The Asymptotic Uniformity of the Output of Convolutional Codes Under Markov Inputs

Patrick Mitran, *Member, IEEE*

***Abstract*— In this letter, we prove a published conjecture on the asymptotic uniformity of the outputs of a convolutional encoder under biased inputs. These results are interesting in light of recent research on joint source-channel coding as well as source coding using turbo codes in which the constituent encoders are convolutional codes. In particular, it is well-known that in many situations a good code should result in a uniform distribution on blocks of consecutive encoded symbols. The results presented here provide insights into the choice of encoders in such scenarios.**

***Index Terms*— Convolutional codes, nonuniform sources, joint source-channel coding.**

I. Introduction

Much recent research in coding theory has been in the area of joint source-channel coding as well as distributed source coding and data compression using methods that are more traditionally associated with channel coding. For example, in [9], non-systematic turbo codes are studied in the context of joint source-channel coding over additive white Gaussian noise (AWGN) channels and Raleigh fading channels while in [8] non-systematic LDPC codes are investigated in the context of AWGN channels. More recently, in [1] error probability bounds are derived for codes in the presence of biased inputs and it was found that a good code should map information sequences with high Hamming distances to codewords with low Hamming distances. In [4] and [5], parallel concatenated turbo codes are employed to compress i.i.d. binary sources and are applied to the Slepian-Wolf distributed source coding problem respectively. In [2], serial turbo codes were applied to the Slepian-Wolf problem.

This work is motivated by an observation in [9]. In particular, it was noted there that one of the limiting factors in approaching the Shannon limit in joint source-channel coding (and likewise for data compression applications) with concatenated convolutional codes is the mismatch between the empirical distribution at the input to the channel and the capacity achieving input distribution. In particular, for AWGN and Raleigh channels, the capacity achieving input distribution is the uniform distribution and a good joint source-channel coding scheme should result in a nearly uniform output distribution (and hence the non-systematic nature of [8], [9]). Indeed, for traditional channel coding it has been shown that for any fixed positive integer $k > 0$, the k-order empirical distribution (see eq. (3) of [7]) of any sequence of good channel codes in the information theoretic sense converges (in divergence) to the capacity achieving input distribution [7].

If the input sequence to a convolutional encoder is biased, no finite length output has a uniform distribution. However if the encoding block lengths are moderately large, then the asymptotic distribution of M consecutive outputs may be a good indicator of channel input matching. In [9], given the feedforward and feedback polynomials $G(D)$ and $F(D)$ in minimal form (i.e., they do not share any common factors), it was conjectured that asymptotically for any bias $0 < \mathbb{P}[U_k = 1] < 1$ on a sequence of i.i.d. input bits, the distribution on M consecutive output bits is asymptotically uniform where M is the degree of $F(D)$. This result was then verified based on extensive simulations for degrees up to $M = 4$.

In this letter, we prove the stronger theorem below which confirms the conjecture in [9] as a special case.

Theorem 1: Suppose a homogeneous Markov source U_k with distribution $0 < \mathbb{P}_{U_{k+1}|U_k}[u_{k+1}|u_k] < 1$ is input into a recursive convolutional encoder with feedback polynomial $F(D)$ and feed-forward polynomial $G(D)$. If $G(D)/F(D)$ is in its minimal form where M is the degree of $F(D)$, then the M-order distribution of M consecutive outputs $(X_n, \ldots, X_{n+M-1})$ is asymptotically uniform for *all* such input processes as $n \to \infty$ while this is not the case for the $(M+1)$-order distribution. Specifically, if $\mathbb{P}^n$ is the distribution of $(X_n, \ldots, X_{n+M-1})$ then $\mathbb{P}^n$ converges to the uniform distribution as $n \to \infty$. □

In related work, Leeper [6] has shown that it is always possible to whiten all the first and second order statistics of an *arbitrary* binary source at the cost of an arbitrarily small error rate ϵ in the reconstruction of the source. In particular, a bound on the order of the feedback polynomial was derived for which it could be guaranteed that an encoder exists for which *all* outputs are first and second order uniform within a given density imbalance δ. By contrast, here, we derive the exact order of the *asymptotic* uniformity for *consecutive* outputs as a function of the feedback and feedforward polynomials for both i.i.d. and Markov inputs.

