Non-Asymptotic Converse Bounds Via Auxiliary Channels

Ioannis Papoutsidakis, Robert J. Piechocki, and Angela Doufexi

Department of Electrical and Electronic Engineering, University of Bristol
Email: {ioannis.papoutsidakis, r.j.piechocki, a.doufexi}@bristol.ac.uk

Abstract—This paper presents a new derivation method of converse bounds on the non-asymptotic achievable rate of discrete weakly symmetric memoryless channels. It is based on the finite blocklength statistics of the channel, where with the use of an auxiliary channel the converse bound is produced. This method is general and initially is presented for an arbitrary weakly symmetric channel. Afterwards, the main result is specialized for the q-ary erasure channel (QEC), binary symmetric channel (BSC), and QEC with stop feedback. Numerical evaluations show identical or comparable bounds to the state-of-the-art in the cases of QEC and BSC, and a tighter bound for the QEC with stop feedback.

Index Terms—Converse bounds, achievability bounds, finite blocklength regime, non-asymptotic analysis, channel capacity, discrete memoryless channels.

I. INTRODUCTION

An important step towards latency mitigation is the study of achievable channel coding rates for finite blocklengths. Shannon proved in [1] that there exists a code that can achieve channel capacity as the blocklength grows to infinity. This is the main reason the majority of conventional communications systems utilize blocks of several thousands of symbols to transmit with a rate approximately close to capacity. In [2], several lower bounds to capacity are established and a clearer image of the achievable coding rates in the finite blocklength regime is provided. Observing the available converse and achievability bounds it is apparent that the channel capacity significantly overestimates the achievable rates in the case of short blocklengths.

The significance of finite blocklength results lies in the following: fewer channel uses prior to decoding allow not only for shorter transmission times but most importantly for less complex encoders and decoders as well as more flexible upper layer protocols. The trade-off between achievable rate and blocklength has been the focus of many researchers in the past decade. Whether the achievability bounds of [2] are optimal remains an open question. This work aims to provide insight on the optimality of the state-of-the-art achievability and converse bounds, along with proposing a different approach of obtaining converse bounds over finite blocklengths.

More specifically, the current paper formulates a novel method for the derivation of converse bounds on the non-asymptotic achievable rate of discrete weakly symmetric memoryless channels. This is accomplished by utilizing the finite statistics of the channel along with an auxiliary channel. The main result considers the average probability of error over arbitrary weakly symmetric memoryless channels and it is presented in Section III. Section II provides a short review on the state-of-the-art. In Sections IV, V, and VI the general method is specialized for the q-ary erasure channel (QEC), binary symmetric channel (BSC), and QEC with stop feedback, respectively.

Notation: Throughout this paper a (n,M,ε) code is a code with blocklength n, codebook size M, and average probability of error ε. Furthermore, the letter q denotes the cardinality of the channel input. Notation Xⁿ denotes the vector X₁,...,Xₙ.

Finally, a (lₙ,M,n,ε) VLSF code is a variable length stop feedback code with average blocklength lₙ, codebook size M, packet size n, and average probability of error ε.

II. PREVIOUS WORK

The best achievability bounds to date for discrete memoryless channels are established in [2]. Random Coding Union (RCU) bound is based on the analysis of error probability of random codes under maximum likelihood decoding [2, Theorem 16]. For an arbitrary Pₓ there exists an (M,ε) code such that,

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

(1)

where Pₓᵧₖ(a,b,c) = Pₓ(a)Pᵧ|ₓ(b|a)Pₓ(c).

A less complex, in terms of computation, bound is the Dependence Testing (DT) bound [2, Theorem 17]. For any distribution Pₓ on A, there exists a code with M codewords and average probability of error not exceeding

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

(2)

Numerical evaluations confirm that RCU bound is tighter for BSC and DT bound for BEC [2].

Regarding the converse bounds, there are several older results from the early days of information theory. For instance, a bound based on Fano’s inequality can be found in [3]. The established sphere-packing bound as well as an improvement is given in [4] and [5], respectively. A special converse bound for the BEC with erasure probability δ is given in [2, Theorem 38],

$$\epsilon \geq \sum_{l=\lfloor n - \log₂ M \rfloor + 1}^{n} \binom{n}{l} \delta'(1 - \delta)^{n-l} \left(1 - \frac{2^{n-l}}{M}\right),$$

(3)
for a \((n, M, \epsilon)\) code. This bound corresponds to the meta-converse bound \([6]\). Similarly, \([2]\) Theorem 35) provides a bound for the BSC that coincides with the meta-converse bound as well as the classical sphere-packing bound.

