On the Typicality of the Linear Code Among the LDPC Coset Code Ensemble

C.-C. Wang, S.R. Kulkarni, and H.V. Poor[1]
Department of Electrical Engineering
Princeton University
Princeton, New Jersey 08544
e-mail: {chihw, kulkarni, poor}@princeton.edu

Abstract — **Density evolution (DE) is one of the most powerful analytical tools for low-density parity-check (LDPC) codes on memoryless binary-input/symmetric-output channels. The case of non-symmetric channels is tackled either by the LDPC coset code ensemble (a channel symmetrizing argument) or by the generalized DE for linear codes on non-symmetric channels. Existing simulations show that the bit error rate performances of these two different approaches are nearly identical. This paper explains this phenomenon by proving that as the minimum check node degree d_c becomes sufficiently large, the performance discrepancy of the linear and the coset LDPC codes is theoretically indistinguishable. This typicality of linear codes among the LDPC coset code ensemble provides insight into the concentration theorem of LDPC coset codes.**

I. Introduction

Low-density parity-check (LDPC) codes [1] have found many applications in cellular networks, magnetic/optical storage devices, and satellite communications, due to their near-capacity performance and the embedded efficient distributed decoding algorithms, namely, the belief propagation (BP) decoder [2]. For binary-input/symmetric-output (BI-SO) channels, the behavior of the BP decoder and the decodable noise threshold can be explained and predicted by the density evolution (DE) method, which traces the evolved distribution on the log-likelihood ratio (LLR) message used in BP [3]. Additional references on LDPC codes can be found in [4].

Although the classical DE does not apply to binary-input/non-symmetric-output (BI-NSO) channels, in practice, LDPC codes are applicable to BI-NSO channels as well and near capacity performance is reported [5]. Rigorous analyses of BI-NSO channels are addressed either by the coset code argument (namely, a channel-symmetrizing argument) [6, 7] or by the generalized DE for linear codes on BI-NSO channels [5].

A coset code consists of all sequences $\mathbf{x}$ of length n satisfying

$$\mathbf{H}\mathbf{x} = \mathbf{s}, \tag{1}$$

for some fixed, coset-defining syndrome $\mathbf{s}$, where $\mathbf{H}$ is a fixed parity-check matrix of dimension $(n(1-R)) \times n$ and R is the rate of this coset code. When $\mathbf{s} = \mathbf{0}$, (1) corresponds to a linear code. It has been shown in [6] that for sufficiently large n, almost all $\mathbf{s} \in \{0,1\}^{n(1-R)}$ and almost all $\mathbf{H}$ drawn from the equiprobable bipartite graph ensemble, the *codeword-averaged* performance can be predicted by the coset-code-based DE within arbitrary precision. If one further assumes that there is a common independent, uniformly distributed bit sequence accessible to both the transmitter and the receiver, then a coset-code-averaged (syndrome-$\mathbf{s}$-averaged) scheme can be obtained as in Fig. 1(a). This coset-code-averaged scheme in Fig. 1(a) is equivalent to a linear LDPC code on the symmetrized channel as demonstrated in Fig. 1(b), of which the error-probability is codeword-independent. On the other hand, the generalized DE in [5] analyzes the *codeword-averaged* performance when linear codes plus BI-NSO channels are considered as in Fig. 1(c). It is shown in [5] that the necessary and sufficient stability conditions in both schemes (Figs. 1(b) and 1(c)) are identical. Monte Carlo simulations based on finite-length codes ($n = 10^4$) [7] further show that the codeword-averaged performance in Fig. 1(c) is nearly identical[1] to the performance[2] of Fig. 1(b) when the same encoder/decoder pair is used. The above two facts suggest close a relationship between linear codes and the coset code ensemble.

This paper addresses this phenomenon by proving that for sufficiently large minimum check node degree d_c, the asymptotic performance (and behavior) of linear LDPC codes is theoretically indistinguishable from that of the LDPC coset code ensemble. In practice, the convergence rate of the thresholds of these two schemes is very fast (with respect to d_c). For moderate $d_c \geq 6$, the discrepancy of the asymptotic thresholds[3] for the linear codes and the coset code ensemble (Figs. 1(b) and 1(c)) is within 0.05%. Our result shows the typicality of linear codes among the coset code ensemble. Besides its theoretical importance, we are then on solid ground when simulating the codeword-averaged linear code performance by assuming the all-zero codeword in the coset code ensemble.

