Detailed Network Measurements Using Sparse Graph Counters: The Theory

Yi Lu, Andrea Montanari and Balaji Prabhakar

Abstract— **Measuring network flow sizes is important for tasks like accounting/billing, network forensics and security. Per-flow accounting is considered hard because it requires that many counters be updated at a very high speed; however, the large fast memories needed for storing the counters are prohibitively expensive. Therefore, current approaches aim to obtain approximate flow counts; that is, to detect large *elephant* flows and then measure their sizes.**

Recently the authors and their collaborators have developed [1] a novel method for per-flow traffic measurement that is fast, highly memory efficient and accurate. At the core of this method is a novel counter architecture called "counter braids." In this paper, we analyze the performance of the counter braid architecture under a Maximum Likelihood (ML) flow size estimation algorithm and show that it is optimal; that is, the number of bits needed to store the size of a flow matches the entropy lower bound. While the ML algorithm is optimal, it is too complex to implement. In [1] we have developed an easy-to-implement and efficient message passing algorithm for estimating flow sizes that is analyzed elsewhere.

I. Introduction

This paper addresses a theoretical problem arising in a novel approach to network traffic measurement the authors and their collaborators have recently developed. We refer the reader to [1] for technological background, motivation, related literature and other details. In order to keep this paper self-contained, we summarize the background and restrict the literature survey to what is relevant for the results of this paper.

Background. Measuring the sizes of network flows on high speed links is known to be a technologically challenging problem [2]. The nature of the data to be measured is as follows: At any given time several 10s or 100s of thousands of flows can be active on core Internet links. Packets arrive at the rate of one in every 40-50 nanoseconds on these links which currently run at 10 Gbps. Finally, flow size distributions are heavy-tailed, giving rise to the well-known decomposition of flows into a large number of short "mice" and a few large "elephants." As a rule of thumb, network traffic follows an "80-20 rule": 80% of the flows are small, and the remaining 20% of the large flows bring about 80% of the packets or bytes.

This implies that measuring flow sizes accurately requires a large array of counters which can be updated at very high speeds, and a good counter management algorithm for updating counts, installing new counters when flows initiate and uninstalling them when flows terminate.

Since high-speed large memories are either too expensive or simply infeasible in the current technology, the bulk of research on traffic measurement has focused on approximate counting methods. These approaches work aim at detecting elephant flows and measure their sizes.

Counter braids. In [1] we develop a novel counter architecture, called "counter braids", which is fast, very efficient with memory use and gives an accurate measurement of *all* flow sizes, not just the elephants. We will briefly review this architecture using the following simple example.

Suppose we are given 5 numbers and are told that four of them are no more than 2 bits long while the fifth can be 8 bits long. We are not told which is which!

Figures 1 and 2 present two approaches for storing the values of the 5 numbers. The first one corresponds to a traditional array of counters, whereby the same number of memory registers is allocated to each measured variable (flow). The structure in Fig. 2 is more efficient in memory, but retrieving the count values is less straightforward, requiring a flow size estimation algorithm.

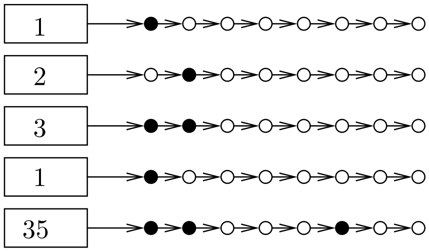

Fig. 1. A simple counter structure: to each flow size we associate its binary representation (filled circle = 1, empty circle = 0).

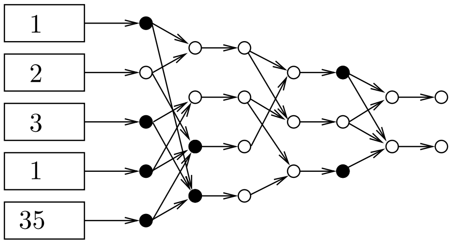

Fig. 2. Counters braid.

Viewed from an information-theoretic perspective, the design of an efficient counting scheme and a good flow size estimation is equivalent to the design of an efficient *source*

Yi Lu is with the Department of Electrical Engineering, Stanford University, yi.lu@stanford.edu. Andrea Montanari is with Departments of Electrical Engineering and Statistics, Stanford University, montanari@stanford.edu. Balaji Prabhakar is with the Department of Electrical Engineering, Stanford University, balaji@stanford.edu.

code [3]. However, the applications we consider impose a stringent constraint on such a code: each time that the size of a flow changes (because a new packet arrives) a small number of operations must be sufficient to update the stored information. This is not the case with standard source codes, where changing a single letter in the source stream may alter completely the compressed version.

In this paper we prove that, under a probabilistic model for the flow sizes (namely that they form a vector of iid random variables), counter braids achieve a compression rate equal to the entropy of the flow sizes distribution, in the large system limit. Namely, for any rate larger than the flow entropy, the flow sizes can be recovered from the counters values, with error probability vanishing in the large system limit. Further, we prove optimal compression can be achieved by using braids that are *sparse*. The result is non-obvious, since counter braids form a pretty restrictive family of architectures.

Our treatment makes use of techniques from the theory of low-density parity check codes, and the whole construction is inspired to LDPC's [4], [5]. These have an analogous in the source coding problem thanks to standard equivalence between coding over discrete memoryless symmetric channels, and compressing iid discrete random variables [6]. However, the key ideas in the present paper have been developed to deal with the problem that the flow sizes are *a priori* unbounded. In the channel coding language, this would be equivalent to use a countable but infinite input alphabet.

Finally, we insist on using sparse braids for two reasons. First, this allows the stored values to be updated with a *small* (typically bounded) number of operations. Second, it is easy to realize that ML decoding of counter braids is NP-hard, since it has ML decoding of linear codes as a special case [7]. However, thanks to the sparseness of the underlying graph, one may use iterative message passing techniques [8]. Indeed, a simple message passing algorithm for estimating flow sizes is described and analyzed using real and synthetic network traces in [1].

II. Counter Braids: Basic definitions

Definition 1. *A* counter braid *is a couple (G,q) where $q \geq 2$ is an integer (register capacity) and G is a directed acyclic graph on vertex sets I (input nodes) and R (registers), with the input nodes having in-degree zero. We write $G = (I,R,E)$, with E the set of directed edges in G.*

For any node $i \in I \cup R$, we will denote by $\partial_+ i \equiv \{j : (i,j) \in E\}$ the set of descendants of i, and by $\partial_- i \equiv \{j : (j,i) \in E\}$ the set of parents of i. Finally $\partial i \equiv \partial_+ \cup \partial_- i$.

