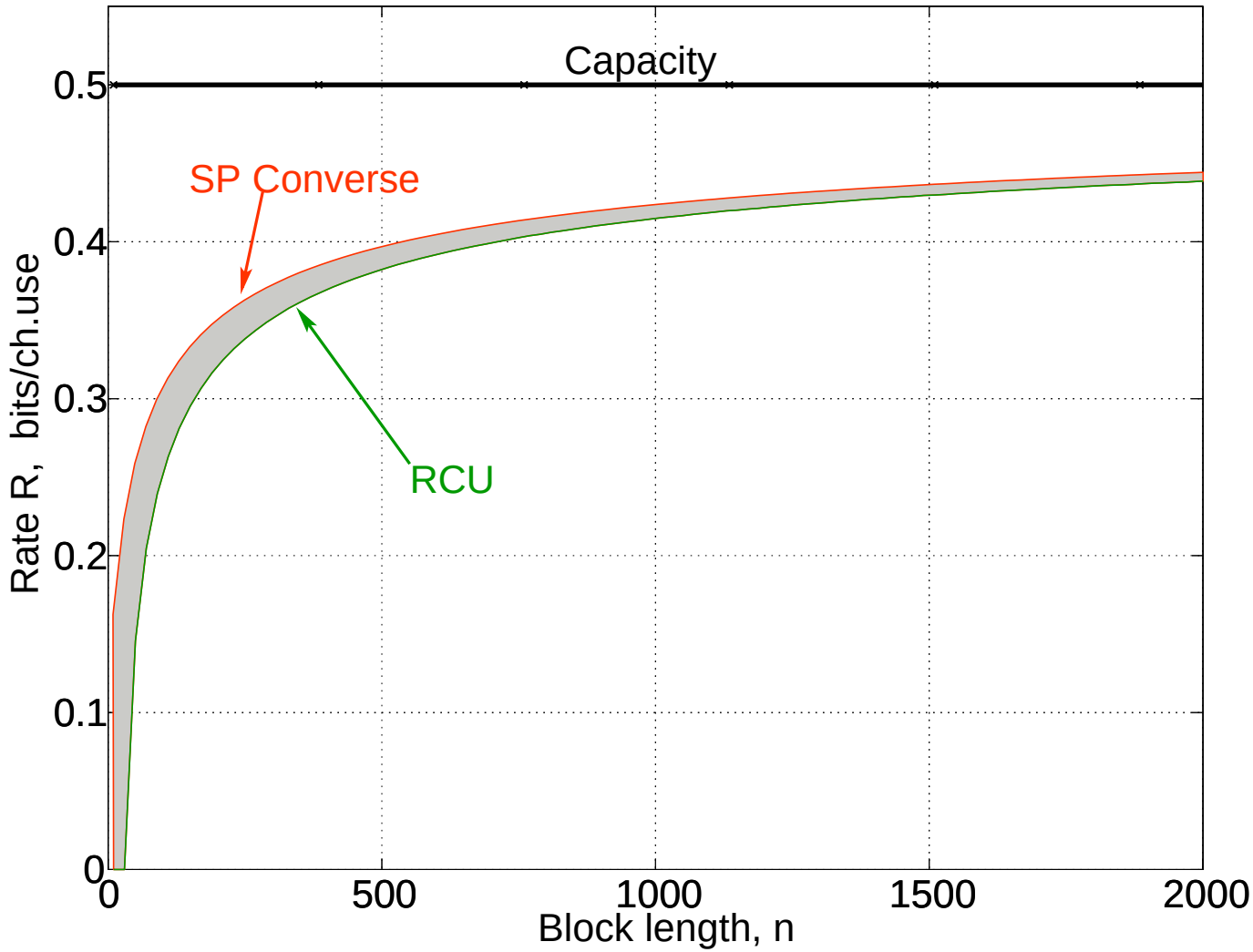
BSC with $\delta = 0.11$; $\epsilon = 10^{-3}$



AWGN with SNR 0 dB; $\epsilon = 10^{-3}$



AWGN with SNR 0 dB; $\epsilon = 10^{-3}$



Discussion

- RCU bound – harder to compute, but usually is the tightest.
- DT bound beats Feinstein-Shannon and Gallager bounds.
- DT bound – good tradeof between tightness and tractability.