III. METHODOLOGY

The stochasticity of a discrete memoryless channel is modelled by the conditional probability mass function of the channel output given the channel input \(P_{Y|X}\) or vice versa \(P_{X|Y}\). The nature of this process in the finite regime results to relative frequencies that do not always follow the probability mass function. For instance, a finite Bernoulli process with \(P(X_i = 1) = p\) has a positive probability of producing sequences where the relative frequency of 1s is higher than \(p\). Motivated by this observation a converse bound for weakly symmetric memoryless channels is developed.

The focus of this work is on weakly symmetric channels because the output of such channel can be defined as the function \(Y^n = f(X^n, Z^n)\), where \(X^n\) is the channel input and \(Z^n\) is a noise vector independent from \(X^n\). Thus, by analysing the non-asymptotic behaviour of \(Z^n\) an auxiliary channel can be produced in order to derive a converse to the non-asymptotic achievable rate of the main channel.

The main channel is depicted in Figure 1 as a channel with random states where the conditional probability of the output given the input depends on the state. The state is not available at neither the encoder nor the decoder, therefore it is treated as part of the noise. The auxiliary channel is presented in Figure 2. The noise is produced similarly to the main channel but in this case, there is side information at the receiver and the transmitter. That is to say, the state of the channel is known before transmission.

Clearly, the auxiliary channel has greater achievable rate than the main one since there is no uncertainty about the state. Hence, a converse bound for the auxiliary channel bounds the achievable rate of the main channel as well.

The question arises of how an auxiliary channel is produced based on the given main channel \(P_{Y^n|X^n}\). A simple yet effective way to classify the channel states is based on the number of erroneous transmissions over one blocklength. In this manner, the states follow the binomial distribution,

\[
P(S = s) = \binom{n}{s} (1 - P(Y = i|X = i))^s \cdot P(Y = i|X = i)^{n-s},
\]

for any \(i\) that belongs to the input alphabet.

At this stage, the capacity \(C_s\) of each state of the auxiliary channel is calculated,

\[
C_s = \max_{P(x_1, \ldots, x_n)} I(X^n; Y^n_1|S = s).
\]

By setting a fixed aggregate rate \(R_a = nR\) over all states, one can derive two different classes of states. One class includes the states with a capacity lower or equal to the rate \(R_a\) and the other class the remaining. The latter class of states denotes cases where the error probability is certainly non-zero since it contradicts the noisy channel coding theorem where \(R < C\) for arbitrarily low probability of error. The probability of erroneous transmission due to rates that are not supported is a lower bound on the average probability of error given the rate \(R_a\) for the auxiliary channel, and therefore for the main one as well.

Formally, this is formulated as a corollary of the noisy channel coding theorem \([1]\).

**Corollary 1.** For a weakly symmetric memoryless discrete channel, the average error probability of a \((n, M, \epsilon)\) code satisfies,

\[
\epsilon \geq \sum_{s \in S} P(\text{error}|S = s)P(S = s),
\]

where,

\[
S = \{ s : C_s < \log_q M = nR \}.
\]

To compute a lower bound on \(P(\text{error}|S = s)\) we can use two different approaches. The first and more general one is the strong converse to the noisy channel coding theorem by Wolfowitz \([7]\).

**Theorem 1.** For an arbitrary discrete memoryless channel of capacity \(C\) nats and any \((n, e^nR, \epsilon)\) code with \(R > C\),

\[
\epsilon \geq 1 - \frac{4A}{n(R-C)^2} e^{-\frac{n(R-C)}{2}},
\]

where \(A\) is a finite positive constant independent of \(n\) or \(R\).

As noted by the authors of \([2]\), Theorem 1 is not useful for finite blocklength analysis, and indeed when used directly to the main channel it produces a converse bound which is looser than capacity. However, when it is combined with the method of auxiliary channels it can provide bounds that are comparable to the state-of-the-art as is demonstrated in the following sections.

