# Weight Distribution of the Syndrome of Linear Codes and Connections to Combinatorial Designs

Christoph Pacher[1], Philipp Grabenweger[1], and Dimitris E. Simos[2]

[1]Digital Safety & Security Department, AIT Austrian Institute of Technology GmbH, Donau-City-Straße 1, 1220 Vienna, Austria
[2]SBA Research, Favoritenstraße 16, 1040 Vienna, Austria

***Abstract*—The expectation and the variance of the syndrome weight distribution of linear codes after transmission of codewords through a binary symmetric channel is derived exactly in closed form as functions of the code's parity-check matrix and of the degree distributions of the associated Tanner graph. The influence of (check) regularity of the Tanner graph is studied. Special attention is payed to Tanner graphs that have no cycles of length four. We further study the equivalence of some classes of combinatorial designs and important classes of LDPC codes and apply our general results to those more specific structures. Simulations are performed to show the validity of the theoretical approach.**

## I. Introduction

In channel coding, channel state information (CSI), i.e. an estimate of the crossover probability or the signal-to-noise ratio, can be used to reduce the error rate of iterative decoding algorithms like the sum-product algorithm. In [1] a maximum likelihood channel estimator that is very useful if low-density parity-check codes are used as error correcting codes, was presented. The only input of the estimator is the syndrome weight at the receiver. Interestingly, the authors took as granted that the syndrome bits are independent and identically distributed. In [2] it was observed that this assumption does not hold in general, a closed form was given and a refined statistical analysis of that estimator has been performed. The current work can be seen as a continuation of the work done in [2]: a detailed statistical (both analytical and numerical) analysis of the syndrome weight distribution of LDPC codes is provided, an optimality condition is derived, and connections to combinatorial design theory are made. Ultimatively, these results should allow to reduce the mean squared error of the aforementioned estimator or to derive an confidence interval estimator.

Despite this special application, we think that the distribution of the syndrome weight is a very fundamental object of any linear code and thus of general interest. Potential fields where this study might be relevant are (iteratice) decoding algorithms and code-based cryptography.

### A. Related Work

There is a lot of work that deals with the weight distribution of the *codewords* of a code (see e.g. [3, Chap. 7] and references therein). However, the weight distribution of syndromes has not been studied in detail yet. In addition to the work already mentioned [1], [2], in [4] the correlation between the error weights and the syndrome weights during the sum product algorithm (SPA) decoding process of LDPC codes has been analyzed.

### B. Outline of the Paper

In Section II we first fix the notation and then calculate in closed form the expectation and variance of the syndrome weight after transmission of codewords through a binary symmetric channel. In Section III we derive relations between the concurrence matrix entries and the degree node distributions and give some connections from families of LDPC codes to different classes of combinatorial designs together with some favorable cases to avoid cycles of length four. Section IV presents a comparison of the theoretical results with simulation results. We conclude in Section V.

## II. Expectation and Variance of the Syndrome Weight after Transmission through a BSC($\rho$)

### A. Preliminaries

First let us fix some notation. Let the binary linear code $\mathcal{C}$ be the nullspace of a binary parity-check (PC) matrix $\mathbf{H}$ of dimension $m \times n$, i.e., $\mathcal{C} := \{\boldsymbol{x} \in \{0,1\}^n : \boldsymbol{x}\mathbf{H}^T = \mathbf{0}\}$, and let $\oplus$ denote summation modulo two. The $i$-th *row* of $\mathbf{H}$ will be denoted as $\boldsymbol{h_i}$. We will make use of the Tanner graph [5] that is associated with the PC matrix $\mathbf{H}$, and say that the code has $n$ variable nodes and $m$ check nodes. The set of variable nodes incident to check node $i$ will be denoted as $\mathcal{V}_i$. The check node degrees $|\mathcal{V}_i|$ (which are the Hamming weights of the rows of $\mathbf{H}$) will be denoted by $d_i$, $i \in \{1, \dots, m\}$, the variable node degrees (which are the Hamming weights of the columns of $\mathbf{H}$) will be denoted by $d_{v_j}$, $j \in \{1, \dots, n\}$. The maximal check (variable) node degree will be denoted by $d_{\max}$ ($d_{v_{\max}}$). If all check (variable) node degrees of a code are identical this node degree will be simply denoted as $d$ ($d_v$).