II. Formulation

A. Code Ensembles

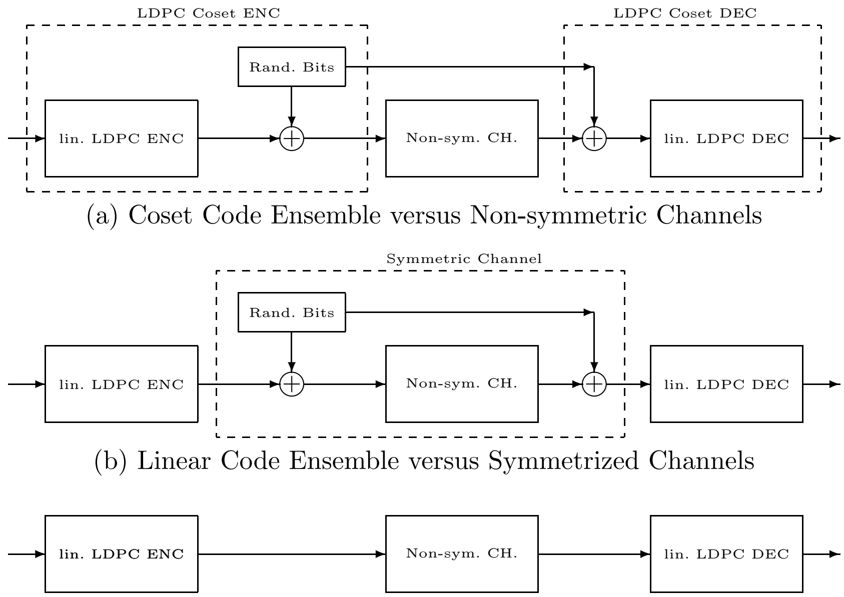
The coset code ensemble is based on (1), where the coset-defining syndrome $\mathbf{s}$ is uniformly distributed in $\{0,1\}^{n(1-R)}$ and $\mathbf{H}$ is from the equiprobable bipartite graph ensemble. To be more explicit, the equiprobable bipartite graph ensemble $\mathcal{C}^n(d_v, d_c)$ is obtained by putting equal probability on each of the possible configurations of the regular bipartite graphs with the variable node degree d_v and the check degree d_c, and by the convention that $H_{j,i}$, the (j,i)-th entry of the parity

[1]That is, it is within the precision of the Monte Carlo simulation.

[2]In Fig. 1(b), the error probability is codeword-independent so that the all-zero codeword can be assumed, which drastically simplifies the computation of taking the average over the entire codebook.

[3]The asymptotic thresholds (with respect to n) are obtained by DE/generalized DE rather than Monte Carlo simulation.

[1]This research was supported in part by the Army Research Laboratory under Contract DAAD-19-01-2-0011.



(a) Coset Code Ensemble versus Non-symmetric Channels

(b) Linear Code Ensemble versus Symmetrized Channels

(c) Linear Code Ensemble versus Non-symmetric Channels

Fig. 1: Comparison of the approaches based on the coset code ensemble and on codeword averaging.

check matrix $\mathbf{H}$, equals one iff there is an *odd* number of edges connecting variable node i and check node j. We can also consider irregular code ensembles $\mathcal{C}^n(\lambda, \rho)$ such that λ and ρ denote the finite order *edge degree distribution* polynomials

$$\begin{aligned} \lambda(z) &= \sum_k \lambda_k z^{k-1} \\ \rho(z) &= \sum_k \rho_k z^{k-1}, \end{aligned}$$

where λ_k or ρ_k is the fraction of edges connecting to a degree k variable or check node, respectively. Further details on the equiprobable bipartite graph ensemble can be found in [3].

If we hard-wire $\mathbf{s} = \mathbf{0}$ and still let $\mathbf{H}$ be drawn from $\mathcal{C}^n(\lambda, \rho)$, we then obtain the traditional linear LDPC code ensemble.

B. The Classical Density Evolutions

In this paper, we consider only the BP decoder such that the passed message m corresponds to the LLR $m = \ln \frac{\mathsf{P}(y|X=0)}{\mathsf{P}(y|X=1)}$. The detailed representation of the variable and check node message maps is as follows.

$$\begin{aligned} m_0 &:= \ln \frac{\mathsf{P}(y|x=0)}{\mathsf{P}(y|x=1)} \\ \Psi_v(m_0, m_1, \cdots, m_{d_v-1}) &:= \sum_{j=0}^{d_v-1} m_j \\ \Psi_c(m_1, \cdots, m_{d_c-1}) &:= \gamma^{-1}\left(\sum_{i=0}^{d_c-1} \gamma(m_i)\right), \end{aligned}$$

where $\gamma : \mathbb{R} \mapsto \mathsf{GF}(2) \times \mathbb{R}^+$ is such that

$$\gamma(m) := \left(1_{\{m\leq 0\}}, \ln\coth\left|\frac{m}{2}\right|\right) = (\gamma_1, \gamma_2) \in \mathsf{GF}(2) \times \mathbb{R}^+.$$

For BI-SO channels, the probability density of the messages in any symmetric message passing algorithm is codeword independent, by which we mean that for different transmitting codewords, the densities of the messages are of the same shape and differ only in parities. Let $P^{(l)}$ denote the density of the LLR messages from variable nodes to check nodes during the l-th iteration given that the all-zero codeword is being transmitted. Similarly, $Q^{(l)}$ denotes the density of the LLR message from check nodes to variable nodes assuming the all-zero codeword. The classical DE [3] derives the iterative functionals on the evolved densities as follows.

$$\begin{aligned} P^{(l)} &= P^{(0)} \otimes \lambda\left(Q^{(l-1)}\right) \\ Q^{(l-1)} &= \Gamma^{-1}\left(\rho\left(\Gamma\left(P^{(l-1)}\right)\right)\right), \end{aligned}$$

where "$\otimes$" denotes the convolution operator and all scalar multiplications in $\lambda(\cdot)$ and $\rho(\cdot)$ are replaced by convolutions as well. The operator Γ transforms the distribution on $\mathbb{R}$ into a distribution on $\mathsf{GF}(2) \times \mathbb{R}^+$ based on the measurable function $\gamma(\cdot)$. Γ^{-1} represents the corresponding inverse transform.