In the following we shall often omit the explicit reference to the register capacity and write G for (G,q). The input size of the braid is $|I| \equiv n$, and its storage size $|R| \equiv m$. An important parameter is its rate, which we measure in bits

$$r = \frac{|R| \log_2 q}{|I|}\,. \tag{1}$$

We will say that a sequence of counters braids $\{G_n = (I_n,R_n,E_n)\}$ is *sparse* if the number of edges per input node $|E_n|/|I_n|$ is bounded.

Definition 2. *A* state *(or* configuration *of the counter braid G_q, with is an assignment (x,y) of non-negative integers to the nodes in G, with $x = \{x_i : i \in I\} \in \mathbb{N}^I$, and $y = \{y_j : j \in R\} \in \mathbb{N}^R$. The state (x,y) is* valid *if $y_j \in \{0,\dots,q-1\}$ for any register $j \in R$.*

Notice that a valid register configuration can be regarded as an element of $(\mathbb{Z}_q)^R$ (where $\mathbb{Z}_q$ is the group of integers modulo q.) We denote by 0 the zero vector in $\mathbb{N}^K$.

We want now to describe the braid behavior when one of the input nodes is incremented by one unity (i.e. when a packet arrives at input node $i \in I$.) Assume the braid (G,q) to be in a valid state (x,y). Given $i \in I$, we define the new state $(x',y') = \mathsf{T}_i(x,y)$ by letting $x'_i = x_i + 1$, $x'_j = x_j$ for any $j \neq i$, and y' be defined by the following procedure. Notice that this definition is ambiguous in that we did not specify which register to pick among the ones with $y_j(t) \geq q$ at step 4 in the registers update routine. However this is not necessary, as stated in the following lemma (the proof is omitted from this extended abstract).

```
REGISTERS UPDATE (INPUT: flow index $i$)
1:   $y_j(0) = y_j$ for $j \notin \partial_+ i$,
       and $y_j(0) = y_j + 1$ otherwise.
2:   Set $t = 0$.
3:   while $y(t)$ is not valid
4:       Let $j \in R$ be such that $y_j(t) \geq q$;
5:       Set $y_j(t+1) = y_j(t) - q$;
6:       For any $l \in \partial_+ j$, set $y_l(t+1) = y_l(t) + 1$;
7:       For any $l \in R \setminus \{j, \partial_+ j\}$, set $y_l(t+1) = y_l(t)$;
8:       Increment $t := t + 1$;
9:   end
10:  return $y(t)$.
```

Lemma 1. *The update procedure above halts after a finite number of steps. Further its output $\mathsf{T}_i(x,y)$ does not depend on the order of update of the registers.*

With an abuse of notation we shall write $x' = \mathsf{T}_i(x)$, $y' = \mathsf{T}_i(y)$, when $(x',y') = \mathsf{T}_i(x,y)$.

When input values x are incremented sequentially, the stored information y is updated according to the above procedure. From now on we shall take a static view and assume a certain input x. The corresponding stored information y is obtained through the mapping defined below.

Definition 3. *Given a counter braid (G,q), the associated* storage function *$\mathsf{F}_G : \mathbb{N}^I \to \mathbb{Z}_q^R$ returns, for any input configuration $x \in \mathbb{N}^I$ a register configuration $y = \mathsf{F}_G(x) \in \mathbb{Z}_q^R$ defined as follows. Let $x^{(0)} = 0$, $x^{(1)},\dots,$ $x^{(N)} = x$ be a sequence of input configurations such that $x^{(s+1)}$ is obtained from $x^{(s)}$ by incrementing its entry $i(s)$. Then*

$$\mathsf{F}_G(x) \equiv \mathsf{T}_{i(N)} \circ \mathsf{T}_{i(N-1)} \circ \cdots \circ \mathsf{T}_{i(1)}(0)\,. \tag{2}$$

We shall drop the subscript G from F_G whenever clear from the context. A priori it is not obvious that the mapping

F_G is well defined. In particular, it is not obvious that it does not depend on the order in which input values are incremented, i.e. on the sequence $\{i(1),\dots,i(N)\}$. This is nevertheless the case (the proof is omitted.)

Definition 4. *Given a counter braid* (G,q), *a* reconstruction *(or* decoding*)* function *is a function* $\widehat{\mathsf{F}}:\mathbb{Z}_q^R\to\mathbb{N}$.

A. Main results

Throughout this paper, we shall model the input values as iid integer random variables $(X_1,\dots,X_n)\equiv X$ (identifying $V=[n]$) with common distribution p. The (binary) entropy of this distribution will be denoted by $H_2(p)\equiv -\sum_x p(x)\log_2 p(x)$.

Definition 5. *A sequence of counters braids* $\{G_n=(I_n,R_n,E_n)\}$, *with* $|I_n|=n$ *has* design rate r *if*

$$r=\lim_{n\to\infty}\frac{|R_n|}{|I_n|}\log_2 q\,. \tag{3}$$

It is reliable *for the distribution* p *if there exists a sequence of reconstruction functions* $\widehat{\mathsf{F}}_n\equiv\widehat{\mathsf{F}}_{G_n}$ *such that, for* X *a random input and* $Y\equiv\mathsf{F}_{G_n}(X)$

$$\mathrm{P}_{\rm err}(G_n,\widehat{\mathsf{F}}_n)\equiv\mathbb{P}\{\widehat{\mathsf{F}}_n(Y)\neq X\}\overset{n}{\to}0\,. \tag{4}$$

Shannon's source coding theorem implies that there cannot exist reliable counter braids with asymptotic rate $r<h_2(p)$. However, the achievability of such rates is far from obvious, since counter braids are a fairly specific compression scheme. The main theorem of this paper establishes achievability, even under the restriction that the braid is sparse.

In order to avoid technical complication, we make two assumptions on the input distribution p:

1) It has *at most power-law tails*. By this we mean that $\mathbb{P}\{X_i\geq x\}\leq Ax^{-\epsilon}$ for some $\epsilon>0$.
2) It has *decreasing digit entropy*. Let $X_i=\sum_{a\geq 0}X_i(a)q^a$ be the q-ary expansion of X_i, and h_l be the q-ary entropy of $X_i(l)$. Then h_l is monotonically decreasing in l for any q large enough.

We call a distribution p with this two properties *admissible*. While this class does not cover all possible distributions, it is likely to include any case of practical interest.

Theorem 1. *For any admissible input distribution* p, *and any rate* $r>H_2(p)$ *there exist a sequence of reliable sparse counter braids with asymptotic rate* r.

