Power Spectra of Finite-Length Constrained Codes With Level-Based Signaling

Jessica Centers*, Xinyu Tan*, Ahmed Hareedy, and Robert Calderbank
Electrical and Computer Engineering Department, Duke University, Durham, NC 27708 USA
{jessica.bilskie, xinyu.tan, ahmed.hareedy, robert.calderbank}@duke.edu

***Abstract*—In various practical systems, certain data patterns are prone to errors if written or transmitted. Constrained codes are used to eliminate error-prone patterns, and they can also achieve other goals. Recently, we introduced efficient binary symmetric lexicographically-ordered constrained (LOCO) codes and asymmetric LOCO (A-LOCO) codes to increase density in magnetic recording systems and lifetime in Flash systems by eliminating the relevant detrimental patterns. Due to their applications, LOCO and A-LOCO codes are associated with level-based signaling. In this paper, we first modify a framework from the literature in order to introduce a method to derive the power spectrum of a sequence of constrained data associated with level-based signaling. We then provide a generalized method for developing the one-step state transition matrix (OSTM) for finite-length codes constrained by the separation of transitions. Via their OSTMs, we devise closed form solutions for the spectra of finite-length LOCO and A-LOCO codes.**

I. Introduction

When Shannon introduced the concept of coding for constrained channels in 1948, he represented a constrained sequence via a finite-state transition diagram (FSTD) and showed how the capacity can be derived [1]. About 20 years later, researchers began using constrained codes in magnetic recording (MR) devices to increase their density [2]. In data storage and transmission, deriving the power spectral density (PSD) of a random stream of data/codewords constrained by forbidding error-prone patterns after signaling is important in the infinite and more importantly the finite-length scenarios. The PSD reveals to the system designer how power is distributed across different frequencies [5], [6]. It identifies properties such as the average power of the stream at DC, stream bandwidth, and any discrete power component resulting from the periodicity of the auto-correlation function [5]. These properties are considered when choosing a particular coding scheme to use.

We develop a method for calculating the PSD of any finite-length binary code constrained by the separation of its bit transitions and associated with level-based (NRZ) signaling. This includes LOCO codes, where separations between transitions are restricted regardless of bit values, as well as A-LOCO codes, where only separations between particular transitions are restricted. LOCO codes find application primarily in MR and communication over transmission lines, where data patterns require sufficiently separated transitions to minimize interference. Whereas, A-LOCO codes find application in two-level (SLC) Flash memory, where data patterns causing high-low-high charge levels on adjacent cells are error-prone.

II. PSD Derivation for Codes With NRZ Signaling

Gallopoulos et al. presented a method to calculate the spectrum of any run-length-limited (RLL) code with non-return-to-zero inverted (NRZI) signaling based on its FSTD [6]. We modify their method to address binary constrained codes associated with non-return-to-zero (NRZ) signaling. A summary for deriving the OSTM is given in the following steps:

1) Derive the FSTD that generates a valid code sequence $\{X_n\}$ of binary data satisfying the system constraints while tracking the probabilities of each transition.
2) Derive the one-step state transition diagram (OSTD) by assigning states only to the vertices where the most recent generated bit is a 1. The edges of the OSTD define the number of transitions it takes to get from one state to the next according to the FSTD. Maintain record of the probabilities associated with these edges.
3) Generate the one-step state transition matrix (OSTM) $\mathbf{G}(D)$ from the known edge probabilities and run lengths defining the OSTD. The general entry $\mathbf{G}_{ij}(D)$ is given by $\mathbf{G}_{ij}(D) = \sum_{t=1}^{\infty} p_{ij}(t)D^t$, where $p_{ij}(t)$ is the probability of transiting from state i to state j in the OSTD and t is the run length of that transition. D is defined as the complex exponential $e^{i2\pi fT}$ with frequency f and bit interval T.

More details can be found in [7].

We adopt the following signal generation scheme:

$$\{X_n\} \xrightarrow[\text{signaling}]{\text{NRZ}} \{Y_n\} \xrightarrow[\text{shaping}]{\text{Pulse}} W(t). \tag{1}$$

Though the write signal $W(t)$ is often the signal of interest, our method also calculates the PSD of the intermediate signals $\{X_n\}$ and $\{Y_n\}$ in case the application considered utilizes the discrete signal prior to or after NRZ signaling. An example of this would be using the PSD of $\{X_n\}$ to evaluate the spectral properties of the written signal in solid-state media.