The second approach is to take into account the specific structure of each channel state. Let \(X^n_1 = x^n_1\) be the decision of the optimal decoder given \(Y^n_1 = y^n_1\). Then,

\[
P(\text{error}|S = s) = 1 - P(X^n_1 = x^n_1|Y^n_1 = y^n_1) = 1 - \frac{P(Y^n_1 = y^n_1|X^n_1 = x^n_1)P(X^n_1 = x^n_1)}{P(Y^n_1 = y^n_1)}
\]

\[
= 1 - \frac{P(Z^n_1 = z^n_1)P(X^n_1 = x^n_1)}{\prod_{k=1}^nP(Y_k = y_k|X^n_1 = x^n_1)}
\]

\[
\geq 1 - \frac{P(Z^n_1 = z^n_1)P(X^n_1 = x^n_1)}{\prod_{k=1}^nP(Y_k = y_k)}. \tag{7}
\]

Remarkably, the technique of using an auxiliary channel to derive results is not new in the field. Several network information theory problems are tackled in this manner. For
instance, the characterisation of the sum-capacity of the Gaussian interference channel with weak interference, where a genie-aided channel is used to produce a converse [8]. There are also some recent results where similar ideas are utilized for finite blocklength analysis [9]. In the following sections, we produce tight and near-tight converse bounds for several channels using this method.

IV. q-ARY ERASURE CHANNEL

Erasure channels are considered as a first illustration of the general method of Section III since they are easy to manipulate in terms of mathematical analysis and intuition. When a q-ary symbol is transmitted over an erasure channel, the receiver obtains either the correct symbol with probability 1 − δ or an unknown symbol with probability δ.

Following the general method, the capacity of a channel with exactly s erasures is computed as follows,

\[ C_s = n - s \text{ q-ary symbols/channel use.} \]  

(8)

The proof can be found in Appendix A.

As mentioned previously, the average probability of error ε is lower bounded by the probability of erroneous transmission due to unsupported rates. In the special case of QEC, the channel state is a linear function of the number of erasures. This allows for an easily established upper bound on the non-asymptotic achievable rate of the QEC, which can be composed as a lower bound to average probability of error.

Theorem 2. For the q-ary erasure channel with erasure probability δ, the average error probability of a \((n,q^nR,\epsilon)\) code satisfies,

\[ \epsilon \geq \sum_{s=n-[nR]+1}^{n} \binom{n}{s} \delta^s (1-\delta)^{n-s} \left( 1-q^{-n-s-nR} \right). \]  

(9)

Proof. The channel states follow the binomial distribution as in [3],

\[ P(S = s) = \binom{n}{s} \delta^s (1-\delta)^{n-s}. \]

For \(nR > C_s\), the probability of error given the state is bounded based on (7) as follows,

\[ P(\text{error}|S = s) \geq 1 - \frac{P(Z_1^n = z_1^n)P(X_1^n = x_1^n)}{\prod_{k=1}^{n} P(Y_k = y_k)} = 1 - \frac{P(Z_1^n = z_1^n)P(X_1^n = x_1^n)}{P(Z_1^n = z_1^n) \prod_{k \in \mathcal{K}} P(X_k = x_k)}, \]

\[ = 1 - \frac{q^{-nR}}{q^{-n-s-nR}} = 1 - q^{n-s-nR}. \]

where \(\mathcal{K}\) is the set of the indices of the unerased symbols.

For the bounds of the summation, it holds from Corollary [1] and [8] that,

\[ nR > C_s = n - s \Rightarrow \]

\[ nR > n - s \Rightarrow \]

\[ s > n - nR \Rightarrow \]

\[ s > [n-nR] \Rightarrow \]

\[ n - [nR] + 1 \leq s \leq n. \]

Theorem 3. For the q-ary erasure channel with erasure probability δ, the average error probability of a \((n,q^nR,\epsilon)\) code satisfies,

\[ \epsilon \geq \sum_{s=n-[nR]+1}^{n} \binom{n}{s} \delta^s (1-\delta)^{n-s} \left( 1-q^{-n-s-nR} \right). \]  

(10)

\[ \text{for any constant } A > 0. \]