C. Generalized DE and the Coset-Code-Based Approach

For BI-NSO channels, the error-protection capability is codeword dependent and we cannot assume that the all-zero codeword is transmitted. This difficulty is circumvented by the codeword-averaged approach in which we trace pairs of evolved densities, $\left((P^{(l)}(0), P^{(l)}(1)\right)$ and $\left((Q^{(l)}(0), Q^{(l)}(1)\right)$, where $P^{(l)}(x)$ denotes the distribution of the LLR message $m = \ln \frac{\mathsf{P}(y|X=x)}{\mathsf{P}(y|X=\bar{x})}$ from the variable node to the check node during the l-th iteration, given that the transmitting bit at the *source* variable node is x. $Q^{(l)}(x)$ denotes the distribution of the LLR message $m = \ln \frac{\mathsf{P}(y|X=x)}{\mathsf{P}(y|X=\bar{x})}$ from the check node to the variable node during the l-th iteration, given that the transmitting bit at the *destination* variable node is x. The generalized DE for linear codes on BI-NSO channels can then be stated as follows.

$$P^{(l)}(x) = P^{(0)}(x) \otimes \lambda\left(Q^{(l-1)}(x)\right) \quad (2)$$

$$Q^{(l-1)}(x) = \Gamma^{-1}\left(\rho\left(\Gamma\left(\frac{P^{(l-1)}(0)+P^{(l-1)}(1)}{2}\right)\right)\right. + \left.(-1)^x \rho\left(\Gamma\left(\frac{P^{(l-1)}(0)-P^{(l-1)}(1)}{2}\right)\right)\right) \quad (3)$$

The proportion of incorrect variable-to-check messages in the l-th iteration can be computed by

$$p_{e,linear}^{(l)} := \int_{m=-\infty}^{0} \left(\frac{P^{(l)}(0)+P^{(l)}(1)}{2}\right)(dm).$$

By iteratively computing $\left(P^{(l)}(0), P^{(l)}(1)\right)$ and checking whether $p_{e,linear}^{(l)}$ converges to zero, we can determine whether the channel of interest is asymptotically decodable when a sufficiently long linear LDPC code is applied. Further discussion on the generalized DE can be found in [5].

Let $\langle\cdot\rangle$ denote the average operator such that $\langle f\rangle := \frac{f(0)+f(1)}{2}$. One can easily show that the DE corresponding to Fig. 1(b) becomes

$$P_{coset}^{(l)} = \left\langle P^{(0)}\right\rangle \otimes \lambda\left(Q_{coset}^{(l-1)}\right) \quad (4)$$

$$Q_{coset}^{(l-1)} = \Gamma^{-1}\left(\rho\left(\Gamma\left(P_{coset}^{(l-1)}\right)\right)\right), \quad (5)$$

and the proportion of incorrect messages is

$$p_{e,coset}^{(l)} := \int_{m=-\infty}^{0} P_{coset}^{(l)}(dm).$$

One can check whether $p_{e,coset}^{(l)}$ converges to zero to determine whether the channel of interest is asymptotically decodable with a sufficiently long coset code ensemble. Further discussion on the coset-code-based approach can be found in [6].

III. Typicality of Linear LDPC Codes

It was conjectured in [7] that the scheme in Fig. 1(c) should have the same/similar codeword-averaged performance as those illustrated by Figs. 1(a) and 1(b). To be more precise, the question is whether for the same channel model (namely, for the same initial distribution pair $\left(P^{(0)}(0), P^{(0)}(1)\right)$) we are able to show

$$\lim_{l\to\infty} p_{e,linear}^{(l)} = 0 \overset{?}{\Longleftrightarrow} \lim_{l\to\infty} p_{e,coset}^{(l)} = 0.$$

This paper is devoted to the above question. We can answer immediately that the performance of the linear code ensemble is very unlikely to be identical to that of the coset code ensemble. However, when the minimum check node degree $d_{c,min} := \{k \in \mathbb{N} : \rho_k > 0\}$ is relatively large, we can prove that their performance discrepancy is theoretically indistinguishable. In practice, the discrepancy of decodable thresholds of the linear and the coset code ensemble is within 0.05% for a moderate $d_{c,min} \geq 6$.

It is clear from (2) that for linear codes, the variable node iteration involves convolution of several densities having the same x value. The difference between $Q^{(l-1)}(0)$ and $Q^{(l-1)}(1)$ is thus amplified after each variable node iteration. It is very unlikely that the decodable threshold of linear codes (obtained from (2) and (3)) and the decodable threshold of coset codes

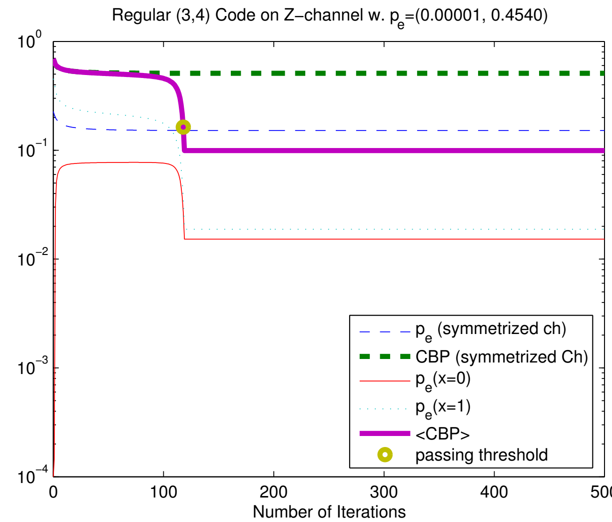


Fig. 2: Density evolution for z-channels with the linear code ensemble and the coset code ensemble.