Theorem 1. *The power spectrum $S_X(D)$ of process $\{X_n\}$ is*

$$S_X(D) = p(1)\boldsymbol{\pi}\left[(\mathbf{I}-\mathbf{G}(D))^{-1} + (\mathbf{I}-\mathbf{G}(D^{-1}))^{-1} - \mathbf{I}\right]\mathbf{u}^{\mathsf{T}}, \tag{2}$$

where $p(1)$ is the equilibrium probability of a 1 in $\{X_n\}$, $\boldsymbol{\pi}$ is the stationary distribution of $\mathbf{G}(1)$, $\mathbf{I}$ is the identity matrix of the same size as $\mathbf{G}(D)$, and $\mathbf{u}$ is an all-one vector.

Note that finite-length constrained codes result in cyclostationary processes (aside from the cyclic properties arising from pulse shaping). This requires the PSD to be found by obtaining its continuous and discrete portions separately. In those cases, this theorem results in the continuous portion. Cyclostationarity discussion and an example are provided in [7].

After NRZ signaling, the modulation sequence is $\{Y_n\}$ with $Y_n = 2X_n - 1$. Therefore, the PSD $S_Y(e^{i2\pi fT})$ of $\{Y_n\}$ is

$$S_Y(e^{i2\pi fT}) = 4S_X(e^{i2\pi fT}) + [1-4p(1)]\delta(f). \tag{3}$$

*These two authors contributed equally to this work.

Because the modulation pulse function is chosen as a rectangular pulse, the PSD $S_W(f)$ of the write signal $W(t)$ is

$$S_W(f) = \mathrm{sinc}^2(\pi f T) T^2 S_Y(e^{i2\pi f T}). \quad (4)$$

For $f = 0$, (4) becomes

$$S_W(0) = T^2 S_Y(1) = T^2 \left(4 S_X(1) + [1 - 4p(1)]\delta(0)\right). \quad (5)$$

We consider f to be the normalized frequency, i.e., T is 1.

For all coding schemes we consider except LOCO codes, we obtain the PSD through (2), (3), (4), and (5). LOCO codes require a three-level waveform because of their bridging [3]. The modified procedure for LOCO codes is discussed in [7].

III. Solutions for the PSD of any $\mathcal{AC}_{m,x}$ or $\mathcal{C}_{m,x}$

To generate the FSTD/OSTM of a finite-length codebook constrained by forbidding the patterns in the set $\mathcal{S}_x = \{010, 101, 0110, 1001, \dots, 0\mathbf{1}^x 0, 1\mathbf{0}^x 1\}$ (LOCO) or the set $\mathcal{A}_x = \{101, 1001, \dots, 1\mathbf{0}^x 1\}$ (A-LOCO), we begin with a grid of states where each column represents a bit position within the codeword and its bridging pattern, and each row represents a possible bit sequence consisting of the previous x bits and the current bit. This results in a maximum of $(m + x)2^{x+1}$ states. The derived OSTM, $\mathbf{G}(D)$, associated with this FSTD is then used in Theorem 1 and its following discussion to get the final PSD of the code. For classes of codes, patterns in their OSTMs in terms of the code parameters can be used to provide a general solution.

The general OSTM solution for any A-LOCO code $\mathcal{AC}_{m,x}$ defined by the parameters m (length of codeword) and x (length of bridging pattern) is the $m + x$ square matrix

$$\mathbf{G}(D) = \begin{bmatrix} & & & 0 & \mathbf{0} \\ & \mathbf{A}_{m\times m} & & \vdots & \vdots \\ & & & 0 & \vdots \\ & & & \zeta D & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & D\mathbf{I}_{x-1} \\ D & 0 & \cdots & \cdots & \mathbf{0} \end{bmatrix}, \quad (6)$$

where $\zeta = \frac{N_2(m,x) + N_3(m,x)}{N(m,x)}$. Here, the notation $N(m,x)$, $N_2(m,x)$, and $N_3(m,x)$ refers to the cardinality of $\mathcal{AC}_{m,x}$, the cardinality of Group 2 in $\mathcal{AC}_{m,x}$, and the cardinality of Group 3 in $\mathcal{AC}_{m,x}$, adopted from [4].