Proof. The proof is similar to the proof of Theorem 2 with different derivation of the lower bound on \(P(\text{error}|S = s)\). For \(nR > C_s\), the probability of error given the state is bounded based on Theorem [1] as follows,

\[ P(\text{error}|S = s) \geq 1 - \frac{4A}{((nR-\epsilon \ln(q))^2 - e^{-\frac{(nR-C_s-\epsilon \ln(q))}{2}} - e^{-\frac{(nR-C_s-\epsilon \ln(q))}{2}}) \cdot \ln(q)}. \]

for any constant \(A > 0\). Note that aggregate rate \(nR\) and state capacity \(C_s\) are converted to nats from q-ary symbols and the blocklength in Theorem [1] is set to 1 since n channel uses at
Remarkably, Theorem 2 is the same as the meta-converse bound [5]. It complies with the Singleton bound [10] and consequently can be achieved by maximum distance separable (MDS) codes. Note also that there are not any non-trivial binary MDS codes that can achieve this bound for the binary erasure channel (BEC) [11]. As a result, it is strictly greater than the achievable rate of BEC in non-trivial settings. Theorem 3 produces a slightly relaxed bound as it is depicted in Figure 3. This is expected since it does not consider the specific structure of each channel state.

V. BINARY SYMMETRIC CHANNEL

Binary symmetric channel is another important binary-input channel where bits are inverted during transmission with probability δ. Naturally, the differentiation of the states is based on the number of errors s. Hence, channel state capacities are measured as follows,

\[ C_s = n - \log_2 \left( \frac{n}{s} \right) \text{ bits/channel use.} \]

(11)

The proof can be found in Appendix B.

By manipulating (11) similarly to (8), the lower bounds to average probability of error are established. Formally,

**Theorem 4.** For the binary symmetric channel with error probability δ, the average error probability of a \((n, 2^{nR}, \epsilon)\) code satisfies,

\[ \epsilon \geq \sum_{s \in \mathcal{S}} \binom{n}{s} \delta^s (1 - \delta)^{n-s} \left( 1 - \left( \frac{n}{s} \right)^{-1} 2^{n-nR} \right), \]

(12)

where,

\[ \mathcal{S} = \left\{ s : \left( \frac{n}{s} \right) > 2^{n-nR} \wedge 0 \leq s \leq n \right\}. \]

**Proof.** The channel states follow the binomial distribution as in (4).

\[ P(S = s) = \binom{n}{s} \delta^s (1 - \delta)^{n-s}. \]

For \( nR > C_s \), the probability of error given the state is bounded based on (7) as follows,

\[ P(\text{error}|S = s) \geq 1 - \frac{1}{\prod_{k=1}^{n} P(Y_k = y_k)} \]

\[ = 1 - \left( \frac{n}{s} \right)^{-1} 2^{-nR} \]

\[ = 1 - \left( \frac{n}{s} \right)^{-1} 2^{n-nR}. \]

For the bounds of the summation, it holds from Corollary 1 and (11) that,

\[ nR > C_s = n - \log_2 \left( \frac{n}{s} \right) \Rightarrow \log_2 \left( \frac{n}{s} \right) > n - nR \Rightarrow \left( \frac{n}{s} \right) > 2^{n-nR}. \]

**Theorem 5.** For the binary symmetric channel with error probability δ, the average error probability of a \((n, 2^{nR}, \epsilon)\) code satisfies,

\[ \epsilon \geq \sum_{s \in \mathcal{S}} \binom{n}{s} \delta^s (1 - \delta)^{n-s} F(n, s, R, A), \]

(13)

\[ F(n, s, R, A) = 1 - \frac{4A}{(nR - n + \log_2 \left( \frac{n}{s} \right) \ln(2))^2} - e^{-\frac{\ln(2)}{2(nR - n + \log_2 \left( \frac{n}{s} \right))}}, \]

where,

\[ \mathcal{S} = \left\{ s : \left( \frac{n}{s} \right) > 2^{n-nR} \wedge 0 \leq s \leq n \right\}, \]

for any constant \( A > 0 \).

