ERC Report - February 19, 2010

Amir Hesam Salavati
E-mail hesam.salavati@epfl.ch

Supervisor Prof. Amin Shokrollahi
E-mail amin.shokrollahi@epfl.ch
Algorithmics Laboratory (ALGO)
Ecole Polytechnique Federale de Lausanne (EPFL)

May 5, 2010

1 Introduction

In the past week, I also worked with Raj on the phase transition problem where I tried to derive the fraction of check nodes with degree 2, i.e. Ω_2. In section III, you will find the details of our approach on finding Ω_2.

2 In Search of Ω_2

In this section, we are going to calculate the fraction of check nodes with degree 2, i.e. Ω_2, in a code whose generator matrix is composed of random k-tuples with weight d. In other words, the $k \times n$ generator matrix G is a random binary matrix whose columns have weight d.

Therefore, one can build G by picking n vectors uniformly at random from the pool of binary vectors with weight d. Note that this is a little bit different from the Kolchin's ensemble.

We will compute Ω_2 as n goes to infinity using two different approaches. Both approaches, however, leads us to the fact that as n grows to infinity, Ω_2 tends to zero.

2.1 Approach 1

To find out Ω_2, we first build the parity check matrix by choosing $n-k$ binary vectors from the set of 2^{n-k} vectors v that lie into left null space of G^T, i.e. satisfy $v.G^T = 0$. Then, we calculate the expected fraction of rows with weight 2 in H which gives us Ω_2. Let ¶ denote the probability of having a full rank $(n-k) \times n$ matrix, given that we sample from a reduced set of 2^{n-k} vectors, with replacement. Then, by multiplying ¶ and the probability of having i rows in H with weight two, we approximately get $E(\Omega_2)$ over all possible and valid choices of H.

Let's denote the probability of having **exactly** i rows with weight 2 in H by P_i. Moreover, let N is the number of vectors h with weight 2 that lie in the left null space of G^T, i.e. $\{h.G^T = 0, |h| = 2\}$. N could be approximated to be $\binom{n}{2}.p$ where p is the probability of having a vector h with weight 2 to lie in the left null space of G^T. Obviously, p is a function of d, the weight of each column of G. Now, the probability of having exactly i rows with weight 2 in a valid parity check matrix, H, is given by equation (1).

$$P_i = \frac{N^i(2^{n-k} - N)^{n-k-i}}{2^{(n-k)^2}} \tag{1}$$

The above equation is derived as follows: the total number of ways in which one can build a binary random $(n-k) \times n$ matrix out of a pool of 2^{n-k} n-tuples is $2^{(n-k)^2}$. Because for each row, we have 2^{n-k} possibilities and we have $n-k$ rows. Likewise, the total number of ways according to which one can choose i vectors out of a pool of N vectors with weight 2 is N^i. The rest of $n-k-i$ vectors can be selected from a pool of $2^{n-k} - N$ vectors in $(2^{n-k} - N)^{n-k-i}$ ways.

As a result of equation (1), $E(\Omega_2)$ is computed as follows:

$$\begin{aligned}
(n-k)E(\Omega_2) &= \sum_{i=0}^{n-k} iP_i¶ = \sum_{i=0}^{n-k} i¶\frac{N^i(2^{n-k} - N)^{n-k-i}}{2^{(n-k)^2}} \\
&\leqslant \sum_{i=0}^{n-k} i\frac{N^i(2^{n-k} - N)^{n-k-i}}{2^{(n-k)^2}} \\
&\leqslant \sum_{i=0}^{n-k} i\frac{N^i(2^{n-k})^{n-k-i}}{2^{(n-k)^2}}
\end{aligned}$$

$$\begin{aligned} &= \sum_{i=0}^{n-k} i \frac{N^i}{2^{i(n-k)}} \\ &= \sum_{i=0}^{n-k} i (\frac{N}{2^{(n-k)}})^i \end{aligned} \qquad (2)$$

By denoting $\frac{N}{2^{(n-k)}}$ with x, the last summation in equation (2) simplifies to $\sum_{i=0}^{n-k} ix^i$ which is equal to $\frac{x}{(x-1)^2}(nx^{n+1} - (n+1)x^n + 1)$. Therefore, we obtain the following upper bound for $E(\Omega_2)$:

$$E(\Omega_2) \leqslant \frac{x}{(n-k)(x-1)^2}(nx^{n+1} - (n+1)x^n + 1) \qquad (3)$$

As n goes to infinity, $N = O(n^2)$ and grows much slower than the exponential factor in denominator of x. Hence, x tends to zero which means that $\boxed{\lim_{n\to\infty} E(\Omega_2) = 0}$

2.2 Approach 2

In another attempt to calculate Ω_2, we focus on the generator matrix, G. Having a row with weight 2 in H is equivalent to having two similar rows in G^T. Therefore, if we have two similar rows in G^T, we can have **at most** one row with weight 2 in H. Hence, by calculating the number of equal rows in G^T, we can drive an *upper bound* for Ω_2. For instance, if we have four equal rows in G^T, we can have either two or three **independent** rows with weight 2 in H. Therefore, we can have at most three rows with weight 2 in H and, as a result, $\Omega_2 \leqslant 3/(n-k)$ in this case.