(obtained from (4) and (5)) will be analytically identical after the amplification during the variable node iterations. Fig. 2 demonstrates the traces of the evolved densities for the regular (3,4) code on z-channels.[4] With the one-way crossover probability being 0.4540, the generalized DE for linear codes is able to converge within 179 iterations, while the coset code ensemble shows no convergence within 500 iterations. This demonstrates the possible performance discrepancy, though we do not have analytical results proving that the latter will not converge after more iterations. Table 1 compares the decodable thresholds such that the density evolution enters the stability region within 100 iterations. We notice that the larger $d_{c,min}$ is, the smaller the discrepancy is. This phenomenon can be characterized by the following theorem.

Theorem 1 *Consider BI-NSO channels and a fixed pair of finite-degree polynomials λ and ρ. The shifted version of the check node polynomial is denoted by $\rho_\Delta = x^\Delta \cdot \rho$ where $\Delta \in \mathbb{N}$. Let $P_{coset}^{(l)}$ denote the evolved density from the coset code ensemble with degrees (λ, ρ_Δ) (obtained from (4) and (5)), and $\langle P^{(l)}\rangle = \frac{1}{2}\sum_{x=0,1} P^{(l)}(x)$ denote the averaged density from the linear code ensemble with degrees (λ, ρ_Δ) (obtained from (2) and (3)). Then, for any $l_0 \in \mathbb{N}$, $\lim_{\Delta\to\infty}\langle P^{(l)}\rangle \overset{\mathcal{D}}{=} P_{coset}^{(l)}$ in distribution for all $l \leq l_0$, with the convergence rate being $\mathcal{O}\left(\mathsf{const}^\Delta\right)$ for some $\mathsf{const} < 1$.*

Corollary 1 (The Typicality on Z-Channels) *Define*

$$p_{1\to0,linear}^{*} := \sup\left\{p_{1\to0} > 0 : \lim_{l\to\infty} p_{e,linear}^{(l)} = 0\right\}$$

$$\text{and } p_{1\to0,coset}^{*} := \sup\left\{p_{1\to0} > 0 : \lim_{l\to\infty} p_{e,coset}^{(l)} = 0\right\}.$$

For any $\epsilon > 0$, there exists a $\Delta \in \mathbb{N}$ such that

$$\left|p_{1\to0,linear}^{*} - p_{1\to0,coset}^{*}\right| < \epsilon.$$

[4] A z-channel is a binary-input/binary-output channel such that only bit value 1 may be contaminated with one-way crossover probability $p_{1\to0}$. Bit value 0 will always be received perfectly.

Table 1: Threshold comparison $p^*_{1\to 0}$ of linear and coset LDPC codes on Z-channels

(λ,ρ)	(x^2,x^3)	(x^2,x^5)	$(x^2,0.5x^2+0.5x^3)$	$(x^2,0.5x^4+0.5x^5)$
Linear	0.4540	0.2305	0.5888	0.2689
Coset	0.4527	0.2304	0.5908	0.2690

Namely, the asymptotic decodable thresholds of the linear and the coset code ensemble are arbitrarily close when the minimum check node degree $d_{c,min}$ is sufficiently large.

Similar corollaries can be constructed for other channel models with different types of noise parameters. For example, the σ^* in the binary-input additive white Gaussian channel, the λ^* in the binary-input Laplace channel, etc. The proof of **Corollary 1** is in APPENDIX A.

Proof of **Theorem 1***:* Since the functionals in (2) and (3) are continuous with respect to convergence in distribution, we only need to show that $\forall l \in \mathbb{N}$,

$$\begin{aligned}
&\lim_{\Delta\to\infty} Q^{(l-1)}(0) \stackrel{\mathcal{D}}{=} \lim_{\Delta\to\infty} Q^{(l-1)}(1)\\
&\stackrel{\mathcal{D}}{=} \Gamma^{-1}\left(\rho\left(\Gamma\left(\frac{P^{(l-1)}(0)+P^{(l-1)}(1)}{2}\right)\right)\right)\\
&= \frac{Q^{(l-1)}(0)+Q^{(l-1)}(1)}{2}, \qquad (6)
\end{aligned}$$

where $\stackrel{\mathcal{D}}{=}$ denotes convergence in distribution. Then by inductively applying this weak convergence argument, for any bounded l_0, $\lim_{\Delta\to\infty}\langle P^{(l)}\rangle \stackrel{\mathcal{D}}{=} P^{(l)}_{coset}$ in distribution for all $l \le l_0$. Without loss of generality,[5] we may assume $\rho_\Delta = x^\Delta$ and prove the weak convergence of distributions on the domain

$$\gamma(m) \quad := \quad \left(1_{\{m\le 0\}}, \ln\coth\left|\frac{m}{2}\right|\right) = (\gamma_1,\gamma_2) \in \mathsf{GF}(2)\times\mathbb{R}^+,$$

on which the check node iteration becomes

$$\gamma_{out,\Delta} = \gamma_{in,1}+\gamma_{in,2}+\cdots+\gamma_{in,\Delta}.$$

Let P'_0 denote the density of $\gamma_{in}(m)$ given that the distribution of m is $P^{(l-1)}(0)$ and let P'_1 similarly correspond to $P^{(l-1)}(1)$. Similarly let $Q'_{0,\Delta}$ and $Q'_{1,\Delta}$ denote the output distributions on $\gamma_{out,\Delta}$ when the check node degree is $\Delta+1$. It is worth noting that any pair of $Q'_{0,\Delta}$ and $Q'_{1,\Delta}$ can be mapped bijectively back to the LLR distributions $Q^{(l-1)}(0)$ and $Q^{(l-1)}(1)$.