- $\log M^*(n, \epsilon)$ is bounded to within a few bits.
- **The gap to capacity is not negligible!**
- Natural question: Is there an easy approximation to the maximal rate for fixed n and ϵ?

Strassen's Theorem

Theorem

For any discrete memoryless channel (DMC) and $\epsilon \leq 1/2$

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

where

$$\begin{aligned} C &= I(X^*;Y^*) \\ V &= \operatorname{Var} i(X^*;Y^*) \\ Q(x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \end{aligned}$$

V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," *Trans. Third Prague Conf. Information Theory*, 1962

Beyond Strassen: using our 3 new bounds

Expansion for the AWGN (and parallel AWGN)

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

$$C = \frac{1}{2}\log(1+P)\,, \quad V = \frac{P}{2}\frac{P+2}{(P+1)^2}\log^2 e$$

Proof:

- Achievability: DT bound with input constraints ($\kappa - \beta$ bound).
- Converse: generalized sphere-packing with X constrained to the power sphere.
- For both: $P_{Y^n} = [\mathcal{N}(0, 1+P)]^n$

Beyond Strassen: using our 3 new bounds

Expansion up to $O(1)$ for the BSC:

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}\,Q^{-1}(\epsilon) + \frac{1}{2}\log n + O(1)$$

$$C = 1 - h(\delta)\,, \quad V = \delta(1-\delta)\log^2\frac{\delta}{1-\delta}$$

Proof:

- Achievability: RCU bound for the average probability of error.
- From average to maximal: random linear code method.
- Converse: sphere-packing bound.

Beyond Strassen: using our 3 new bounds

Expansion up to $O(1)$ for the BEC:

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

$$C = 1-\delta\,, \quad V = \delta(1-\delta)$$

Proof:

- Achievability: DT bound for maximal probability of error.
- Converse: special converse for the BEC (*conditional channel converse*).