As stressed above, we insist on the braid being sparse for two reasons: (i) It allows to update the registers content y with a small number of operations, whenever one entry of x is incremented (i.e. the storage function can be efficiently recomputed); (ii) It opens the way to using low-complexity message passing algorithms for estimating the input vector x, given the stored information (i.e. for evaluating the recovery function $\widehat{\mathsf{F}}_G$).

III. The architecture

A. Layering

We will consider *layered* architectures. By this we mean that the set of register is the disjoint union of L layers $R=R^1\cup R^2\cup\cdots\cup R^L$ and that directed edges are either from I to R^1 or from R^l to R^{l+1} for some $l\in\{1,\dots,L-1\}$ (we shall sometimes adopt the convention $R^0\equiv I$). We denote by $y^{(l)}=\{y_i:\,i\in R^l\}$ the vector of register values in layer l. We further let $m_l\equiv|R^l|$ denote the size of the l-th layer (with $m_0\equiv n$).

The graph structure is conveniently encoded in L matrices $\mathbb{H}_1,\dots\mathbb{H}_L$, whereby $\mathbb{H}_l$ is the $m_l\times m_{l-1}$ adjacency matrix of the subgraph induced by $R^l\cup R^{l-1}$. We further let $\mathbb{H}^l=\mathbb{H}_l\cdot\mathbb{H}_{l-1}\cdots\mathbb{H}_1$. The storage function F can be characterized as follows.

Lemma 2. *Consider an* L*-layers counters braid, let* x *be its input, and define the sequence of vectors* $z^{(l)}\in\mathbb{N}^{R^l}$, *by* $z^{(0)}=x$ *and*

$$z^{(l)}=\lfloor(\mathbb{H}_l z^{(l-1)})/q\rfloor\,. \tag{5}$$

(the division and floor operation being component-wise on the vector $\mathbb{H}_l z^{(l-1)}$.*) Then, the register values are* $y^{(l)}=\mathbb{H}_l z^{(l-1)}\mod q$.

B. Recovery function

We now describe the recovery function $\widehat{\mathsf{F}}$. Since in this paper we are only interested in achievability, we will neglect complexity considerations.

1) One layer: Let us start from a one-layer braid and assume the inputs to be iid with common distribution p_* supported on $\{0,\dots q-1\}\ni x_i$. Then the register values are $y=\mathbb{H}x\mod q$, where $\mathbb{H}$ is the adjacency matrix of the braid. Fix $\gamma\in(0,1)$. We say that the input $x\in\{0,\dots,q-1\}^n$ is *typical*, and write $x\in\mathsf{T}_n(p_*)$ if its type θ_x satisfies $D(\theta_x||p_*)\leq n^{-\gamma}$ (here the Kullback-Leibler divergence is computed in natural base). Denote by $\mathsf{T}_n(p_*;y)$ the set of input vectors that are typical and such that $\mathbb{H}x=y,\mod q$. The 'typical set decoder' returns a vector $\widehat{x}$ if this is the the unique element in $\mathsf{T}_n(p_*;y)$ and a standard error message otherwise. In formulae

$$\widehat{\mathsf{F}}(y)=\left\{\begin{array}{ll}\widehat{x} & \text{if } \mathsf{T}_n(p_*;y)=\{\widehat{x}\},\\ * & \text{if } |\mathsf{T}_n(p_*;y)|\neq 1.\end{array}\right. \tag{6}$$

2) Multi-layer: Consider now a multi-layer braid and $x\in\mathbb{N}^I$ (inputs not restricted to be smaller than q) with x_i's distributed independently according to p. It is convenient to write the input vector in base q

$$x=\sum_{a\geq 0}x(a)\,q^a\,. \tag{7}$$

where $x(a)=\{x_i(a):\,i\in V\}$ with $x_i(a)\in\{0,\dots,q-1\}$. Notice that, for each $a\geq 0$, the vector $x(a)$ has iid entries. Let p_a be the distribution on $x_i(a)$ when x_i has distribution p.

We'll apply typical set decoding recursively, determining the q-ary vectors $x(0)$, $x(1)$, $x(2),\dots$ in this order. Consider first $x(0)$. It is clear from Lemma 2 that $y^{(1)} = \mathbb{H}_1 x(0) = \mathbb{H}^1 x(0) \mod q$. We then apply typical set decoding to the determination of $x(0)$. More precisely, we look for a solution of $\mathbb{H}^1 x = y^{(1)} \mod q$ that is typical under distribution p_0. If there is a unique such solution, we declare it our estimate of $x(0)$ and denote by $\widehat{x}(0)$. Otherwise we declare an error.

Consider now the determination of $x(l)$ and assume the lower order terms in the expansion (7) have already been estimated to be $\widehat{x}(0)$, $\widehat{x}(1)$, $\dots$, $\widehat{x}(l-1)$. Let $\widehat{z}^{(0)} \equiv \sum_{a=0}^{l-1} \widehat{x}(a)\, q^a$, and $\widehat{z}^{(a)}$, $a \ge 1$ be determined through the same recursion as in Eq. (5). Further let $\widehat{y}^{(a)} = \mathbb{H}_a \widehat{z}^{(a-1)} \mod q$ (this are nothing but the register values on input $\widehat{z}^{(0)}$).

Assume the estimates $\widehat{x}(0)$, $\widehat{x}(1)$, $\dots$, $\widehat{x}(l-1)$ to be correct. It is then easy to realize that $\widehat{y}^{(a)} = y^{(a)}$ for $a = 1, \dots l$. Further $z^{(l)} = \widehat{z}^{(l)} + \mathbb{H}^l x(l) \mod q$, hence

$$y^{(l+1)} = \widehat{y}^{(l+1)} + \mathbb{H}^{l+1} x(l) \mod q\,. \tag{8}$$

We therefore proceed to compute $y^{(l+1)} - \widehat{y}^{(l+1)} \mod q$. If the linear system $\mathbb{H}^{l+1} x(l) = y^{(l+1)} - \widehat{y}^{(l+1)} \mod q$ admits more than one or no solution that is typical with respect to the distribution p_l, an error is returned. Otherwise, the next term in the expansion (7) is estimated through the unique typical solution of such linear system.