We will denote the binary symmetric channel with crossover probability $\rho$ as BSC($\rho$). Let us assume, that we transmit a code word $\boldsymbol{x}$ over a BSC($\rho$) with $0 < \rho < 1/2$ and that the receiver obtains $\boldsymbol{y} = \boldsymbol{x} \oplus \boldsymbol{e}$, where $\boldsymbol{e}$ is the error word. The syndrome of $\boldsymbol{y}$ is given by $\boldsymbol{s} = \boldsymbol{y}\mathbf{H}^T = \boldsymbol{e}\mathbf{H}^T$. Let

$$s_i = \bigoplus_{v \in \mathcal{V}_i} e_v, \qquad (1)$$

be the $i$-th syndrome bit, $w := \sum_{i=1}^{m} s_i$ the syndrome weight and let us define the function

$$f_d(\rho) := \frac{1-(1-2\rho)^d}{2}, \tag{2}$$

that fulfills the identities $f_d(\rho)(1-f_d(\rho)) = \frac{1}{2} f_{2d}(\rho)$, $f_0(\rho) = 0$ and for $d > 0$ we have $f_d(\frac{1}{2}) = \frac{1}{2}$.

In [2] it was shown that the maximum likelihood estimate $\hat{\rho}$ of the crossover probability $\rho$ is given by

$$\hat{\rho}(w) = \begin{cases} f_{1/d}\left(\frac{w}{m}\right) & ; \quad \frac{w}{m} \leq 1/2 \\ \frac{1}{2} & ; \quad \frac{w}{m} > 1/2 \end{cases} . \tag{3}$$

Since the estimate is derived from the weight $w$ of the syndrome, it is a natural question to study the statistical properties of $w$.

### *B. General linear codes*

In the following we want to derive the expectation value and the variance of $w$, i.e. the weight of $\boldsymbol{s}$, after a codeword has been transmitted over a BSC($\rho$). The probability that a syndrome bit is one is given by

$$\Pr\left[s_i = 1\right] = \sum_{\substack{k \in \mathbb{N}_{\text{odd}} \\ 0 \leq k \leq d_i}} \binom{d_i}{k} \rho^k (1-\rho)^{d_i - k} = f_{d_i}(\rho). \tag{4}$$

Consequently, the expectation value of the weight $w$ of the syndrome is given by

$$\mathbb{E}\left[w\right] = \sum_{i=1}^{m} \mathbb{E}\left[s_i\right] = \sum_{i=1}^{m} \Pr\left[s_i = 1\right] = \sum_{i=1}^{m} f_{d_i}(\rho). \tag{5}$$

The expectation value of $w^2$ is given by

$$\mathbb{E}\left[w^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{m} s_i\right)\left(\sum_{j=1}^{m} s_j\right)\right] = \sum_{i,j=1}^{m} \mathbb{E}\left[s_i s_j\right]. \tag{6}$$

In case $i = j$ we obtain $\mathbb{E}\left[s_i s_j\right] = \mathbb{E}\left[s_i^2\right] = \mathbb{E}\left[s_i\right] = \Pr\left[s_i = 1\right] = f_{d_i}(\rho)$. In case $i \neq j$ we have $\mathbb{E}\left[s_i s_j\right] = \Pr\left[s_i = 1 \wedge s_j = 1\right]$. In case that $\mathcal{V}_i$ and $\mathcal{V}_j$ have variable nodes in common, $s_i$ and $s_j$ are correlated. Thus, to calculate $\Pr\left[s_i = 1 \wedge s_j = 1\right]$ we write the two potentially intersecting sets $\mathcal{V}_i$ and $\mathcal{V}_j$ in terms of three disjoint sets $\mathcal{A}_{ij} := \mathcal{V}_i \backslash \mathcal{V}_j$, $\mathcal{B}_{ij} := \mathcal{V}_j \backslash \mathcal{V}_i$, and $\mathcal{C}_{ij} := \mathcal{V}_i \cap \mathcal{V}_j$: $\mathcal{V}_i = \mathcal{A}_{ij} \cup \mathcal{C}_{ij}$, and $\mathcal{V}_j = \mathcal{B}_{ij} \cup \mathcal{C}_{ij}$. ($\mathcal{C}_{ij}$ is not related to the codebook $\mathcal{C}$.)