Let $\Phi_{P'}(k,r) := \mathsf{E}_{P'}\left\{(-1)^{k\gamma_1}e^{ir\gamma_2}\right\}, \forall k\in\mathbb{N}, r\in\mathbb{R}$, denote the Fourier transform of the density P'. Proving (6) is equivalent to showing that

$$\forall k\in\mathbb{N}, r\in\mathbb{R},\quad \lim_{\Delta\to\infty}\Phi_{Q'_{0,\Delta}}(k,r) = \lim_{\Delta\to\infty}\Phi_{Q'_{1,\Delta}}(k,r).$$

However, to deal with the strictly growing average of the limit distribution on the second component of $\gamma_{out,\Delta}$, we concentrate instead on the distribution of the normalized output $\left(\gamma_{1,out,\Delta}, \frac{\gamma_{2,out,\Delta}}{\Delta}\right)$. We then need to prove that

$$\forall k\in\mathbb{N}, r\in\mathbb{R},\quad \lim_{\Delta\to\infty}\Phi_{Q'_{0,\Delta}}(k,\frac{r}{\Delta}) = \lim_{\Delta\to\infty}\Phi_{Q'_{1,\Delta}}(k,\frac{r}{\Delta}).$$

[5] We also need to assume that $\forall x, P^{(l-1)}(x)(m=0)=0$ so that $\ln\coth\left|\frac{m}{2}\right| \in \mathbb{R}^+$ almost surely. This assumption can be relaxed by separately considering the event that $m_{in,i}=0$ for some $i\in\{1,\cdots,d_c-1\}$.

We first note that $Q'_{0,\Delta}$ is the averaged distribution of $\gamma_{out,\Delta}$ when the inputs $\gamma_{in,i}$ are governed by $P^{(l-1)}(x_i)$ with $\sum_{i=1}^{\Delta} x_i = 0$. Similarly $Q'_{1,\Delta}$ is the averaged distribution of $\gamma_{out,\Delta}$ when the inputs $\gamma_{in,i}$ are governed by $P^{(l-1)}(x_i)$ with $\sum_{i=1}^{\Delta} x_i = 1$. From the above observation, we can derive the following iterative equations: $\forall\Delta\in\mathbb{N}$,

$$\begin{aligned}
\Phi_{Q'_{0,\Delta}}(k,\frac{r}{\Delta}) &= \frac{1}{2}\sum_{x=0,1}\Phi_{Q'_{x,\Delta-1}}(k,\frac{r}{\Delta})\Phi_{P'_x}(k,\frac{r}{\Delta})\\
\Phi_{Q'_{1,\Delta}}(k,\frac{r}{\Delta}) &= \frac{1}{2}\sum_{x=0,1}\Phi_{Q'_{x,\Delta-1}}(k,\frac{r}{\Delta})\Phi_{P'_{\bar{x}}}(k,\frac{r}{\Delta}).
\end{aligned}$$

By induction, the difference thus becomes

$$\begin{aligned}
&\Phi_{Q'_{0,\Delta}}(k,\frac{r}{\Delta}) - \Phi_{Q'_{1,\Delta}}(k,\frac{r}{\Delta})\\
&= \left(\Phi_{Q'_{0,\Delta-1}}(k,\frac{r}{\Delta}) - \Phi_{Q'_{1,\Delta-1}}(k,\frac{r}{\Delta})\right)\\
&\quad\cdot\left(\frac{\Phi_{P'_0}(k,\frac{r}{\Delta})-\Phi_{P'_1}(k,\frac{r}{\Delta})}{2}\right)\\
&= 2\left(\frac{\Phi_{P'_0}(k,\frac{r}{\Delta})-\Phi_{P'_1}(k,\frac{r}{\Delta})}{2}\right)^{\Delta}. \qquad (7)
\end{aligned}$$

By Taylor's expansion and the channel decomposition argument in [8], we can show that for all $k\in\mathbb{N}$, $r\in\mathbb{R}$, and for all possible P'_0 and P'_1, the quantity in (7) converges to zero with convergence rate $\mathcal{O}\left(\mathsf{const}^\Delta\right)$ for some $\mathsf{const} < 1$. A detailed derivation of the convergence rate is given in APPENDIX B. Since the limit of the right-hand side of (7) is zero, the proof of weak convergence is complete. The exponentially fast convergence rate $\mathcal{O}\left(\mathsf{const}^\Delta\right)$ also justifies the fact that even for moderate $d_{c,min}$ (e.g. $d_{c,min}\ge 6$), the performances of linear and coset LDPC codes are very close. ■

