15:848: Practical Information and Coding Theory for Computer Systems
Fall 2018

Lecture 7

September 26 2019

Lecturer: Rashmi Vinayak *Scribe: Hojin Park*

7.1 Overview

In this lecture, we covered the following topics:

1. Convergence of error correcting algorithm in LPDC codes
2. Tornado codes
3. Fountain codes - LT codes

7.2 LPDC code's decoding algorithm (slide 9, 10)

Let's show that for (α, $3d/4$) expansion, if number of error bits is less than $\alpha n/4$, the simple decoding algorithm of LPDC codes converges.

Let u_i: number of unsatisfied check bits on step i, r_i: number of corrupt code bits on step i, and s_i: number of satisfied check bits with corrupt neighbors on step i.
If we show that for $\forall i, r_i < \alpha n/2$, the statement is proved, because from the last lecture, we learned that for (α, $3d/4$) expansion, the distance of code is greater than αn. You may easily get the intuition from Figure 7.1.

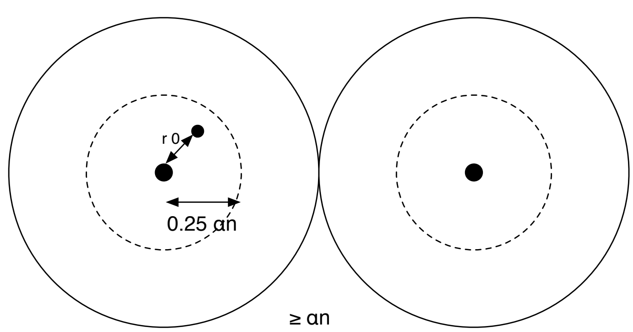

Figure 7.1. Distance between two LPDC codes and r_0

Since the theorem says that $r_0 < \alpha n/4$, it is enough to show that $\forall i, r_i < 2r_0$. From the algorithm, it is trivial that u_i decrements on each step, which is (1) $u_i < u_0$. Also, by the expansion assumption, (2) $u_i + s_i > 3/4 * dr_i$ because the unsatisfied check bits and the satisfied check bits with corrupt neighbors are all connected with corrupt code bits. In addition, if we count the number of all the out going edges of the corrupt code bits, it is dr_0 because it's d-regular LPDC. From the edge counting, dr_0 should be greater than or equal to $2s_i + u_i$ ($\because$ the unsatisfied check bits have at least one corrupt neighbor and the satisfied check bits with corrupt neighbors have at lease two corrupt neighbors to be satisfied). Thus, (3) $2s_i + u_i \leq dr_i$. From (1), (2), (3), $3/2 * dr_i - 2 * u_i < 2s_i \leq dr_i - u_i \Rightarrow 1/2 * dr_i < u_i < u_0 \leq dr_0 \Rightarrow r_i < 2r_0$.

7.3 Probabilistic channel model (slide 11)

Binary symmetric channel (BSC) is a channel that converts 0 to 1 or 1 to 0 with probability p while preserving the original value with probability *1-p*. In hard decision decoding, a 0-1 value is computed for each bit. On the other hand, in soft-decision decoding, typically, a probabilistic channel model is used (e.g., the BSC), and instead of a 0-1 value, probabilities are computed for each of the bit values.

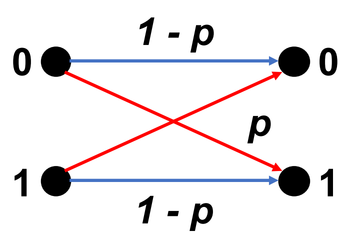

Figure 7.2. Binary symmetric channel

7.4 Encoding LPDC codes (slide 13)

Question. Can we use matrix multiplication of generating matrix G from Y for encoding?

Answer. The density of parity check matrix H is very low because we assumed a small degree bipartite graph. It guarantees efficient decoding. However, it does not guarantee a low density of a generator matrix G, which results in an inefficient encoding.

7.5 The random erasure model (slide 16)

Question. Assume a random erasure model where each bit is erased with probability p. Then, would a random linear code with rate less than *1-p* work?

Answer. Yes. Recall that the definition of the rate of a linear code is $\frac{k}{n}$. Thus, despite the fact that bits are erased with probability p, k bits are still available with high probability. Note that among the n bits of the codewords, a (1-p) fraction of them will arrive, and we call the (1-p) *Channel Capacity.*

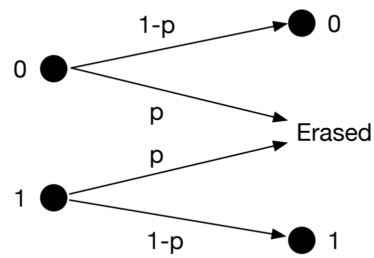

Figure 7.3. Random Erasure Model

7.6 Encoding tornado codes (slide 19)

Question. Why is encoding time of tornado codes linear?

Answer. We can compute parity bits of a message by counting the number of its neighbors and applying modulo 2 operation. This is the same complexity with the number of edges in a d-left-regular bipartite graph, $k \times d$, which is a linear with a small d value.

7.7 Fountain codes (slide 28)

Question. What is the reception overhead and the decoding failure probability of RS codes?

Answer. If there are k received symbols, it is always possible to decode encoded RS codewords. Thus, both the reception overhead and the probability of failure to decode are 0.

7.8 LT codes: Decoding (slide 32)

Question. What does $degree = 1$ mean in a decoding algorithm of LT codes?

Answer. If you think about the encoding steps, a degree means the number of original message symbols that are connected to the coded symbol. So if the coded symbol has a $degree = 1$, the number of message symbols that are linked with the coded symbol is 1, meaning that the linked message symbol has the same symbol as the coded symbol.

7.9 Degree distribution (slide 36)

Question. What is the expected reception overhead of LT codes, if we use a "one-by-one" distribution as a degree distribution?

Answer. The expected reception overhead is $k \cdot ln(k)$. You can associate this problem with the "coupon collector's problem." If we use a "one-by-one" distribution for encoding, to decode the codewords into the original message symbols, all the original message symbols should appear at least once in the encoded codewords. Thus the expected reception overhead of LT codes can be interpreted as the number of trials randomly choosing a symbol from the original message symbols until all the original message symbols appear at least once, which is the same problem with the "coupon collector's problem."