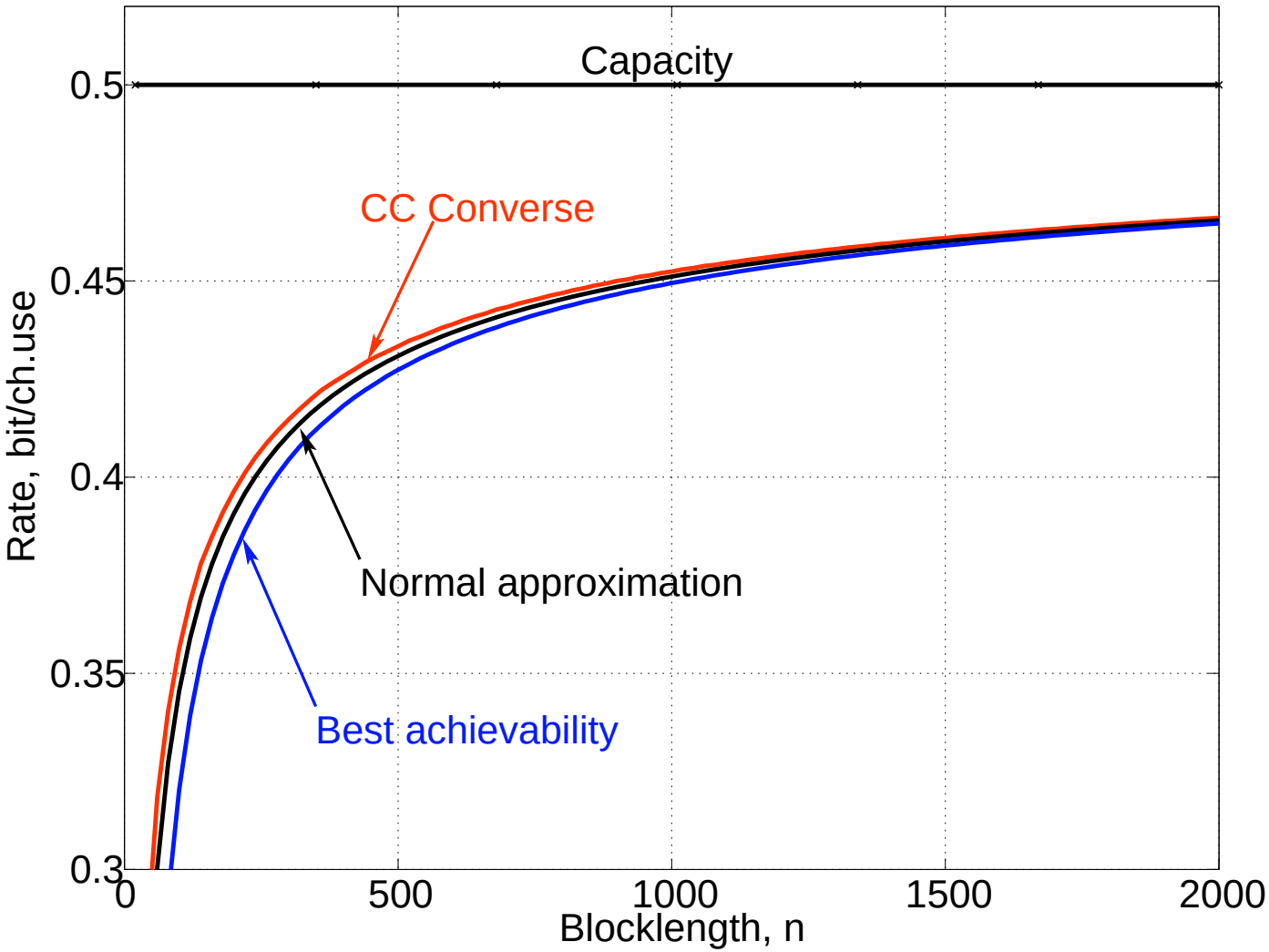
Normal approximation

$$\frac{\log M^*(n,\epsilon)}{n} \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)\,,$$