The entries of the submatrix $\mathbf{A}_{m\times m}$, which are $\mathbf{A}_{i,j}$, $1 \leq i, j \leq m$, can then be found using a few rules. If $j = i$, then $\mathbf{A}_{i,j} = \beta_{1,m+x}$. If $j = i + 1$ or $j > i + 1 + x$, then $\mathbf{A}_{i,j} = \lambda_{m+1-i,j-i-1} D^{j-i} + \beta_{\lambda_{m+1-i,j-i-1},m-i+j+x}$. If $i + 2 \leq j \leq i + 1 + x$, then $\mathbf{A}_{i,j} = \beta_{\lambda_{m+1-i,j-i-1},m-i+j+x}$. If $j < i$, then $\mathbf{A}_{i,j} = \beta_{\frac{1}{\lambda_{m+1-j,i-j-1}},m-i+j+x}$. If $i = m$ and $j = 1$, then $\mathbf{A}_{i,j} = \beta_{\zeta,m+2x+1}$. Here, $\beta_{a,b} = \frac{aD^b}{N(m,x) - D^{m+x}}$ and $\lambda_{d,g} = \prod_{k=0}^{g}(1 - \frac{N_3(d-k,x)}{N_2(d-k,x) + N_3(d-k,x)})$.

Since A-LOCO codes are asymmetric, and therefore have cyclostationary processes with non-zero means, there is also a discrete component to their PSDs (not shown in Fig. 1) occurring at DC and integer multiples of the codeword normalized frequency, $\frac{1}{m+x}$. See also [7] for more results.

We also derive a generalized solution for the OSTM for LOCO codes. A few examples of LOCO PSDs are provided in Fig. 2. Because LOCO codes are symmetric, they have no discrete PSD component. More results are provided in [7].

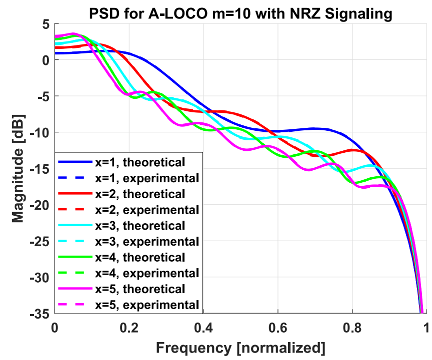

Fig. 1: The continuous component of various $\mathcal{AC}_{m,x}$ spectra.

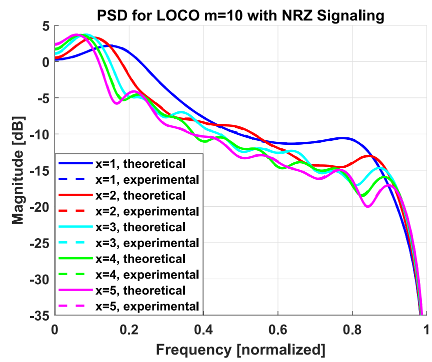

Fig. 2: The full spectra of various $\mathcal{C}_{m,x}$.

IV. Conclusion

We introduced a method for deriving the power spectra of constrained codes associated with level-based signaling. Though our method can be applied to infinite-length codes as shown in [7], we focused on generating the FTSD and OSTM for finite-length codes so that the PSD can be calculated. We applied our method to finite-length A-LOCO and LOCO codes. Our theoretical PSD results perfectly match the experimental PSD results for all codes. We suggest that our method and ideas can provide helpful insights to data storage and data transmission engineers regarding different classes of constrained codes they plan to employ.

References

[1] C. E. Shannon, "A mathematical theory of communication," Bell Sys. Tech. J., vol. 27, Oct. 1948.

[2] D. T. Tang and R. L. Bahl, "Block codes for a class of constrained noiseless channels," *Inf. and Control*, vol. 17, no. 5, pp. 436–461, 1970.

[3] A. Hareedy and R. Calderbank, "LOCO codes: Lexicographically-ordered constrained codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3572–3589, Jun. 2020.

[4] A. Hareedy and R. Calderbank, "Asymmetric LOCO codes: Constrained codes for Flash memories," in *Proc. 57th Annual Allerton Conf. Commun., Control, and Computing*, Monticello, IL, USA, Sep. 2019, pp. 124–131.

[5] K. A. S. Immink, *Codes for Mass Data Storage Systems.* Shannon Foundation Publishers, 2004.

[6] A. Gallopoulos, C. Heegard, and P. H. Siegel, "The power spectrum of run-length-limited codes," *IEEE Trans. Commun.*, vol. 37, no. 9, pp. 906–917, Sep. 1989.

[7] J. Centers, X. Tan, A. Hareedy, and R. Calderbank, "Power Spectra of Constrained Codes with Level-Based Signaling: Overcoming Finite-Length Challenges," Oct. 2020. [Online]. Available: https://arxiv.org/abs/2010.04878