II. Definitions and Proof

We consider the traditional structure of a convolutional encoder over $\mathbb{F} = GF(2)$ as illustrated in Fig. 1. If $H(D)$ denotes the D-transform of a one-sided binary sequence $\mathbf{h} = h_0, h_1, h_2, \ldots$, i.e., $H(D) = h_0 + h_1 D + h_2 D^2 + \cdots$, then it is well-known that the state sequence $r_n; k = 0, 1, 2, \ldots$, and output sequence $x_n; k = 0, 1, 2, \ldots$, are related to the input sequence $u_n; n = 0, 1, 2, \ldots$, by the following relations

$$R(D) = U(D)/F(D) \tag{1}$$

$$X(D) = U(D)G(D)/F(D), \tag{2}$$

where $F(D)$ and $G(D)$ are the D-transforms of the $f_0, f_1, \ldots, f_M$ and $g_0, g_1, \ldots, g_M$ sequences respectively (with the convention that $f_0 = 1$).

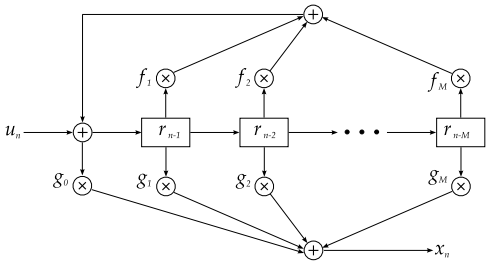

Fig. 1. The structure of a convolutional encoder.

If $F(D) = 1$ then clearly not even the 1st order distribution is in general asymptotically uniform since then by (2), x_n is only a finite sum of biased inputs u_k. Therefore we only consider feedback polynomials $F(D) \neq 1$.

In this case, the impulse response $H(D) := G(D)/F(D)$ is infinite and except for an initial transient of length $T < 2^M$, is periodic with repetitions of the fundamental sequence $\mathbf{c} = (c_0, c_1, \ldots, c_{L-1}) = (h_T, h_{T+1}, \ldots, h_{T+L-1})$ where L is the period of the tail. Since multiple concatenations of the sequence $\mathbf{c}$ is also a periodic sequence in $H(D)$, without loss of generality we assume $L \geq M + 1$.

Also, we denote by $\mathbf{w}$ the infinite vector obtained by repeatedly concatenating $\mathbf{c}$ and we denote by S the operation of cyclically left shifting a sequence, i.e., $\mathsf{S}\mathbf{c} = (c_1, c_2, \ldots, c_{L-1}, c_0)$. Given a vector $\mathbf{x}$, $\mathbf{x}_n^{(m)}$ denotes the first m entries starting at index n, i.e, $\mathbf{x}_n^{(m)} := (x_n, x_{n+1}, \ldots, x_{n+m-1})$.

In particular, we have that $\mathbf{w}_0^{(m)} = (c_0, \ldots, c_{m-1}) = \mathbf{w}_L^{(m)}$, $\mathbf{w}_1^{(m)} = (c_1, \ldots, c_m)$, …, $\mathbf{w}_{L-1}^{(m)} = (c_{L-1}, c_0, \ldots, c_{m-2})$. In this paper, all vector additions are understood to be modulo 2.

For simplicity of exposition, we first prove the following lemma about the dimension of the space spanned by these L m-windowed sequences in the i.i.d. case which illustrates the key ideas. This is then generalized to the Markov case by highlighting the key differences.