Remark 1: Consider any non-perfect message distribution, namely, $\exists x_0\in\{0,1\}$ such that $P^{(l-1)}(x_0)\neq\delta_\infty$, where δ_{m_0} is the Dirac delta measure centered on m_0. A persistent reader may notice that $\forall x, \lim_{\Delta\to\infty} Q^{(l-1)}(x) \stackrel{\mathcal{D}}{=} \delta_0$. That is, as Δ becomes large, all information is erased after passing a check node of degree Δ. If this convergence (erasure effect) occurs earlier than the convergence of $Q^{(l-1)}(0)$ and $Q^{(l-1)}(1)$, the performances of linear and coset LDPC codes are "close" only when the corresponding codes are "useless."[6] To quantify the convergence rate, we consider again the distributions on γ and their Fourier transforms. For the average of the output

[6] To be more precise, they correspond to extremely high-rate codes and the information is erased after every check node iteration.

distributions $\left\langle Q^{(l-1)} \right\rangle$, we have

$$\begin{aligned}
&\frac{\Phi_{Q'_{0,\Delta}}(k, \frac{r}{\Delta}) + \Phi_{Q'_{1,\Delta}}(k, \frac{r}{\Delta})}{2} \\
&= \left(\frac{\Phi_{Q'_{0,\Delta-1}}(k, \frac{r}{\Delta}) + \Phi_{Q'_{1,\Delta-1}}(k, \frac{r}{\Delta})}{2} \right) \\
&\quad \cdot \left(\frac{\Phi_{P'_0}(k, \frac{r}{\Delta}) + \Phi_{P'_1}(k, \frac{r}{\Delta})}{2} \right) \\
&= \left(\frac{\Phi_{P'_0}(k, \frac{r}{\Delta}) + \Phi_{P'_1}(k, \frac{r}{\Delta})}{2} \right)^{\Delta}. \qquad (8)
\end{aligned}$$

By Taylor's expansion and the channel decomposition argument, one can show that the limit of (8) exists and the convergence rate is $\mathcal{O}(\Delta^{-1})$. (A detailed derivation is included in Appendix B.) This convergence rate is much slower than the exponential rate $\mathcal{O}\left(\mathsf{const}^{\Delta}\right)$ in the proof of **Theorem 1**. Therefore, we do not need to worry about the case in which the required Δ for the convergence of $Q^{(l-1)}(0)$ and $Q^{(l-1)}(1)$ is excessively large so that $\forall x \in \mathsf{GF}(2), Q^{(l-1)}(x) \overset{\mathcal{D}}{\approx} \delta_0$.

Remark 2: The intuition behind **Theorem 1** is that when the minimum d_c is sufficiently large, the parity check constraint becomes relatively less stringent. Thus we can approximate the density of the outgoing messages for linear codes by assuming all bits $\{x_i\}_{i\in\{1,\cdots,d_c-2\}}$ involved in that particular parity check equation are "independently" distributed in $\{0,1\}$ rather than satisfying $\sum x_i = x$, which leads to the formula for the coset code ensemble. On the other hand, extremely large d_c is required for a check node iteration to completely destroy all information coming from the previous iteration. This explains the difference between their convergence rates: $\mathcal{O}\left(\mathsf{const}^{\Delta}\right)$ versus $\mathcal{O}(\Delta^{-1})$.

Fig. 3 illustrates the weak convergence predicted by **Theorem 1** and depicts the convergence rates of $Q^{(l-1)}(0) \longrightarrow Q^{(l-1)}(1)$ and $\frac{Q^{(l-1)}(0)+Q^{(l-1)}(1)}{2} \to \delta_0$.

IV. Conclusions

The typicality of the linear LDPC code ensemble has been proven by the weak convergence (w.r.t. d_c) of the evolved densities in our codeword-averaged density evolution. Namely, when the check node degree is sufficiently large (e.g. $d_c \geq 6$), the performance of the linear LDPC code ensemble is very close to (e.g. within 0.05%) the performance of the LDPC coset code ensemble. This result can be viewed as a complementing theorem of the concentration theorem in [*Corollary 2.2* of [6]], where a constructive method of finding a typical coset-defining syndrome $\mathbf{s}$ is not specified.[7]

Besides the theoretical importance, we are then on a solid basis to interchangeably use the linear LDPC codes and the LDPC coset codes when the check node degree is of moderate size. For instance, from the implementation point of view, the hardware uniformity of linear codes makes them a superior choice compared to any other coset code. We can then use fast density evolution [9] plus the coset code ensemble to optimize the degree distribution for the linear LDPC codes. Or instead of simulating the codeword-averaged performance of linear LDPC codes, we can simulate the error probability of the all-zero codeword in the coset code ensemble, in which the efficient LDPC encoder [10] is not necessary.

[7] Our result shows the typicality of the all-zero $\mathbf{s}$ when d_c is sufficiently large. The results in [6], on the other hand, prove that a very large proportion of $\mathbf{s}$ is typical but $\mathbf{0}$ may or may not be one of them.

Appendices

A. Proof of **Corollary 1**

We prove one direction that

$$p^*_{1\to 0,linear} \quad > \quad p^*_{1\to 0,coset} - \epsilon.$$

The other direction that $p^*_{1\to 0,coset} > p^*_{1\to 0,linear} - \epsilon$ can be easily obtained by symmetry. One prerequisite of the following proof is that both the linear code and the coset code have the same stability region [5].