The recovery algorithm is summarized below, with one improvement with respect to the description above. Instead of recomputing $\widehat{z}^{(0)}, \dots, \widehat{z}^{(l)}$, at stage l we only compute the vector that $\widehat{z}^{(l)}$ that is needed at the present stage.

```
RECOVERY (INPUT: register values $y$)
Initialize $\widehat{z}^{(a)} = 0$ for $a \ge 0$;
for $l \in \{0, \dots L\}$
    Set $\widehat{y}(l+1) = \mathbb{H}_{l+1} \widehat{z}^{(l)} \mod q$;
    Let $\mathsf{T}_l$ be the set of $p_l$-typical
      solutions of $\mathbb{H}^{l+1} \widehat{x} = y^{(l+1)} - \widehat{y}^{(l+1)}, \mod q$;
    If $\mathsf{T}_l = \{\widehat{x}\}$ let $\widehat{x}(l) = \widehat{x}$
      otherwise if $|\mathsf{T}_l| \neq 1$ return error;
    Set $\widehat{z}^{(l+1)} = \lfloor \{\mathbb{H}^{l+1} \widehat{z}^{(l)} + \mathbb{H}_{l+1} \widehat{x}(l)\}/q \rfloor$;
end
return $\widehat{x} = \sum_i \widehat{x}(i)\, q^i$.
```

C. Sparse graph ensemble and choice of the parameters

The optimal compression rate in Theorem 1 is achieved with the following random sparse graph construction. Fix the registers capacity q and an integer $k \ge 2$. Then for $l = 1, \dots, L_0$ the graph induced by vertices $R_{l-1} \cup R_l$ has a random edge set that is sampled by connecting each $i \in R_{l-1}$ to k iid uniformly random vertices in R_l (all edges being directed from R_{l-1} to R_l). In other words, the $m_l \times m_{l-1}$, $0-1$ matrix $\mathbb{H}_l$ has independent columns, each sampled by incrementing k iid positions.

The choice of this ensemble is motivated by implementation concerns. In the flow counting problem, we do not know a priori the exact number of flows that needs to be stored. The above structure, this number can be changed without modifying existing links. Further, for each new flow, the subset of k registers it is connected to can be chosen through a simple hash function.

To these L_0 stages, we add further L_1 stages, all of the same size $m_{L_0+1} = \cdots = m_{L_0+L_1} = m_*$, with edges connecting each node in R_{l-1} to a different node in R_l. Equivalently, we take $\mathbb{H}_l$ to be the identity matrix in these stages.

It remains to specify the number of stages L_0, L_1 and their sizes $m_1, \dots, m_{L_0}$. Let p_l be the distribution of the l-th least significant digit in the q-ary expansion of X_i. Recall that we defined h_l to be the q-ary entropy of the distribution p_l, i.e.

$$h_l \equiv - \sum_{x=0}^{q-1} p_l(x) \log_q p_l(x)\,. \tag{9}$$

Finally, in the achievability proof, we shall assume that q is a prime number, large enough for h_l to be monotonically decreasing.

Lemma 3. *Assume $\mathbb{P}\{X_1 \ge x\} \le A\, x^{-\epsilon}$. Then there exists constants B, C that only depend on A, ϵ, such that for all $l \ge 1$, and all q large enough*

$$h_l \le B\, l\, q^{-l\epsilon}\,, \tag{10}$$

$$\Big| h_2(p) - \sum_{l \ge 0} h_l \log_2 q \Big| \le C\, q^{-\epsilon} (\log_2 q)^2\,. \tag{11}$$

The proof of this simple Lemma is deferred to Section VI.

Lemma 4. *Let p_* be a distribution over $\{0, \dots, q\}$, with q-ary entropy $H(p_*)$, and $\mathsf{T}_n(p_*)$ be the set of p_*-typical vectors defined as in Sec . Let $|\mathsf{T}_n(p_*)|$ be the size of this set. Recall that $x \in \mathsf{T}_n(p_*)$ if its type θ_x satisfies $D(\theta_x || p_*) \le n^{-\gamma}$. Then, for any $\beta \in (1 - \gamma/2, 1)$, there exists $A = A(\beta, \gamma, q)$ such that*

$$|\mathsf{T}_n(p_*)| \le q^{nH(p_*) + A\, n^{\beta}}\,,\,.$$

Further, if $X = (X_1, \dots, X_n)$ is a vector with iid entries with common distribution p_*

$$\mathbb{P}\{X \notin \mathsf{T}_n(p_*)\} \le (n+1)^q\, e^{-n^{1-\gamma}}\,. \tag{12}$$

In the following we will consider γ and β fixed once and for all, for instance by $\gamma = 1/2$ and $\beta = 7/8$.

Fix some $\delta > 0$, and let $A(q)$ be a suitably large constant, we let, for $l = 1, \dots, L_0$,

$$m_l \equiv \max\{\underline{m}_l, \lceil \delta\, m_{l-1} \rceil\}\,, \tag{13}$$

$$\underline{m}_l \equiv (1+\delta) \left\lceil n h_{l-1} (1+\delta) + A(q) n^{\beta} \right\rceil\,. \tag{14}$$

The number of stages is such that

$$\underline{m}_l \le n (\log n)^{-2} \text{ for any } l \ge L_0\,. \tag{15}$$

This implies, by Lemma 3, $L_0 = O(\log \log n)$. To this we add $L_1 = (\log n)^{3/2}$ stages within the second group, of size $m_* = m_{L_0} \le n(\log n)^{-2}$. The total number of registers is therefore upper bounded as $|R| \le n(1+\delta) \sum_{l \ge 0} (h_l +$

$An^{\beta-1})(\sum_{i\geq 0}\delta^i)+n(\log n)^{-1/2}$, and therefore the asymptotic rate of this architecture

$$r \leq \frac{1+\delta}{1-\delta}\sum_{l\geq 0} h_l \log_2 q\,. \tag{16}$$

Since the right hand side can be made arbitrarily close to $h(p)$ by Lemma 3, Theorem 1 follows from the following.

Theorem 2. *For any input distribution p with at most power-law tails and any choice of $q \geq 2$ and $\delta > 0$, there exists $k \geq 2$ such that the multi-layer braid described above is reliable.*

IV. Analysis of one-layer architectures

In order to prove our main Theorem or, equivalently, Theorem 2, we need first to prove a few preliminary results concerning a one-layer architecture. The proof here follows the technique of [9], the main tool being an estimate of the distance enumerator as in [4], [10], [11]. Distance enumerators for non-binary LDPC codes have been estimated in [12]. Unhappily we cannot here limit ourselves to citing these works, because the graph ensemble is different from the regular ones treated there.