Lemma 2: The output of a convolutional encoder is asymptotically uniform with order m for all i.i.d. input processes $0 < \mathbb{P}_{U_k}[u_k = 1] < 1$ if and only if the vectors $\mathbf{w}_L^{(m)} = \mathbf{w}_0^{(m)}, \ldots, \mathbf{w}_{L-1}^{(m)}$ have dimension m. □

Proof: Given $n = T + kL + r$ with $0 \leq r < L$ then as shown in the Appendix, the output $\mathbf{X}$ of the convolutional encoder satisfies

$$\mathbf{X}_n^{(m)} = \mathbf{V}_1 + \mathbf{V}_2 + \mathbf{V}_3, \tag{3}$$

where,

$$\mathbf{V}_1 = \sum_{p=-m+1}^{T} U_{n-p} \mathbf{h}_p^{(m)} \tag{4}$$

$$\mathbf{V}_2 = \left[\sum_{p=0}^{k-1} \mathbf{U}_{r+pL}^{(L)} \right] W^{(m)} \tag{5}$$

$$\mathbf{V}_3 = \sum_{p=0}^{r-1} U_p \mathbf{w}_{r-p}^{(m)}, \tag{6}$$

and

$$W^{(m),T} = \left[\begin{array}{cccc} \mathbf{w}_L^{(m),T} & \cdots & \mathbf{w}_2^{(m),T} & \mathbf{w}_1^{(m),T} \end{array} \right]^T. \tag{7}$$

In the above, $\mathbf{V}_1$ is the effect of the transient behavior of the impulse response, $\mathbf{V}_2$ is the effect due to full repetitions of the fundamental sequence $\mathbf{c}$ and $\mathbf{V}_3$ is the effect of a partially complete sequence $\mathbf{c}$.

In the limit that $k \to \infty$, each entry in $\tilde{\mathbf{U}} := \sum_{p=0}^{k-1} \mathbf{U}_{r+pL}^{(L)}$ is the modulo sum of infinitely many i.i.d. biased binary random variables. This latter is known to converge to a uniform binary random variable [3]. Furthermore, since each entry of $\mathbf{U}_{r+pL}^{(L)}$ is independent of every other entry of $\mathbf{U}_{r+\tilde{p}L}^{(L)}$, then the entries in the vector $\tilde{\mathbf{U}}$ are independent, i.e., as $k \to \infty$, $\tilde{\mathbf{U}}$ is i.i.d. and asymptotically uniform.

As the $W^{(m)}$ matrix is assumed to have rank m, it follows that the right side of (5) results in a uniform distribution on the vector $\mathbf{V}_2$, i.e., each of the 2^m binary sequences of $\mathbf{V}_2$ are asymptotically equally likely.

As the vectors $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ are independent, then the sum $\mathbf{X}_n^{(m)}$ also has a uniform distribution in the limit as $k \to \infty$ or equivalently, as $n \to \infty$.

To prove the converse, suppose that $\mathbf{z}^T \neq 0$ is in the null space of the matrix $W^{(m)}$. Then $\mathbf{X}_n^{(m)} \mathbf{z}^T = \mathbf{V}_1 \mathbf{z}^T + \mathbf{V}_2 \mathbf{z}^T + \mathbf{V}_3 \mathbf{z}^T = \mathbf{V}_1 \mathbf{z}^T$. Since $\mathbf{V}_1 \mathbf{z}^T$ is the modulo sum of at most $2^M + m$ binary random variables, the distribution is not asymptotically uniform as $n \to \infty$ for any i.i.d. biased input process and hence the distribution on $\mathbf{X}_n^{(m)}$ cannot be asymptotically uniform as $n \to \infty$. ■

The following is a generalization of Lemma 2 to the Markov case where for clarity, we focus on the key differences.

Lemma 3: The output of a convolutional encoder is asymptotically uniform with order m for all binary homogeneous Markov input process with $0 < \mathbb{P}_{U_{k+1}|U_k}[u_{k+1}|u_k] < 1$ if and only if the vectors $\mathbf{w}_0^{(m)}, \ldots, \mathbf{w}_{L-1}^{(m)}$ have dimension m. □

Proof: All the steps in the proof for the i.i.d. case hold expect that we must show that $\tilde{\mathbf{U}}$ is asymptotically uniform under the Markov assumption and while $\mathbf{V}_2$ is not independent of $\mathbf{V}_1$ and $\mathbf{V}_3$ for any fixed n, we will show that in the limit as $n \to \infty$ it is.