**Proof.** The proof is similar to the proof of Theorem 4 with different derivation of the lower bound on \( P(\text{error}|S = s) \). For \( nR > C_s \), the probability of error given the state is bounded based on Theorem 4 as follows,

\[ P(\text{error}|S = s) \geq 1 - \frac{4A}{(nR - C_s) \ln(2))^2} - e^{-\frac{(nR - C_s) \ln(2)}{2}} \]

\[ = 1 - \frac{4A}{(nR - n + \log_2 \left( \frac{n}{s} \right) \ln(2))^2} - e^{-\frac{\ln(2)}{2(nR - n + \log_2 \left( \frac{n}{s} \right))}}, \]

for any constant \( A > 0 \). Note that aggregate rate \( nR \) and state capacity \( C_s \) are converted to nats from bits and the blocklength in Theorem 4 is set to 1 since \( n \) channel uses at the main channel are equivalent to one channel use for the auxiliary. ■
Hence, channel state capacity (8) stands. Since (8) is always an integer, the criterion for the possibly supported rates can be the following.

\[ R_a \leq C_s \Rightarrow \lceil k \rceil = j - s. \]  (15)

If the \([k]\)th unerased symbol is received at the \(j\)th transmission, then the rate is supported and successful decoding might be possible. This process is described by the negative binomial distribution,

\[ P(j; \delta, \lceil k \rceil) = \binom{j - 1}{\lceil k \rceil - 1} \delta^{\lceil k \rceil} (1 - \delta)^{j - \lceil k \rceil}. \]  (16)

The probability of successful decoding during the \(m\)th packet is the following,

\[ P(m; \delta, n, \lceil k \rceil) = \sum_{j = mn - n + 1}^{mn} P(j; \delta, \lceil k \rceil) \]

\[ = \sum_{j = mn - n + 1}^{mn} \left( \frac{j - 1}{\lceil k \rceil - 1} \right) \delta^{\lceil k \rceil} (1 - \delta)^{j - \lceil k \rceil}. \]  (17)

Hence, the average number of required packets for successful transmission is bounded as follows,

\[ w_a \geq E[m] = \sum_{m \in \mathbb{Z}^+} m P(m; \delta, n, \lceil k \rceil) \Rightarrow \]

\[ l_a \geq n \sum_{m \in \mathbb{Z}^+} m P(m; \delta, n, \lceil k \rceil) \]  (18)

In [13], several bounds for the BEC with feedback are presented. The authors discuss the gap between their achievability [13 Theorem 7, Theorem 9] and converse [13 Corollary 6] bounds which increases for very short average blocklengths. By comparing their results with Theorem 6 in Figure 5 it is becoming apparent that this gap is fundamental since Theorem 6 coincides with the achievability bound [13 Theorem 9]. Additionally, Theorem 6 is generally tighter than converse [13 Corollary 6].

VII. Conclusions

The current paper presents novel results towards the characterization of the non-asymptotic achievable rate of memoryless discrete channels. A general derivation method of converse bounds is described for weakly symmetric channels and then it is particularized for the \(q\)-ary erasure channel, the binary symmetric channel, and the \(q\)-ary erasure channel with stop feedback. In the case of QEC, one of the derived bounds is identical to the meta-converse bound. Numerical evaluations for BSC show quick convergence to other state-of-the-art converse bounds, namely the meta-converse and sphere-packing bounds. For the QEC with stop feedback, a bound that improves on the state-of-the-art is produced and it coincides with an achievable bound in certain settings. The extension

The average blocklength of a \((l_a, q^k, n, 0)\) VLSF code over a \(q\)-ary erasure channel with erasure rate \(\delta\) is bounded as follows,

\[ l_a \geq n \sum_{m \in \mathbb{Z}^+} m \sum_{j = mn - n + 1}^{mn} \left( \frac{j - 1}{\lceil k \rceil - 1} \right) \delta^{\lceil k \rceil} (1 - \delta)^{j - \lceil k \rceil}. \]  (14)

Proof. Similarly to previous Sections, channel states are differentiated based on the number of erroneous transmission. Hence, channel state capacity (8) stands. Since (8) is always an integer, the criterion for the possibly supported rates can be the following.

\[ R_a \leq C_s \Rightarrow \lceil k \rceil = j - s. \]  (15)

Fig. 4. Converse and achievability bounds on the non-asymptotic achievable rate of BSC(0.05) for average error probability \(\epsilon = 10^{-3}\).
of this result to arbitrary and asymmetric discrete memoryless channels is particularly interesting. Additionally, other settings with variable length coding with feedback can be explored. These coding schemes are very promising in the aspect of low-latency communications, however complex bounds are not easily specialized. Auxiliary channels with feedback could be a valuable tool for a better characterization of their achievable rates.