In terms of the cardinalities of the new sets we obtain

$$\begin{aligned} \mathbb{E}\left[s_i s_j\right] \quad = \quad & f_{|\mathcal{A}_{ij}|}(\rho) f_{|\mathcal{B}_{ij}|}(\rho) + \\ & f_{|\mathcal{C}_{ij}|}(\rho)\left[1 - f_{|\mathcal{A}_{ij}|}(\rho) - f_{|\mathcal{B}_{ij}|}(\rho)\right]. \end{aligned} \tag{7}$$

The number of variable nodes that are incident to check nodes $i$ and $j$, let us call it $\lambda_{ij}$, can be easily obtained by taking the standard inner product of the corresponding rows of the binary parity check matrix $\mathbf{H}$

$$\lambda_{ij} := |\mathcal{V}_i \cap \mathcal{V}_j| = \boldsymbol{h_i} \boldsymbol{h_j^T} = \lambda_{ji}. \tag{8}$$

Note, that the inner product of $\boldsymbol{h_i}$ with itself is the weight of $\boldsymbol{h_i}$, i.e. the check node degree $d_i$:

$$\lambda_{ii} = d_i. \tag{9}$$

The matrix $\mathbf{\Lambda} := \mathbf{H}\mathbf{H}^T$ is a symmetric matrix composed of all values $\lambda_{ij}$ (see also Section III). In terms of these $\lambda_{ij}$ we obtain

$$\begin{aligned} \mathbb{E}\left[s_i s_j\right] \quad = \quad & f_{d_i - \lambda_{ij}}(\rho) f_{d_j - \lambda_{ij}}(\rho) + \\ & f_{\lambda_{ij}}(\rho)\left[1 - f_{d_i - \lambda_{ij}}(\rho) - f_{d_j - \lambda_{ij}}(\rho)\right]. \end{aligned} \tag{10}$$

Equations (5), (6), (8), and (10) allow us to calculate the expectation value $\mathbb{E}\left[w\right]$ and the variance

$$\mathbb{V}\left[w\right] = \mathbb{E}\left[w^2\right] - \left(\mathbb{E}\left[w\right]\right)^2 \tag{11}$$

of the distribution of the syndrome weight $w$ of a linear block code after a code word (or the zero vector) has been transmitted over a BSC($\rho$). The expectation value Eq. (5) is a function of only $\rho$ and the check node degrees $d_i$ of the Tanner graph. Due to Eq. (10) the variance is a function of $\rho$, the check node degrees and the values $\lambda_{ij}$ defined in Eq. (8).[1]

### *C. Check-regular codes*

In the following we want to consider only codes that have constant check node degree $d$ and give simpler equations for $\mathbb{E}\left[w\right]$ and $\mathbb{V}\left[w\right]$. We obtain

$$\mathbb{E}\left[w\right] = m f_d(\rho), \tag{12}$$

$$\mathbb{E}\left[s_i s_j\right] = f_{d-\lambda_{ij}}^2(\rho) + f_{\lambda_{ij}}(\rho)\left(1 - 2 f_{d-\lambda_{ij}}(\rho)\right). \tag{13}$$

Using the definition of $f_d(\rho)$ we can simplify the last equation to

$$\mathbb{E}\left[s_i s_j\right] = f_d(\rho) - \frac{1}{2} f_{2(d-\lambda_{ij})}(\rho). \tag{14}$$

Using Equations (11), (6), (14), and (12) we obtain for the variance of $w$

$$\mathbb{V}\left[w\right] = \frac{1}{2}\left(m^2 f_{2d}(\rho) - \sum_{i \neq j} f_{2(d-\lambda_{ij})}(\rho)\right). \tag{15}$$