Throughout this Section the source is a vector $X = (X_1,\dots,X_n)$ with iid entries taking values in $\{0,\dots,q-1\}$ and distribution p_* (in the application to multi-layer schemes p_* will coincide with p_l for some $l \geq 0$). We let $\mathbb{H}$ be an $m\times n$ matrix whose columns are independent vectors with integer entries. Each column is obtained by choosing k positions independently and uniformly at random (eventually with repetition) and incrementing the corresponding entry by one. In other words, $\mathbb{H}$ is distributed as the adjacency matrix of a given layer in the multi-layer architecture.

Our first result is a simple combinatorial calculation. Let $\vec{\lambda} = \{\lambda_z\,:\, z=1,\dots,q-1\}$ be a vector in $\mathbb{R}_+^{q-1}$. It is convenient to introduce the random variable $\vec{W} = \{W_z\,:\,z=1,\dots,q-1\}$ taking values in $\mathbb{N}^{q-1}$. The joint distribution of $(W_1,\dots,W_{q-1})$ (to be denoted by $\mathbb{P}_{\vec{\lambda}}$) is the one of $q-1$ Poisson random variables with means (respectively) $\lambda_1,\dots,\lambda_{q-1}$, conditioned on $\sum_{z=1}^{q-1} zW_z = 0$, $\mod q$.

Lemma 5. *Let $x\in\{0,\dots,q-1\}^n$ be an input vector with n_z entries equal to z, for $z=0,\dots,q-1$, and $\mathbb{H}$ be a random matrix as above. Define $\vec{n} = \{n_z\,:\,z=1,\dots,q-1\}$. For any $\vec{\lambda}\in\mathbb{R}_+^{q-1}$, let $\vec{W}_1$,... $\vec{W}_m$ be m iid vectors with distribution $\mathbb{P}_{\vec{\lambda}}$. Then the probability that $\mathbb{H}x = 0 \mod q$ is*

$$\mathbb{P}\{\mathbb{H}x=0\} = \prod_{z=1}^{q-1}\frac{(kn_z)!\,e^{m\lambda_z}}{(m\lambda_z)^{kn_z}}\,Q(\vec{\lambda})^m\,\mathbb{P}_{\vec{\lambda}}\left\{\sum_{i=1}^m \vec{W}_i = k\vec{n}\right\}\,, \tag{17}$$

where $Q(\vec{\lambda})$ is the probability that $\sum_{z=1}^{q-1} z\,U_z = 0, \mod q$ for independent Poisson random variables with means λ_z.

Further, for some universal constant C, and $D_q = 1-\cos 2\pi/q$, and $n_* = \sum_z n_z$

$$\mathbb{P}\{\mathbb{H}x=0\} \;\leq\; (Ckn_*)^{\frac{q-1}{2}}Q(k\vec{n}/m)^m R(m,\frac{k}{q}\sum_{z=1}^{q-1} zn_z) \tag{18}$$

$$Q(k\vec{n}/m) \;\leq\; \frac{1}{q}\left[1+(q-1)\,e^{-D_q\frac{kn_*}{m}}\right]\,, \tag{19}$$

$$R(m,N) \;=\; \min\left\{1,(Cq^2N/m)^N\right\} \tag{20}$$

Proof. Due to the symmetry of the distribution of $\mathbb{H}$ with respect to permutation in its columns, $\mathbb{P}\{\mathbb{H}x = 0\}$ does depend on x only through the number of ones, twos, etc. Without loss of generality we can assume the first n_1 coordinates to be ones, the next n_2 to be twos, and so on, and neglect the last $n-\sum_z n_z$ columns, corresponding to zeros. Think now of filling the matrix, by choosing its non-zero entries (edges in the associated graph). If we associate to each such entry the value of the corresponding coordinate in x, we want the probability for the sum of labels on each row to be $0 \mod q$. Since entries are independent and uniformly random, this is equal to the probability that each of m urns is filled with balls whose labels add to 0, when we throw kn_1 balls labeled with 1, kn_2 labeled with 2, and so on. It is an exercise in combinatorics to show that this is

$$\prod_{z=1}^{q-1}\frac{(kn_z)!}{m^{kn_z}}\,\mathsf{coeff}\left\{P(\xi_1,\dots,\xi_{q-1})^m,\xi_1^{kn_1}\cdots\xi_{q-1}^{kn_{q-1}}\right\}\,,$$

$$P(\cdots)\equiv\sum_{l_1\dots l_{q-1}}\frac{\xi_1^{l_1}}{l_1!}\cdots\frac{\xi_{q-1}^{l_{q-1}}}{l_{q-1}!}\,\mathbb{I}\left\{\sum_{z=1}^{q-1} z\,l_z = 0\right\}\,.$$

Equation (17) is then obtained by evaluating $\mathbb{P}_{\vec{\lambda}}$ and showing that it yields the above combinatorial expression.

In order to get Eq. (18), we denote $\mathbb{P}_{\vec{\lambda}}\{\cdots\}$ by R, and use $\lambda_z = kn_z/m$, thus leading to

$$\mathbb{P}\{\mathbb{H}x=0\} = \prod_{z=1}^{q-1}\frac{(kn_z)!\,e^{kn_z}}{(kn_z)^{kn_z}}\,Q(k\vec{n}/m)^m R\,.$$

Equation (18) follows from the observation that $N! \leq \sqrt{CN}\,(N/e)^N$ for some universal constant C.

In order to prove Eq. (19), notice that, by discrete Fourier transform

$$\begin{aligned} Q(\vec{\lambda}) &= \frac{1}{q}\sum_{\ell=0}^{q-1}\mathbb{E}\left\{e^{\frac{2\pi i\ell}{q}\sum_{z=1}^{q-1} zU_z}\right\} \\ &= \frac{1}{q}\sum_{\ell=0}^{q-1}\exp\left\{-\sum_{z=1}^{q-1}\lambda_z(1-e^{\frac{2\pi i\ell z}{q}})\right\}\,. \end{aligned}$$

The claim is proved by singling out the $\ell = 0$ term and bounding the others using $\mathrm{Re}(1-e^{\frac{2\pi i\ell z}{q}}) \geq D_q$.