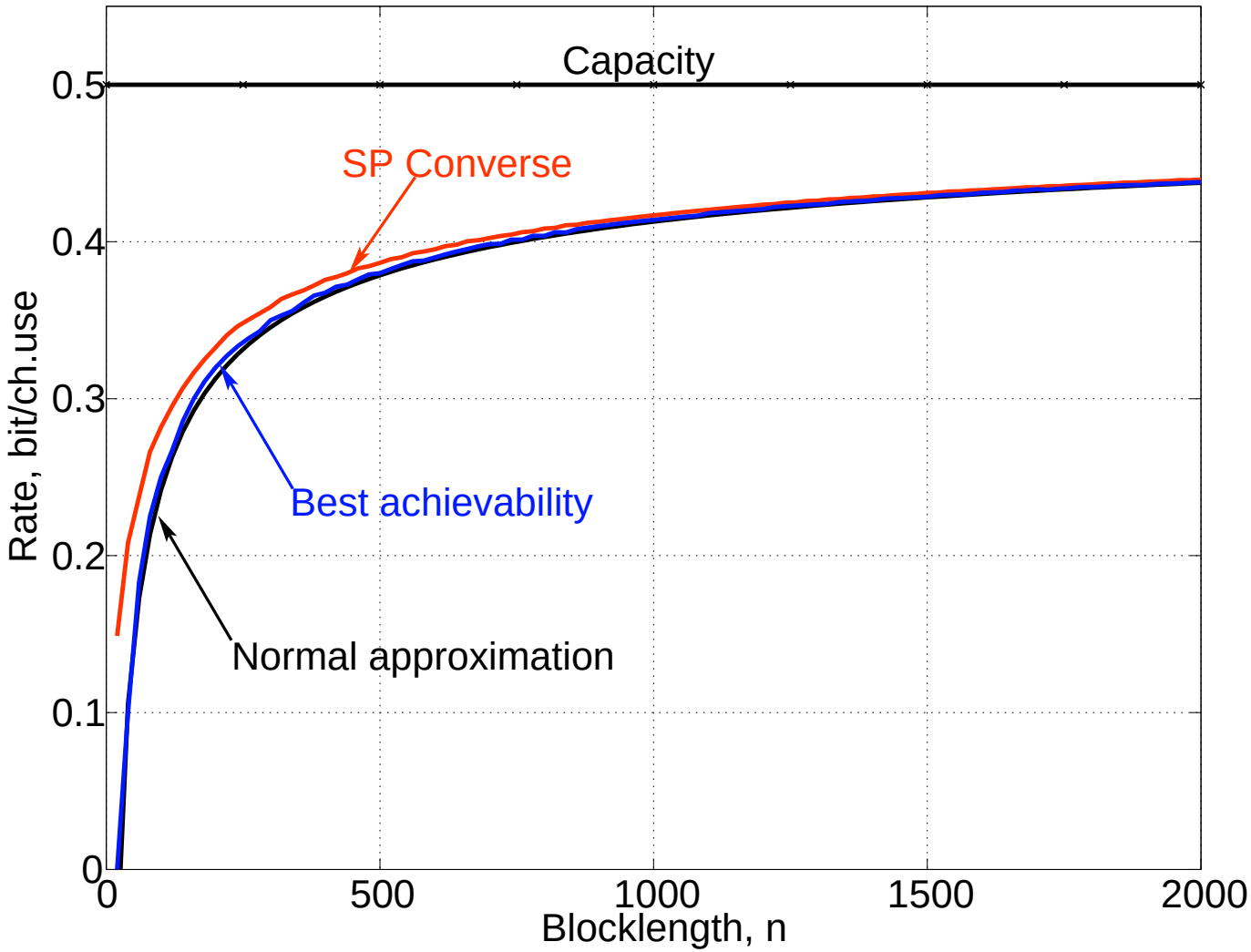
$C =$ the channel capacity

$V =$ the channel dispersion

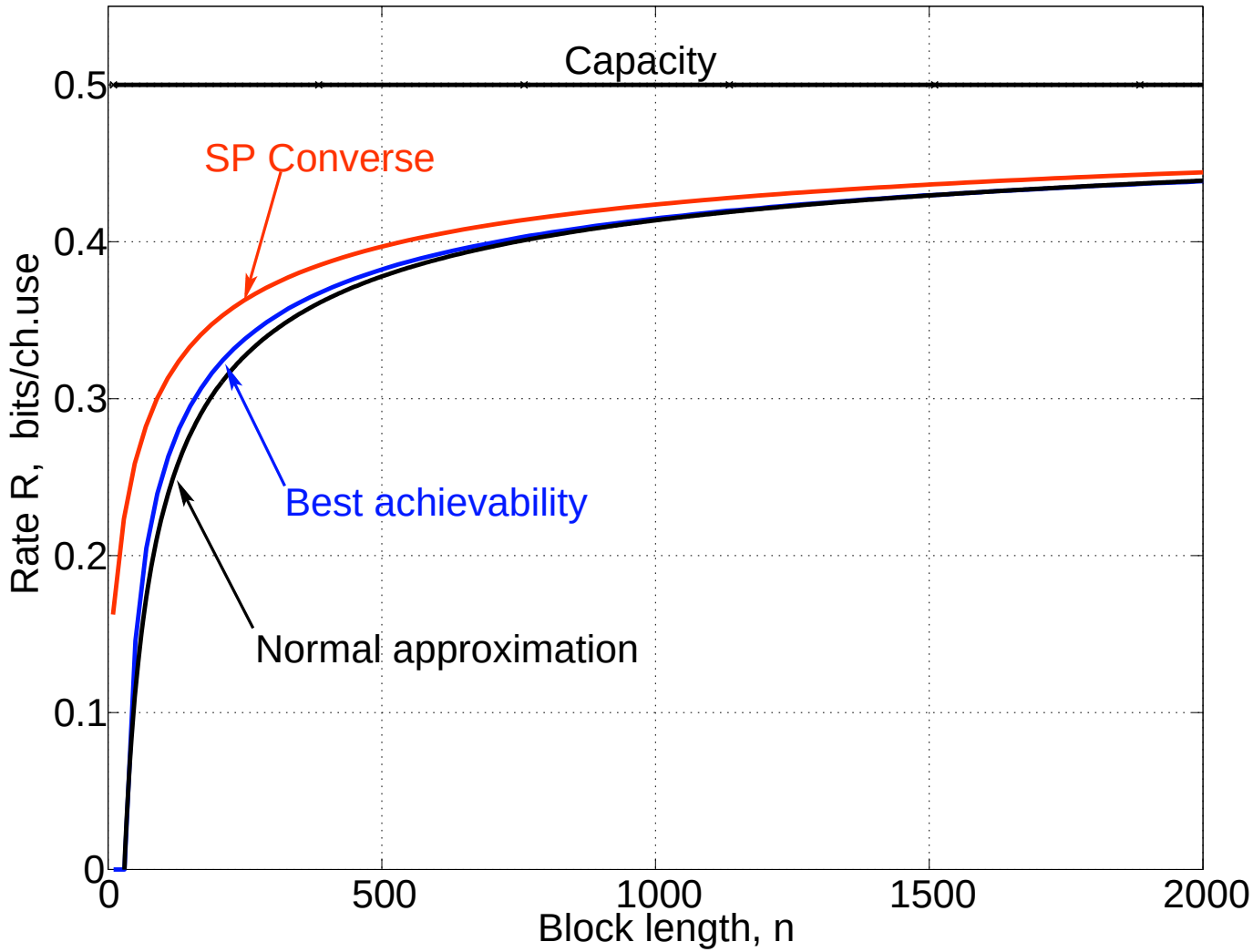
BEC with $\delta = 0.5$; $\epsilon = 10^{-3}$