Let

$$g_\lambda := \left|\{(i,j) \in \{1, \ldots, m\}^2 | \lambda_{ij} = \lambda\}\right| \tag{16}$$

be the number of all ordered pairs of check nodes which have exactly $\lambda$ variable nodes in common. If we further assume that all rows of $\mathbf{H}$ are pairwise distinct we obtain that only the inner product of each row of $\mathbf{H}$ with itself is identical to the check node degree $d$, i.e. $\boldsymbol{h_i} \boldsymbol{h_j^T} = d \iff i = j$, thus

$$g_d = m. \tag{17}$$

Counting all elements of $\mathbf{\Lambda}$ we get $\sum_{\lambda=0}^{d} g_\lambda = m^2$.

[1] The matrix $\mathbf{\Lambda}$ fully defines the first and (central) second moment of the syndrome weight distribution. For the $p$-th moment we need to calculate $|\mathcal{V}_{i_1} \cap \cdots \cap \mathcal{V}_{i_p}|$ which needs a more complicated case analysis.

If we express the variance with the help of $g_\lambda$ we obtain

$$\begin{aligned}\mathbb{V}\left[w\right] &= \frac{1}{2}\left(m^2 f_{2d}(\rho) - \sum_{\lambda=0}^{d} g_\lambda f_{2(d-\lambda)}(\rho)\right) \\ &= \frac{m}{2} f_{2d}(\rho) + \frac{1}{2}\sum_{\lambda=1}^{d-1} g_\lambda \left(f_{2d}(\rho) - f_{2(d-\lambda)}(\rho)\right)\end{aligned}\tag{18}$$

Note, that the first summand in the second line is the variance of $m$ i.i.d. syndrome bits:

$$\mathbb{V}\left[w\right]_{\text{i.i.d.}} = m f_d(\rho)(1-f_d(\rho)) = \frac{m}{2} f_{2d}(\rho).\tag{19}$$

The second summand corresponds to the increase of the variance due to the pairwise correlation between syndrome bits:

$$\mathbb{V}\left[w\right]_{\text{corr}} = \frac{1}{2}\sum_{\lambda=1}^{d-1} g_\lambda \left(f_{2d}(\rho) - f_{2(d-\lambda)}(\rho)\right) \geq 0.\tag{20}$$

Using the inequality $f_{i+1}(\rho) > f_i(\rho)$ that holds for $i \geq 0$ and $0 < \rho < \frac{1}{2}$ and the equality $\sum_\lambda g_\lambda = m^2$ it is easy to see that the minimum of the variance is achieved if $g_\lambda = 0$ for $1 \leq \lambda \leq d-1$. Under the condition that $g_1 > 0$ is fixed the minimum of the variance is achieved if $g_\lambda = 0$ for $2 \leq \lambda \leq d-1$.

## III. Relations between the concurrence matrix entries and the degree node distributions

In design theory the matrix $\mathbf{\Lambda} = \mathbf{H}\mathbf{H}^T$ is known as concurrence matrix. It has the following nice properties (valid in the most general case, i.e. the case with irregular variable and check node degree distribution):

$$\sum_{i,j} \lambda_{ij} = \sum_{i,j} \boldsymbol{h_i}\boldsymbol{h_j^T} = \left(\sum_{i=1}^{m} \boldsymbol{h_i}\right)\left(\sum_{j=1}^{m} \boldsymbol{h_j^T}\right) =$$

$$(d_{v1}, d_{v2}, \ldots, d_{vn}) \cdot (d_{v1}, d_{v2}, \ldots, d_{vn})^T = \sum_{i=1}^{n} d_{v_i}^2,\tag{21}$$

$$\sum_{i\neq j} \lambda_{ij} = \sum_{i,j} \lambda_{ij} - \sum_{i=1}^{m} \lambda_{ii} = \sum_{i=1}^{n} d_{v_i}\left(d_{v_i} - 1\right).\tag{22}$$