First, observe that $\mathbf{V}_2$ is correlated to $\mathbf{V}_3$ and $\mathbf{V}_1$ only through U_r and U_{r+kL}.

Let $S_\ell = U_{r+\ell L}$ for $\ell = 0, ..., k$. Then clearly $S_0, S_1, \ldots, S_k$ is an aperiodic irreducible Markov chain. Let $\mathbb{P}_S$ denote its unique invariant distribution.

Similarly, let $\tilde{\mathbf{U}}_\ell = \sum_{p=0}^{\ell-1} \mathbf{U}_{r+pL}^{(L)}$ with the convention that $\tilde{\mathbf{U}}_0 = 0$. Then we also have the Markov relation

$$\mathbf{V}_3 \to (S_0, \tilde{\mathbf{U}}_0) \to (S_1, \tilde{\mathbf{U}}_1) \to \cdots \to (S_k, \tilde{\mathbf{U}}_k) \to \mathbf{V}_1. \tag{8}$$

Now, the time homogeneous Markov chain $(S_\ell, \tilde{\mathbf{U}}_\ell) \to (S_{\ell+1}, \tilde{\mathbf{U}}_{\ell+1})$ is aperiodic and irreducible. Thus, the joint distribution at time k, $p^k_{S,\tilde{\mathbf{U}}}$, converges to the unique invariant distribution $\mathbb{P}_{S,\tilde{\mathbf{U}}}$ of the Markov chain. We claim that this is in fact the product $\mathbb{P}_{S,\tilde{\mathbf{U}}} = \mathbb{P}_S \times \mathbb{P}_{\tilde{\mathbf{U}}}$ where $\mathbb{P}_{\tilde{\mathbf{U}}}$ is uniform over the set of 2^L binary vectors of length L.

If this is the case, since $\tilde{\mathbf{U}}_k = \mathbf{V}_2$, then $(S_k, \mathbf{V}_2)$ is asymptotically uniform and independent of $(S_0 = U_r, \tilde{\mathbf{U}}_0 = 0)$ and by the Markov property, independent of $\mathbf{V}_3$. Furthermore,

$\tilde{\mathbf{U}}_k = \mathbf{V}_2$ is asymptotically uniform and independent of $S_k = U_{r+kL}$ and thus, independent of $\mathbf{V}_1$.

Thus, it remains to show that $\mathbb{P}_S \times \mathbb{P}_{\tilde{\mathbf{U}}}$ is an invariant distribution of the Markov chain. First, note that

$$\mathbb{P}_{S_{\ell+1},\tilde{\mathbf{U}}_{\ell+1}|S_\ell,\tilde{\mathbf{U}}_\ell} = \mathbb{P}_{S_{\ell+1}|S_\ell}\mathbb{P}_{\tilde{\mathbf{U}}_{\ell+1}|S_{\ell+1},S_\ell,\tilde{\mathbf{U}}_\ell} \tag{9}$$