By definition, for any $\epsilon > 0$, we can find a sufficiently large $l_0 < \infty$ such that for the one-way crossover probability $p_{1\to 0} := p^*_{1\to 0,coset} - \epsilon$, $P^{(l_0)}_{coset}$ is in the interior of the stability region. We first note that the stability region depends only on the Bhattacharyya noise parameter [3], which is a continuous function with respect to convergence in distribution. Therefore, by **Theorem 1**, there exists a $\Delta \in \mathbb{N}$ such that $\left\langle P^{(l_0)} \right\rangle$ is also in the stability region. By the definition of the stability region, we have $\lim_{l\to\infty} p^{(l)}_{e,linear} = 0$, which implies $p^*_{1\to 0,linear} \geq p_{1\to 0}$. The proof is thus complete.

B. The Convergence Rates of (7) and (8)

For (7), we will consider the cases $k = 0$ and $k = 1$ separately. By the binary asymmetric channel (BASC) decomposition argument, namely, all binary-input non-symmetric channels can be decomposed as the probabilistic combination of many BASCs, we can limit our attention to simple BASCs rather than general BI-NSO channels. Suppose $\left(P^{(l-1)}(0), P^{(l-1)}(1)\right)$ corresponds to a BASC with crossover probabilities ϵ_0 and ϵ_1. Without loss of generality, we may assume $\epsilon_0 + \epsilon_1 < 1$ because of the previous assumption that $\forall x \in \mathsf{GF}(2), P^{(l-1)}(x)(m = 0) = 0$. We then have

$$\Phi_{P'_0}(k, \frac{r}{\Delta}) = (1-\epsilon_0) e^{i\frac{r}{\Delta}\ln\frac{1-\epsilon_0+\epsilon_1}{1-\epsilon_0-\epsilon_1}} + (-1)^k \epsilon_0 e^{i\frac{r}{\Delta}\ln\frac{1+\epsilon_0-\epsilon_1}{1-\epsilon_0-\epsilon_1}}$$

$$\Phi_{P'_1}(k, \frac{r}{\Delta}) = (1-\epsilon_1) e^{i\frac{r}{\Delta}\ln\frac{1+\epsilon_0-\epsilon_1}{1-\epsilon_0-\epsilon_1}} + (-1)^k \epsilon_1 e^{i\frac{r}{\Delta}\ln\frac{1-\epsilon_0+\epsilon_1}{1-\epsilon_0-\epsilon_1}}.$$

By Taylor's expansion, for $k = 0$, (7) becomes

$$\begin{aligned}
&2\left(\frac{\Phi_{P'_0}(0, \frac{r}{\Delta}) - \Phi_{P'_1}(0, \frac{r}{\Delta})}{2}\right)^{\Delta} \\
&= 2\left(i\left(\frac{1-\epsilon_0-\epsilon_1}{2}\right)\left(\frac{r}{\Delta}\right)\ln\left(\frac{1-\epsilon_0+\epsilon_1}{1+\epsilon_0-\epsilon_1}\right) + \mathcal{O}\left(\left(\frac{r}{\Delta}\right)^2\right)\right)^{\Delta},
\end{aligned}$$

which converges to zero with convergence rate $\mathcal{O}\left(\mathcal{O}(\Delta)^{-\Delta}\right)$. For $k = 1$, we have

$$\begin{aligned}
&2\left(\frac{\Phi_{P'_0}(1, \frac{r}{\Delta}) - \Phi_{P'_1}(1, \frac{r}{\Delta})}{2}\right)^{\Delta} \\
&= 2\left((\epsilon_1 - \epsilon_0) + \frac{i}{2}\left(\frac{r}{\Delta}\right) f_-(\epsilon_0, \epsilon_1) + \mathcal{O}\left(\left(\frac{r}{\Delta}\right)^2\right)\right)^{\Delta}, \qquad (9)
\end{aligned}$$

where

$$\begin{aligned}
&f_-(\epsilon_0, \epsilon_1) \\
&:= \left((1-\epsilon_0+\epsilon_1)\ln\frac{1-\epsilon_0+\epsilon_1}{1-\epsilon_0-\epsilon_1} - (1+\epsilon_0-\epsilon_1)\ln\frac{1+\epsilon_0-\epsilon_1}{1-\epsilon_0-\epsilon_1}\right).
\end{aligned}$$

