Good Codes Singleton Bound

David Ponarovsky

October 7, 2022

Abstract

We propose a new asymptotic upper bound on the trade-off between the rate and the distance of good error correction codes.

1 Preambles

Coding theory has emerged by the need to transfer information in noisy communication channels. By embedding a message in higher dimension space, one can guarantee robustness against possible faults. The ratio of the original content length to the passed message length is the rate of the code, and it measures how consuming our communication protocol is. Furthermore, the distance of the code quantifies how many faults the scheme can absorb such that the receiver could recover the original message.

Even though it is obvious that any construction resilient to a large number of faults should have a complexity price, The exact relation between the rate and the code distance is still unknown. However, we do know un-tight upper bounds. The first one was proved by Singelton and set the linear constraint: $\rho + \delta \leq 1 - \frac{1}{\Delta}$ for any $[\Delta, \rho\Delta, \delta\Delta]$ linear code [Sin64].

Non-formally, we say that code is good if its distance and rate are scaled linearly in the encoded message length. Besides the fact that good codes are considered efficient in terms of robustness and overhead, they are also vital components in establishing secure multiparty computation [BGW19] and have a deep connection to probabilistic proofs.

In this work, we show a new upper bound $\rho + \frac{5}{16}\delta \leq \frac{3}{4} + \frac{1}{16}$ which tighter than the Singleton bound and holds for any Good code. First, we state the notations, definitions, and formal theorem in section 2. Then in sections 3 and 4, we review past results and provide their proofs in order to make this paper self-contained. Readers familiar with the basic concepts of LDPC codes and the Tanner code construction should consider skipping directly to section 5, in which we provide our proof.

2 Notations, Definitions, And Our Contribution

Here we focus only on linear binary codes, which one could think about as linear subspaces of $\mathbb{F}_2^n$. A common way to measure resilience is to ask how many bits an evil entity needs to flip such that the corrupted vector will be closer to another vector in that space than the original one. Those ideas were formulated by Hamming [Ham50], who presented the following definitions.

Definition. Let $n \in \mathbb{N}$ and $\rho, \delta \in (0,1)$. We say that C is a *binary linear code* with parameters $[n, \rho n, \delta n]$. If C is a subspace of $\mathbb{F}_2^n$, and the dimension of C is at least ρn. In addition, we call the vectors belong to C *codewords* and define the distance of C to be the minimal number of different bits between any codewords pair of C.

From now on, we will use the term code to refer to linear binary codes, as we don't deal with any other types of codes. Also, even though it is customary to use the above parameters to analyze codes, we will use their percent forms called the relative distance and the rate of code, matching δ and ρ correspondingly.

Definition. A *family of codes* is an infinite series of codes. Additionally, suppose the rates and relative distances converge into constant values ρ, δ. In that case, we abuse the notation and call that family of codes a code with $[n, \rho n, \delta n]$ for fixed $\rho, \delta \in [0,1)$, and infinite integers $n \in \mathbb{N}$.

Notice that the above definition contains codes with parameters attending to zero. From a practical view, it means that etiher we send too many bits, more than a constant amount, on each bit in the original message. Or that for big enough n, adversarial, limited to changing only a constant fraction of the bits, could disrupt the transmission. That distinction raises the definition of good codes.

Definition. We will say that a family of codes is a *good code* if its parameters converge into positive values.

Nowadays, we are aware of a wide range of constructions yield good codes, including the expander codes of Sipser and Spilman [SS96] and the LTC codes of Dinur [Din+22], [PK21], [LZ22]. Thus if, a decade ago, the main question was how to construct a good code, then now it is replaced by what is the best code one could expect.

To get a feeling of the behavior of the distance-rate tradeof, Let us consider the following two codes; each demonstrates a different extreme case. First, define the repetition code $C_r \subset \mathbb{F}_2^{n \cdot r}$, In which, for a fixed integer r, any bit of the original string is duplicated r times. Second, consider the parity check code $C_p \subset \mathbb{F}_2^{n+1}$, which it's codewords are only the vectors with even parity. Let us analyze the repetition code. Clearly, any two n-bits different messages must have at least a single different bit. Therefore their corresponding encoded codewords have to differ in at least r bits. Hence, by scaling r, one could achieve a higher distance as he wishes. Sadly the rate of the code decays as $n/nr = 1/r$. In contrast, the parity check code adds only a single extra bit for the original message. Therefore scaling n gives a family which has a rate attends to $\rho \to 1$. However, flipping any two different bits of a valid codeword is conversing the parity and, as a result, leads to another valid codeword.