Let us finally prove Eq. (20). Obviously $R \leq 1$ since it is an upper bound on the probability $\mathbb{P}_{\vec{\lambda}}\{\cdots\}$. If we let $N \equiv \frac{k}{q}\sum_{z=1}^{q-1} zn_z$, we can therefore assume, without loss of generality, that N is an integer with $N/m \leq 1/q$. Let V_i

be distributed as $\sum_{z=1}^{q-1} W_{i,z} z/q$ conditioned on V_i being an integer. Then the probability $\mathbb{P}_{\vec{\lambda}}\{\cdots\}$ is upper bounded by

$$\begin{aligned}\mathbb{P}\left\{\sum_{i=1}^{m} V_i \geq N\right\} &\leq \binom{m}{N}\mathbb{P}\{V_i \geq 1\}^N \\ &\leq \left(\frac{Cm}{N}\,\mathbb{P}\{V_i \geq 1\}\right)^N .\end{aligned}$$

Recalling the definition of V_i, we have

$$\begin{aligned}\mathbb{P}\{V_i \geq 1\} &= \mathbb{P}\Big\{\sum_{z=1}^{q-1} zU_z \geq q\Big|\sum_{z=1}^{q-1} zU_z = 0 \mod q\Big\} \\ &\leq e^{\sum_{z=1}^{q-1}\lambda_z}\mathbb{P}\Big\{\sum_{z=1}^{q-1} zU_z \geq q\Big\} .\end{aligned}$$

But $\sum_{z=1}^{q-1}\lambda_z = kn_*/m \leq Nq/m \leq 1$. Further, $\sum_{z=1}^{q-1} zU_z \geq q$ only if $\sum_{z=1}^{q-1} U_z \geq 2$. Therefore we get

$$\begin{aligned}\mathbb{P}\{V_i \geq 1\} &\leq e\mathbb{P}\Big\{\sum_{z=1}^{q-1} U_z \geq 2\Big\} \\ &\leq C\Big(\sum_{z=1}^{q-1}\lambda_z\Big)^2 \leq C(kn_*/m)^2 .\end{aligned}$$

The proof is completed by noticing that $(kn_*/m) \leq (Nq/m)$. □

In the following, given a vector $x = (x_1, \dots, x_n)$, we shall denote by $||x||_0$ is number of non-zero entries.

Lemma 6. *Let $\mathbb{H}$ be a random $m \times n$ matrix distributed as above, with column weight k. Assume k not to be a multiple of q, $m \leq n$, $(m/nk)^{1/k} \geq \Delta > 0$ and $3 \leq k \leq m/\log m$. Then, there exists a constant $B = B(q,\Delta)$, $C = C(q,\Delta)$, such that, if*

$$E \leq \frac{C\,m}{\log(nk/m)}\,, \tag{21}$$

then

$$\mathbb{P}\Big\{\exists ||z||_0 \leq E\,:\,\mathbb{H}z = 0 \mod q\Big\} \leq n^2\left(\frac{Bk}{m}\right)^{\frac{k}{q}} \tag{22}$$

(where it is understood that $z \in \{0, \dots, q-1\}^n$.)

Proof. Throughout the proof, A will denote a generic constant depending only on q that can be chosen large enough to make the inequalities below hold.

Let $z \in \{0, \dots, q-1\}^n$ be such that $||z||_0 = \ell$. We will upper bound the probability that $\mathbb{H}z = 0 \mod q$ in different ways depending whether $\ell \leq E_0$ or $\ell > E_0$, where

$$E_0 = \rho(q)\,\frac{m}{k}\,\left(\frac{m}{nk}\right)^{2/(k-2)}\,, \tag{23}$$

with $\rho(q)$ a function to be determined. Notice that, under our hypotheses,

$$\frac{kE_0}{m} \geq \rho(q)\,\Delta^{2k/(k-2)} \tag{24}$$

is bounded away from 0 (as $2 < 2k/(k-2) \leq 6$ for $k \geq 3$.)

For $||z||_0 = \ell \leq E_0$ (and $z \neq 0$) we use Lemma 5, Eq. (18), where we set $Q(\cdots) \leq 1$, $n_* = \ell$ and $\frac{k\ell}{q} \leq \frac{k}{q}\sum_{z=1}^{q-1} zn_z \leq k\ell$. Further we assumed $Ak\ell/m \leq 1$, which holds without loss of generality if we take $\rho(q) \leq 1/A\Delta^6 \leq 1/A\Delta^{2k/(k-2)}$ in Eq. (23), thus getting

$$\mathbb{P}\{\mathbb{H}z = 0\} \leq (Ak\ell)^{\frac{q-1}{2}}(Ak\ell/m)^{k\ell/q}\,. \tag{25}$$

Since $(k\ell)^{(q-1)/2} \leq A^{k\ell/q}$, we have (by properly adjusting A)

$$\mathbb{P}\{\mathbb{H}z = 0\} \leq (Ak\ell/m)^{k\ell/q}\,. \tag{26}$$

For $||z|| > E_0$, we use Eq. (18) with $R(\cdots) \leq 1$. Since $k\ell/m > kE_0/m$ is bounded away from 0 by Eq. (24), we have $Q(\cdots) \leq e^{-C}$ for some $C = C(\Delta, q) > 0$ and therefore

$$\mathbb{P}\{\mathbb{H}z = 0\} \leq (Ak\ell)^{(q-1)/2}\,e^{-Cm}\,. \tag{27}$$

There are at most $\binom{n}{\ell}(q-1)^\ell \leq \left(\frac{An}{\ell}\right)^\ell$ vectors z with $||z||_0 = \ell$. If we denote by $\mathbb{P}_{E_1,E_2}$ the probability of the event $\Big\{\exists z\,:\,E_1 \leq ||z||_0 \leq E_2\,,\,\mathbb{H}z = 0 \mod q\Big\}$, the probability in Eq. (22) is upper bounded by $\mathbb{P}_{2,E_0} + \mathbb{P}_{E_0,E}$ (notice that if k is not a multiple of q, $\mathbb{H}z = 0$ is impossible for $||z||_0 = 1$). By union bound we have

$$\begin{aligned}\mathbb{P}_{2,E_0} &\leq \sum_{\ell=2}^{E_0}\left(\frac{An}{\ell}\right)^\ell\left(\frac{Ak\ell}{m}\right)^{k\ell/q} \\ &\leq \left(\frac{An}{2}\right)^2\left(\frac{2Ak}{m}\right)^{2k/q}\sum_{\ell=2}^{E_0}\xi(\ell)^{\ell-2}\,,\end{aligned}$$

where (using $(\ell/2)^{2k/q-2} \leq A^{k(\ell-2)/q}$ and eventually adjusting the constant A)

$$\xi(\ell) \equiv \frac{n}{\ell}\left(\frac{Ak\ell}{m}\right)^{k/q}\,.$$

For $\ell \leq E_0$, and choosing $\rho(q)$ small enough in Eq. (23), we obtain $\xi(\ell) \leq 1/2$ thus leading to $\mathbb{P}_{2,E_0} \leq n^2(Ak/m)^{k/q}$.