ACKNOWLEDGMENT

This work is supported by the Engineering and Physical Sciences Research Council (EP/L016656/1); and the University of Bristol.

APPENDIX A

PROOF OF (11)

The random variable $S$ denotes the number of errors that occur after $n$ transmissions.

$$C_s = \max_{p(x_1, \ldots, x_n)} I(X_1^n; Y_1^n | S = s)$$

$$= \max_{p(x_1, \ldots, x_n)} (H(Y_1^n | S = s) - H(Y_1^n | X_1^n, S = s))$$

$$= \max_{p(x_1, \ldots, x_n)} (H(X_1^n + Z_1^n) - H(X_1^n + Z_1^n | X_1^n))$$

$$= \max_{p(x_1, \ldots, x_n)} (H(X_1^n + Z_1^n) - H(Z_1^n))$$

$$= n - \log_2 \left( \frac{n}{s} \right) \text{ bits/channel use.}$$

Vector $Z_1^n$ denotes a binary random vector of Hamming weight $s$. Generally, if $X_i$ follows a uniform distribution then $X_i + Z_i$ also follows a uniform distribution and its entropy is maximised.

REFERENCES

[1] C. E. Shannon, “A mathematical theory of communication,” Bell system technical journal, vol. 27, no. 3, pp. 379–423, 1948.
[2] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Channel coding rate in the finite blocklength regime,” IEEE Transactions on Information Theory, vol. 56, pp. 2307–2359, May 2010.
[3] J. Wolfowitz, “Notes on a general strong converse,” Information and Control, vol. 12, no. 1, pp. 1–4, 1968.
[4] C. Shannon, R. Gallager, and E. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels. I,” Information and Control, vol. 10, no. 1, pp. 65–103, 1967.
[5] G. Wiechman and I. Sason, “An improved sphere-packing bound for finite-length codes over symmetric memoryless channels,” IEEE Transactions on Information Theory, vol. 54, pp. 1962–1990, May 2008.
[6] Y. Polyanskiy, “Saddle point in the minimax converse for channel coding,” IEEE Transactions on Information Theory, vol. 59, no. 5, pp. 2576–2595, 2013.
[7] R. G. Gallager, Information Theory and Reliable Communication. USA: John Wiley & Sons, Inc., 1968.
[8] A. E. Gamal and Y.-H. Kim, Network Information Theory. USA: Cambridge University Press, 2012.
[9] J. Liu, “Dispersion bound for the Wyner-Ahlswede-Körner network via a semigroup method on types,” IEEE Transactions on Information Theory, vol. 67, no. 2, pp. 869–885, 2021.
[10] R. Singleton, “Maximum distance q-nary codes,” IEEE Transactions on Information Theory, vol. 10, pp. 116–118, April 1964.
[11] L. R. Vermani, Elements of Algebraic Coding Theory. Chapman & Hall Mathematics Series, Springer US, 1996.
[12] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Feedback in the non-asymptotic regime,” IEEE Transactions on Information Theory, vol. 57, no. 8, pp. 4903–4925, 2011.
[13] R. Devassy, G. Durisi, B. Lindqvist, W. Yang, and M. Dalai, “Nonasymptotic coding-rate bounds for binary erasure channels with feedback,” in 2016 IEEE Information Theory Workshop (ITW), pp. 86–90, 2016.

APPENDIX B

PROOF OF (11)

The random variable $S$ denotes the number of errors that occur after $n$ transmissions.

$$C_s = \max_{p(x_1, \ldots, x_n)} I(X_1^n; Y_1^n | S = s)$$

$$= \max_{p(x_1, \ldots, x_n)} (H(Y_1^n | S = s) - H(Y_1^n | X_1^n, S = s))$$

$$= \max_{p(x_1, \ldots, x_n)} (H(X_1^n + Z_1^n) - H(X_1^n + Z_1^n | X_1^n))$$

$$= \max_{p(x_1, \ldots, x_n)} (H(X_1^n + Z_1^n) - H(Z_1^n))$$

$$= n - \log_2 \left( \frac{n}{s} \right) \text{ bits/channel use.}$$

Vector $Z_1^n$ denotes a binary random vector of Hamming weight $s$. Generally, if $X_i$ follows a uniform distribution then $X_i + Z_i$ also follows a uniform distribution and its entropy is maximised.