Now let's calculate the probability of having at most i independent rows with weight 2, denoted by N_i. For $i = 1$, this is equivalent to having **exactly** two equal rows in G^T. Therefore, we need to compute the probability of having exactly two equal rows in G^T. To find out this number, we fix the first row and note that we have $\binom{k}{d}$ possibilities for this row (recall the ensemble we are using as explained in the introduction). Then, we can choose any one of the $n-1$ remaining rows to be equal to this row. Suppose we have selected the second row to be equal to the first one. Thus, we have $\binom{k}{d} - 1$ possibilities for the third row, $\binom{k}{d} - 2$ possibilities for the fourth row and so on. Hence, the probability of having exactly two equal rows in G^T is[1]

$$P_2^1 = \frac{(n-1)(\binom{k}{d})(\binom{k}{d}-1)\dots(\binom{k}{d}-(n-2))}{\binom{k}{d}^n} \tag{4}$$

Now, what is the probability of having at most two rows with weight two in H? To find out, we must calculate the probability of having either three equal rows or two pairs of equal rows in G^T. The first probability accounts for rows of H of the form $(11000);(10100)$ -meaning that the first, second and third rows in G^T are equal, and the second one accounts for the vectors of the form $(11000);(00110)$, which means that the first and second rows are equal as well as the third and the fourth rows while these two pairs are different from each other.

The number of ways in which we can have three equal rows in G^T is calculated as follows: we fix the first row as usual for which we have $\binom{k}{d}$ possibilities. Then we choose the other two rows in $\binom{n-1}{2}$ ways and make them equal to the first row. For the rest of the rows we have $(\binom{k}{d}-1)(\binom{k}{d}-2)\dots(\binom{k}{d}-(n-3))$ possibilities. Hence the probability of having three equal rows, P_3^1, is:

$$P_3^1 = \frac{\binom{n-1}{2}(\binom{k}{d})(\binom{k}{d}-1)\dots(\binom{k}{d}-(n-3))}{\binom{k}{d}^n} \tag{5}$$

All remains to do now is to find out the probability of having two pairs of equal rows in G^T, P_2^2. As usual, we fix the first row for which we have $\binom{k}{d}$ options. We also fix the second row to be unequal to the first one, with $(\binom{k}{d}-1)$ possible choices. Now, we select another row with $(n-2)$ possibilities and equate it to the first row. We do the same for the second row with $(n-3)$ options. Hence, the probability P_2^2 is given by the following equation:

$$P_2^2 = \frac{(n-2)(n-3)(\binom{k}{d})(\binom{k}{d}-1)\dots(\binom{k}{d}-(n-3))}{\binom{k}{d}^n} \tag{6}$$

Based on equations (5) and (6), the probability of having at most two rows with weight two in H is

$$P_2 = P_3^1 + P_2^2 = \frac{(\binom{n-1}{2}+2\binom{n-2}{2})(\binom{k}{d})(\binom{k}{d}-1)\dots(\binom{k}{d}-(n-3))}{\binom{k}{d}^n} \tag{7}$$

[1] Notation: P_i^j indicates the probability of having j pairs of i equal rows. For example, P_2^1 is the probability of having two equal rows and P_3^1 is the probability of having three equal rows. On the other hand, P_2^2 is the probability of having two pairs of equal rows.

Note that as n goes to infinity, the P_1 is of the form of $O(n)\Pi_{j=0}^{n-2}(\binom{k}{d}-j)/\binom{k}{d}^n$ and P_2 is in the form of $O(n^2)\Pi_{j=0}^{n-3}(\binom{k}{d}-j)/\binom{k}{d}^n$.

We can continue this way to calculate $P_3, P_4, \ldots, P_n$. We can show that P_i have the form of $O(n^i)\Pi_{j=0}^{n-i-1}(\binom{k}{d}-j)/\binom{k}{d}^n$. We can replace $O(n^i)$ with $\alpha_i n^i$ where α_i is a constant. Having done that, we can calculate the upper bound on the expected number of rows with 2 in H as follows:

$$\begin{aligned}(n-k).\Omega_2 &\leqslant \sum_{i=0}^{n} iP_i \simeq \sum_{i=0}^{n} i\frac{\alpha_i n^i \Pi_{j=0}^{n-i-1}(\binom{k}{d}-j)}{\binom{k}{d}^n} \\ &\leqslant \alpha\sum_{i=0}^{n} i\frac{n^i \Pi_{j=0}^{n-i-1}(\binom{k}{d}-j)}{\binom{k}{d}^n} \end{aligned} \tag{8}$$

where $\alpha = \max_i \alpha_i$. Now we have:

$$\begin{aligned}(n-k).\Omega_2 &\leqslant \alpha\sum_{i=0}^{n} i\frac{n^i \Pi_{j=0}^{n-i-1}(\binom{k}{d})}{\binom{k}{d}^n} \\ &= \alpha\sum_{i=0}^{n} i\frac{n^i \binom{k}{d}^{n-i}}{\binom{k}{d}^n} = \sum_{i=0}^{n} i(\frac{n}{\binom{k}{d}})^i \end{aligned} \tag{9}$$

Denoting $\frac{n}{\binom{k}{d}}$ with x, the above sum simplifies to:

$$(n-k).\Omega_2 \leqslant \alpha\frac{x}{(x-1)^2}(nx^{n+1}-(n+1)x^n+1) \tag{10}$$

Therefore, we get the following upper bound for Ω_2:

$$\Omega_2 \leqslant \frac{\alpha}{n-k}\frac{x}{(x-1)^2}(nx^{n+1}-(n+1)x^n+1) \tag{11}$$

As n goes to infinity, and by assuming $k = rn$, where r is the code rate, then x goes to zero for $d > 1$. Therefore, the upper bound in equation (11) also vanishes. This means that: $\boxed{\lim_{n\to\infty}\Omega_2 = 0}$