Finally consider the contribution of vectors $||z||_0 \geq E_0$. Proceeding as above, we have

$$\begin{aligned}\mathbb{P}_{E_0,E} &\leq \sum_{\ell=E_0}^{E}\left(\frac{An}{\ell}\right)^\ell(Ak\ell)^{(q-1)/2}\,e^{-Cm} \\ &\leq E\left(\frac{An}{E}\right)^E(AkE)^{(q-1)/2}\,e^{-Cm}\,.\end{aligned}$$

Here we bounded $(An/\ell)^\ell = [(An/\ell)^{\ell/An}]^{An} \leq (An/E)^E$, using the fact that x^{-x} is an increasing function of x for $x \leq e^{-1}$, and that $E/An = Cm/An\log(nk/m) \leq Cm/An$ is smaller than e^{-1} for C small enough.

Finally we bound $E^{(q+1)/2} \leq A^E$ and $k^{(q-1)/2} \leq k^E$ (which holds for m large enough), thus getting

$$\mathbb{P}_{E_0,E} \leq \left(\frac{nkA}{m}\right)^E\,e^{-Cm}\,.$$

If we take $E = Cm/2\log(nkA/m)$, we get $\mathbb{P}_{E_0,E_1} \leq e^{-Cm/2}$, which is smaller than $(Bk/m)^{\frac{k}{q}}$ for a properly chosen constant B and $k \leq m/\log m$ (indeed $k \leq m\varepsilon_m$ would be enough for any $\varepsilon_m \downarrow 0$.) □

V. Analysis of multi-layer architectures and proof of Theorem 1

Proof. Let $\mathrm{P}_{\mathrm{err}}^{(l)}$ denote the probability that l-th term in the q-ary expansion of x is decoded incorrectly by the decoder in Section III-B (i.e. that $\widehat{x}(l) \neq x(l)$) given that $x(0)$, ..., $x(l-1)$ have been correctly recovered. We will prove that $\mathrm{P}_{\mathrm{err}}^{(l)} = O(n^{-A})$ for some $A > 0$. Since the multi-layer architecture involves at most $C(\log n)^{\frac{3}{2}}$ layers, this implies the thesis. Further, we shall consider only the first L_0 layers, since it will be clear from the derivation below that the error probability is decreasing for the last L_1 layers.

Let x be the input. Since we are focusing on the l-th term in the q-ary expansion of the input, we will drop the index l, and take $x \in \{0, \dots, q-1\}^n$. This is just a vector whose entries are iid with distribution p_l.

The error probability $\mathrm{P}_{\mathrm{err}}^{(l)}$ is upper bounded by the probability that $x \notin \mathsf{T}_n(p_l)$ plus the probability that there exists $x' \neq x$ with $\mathbb{H}^l x' = \mathbb{H}^l x \mod q$. The first contribution is bounded by Lemma 4, and we can therefore neglect it. Denoting the second contribution as $\mathrm{P}_{\mathrm{err}}^{(l,*)}$, and writing $\mathbb{E}_x$, $\mathbb{P}$ for (respectively) expectation with respect to x and probability with respect to the matrices $\mathbb{H}_1, \dots \mathbb{H}_l$, we have (matrix multiplications below are understood to be modulo q)

$$\begin{aligned}
\mathrm{P}_{\mathrm{err}}^{(l,*)} &= \mathbb{E}_x \mathbb{P}\left\{\exists x' \in \mathsf{T}_n(p_l) \setminus \{x\} \text{ s.t. } \mathbb{H}^l x' = \mathbb{H}^l x\right\} \\
&= \sum_{t=1}^{l} Q_t^{(l)}\,, \\
Q_t^{(l)} &\equiv \mathbb{E}_x \mathbb{P}\big\{\exists x' \in \mathsf{T}_n(p_l) \setminus \{x\} \text{ s.t.} \\
&\qquad\qquad \mathbb{H}^t x' = \mathbb{H}^t x,\, \mathbb{H}^{t-1} x' \neq \mathbb{H}^{t-1} x\big\}\,.
\end{aligned}$$

Since, $l \leq L = O(\log n)$, it is sufficient to show $Q_t^{(l)} = O(n^{-A})$. In $Q_t^{(l)}$ we can separate error events due to input x' such that $d_t \equiv d(\mathbb{H}^{t-1}x', \mathbb{H}^{t-1}x) \leq E$ and the other ones. As a consequence $Q_t^{(l)}$ is upper bounded by

$$\begin{aligned}
&\mathbb{E}_x \mathbb{P}\left\{\exists x' \in \mathsf{T}_n(p_l),\ \text{s.t. } 1 \leq d_t \leq E,\ \mathbb{H}^t x' = \mathbb{H}^t x\right\} + \\
&+ \mathbb{E}_x \mathbb{P}\left\{\exists x' \in \mathsf{T}_n(p_l) \text{ s.t. } E < d_t,\, \mathbb{H}^l x' = \mathbb{H}^l x\right\} \leq \\
&\leq \mathbb{P}\{\exists z \text{ s.t. } ||z||_0 \leq E\,, \mathbb{H}_t z = 0\} + \\
&\quad + |\mathsf{T}_n(p_l)| \sup\left\{\mathbb{P}\{\mathbb{H}_t z = 0\}\ :\ ||z||_0 > E\right\}.
\end{aligned}$$

Here z is understood to be a m_{t-1} dimensional vector with entries in $\{0, \dots, q-1\}$.

Notice that $(m_t/km_{t-1})^{1/k} \geq (\delta/k)^{1/k} \geq \delta$. Next we choose $E = C(q, \Delta = \delta)m_t/\log(m_{t-1}k/m_t)$ with $C(q, \Delta)$ as in the statement of Lemma 6. As a consequence the first term above is upper bounded by

$$m_{t-1}^2 \left(\frac{Bk}{m_t}\right)^{\frac{k}{q}} \leq (Bk)^{\frac{k}{q}} \delta^{-2} m_t^{-\frac{k}{q}+2} \leq C(\log n)^{\frac{k}{q}-2} n^{-\frac{k}{q}+2}\,,$$

where we used $m_{t-1} \leq m_t/\delta$ and $m_t \geq n/(\log n)^2$. The constant C that depends uniquely on q, k, δ, but not on n.