Therefore,

$$\begin{aligned}
&\sum_{s_\ell}\sum_{\mathbf{u}_\ell} \mathbb{P}_{S_{\ell+1},\tilde{\mathbf{U}}_{\ell+1}|S_\ell,\tilde{\mathbf{U}}_\ell}[s_{\ell+1},\tilde{\mathbf{u}}_{\ell+1}|s_\ell,\tilde{\mathbf{u}}_\ell]\mathbb{P}_S[s_\ell]\mathbb{P}_{\tilde{\mathbf{U}}}[\tilde{\mathbf{u}}_\ell] \\
&\quad= \sum_{s_\ell}\Big[\mathbb{P}_{S_{\ell+1}|S_\ell}[s_{\ell+1}|s_\ell]\mathbb{P}_S[s_\ell]\times \\
&\qquad \sum_{\mathbf{u}_\ell}\mathbb{P}_{\tilde{\mathbf{U}}_{\ell+1}|S_{\ell+1},S_\ell,\tilde{\mathbf{U}}_\ell}[\tilde{\mathbf{u}}_{\ell+1}|s_{\ell+1},s_\ell,\tilde{\mathbf{u}}_\ell]\mathbb{P}_{\tilde{\mathbf{U}}}[\tilde{\mathbf{u}}_\ell]\Big] \quad (10)\\
&\quad= \sum_{s_\ell}\mathbb{P}_{S_{\ell+1}|S_\ell}[s_{\ell+1}|s_\ell]\mathbb{P}_S[s_\ell]\mathbb{P}_{\tilde{\mathbf{U}}}[\tilde{\mathbf{u}}_{\ell+1}] \quad (11)\\
&\quad= \mathbb{P}_S[s_{\ell+1}]\mathbb{P}_{\tilde{\mathbf{U}}}[\tilde{\mathbf{u}}_{\ell+1}], \quad (12)
\end{aligned}$$

where (11) follows because given a uniform random binary vector $\mathbf{Y}$ independent of a (not necessarily uniform) binary random vector $\mathbf{Z}$, $\mathbf{Y}+\mathbf{Z}$ is uniform. ■

Hence, we have derived a equivalent condition for asymptotic uniformity in terms of the $\mathbf{w}_i^{(m)}$. We now relate this condition to the feedforward and feedback polynomials.

Lemma 4: Given a recursive convolutional encoder with feedback polynomial $F(D)$ and feedforward polynomial $G(D)$, if $G(D)/F(D)$ is in its minimal form where M is the degree of $F(D)$ then the vectors $\mathbf{w}_L^{(M)} = \mathbf{w}_0^{(M)}, \mathbf{w}_1^{(M)}, \ldots, \mathbf{w}_{L-1}^{(M)}$ span a linear space of dimension M. Furthermore, the vectors $\mathbf{w}_L^{(M+1)} = \mathbf{w}_0^{(M+1)}, \mathbf{w}_1^{(M+1)}, \ldots, \mathbf{w}_{L-1}^{(M+1)}$ do not span a space of dimension $M+1$. □

Proof: Consider the matrix A_k formed by the first k cyclic shifts of the fundamental sequence $\mathbf{c}$,

$$A_k^T = \begin{bmatrix} \mathbf{c}^T & \mathsf{S}\mathbf{c}^T & \ldots & \mathsf{S}^{k-1}\mathbf{c}^T \end{bmatrix}^T \tag{13}$$

We first claim that A_M and A_{M+1} both have rank M. To see this, note that given any row vector $\mathbf{z} \in \{0,1\}^M$ with corresponding $Z(D)$ such that $\mathbf{z}A_M = 0$, we must have that $G(D)/F(D) \times Z(D)$ has an expansion as a finite polynomial since the former implies that an appropriate linear combination of at most M shifts of the infinite sequence $G(D)/F(D)$ zeros out all but a finite number of terms. However, since $G(D)$ and $F(D)$ are relatively prime, we must therefore have that $F(D)$ divides $Z(D)$. Since $M-1 = \deg Z(D) < \deg F(D) = M$, this implies that $Z(D) = 0$. Hence, $\mathbf{z} = 0$ which implies that A_M has rank M.

Now, consider the matrix A_{M+1}. The first M rows taken as a submatrix is A_M which has rank M. To show that A_{M+1} does not have rank $M+1$, consider the non-zero row vector $\mathbf{z} \in \{0,1\}^{M+1}$ corresponding to $Z(D) = F(D)$. Then clearly $\mathbf{z}A_{M+1} = 0$ since $G(D)/F(D) \times Z(D)$ has a finite expansion.