For the last equality we used that $\sum_{i=1}^{m} d_i = \sum_{j=1}^{n} d_{v_j}$ is the total number of edges of the graph, i.e. the number of 1's in the PC matrix. Denoting the number of variable nodes of degree $i$ with $\Lambda_i$, we find a new connection to the formal $\Lambda$-polynomial $\Lambda(x) = \sum_{i=1}^{d_{v_{\max}}} \Lambda_i x^i$ that describes the variable node degree distribution from a node perspective[2] [6, Chap. 3.4]:

$$\sum_{i\neq j} \lambda_{ij} = \sum_{i=1}^{n} d_{v_i}\left(d_{v_i} - 1\right) = \sum_{j=2}^{d_{v_{\max}}} \Lambda_j j(j-1) = \Lambda''(1),\tag{23}$$

where the double-prime denotes the second derivative with respect to $x$. As a side note, if we define $\tilde{\mathbf{\Lambda}} := \mathbf{H}^T\mathbf{H}$, as the concurrence matrix of the transpose of $\mathbf{H}$, we get

$$\sum_{i,j} \tilde{\lambda}_{ij} = \sum_{i=1}^{m} d_i^2,\tag{24}$$

$$\sum_{i\neq j} \tilde{\lambda}_{ij} = \sum_{i=1}^{m} d_i\left(d_i - 1\right).\tag{25}$$

The value $\tilde{\lambda}_{ij}$ is the number of check nodes that are incident to both variable nodes $i$ and $j$. Denoting the number of check nodes of degree $i$ with $P_i$, we find a new similar connection to the formal $P$-polynomial $P(x) = \sum_{i=1}^{d_{\max}} P_i x^i$ that describes the check node degree distribution from a node perspective[6, Chap. 3.4]:

$$\sum_{i\neq j} \tilde{\lambda}_{ij} = \sum_{i=1}^{m} d_i\left(d_i - 1\right) = \sum_{j=2}^{d_{\max}} P_j j(j-1) = P''(1).\tag{26}$$

By simply counting and summing up the entries of the concurrence matrix $\mathbf{\Lambda}$ we get (again valid in the most general case, i.e. the case with irregular variable and check node degree distribution):

$$\sum_{\lambda=0}^{d_{\max}} g_\lambda = m^2,\tag{27}$$

$$\sum_{\lambda=1}^{d_{\max}} \lambda g_\lambda = \sum_{ij} \lambda_{ij}.\tag{28}$$

### A. PBDs and irregular LDPC codes

It is known that incidence matrices of a *pairwise balanced design* (PBD) are generally good candidates for high speed information transmission [7]. Let $K$ be a subset of positive integers. In particular, a PBD of order $v$ and index $\lambda$ with block sizes from $K$, denoted by PBD$(v, K, \lambda)$, is an ordered pair $(V, \mathcal{B})$, where $V$ is a nonempty finite set of $v$ elements, called points, and $\mathcal{B}$ is a set of subsets of $V$, called blocks, that satisfies the following two conditions:

(i) each unordered pair of distinct elements of $V$ appears in exactly $\lambda$ blocks of $\mathcal{B}$,

(ii) for every block $B \in \mathcal{B}$ the cardinality $|B| \in K$.

An incidence matrix of a PBD having a pair $(V, \mathcal{B})$ with $|V| = v$ points and $|\mathcal{B}| = b$ blocks is a binary $v \times b$ matrix $\mathbf{H} = (h_{i,j})$ with rows indexed by points, columns indexed by blocks, and $h_{i,j} = 1$ if the $i$-th point is contained in the $j$-th block, and $h_{i,j} = 0$ otherwise. Such incidence matrices can be used as parity-check matrices of irregular LDPC codes, where $K = \{k_i\}$ relates to the irregular variable node degree distribution by the following relation: $\Lambda_{k_j} = |\{i : d_{v_i} = k_j\}|$.