It remains to bound the second contribution, due to inputs x' with $d(x', x) > E$. Using Lemma 4 (to bound $\mathsf{T}_n(p_l)$) and 5 (to bound $\mathbb{P}\{\mathbb{H}_t z = 0\}$ for $||z||_0 > E$)

$$\begin{aligned}
&\mathbb{E}_x \mathbb{P}\left\{\exists x' \in \mathsf{T}_n(p_l) \text{ s.t. } E < d_t,\, \mathbb{H}^l x' = \mathbb{H}^l x\right\} \leq \\
&\leq q^{nh_l + An^\beta} (Ckn)^{\frac{q-1}{2}} \left\{\frac{1}{q}[1 + (q-1)e^{-DkE/m_t}]\right\}^{m_t}\,,
\end{aligned}$$

By eventually enlarging the constant A (in a way that depends on q), we can get rid of the term $(Cn)^{\frac{q-1}{2}}$. By further using $(1+x) \leq q^{x/\log q}$ we can upper bound the above by $k^{q-1/2}q^{\Phi}$ where

$$\Phi = nh_l + A(q)n^\beta - m_t + A'(q)m_t e^{-D(q)kE/m_t}$$

with $A'(q) = (q-1)/\log q$. Notice that $kE/m_t = C(q,\delta)k/\log(km_{t-1}/m_t)$ can be made arbitrarily large by taking k large enough. In particular, we can choose $k_*(q,\delta)$ such that $A'(q)e^{-D(q)kE/m_t} \leq \delta/3$ for any $k \geq k_*$. For such k, and using the fact that $m_t \geq m_l = [nh_l + A(q)n^\beta](1+\delta)$

$$\Phi \leq nh_l + A(q)n^\beta - m_l(1 - \delta/3) \leq -\frac{1}{3}\delta[nh_l + A(q)n^\beta]\,.$$

Summing the various contributions, we obtain, for any $k \geq k_*(q,\delta)$

$$Q_t^{(l)} \leq C(q,k,\delta)(\log n)^{\frac{k}{q}-2} n^{-\frac{k}{q}+2} + k^{\frac{q-1}{2}} q^{-\delta(A(q)n^\beta + nh_l)/3}\,, \quad (28)$$

which proves the thesis. □

VI. Some auxiliary results

Proof: Lemma 3. First consider Eq. (10). Let X_1 be an integer random variable with distribution p, $X_1(l)$ its l-th least significant q-ary digit and Z the indicator function on $X_1(l) \geq 0$. From $H(X_1(l)) = H(Z) + H(X_1(l)|Z)$ it follows that, for $\overline{p}_l = \mathbb{P}\{X_1 > q^l\}$:

$$h_l \leq \overline{p}_l \log_q(q-1) - \overline{p}_l \log_q \overline{p}_l - (1 - \overline{p}_l)\log_q(1 - \overline{p}_l)\,.$$

Choosing q large enough so that $\overline{p}_l \leq A\,q^{-\epsilon} \leq 1/2$ for all $l \geq 1$, we can upper bound $-(1-\overline{p}_l)\log_q(1-\overline{p}_l)$ by $2\overline{p}_l$, thus getting

$$h_l \leq 3\overline{p}_l - \overline{p}_l \log_q \overline{p}_l\,,$$

which implies Eq. (10) for $\overline{p}_l \leq A\,q^{-l\epsilon}$.)

In order to prove Eq. (11), first notice that $H(X_1) = H(\{X_1(l)\}) \leq \sum_{l\geq 0} H(X_1(l))$ whence $h_2(p) \leq \sum_{l\geq 0} h_l \log_2 q$. By the same argument $h_2(p) \geq h_0 \log_2 q$. The thesis follows by bounding $\sum_{l\geq 0} h_l$ using Eq. (10). □

Proof: Lemma 4.. The number of vectors with type θ is upper bounded by $q^{nH(\theta)}$. Since there are at most $(n+1)^q$ distinct types, $|\mathsf{T}_n(p_*)| \leq q^{nH(p_*)+nK_n}$ where

$$K_n \equiv \sup_\theta\{H(\theta) - H(p_*) : D(\theta||p_*) \leq n^{-\gamma}\} + \frac{\log_q(n+1)^q}{n}\,.$$

The bound $H(\theta) - H(p_*) \leq ||\theta - p_*||_1 \log(q/||\theta - p_*||)$ and $||\theta - p_*|| \leq \sqrt{2D(\theta||p_*)}$ [3].

Equation (12) is just Sanov Theorem. □

VII. Acknoledgments

Yi Lu is supported by the Cisco Stanford Graduate Fellowship.

References

[1] S. Dharmapurikar, A. Kabbani, Y. Lu, A. Montanari and B. Prabhakar. “Passing Messages Through Counter Braids: A Novel Approach to Traffic Measurement.” Technical Report, TR06-ISL012201, March, 2007.

[2] C. Estan and G. Varghese. “New Directions in Traffic Measurement and Accounting: Focusing on the Elephants, Ignoring the Mice.” *ACM Trans. on Comp. Syst.*, 21:270–313, 2003.

[3] T. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Interscience, New York, 1991

[4] Robert G. Gallager. *Low-Density Parity-Check Codes*. MIT Press, Cambridge, Massachussetts, 1963.

[5] T. Richardson and R. Urbanke, *Modern Coding Theory*, draft available at `http://lthcwww.epfl.ch/mct/index.php`

[6] G. Caire, S. Shamai, and S. Verdú. “Noiseless data compression with low density parity check codes.” In P. Gupta and G. Kramer, editors, *Dimacs Series in Mathematics and Theoretical Computer Science*, pages 224–235. AMS, 2004.

[7] E. Berlekamp, R. J. McEliecee, and H. C.A. van Tilborg. “On the inherent intractability of certain coding problems.” *IEEE Trans. Inform. Theory*, IT-29:384–386, 1978.

[8] F. R. Kschischang, B. J. Frey and H-A. Loeliger, “Factor graphs and the sum-product algorithm” (2001), *IEEE Trans. Inform. Theory* **47**, 498-519.

[9] S. Aji, H. Jin, A. Khandekar, D. J.C. MacKay, and R. J. McEliece. “BSC Thresholds For Code Ensembles Based on ‘Typical Pairs’ Decoding.” In Brian Marcus and Joachim Rosenthal, editors, *Codes, Systems and Graphical Models*, pages 195–210. Springer, 2001.

[10] G. Miller and D. Burshtein, “Asymptotic enumeration method for analyzing LDPC codes,” *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1115–1131, June 2004.

[11] S. L. Litsyn and V. S. Shevelev, “On ensembles of low-density parity-check codes: asymptotic distance distributions,” *IEEE Trans. Inform. Theory*, vol. IT–48, pp. 887 –908, Apr. 2002.

[12] A. Bennatan and D. Burshtein. “On the application of LDPC Codes to Arbitrary Discrete-Memoryless Channels.” *IEEE Trans. Inform. Theory*, 50:417–438, 2004.