Finally, observe that we also have that

$$A_k^T = \begin{bmatrix} \mathbf{w}_0^{(k),T} & \mathbf{w}_1^{(k),T} & \ldots & \mathbf{w}_{L-1}^{(k),T} \end{bmatrix} \tag{14}$$

Since the rank of a matrix is also the dimension of its column space, it follows that the vectors $\mathbf{w}_L^{(M)} = \mathbf{w}_0^{(M)}, \mathbf{w}_1^{(M)}, \ldots, \mathbf{w}_{L-1}^{(M)}$ have dimension M while $\mathbf{w}_L^{(M+1)} = \mathbf{w}_0^{(M+1)}, \mathbf{w}_1^{(M+1)}, \ldots, \mathbf{w}_{L-1}^{(M+1)}$ also have dimension M. ■

The theorem then follows by applying Lemmas 2 and 4 in the i.i.d. case and Lemmas 3 and 4 in the more general Markov case.

III. Conclusion

We have derived the exact order of asymptotic uniformity on the outputs of a convolutional code under both i.i.d. and Markov binary input processes. The methods employed here also generalize to shift register structures over arbitrary finite fields. These results are of significance to the choice of constituent convolutional encoders when employed in joint source-channel coding as well as source coding applications.

Appendix

Extended justifications for (3)-(7):

$$\begin{aligned}
\mathbf{X}_n^{(m)} &= \sum_{p=0}^{n+m-1} U_p \mathbf{h}_{n-p}^{(m)} \\
&= \sum_{p=0}^{r-1} U_p \mathbf{h}_{n-p}^{(m)} + \sum_{p=r}^{r+kL-1} U_p \mathbf{h}_{n-p}^{(m)} + \sum_{p=r+kL}^{n+m-1} U_p \mathbf{h}_{n-p}^{(m)} \\
&= \sum_{p=0}^{r-1} U_p \mathbf{w}_{r-p}^{(m)} + \sum_{p=0}^{k-1}\left[\sum_{q=0}^{L-1} u_{r+pL+q}\mathbf{w}_{L-q}^{(m)}\right] \\
&\qquad + \sum_{p=-m+1}^{T} U_{n-p}\mathbf{h}_p^{(m)} \\
&= \sum_{p=0}^{r-1} U_p \mathbf{w}_{r-p}^{(m)} + \sum_{p=0}^{k-1}\left[\mathbf{U}_{r+pL}^{(L)}\right] W^{(m)} \\
&\qquad + \sum_{p=-m+1}^{T} U_{n-p}\mathbf{h}_p^{(m)}
\end{aligned}$$

References

[1] A. Abrardo, "Performance bounds and codes design criteria for channel decoding with a-priori information," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 608–612, June 2009.

[2] F. F. Daneshgaran, M. Laddomada, and M. Mondin, "Iterative joint channel decoding of correlated sources employing serially concatenated convolutional codes," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2721–2731, July 2005.

[3] I. J. Fair, V. K. Bhargava, and Q. Wang, "On the power spectral density of self-synchronizing scrambled sequences," *IEEE Trans. Inform. Theory*, vol. 44, no. 4, pp. 1687–11 693, July 1998.

[4] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, no. 10, pp. 417–419, Oct. 2001.

[5] J. Garcia-Frias and Y. Zhao, "Compression of binary memoryless sources using punctured turbo codes," *IEEE Commun. Lett.*, vol. 6, no. 9, pp. 394–396, Oct. 2002.

[6] D. G. Leeper, "A universal digital data scrambler," *Bell Syst. Tech. J.*, vol. 52, no. 10, pp. 1851–1865, Dec. 1973.

[7] S. Shamai and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. Inform. Theory*, vol. 43, no. 3, pp. 836–846, May 1997.

[8] G. I. Shamir and J. J. Boutros, "Non-systematic low-density parity-check codes for nonuniform sources," in *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, WA, June 2005, pp. 1898 – 1902.

[9] G.-C. Zhu, F. Alajaji, J. Bajcsy, and P. Mitran, "Transmission of nonuniform memoryless sources via nonsystematic turbo codes," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1344–1353, Aug. 2004.