### B. IBDs and regular LDPC codes

The parity-check matrix of a regular LDPC code can also be constructed from a sparse incidence matrix of an *incomplete block design* (IBD) due to their good error tolerance at relatively short lengths [8]. An IBD of size $(v, k, r)$ is an

[2] To avoid confusion, we stress that the concurrence matrix $\mathbf{\Lambda}$ has entries $\lambda_{ij}$, and is not related to the polynomial $\Lambda(x)$ which has coefficients $\Lambda_i$.

arrangement of $v$ points set out in blocks of size $k$ $(< v)$ such that each point occurs in exactly $r$ blocks. The number of blocks will be $b$, where $bk = vr$. The incidence matrix of an $\mathrm{IBD}(v, k, r)$ has size $v \times b$ and constant row and column weights equal to $r$ and $k$, respectively. In that case, the blocks of the design form the columns of the $v \times b$ parity-check matrix $\mathbf{H}$ of a regular LDPC code with check and variable node degree, $d = r$ and $d_v = k$, respectively. In particular, blocks correspond to variable nodes, while points correspond to check nodes. Therefore, in the regular case we get the following simplified equations:

$$\sum_{\lambda=0}^{d} \lambda g_\lambda = \sum_{ij} \lambda_{ij} = n d_v^2, \tag{29}$$

$$\sum_{\lambda=0}^{d-1} \lambda g_\lambda = \sum_{i \neq j} \lambda_{ij} = n d_v (d_v - 1). \tag{30}$$

### C. Block designs and check regular LDPC codes

In the check regular case (but for general variable node degrees) we get

$$\sum_{i \neq j} \lambda_{ij} = \sum_{i=1}^{n} d_{v_i}^2 - md. \tag{31}$$

We can also view the parity-check matrix $\mathbf{H}$ of a check regular LDPC code as the incidence matrix of an IBD where we relax the property of all blocks to be of equal size.

### D. Designs and LDPC codes without cycles of length four

In a block design the concurrence $\lambda_{ij}$ of points $i$ and $j$ is $r$ if $i = j$ and otherwise is the number of blocks in which $i$ and $j$ both occur. A regular graph design (RGD) [9] is an $\mathrm{IBD}(v, k, r)$ where any two points belong to either $\lambda$ or $\lambda + 1$ common blocks, for some constant $\lambda$ and is denoted as $\mathrm{RGD}_\lambda(v, k, r)$.

If two check nodes share two variable nodes these four nodes form a 4-cycle. In particular, this is equivalent to a $2 \times 2$ all-one submatrix in a parity-check matrix which implies that the corresponding IBD has a concurrence greater or equal to 2. An RGD with $\lambda = 0$ has the property that any two points occur in at most one block, which implies that the corresponding Tanner graph of the code is thus without 4-cycles. Therefore any code with *minimal* variance $\mathbb{V}[w]$ that has $g_\lambda = 0$ for $2 \leq \lambda \leq d - 1$ must be free of 4-cycles and in the case of a regular LDPC code is equivalent to an RGD with $\lambda = 0$. Note that a similar connection has been obtained from Steiner 2-designs (see [10] for undefined terms and references therein), however these results apply only to a specific algebraic construction while ours are applicable in the general case of a regular LDPC code. Taking for the parity-check matrix of an irregular LDPC code a PBD of index $\lambda = 1$ ensures that no pair of points appears twice, hence there exist no 4-cycles in the Tanner graph. In addition, if the Tanner graph does not contain 4-cycles, only $g_0$ (disjoint sets of variable nodes), $g_1$ (one common variable node adjacent to both check nodes) and $g_d$ (all variable nodes in common) can be non-zero.

In case of a code without 4-cycles (and with pairwise different check rows), we obtain a very nice result for $g_1$. Consider that for four-cycle free codes $g_\lambda = 0$ for $2 \leq \lambda \leq d - 1$. That means that the off-diagonal elements of $\mathbf{\Lambda}$ are either 0 or 1. Combining the fact that the number of 1s is by definition equal to $g_1$ with Eq. (23) we immediately get that

$$g_1 = \sum_{i \neq j} \lambda_{ij} = \Lambda''(1). \tag{32}$$

For a regular code, where all $n$ variable nodes have degree $d_v$, we have $g_1 = n d_v (d_v - 1)$.
Thus we obtain for codes without 4-cycles the following result:

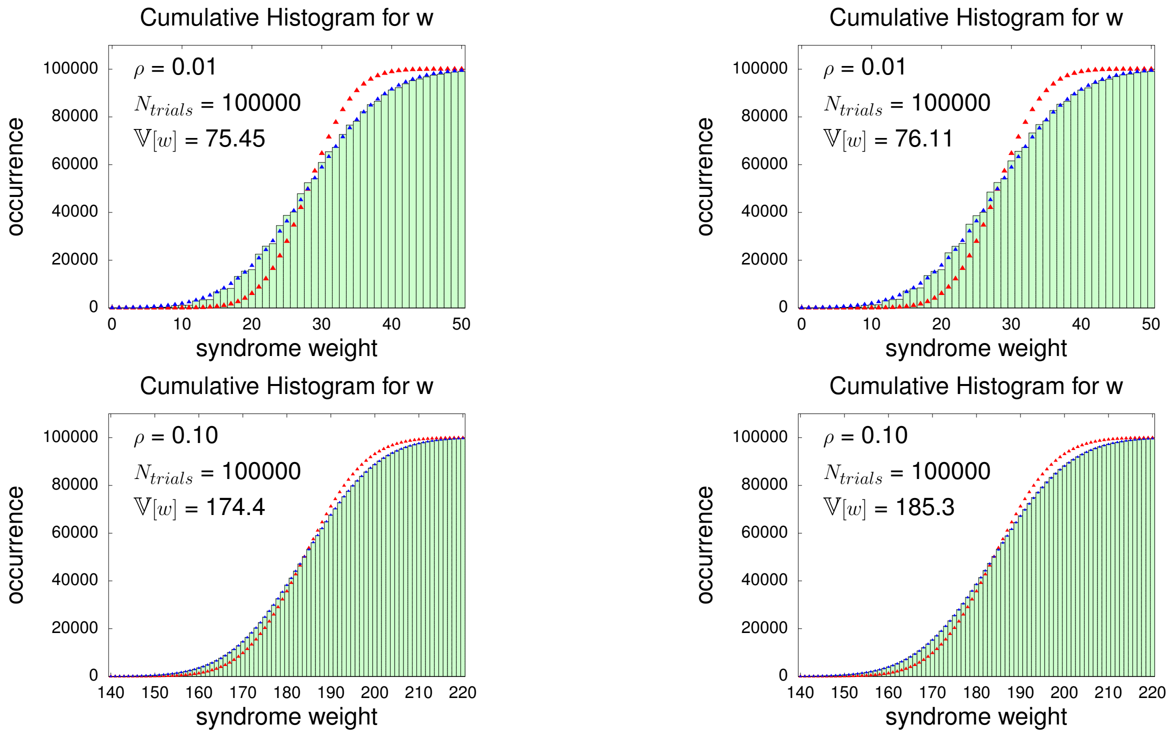
$$\mathbb{V}[w] \quad = \quad \frac{m}{2} f_{2d}(\rho) + \frac{\Lambda''(1)}{2} \left( f_{2d}(\rho) - f_{2d-2}(\rho) \right). \tag{33}$$

The variance of the syndrome weight is the same for all codes of a code ensemble defined by its degree distributions.

## IV. Comparison of theory with simulation results

In addition to the analytical results presented so far, we performed numerical simulations for two different regular LDPC codes with rate $1/2$. We determined two different parity check matrices, such that one code has no cycles of length four, while the other code was deliberately constructed to have 2000 cycles of length four. Both codes share the following parameters: number of variable nodes $n = 1000$, number of check nodes $m = 500$, check node degree $d = 6$, and variable node degree $d_v = 3$. The code without four-cycles corresponds to an $\mathrm{RGD}_0(500, 3, 6)$, the other code corresponds to an $\mathrm{IBD}(500, 3, 6)$. We simulated the transmission of $10^5$ all-zero codewords over a $\mathrm{BSC}(\rho)$ for $\rho \in \{0.01, 0.1\}$ and calculated and recorded the weights of the syndromes of both codes. Figure 1 shows the cumulative histogram for the simulated syndrome weights (green bars) and for two discrete normal distributions with different variances. The red triangles show the distribution with the variance of uncorrelated syndrome bits, while the blue triangles correspond to the variance when the correlation between any two syndrome bits is taken into account. In the figure we can observe several interesting properties:

We see that the cumulative distribution function of the simulated syndromes is very well approximated by the cumulative distribution function of a normal distribution using the analytical values for the mean and the exact variance that we have derived. Using the variance that corresponds to the i.i.d. approximation (which neglects the correlation between syndrome bits) results in a bad fit of the simulated results, especially so for small values of the crossover probability $\rho$. However, for larger values of $\rho$ the i.i.d. approximation becomes better. An increase of $\rho$ reduces also the small deviation between simulation results and the normal distribution using the exact variance even further. We also observe, that the variance of the weight increases with increasing crossover probability. Interestingly, if we compare the figures on the left



(a) LDPC without four-cycles.