BSC with $\delta = 0.11$; $\epsilon = 10^{-3}$



AWGN with SNR 0 dB; $\epsilon = 10^{-3}$



Discussion

- Normal approximation: simple, closed form, very tight!
- Ties together delay, reliability and rate $\Longrightarrow$ many applications!
- Claim: "for large n capacity can be achieved".
  Question: How large?

$$C \gg \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) \quad \Longrightarrow \quad \boxed{n \gtrsim 10^4 \frac{V}{C^2}}$$

($\epsilon \sim 10^{-3}$, to 3% of capacity)

AWGN 0 dB : $n \gtrsim 15000$
AWGN 20 dB : $n \gtrsim 500$

- Another example: maximize throughput in an ARQ system.

$$\epsilon^* \approx 10^{-2} \sim 10^{-3}$$

for most practical packet sizes and channels!

ARQ feedback

- Want: use WiFi and download MP3 from iTunes faster!
- Typically: higher level protocols use CRC to retransmit the packet until correctly delivered (ARQ feedback).
- **Tradeoff:** high rate code $\Longrightarrow$ many retransmissions!
  low rate code $\Longrightarrow$ bad throughput!
- Question: What is the optimum?
  Answer: maximize average throughput!

$$\text{average throughput} = \text{Rate} \times (1 - \mathbb{P}[\text{error}])$$

- Alternatively: fix data payload K, minimize avg. delay!

ARQ feedback

Goal:

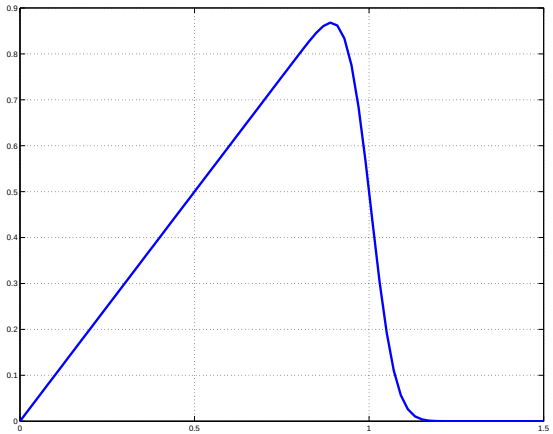
$$\max\left[\text{Rate} \times (1 - \mathbb{P}[\text{error}])\right]$$

Normal approximation:

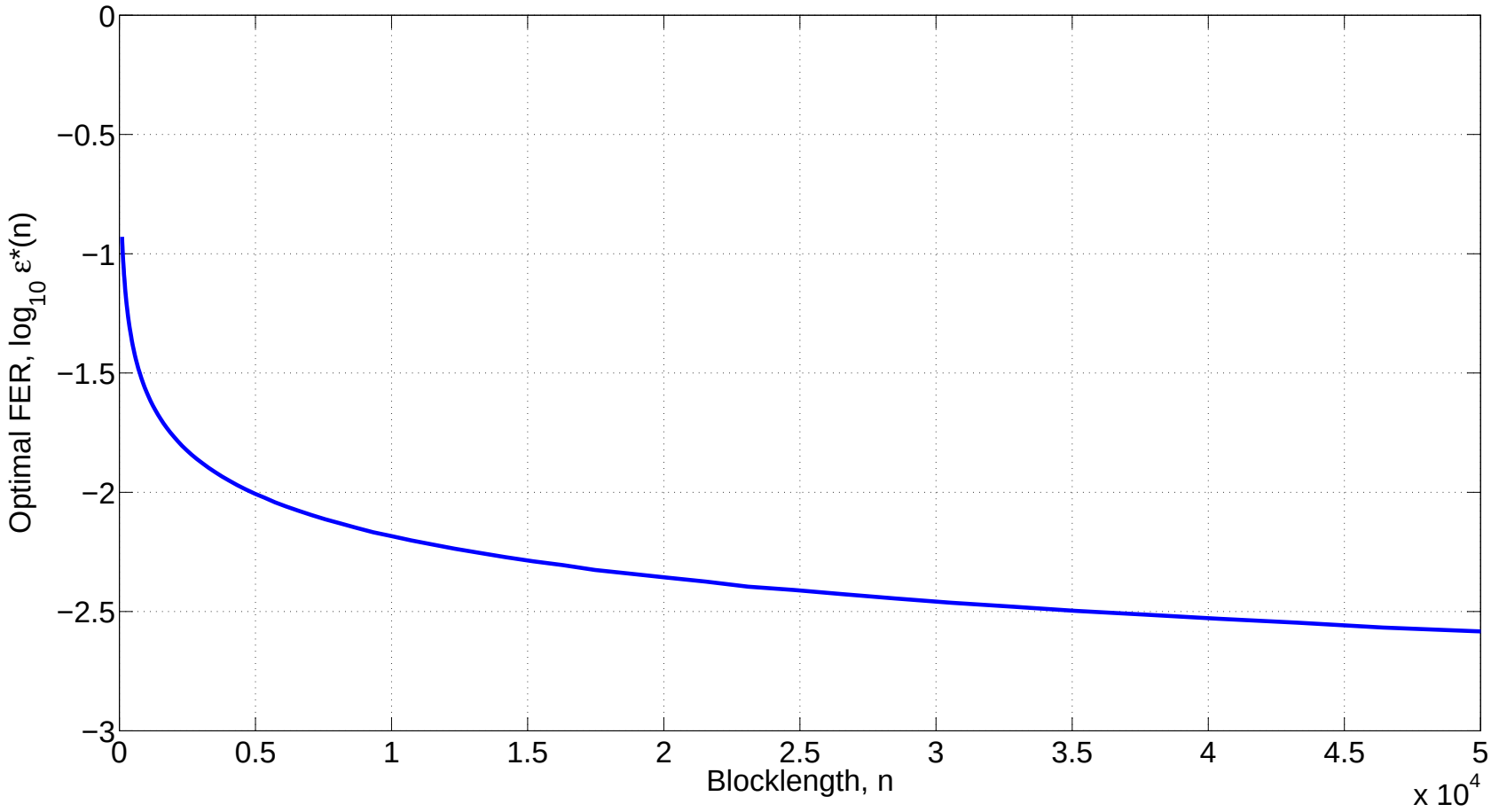
$$R \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) \iff 1 - \epsilon \approx Q\left(\sqrt{\frac{nC^2}{V}}\left\{\frac{R}{C} - 1\right\}\right)$$

For each n there is an optimal $\epsilon^*(n)$ maximizing throughput

$$\begin{aligned} \epsilon^*(n) &= 1 - Q\left(\sqrt{\frac{nC^2}{V}}(x^* - 1)\right) \\ x^*(n) &= \operatorname*{argmax}_{x \geq 0}\left[xQ\left(\sqrt{\frac{nC^2}{V}}(x - 1)\right)\right] \end{aligned}$$



ARQ feedback: AWGN 0 dB



- weak dependence on n
- . . . and the channel!
- Overall: $\epsilon = 10^{-3}$ – reasonable choice!

Summary

Conceptually:

- Analyzed $\log M^*(n, \epsilon)$ directly
- For practical n, ϵ computed upto a few bits
- ... and approximated analytically!

Theoretically:

- New bounds that surpass classical ones
- Normal approximation for the AWGN

Practically:

- Guidelines for choosing block length