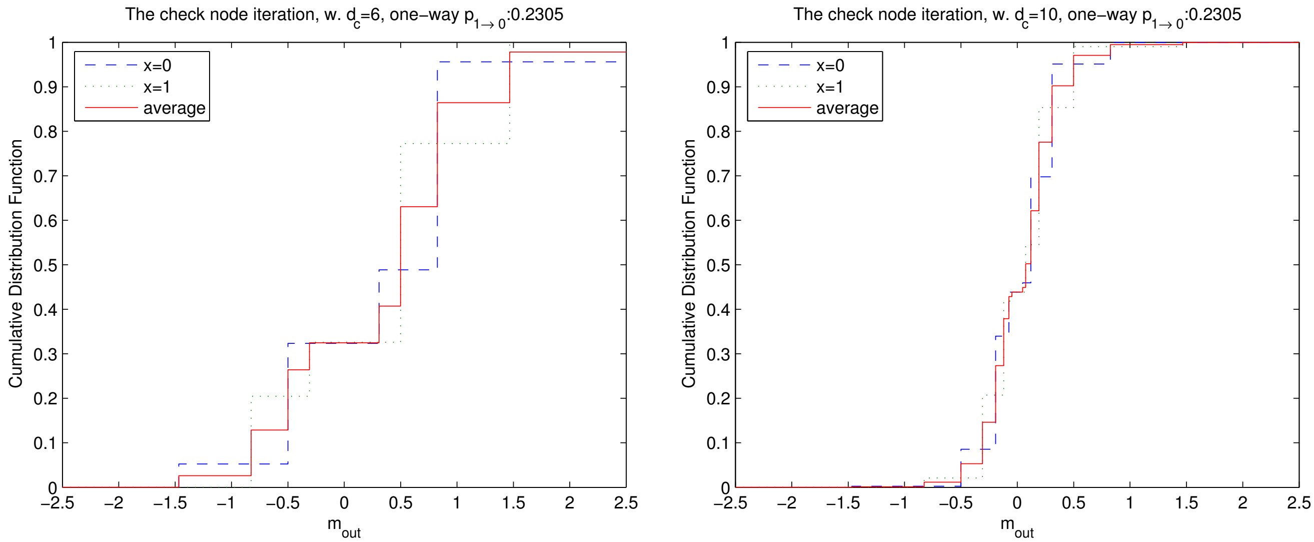


Fig. 3: Illustration of the weak convergence of $Q^{(l-1)}(0)$ and $Q^{(l-1)}(1)$. One can see that the convergence of $Q^{(l-1)}(0)$ and $Q^{(l-1)}(1)$ is faster than the convergence of $\frac{Q^{(l-1)}(0)+Q^{(l-1)}(1)}{2}$ and δ_0.

(9) converges to zero with convergence rate $\mathcal{O}(\mathsf{const}^{\Delta})$, where const satisfies $|\epsilon_1 - \epsilon_0| < \mathsf{const} < 1$. Since the convergence rate is determined by the slower of the above two, we have proven that (7) converges to zero with rate $\mathcal{O}(\mathsf{const}^{\Delta})$ for some $\mathsf{const} < 1$.

Consider (8). By the assumption that the input is not perfect, we have $\max(\epsilon_0, \epsilon_1) > 0$. For $k = 0$, by Taylor's expansion, we have

$$\left(\frac{\Phi_{P_0'}(0, \frac{r}{\Delta}) + \Phi_{P_1'}(0, \frac{r}{\Delta})}{2}\right)^{\Delta}$$
$$= \left(1 + \frac{i}{2}\left(\frac{r}{\Delta}\right) f_+(\epsilon_0, \epsilon_1) + \mathcal{O}\left(\left(\frac{r}{\Delta}\right)^2\right)\right)^{\Delta}, \quad (10)$$

where

$$f_+(\epsilon_0, \epsilon_1)$$
$$:= \left((1 - \epsilon_0 + \epsilon_1) \ln \frac{1 - \epsilon_0 + \epsilon_1}{1 - \epsilon_0 - \epsilon_1} + (1 + \epsilon_0 - \epsilon_1) \ln \frac{1 + \epsilon_0 - \epsilon_1}{1 - \epsilon_0 - \epsilon_1}\right).$$

The quantity in (10) converges to

$$e^{i\left(\frac{r}{2}\right) f_+(\epsilon_1, \epsilon_2)}$$

with rate $\mathcal{O}\left(\Delta^{-1}\right)$. For $k = 1$, we have

$$\left(\frac{\Phi_{P_0'}(1, \frac{r}{\Delta}) + \Phi_{P_1'}(1, \frac{r}{\Delta})}{2}\right)^{\Delta}$$
$$= \left((1 - \epsilon_0 - \epsilon_1)\left(\frac{e^{i\frac{r}{\Delta}\ln\frac{1-\epsilon_0+\epsilon_1}{1-\epsilon_0-\epsilon_1}} + e^{i\frac{r}{\Delta}\ln\frac{1+\epsilon_0-\epsilon_1}{1-\epsilon_0-\epsilon_1}}}{2}\right)\right)^{\Delta},$$

which converges to zero with rate $\mathcal{O}\left((1 - \epsilon_0 - \epsilon_1)^{\Delta}\right)$. Since the overall convergence rate is the slower of the above two, we have proven that the convergence rate is $\mathcal{O}\left(\Delta^{-1}\right)$.

References

[1] R. G. Gallager, *Low-Density Parity-Check Codes*, Number 21 in Research Monograph Series. MIT Press, Cambridge, MA, 1963.

[2] R. J. McEliece, D. J. C. Mackay, and J. F. Cheng, "Turbo decoding as an instance of Pearl's "Belief Propagation" algorithm," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 140–152, Feb. 1998.

[3] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[4] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[5] C. C. Wang, S. R. Kulkarni, and H. V. Poor, "Density evolution for asymmetric memoryless channels," in *Proc. Int'l. Symp. Turbo Codes & Related Topics*. Brest, France, 2003, pp. 121–124.

[6] A. Kavčić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution and code performance bound," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1636–1652, July 2003.

[7] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 49, no. 9, pp. 2141–2155, Sept. 2003.

[8] C. C. Wang, S. R. Kulkarni, and H. V. Poor, "On finite-dimensional bounds for LDPC-like codes with iterative decoding," in *Proc. Int'l Symp. Inform. Theory & its Applications*. Parma, Italy, Oct. 2004.

[9] H. Jin and T. J. Richardson, "Fast density evolution," in *Proc. 38th Conf. Inform. Sciences and Systems*. Princeton, NJ, USA, 2004.

[10] T. J. Richardson and R. L. Urbanke, "Efficient encoding of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 638–656, Feb. 2001.