(b) LDPC with 2000 four-cycles.

Fig. 1: Simulations of syndrome weights (green bars) and comparisons with normal distributions with variance $\mathbb{V}\left[w\right]_{\text{i.i.d.}}$, Eq. (19) (red triangles), and with exact variance $\mathbb{V}\left[w\right]$, Eq. (18) (blue triangles). The number of simulated error words and thus syndromes is $N = 10^5$ in all four cases. Both top figures show simulated transmission over a BSC(0.01), the bottom figures over a BSC(0.1).

with the figures on the right, we see that the increase of the variance due to even a large number (2000) of four-cycles is almost negligible (around 1%).

## V. Conclusions

Exact analytical expressions in closed form for the mean and the variance of the syndrome weight distribution of linear codes in general and LDPC codes in particular have been derived. Codes with irregular and regular check node distributions have been considered separately. The concurrence matrix $\mathbf{\Lambda} = \mathbf{H}\mathbf{H}^T$ plays a very important role throughout the analysis. An interesting connection between the matrix $\mathbf{\Lambda}$ and the variable node distribution of the Tanner graph was found. Relations between different types of LDPC codes and different combinatorial designs have been established. From all codes with a given variable node distribution, those codes that have a Tanner graph without cycles of length four show the minimum possible variance of the syndrome weight. We performed numerical simulations of error words and corresponding syndrome weights to check our analytical results. The simulations also show that the syndrome weight distribution is close to a discrete normal distribution.

## Acknowledgements

This research has been funded in part by the Austrian COMET Program (FFG).

## References

[1] V. Toto-Zarasoa, A. Roumy, and C. Guillemot, "Maximum Likelihood BSC Parameter Estimation for the Slepian-Wolf Problem," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 232–234, Feb. 2011.

[2] G. Lechner and C. Pacher, "Estimating channel parameters from the syndrome of a linear code," *Communications Letters, IEEE*, vol. 17, no. 11, pp. 2148–2151, November 2013.

[3] W. C. Huffman and V. Pless, *Fundamentals of Error-Correcting Codes*. Cambridge University Press, 2003, cambridge Books Online. [Online]. Available: http://dx.doi.org/10.1017/CBO9780511807077

[4] I. Adjudeanu, J.-Y. Chouinard, and P. Fortier, "On the correlation between error weights and syndrome weights for belief propagation decoding of ldpc codes," in *Information Theory, 2009. CWIT 2009. 11th Canadian Workshop on*, May 2009, pp. 36–41.

[5] R. Tanner, "A recursive approach to low complexity codes," *Information Theory, IEEE Transactions on*, vol. 27, no. 5, pp. 533–547, Sep 1981.

[6] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008, cambridge Books Online.

[7] Y. Fujiwara, A. Gruner, and P. Vandendriessche, "High-rate quantum low-density parity-check codes assisted by reliable qubits," *Information Theory, IEEE Transactions on*, vol. 61, no. 4, pp. 1860–1878, April 2015.

[8] B. Ammar, B. Honary, Y. Kou, J. Xu, and S. Lin, "Construction of low-density parity-check codes based on balanced incomplete block designs," *Information Theory, IEEE Transactions on*, vol. 50, no. 6, pp. 1257–1267, 2004.

[9] W. D. Wallis, "Regular graph designs," *Journal of Statistical Planning Inference*, vol. 51, pp. 273–281, 1996.

[10] S. Johnson and S. R. Weller, "Resolvable 2-designs for regular low-density parity-check codes," *Communications, IEEE Transactions on*, vol. 51, no. 9, pp. 1413–1419, Sept 2003.