To summarize the above, we have that, using a simple construction, one could construct the codes $[r, 1, r]$, $[r, r - 1, 2]$. Each has a single perfect parameter while the other decays to the worst. In the next section, we will review the Singleton bound, which states that for any code (not necessarily good), there must be a zero-sum game between the relative distance and the rate. Now, we are ready to formulate our contribution.

Theorem 1 Let C be a family of good codes, with parameters $[n, \rho n, \delta n]$. Then:

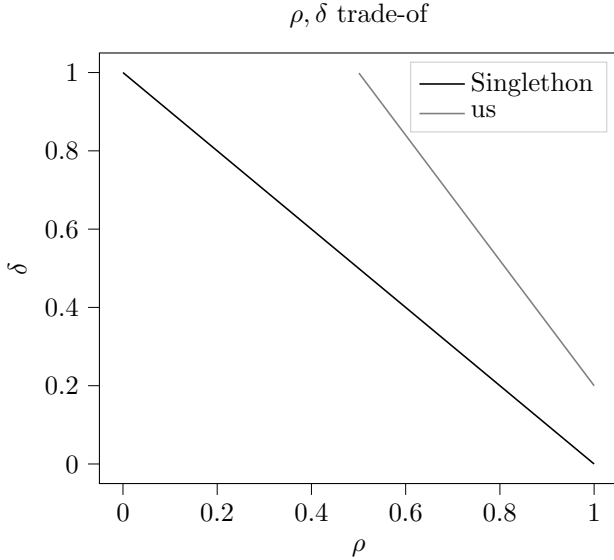$$\rho + \frac{5}{16}\delta \le \frac{3}{4} + \frac{1}{16}$$

Figure 1: Comparing our bound, in black, to Singleton bound, in gray.

3 Singleton Bound

Besides of been the first bound, Singleton bound demonstrates how one could get results by using relatively simple elementary arguments. It is also engaging to ask why the proof yields a bound that, empirically, seems far from being tight.

Theorem, Singleton Bound.   For any linear code with parameter $[n, k, d]$, the following inequality holds:

$$k + d \le n + 1$$

Proof: Since any two codewords of C differ by at least d coordinates, we know that by ignoring the first $d - 1$ coordinate of any vector, we obtain a new code with one-to-one corresponding to the original code. In other words, we have found a new code with the same dimension embedded in $\mathbb{F}_2^{n-d+1}$. Combine the fact that dimension is, at most, the dimension of the container space, we get that:

$$\dim C = 2^k \le 2^{n-d+1} \Rightarrow k + d \le n + 1$$

$\square$

It is also well known that the only binary codes that reach the bound are: $[n, 1, n]$, $[n, n - 1, 2]$, $[n, n, 1]$ [AF22]. In particular, there are no good binary codes that obtain equality.

4 Tanner Code

5 Construction

6 Comparing To Known Bounds In A Variety Of Regimes

References

[Ham50]   R. W. Hamming. "Error detecting and error correcting codes". In: *The Bell System Technical Journal* 29.2 (1950), pp. 147–160. DOI: 10.1002/j.1538-7305.1950.tb00463.x.

[Sin64]   R. Singleton. "Maximum distanceq-nary codes". In: *IEEE Transactions on Information Theory* 10.2 (Apr. 1964), pp. 116–118. ISSN: 1557-9654. DOI: 10.1109/TIT.1964.1053661.

[SS96]   M. Sipser and D.A. Spielman. "Expander codes". In: *IEEE Transactions on Information Theory* 42.6 (1996), pp. 1710–1722. DOI: 10.1109/18.556667.

[BGW19]   Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. "Completeness Theorems for Non-Cryptographic Fault-Tolerant Distributed Computation". In: *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 351–371. ISBN: 9781450372664. URL: https://doi.org/10.1145/3335741.3335756.

[PK21]   Pavel Panteleev and Gleb Kalachev. *Asymptotically Good Quantum and Locally Testable Classical LDPC Codes*. 2021. DOI: 10.48550/ARXIV.2111.03654. URL: https://arxiv.org/abs/2111.03654.

[AF22]   "Maximum distance separable (MDS) code". In: *The Error Correction Zoo*. Ed. by Victor V. Albert and Philippe Faist. 2022. URL: https://errorcorrectionzoo.org/c/mds.

[Din+22]   Irit Dinur et al. *Good Locally Testable Codes*. 2022. DOI: 10.48550/ARXIV.2207.11929. URL: https://arxiv.org/abs/2207.11929.

[LZ22]   Anthony Leverrier and Gilles Zémor. *Quantum Tanner codes*. 2022. arXiv: 2202.13641 [quant-ph].