

Power Spectra of Constrained Codes with Level-Based Signaling: Overcoming Finite-Length Challenges

Jessica Centers*, Xinyu Tan*, Ahmed Hareedy, *Member, IEEE*,
and Robert Calderbank, *Fellow, IEEE*

Abstract

In various practical systems, certain data patterns are prone to errors if written or transmitted. In magnetic recording and communication over transmission lines, data patterns causing consecutive transitions that are not sufficiently separated are prone to errors. In Flash memory with two levels per cell, data patterns causing high–low–high charge levels on adjacent cells are prone to errors. 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 application, LOCO and A-LOCO codes are associated with level-based signaling. Studying the power spectrum of a random signal with certain properties is principal for any storage or transmission system. It reveals important properties such as the average signal power at DC, the bandwidth of the signal, and whether there are discrete power components at certain frequencies. 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 apply our method to infinitely long sequences satisfying symmetric and asymmetric constraints. Next, we show how to generalize the method such that it works for a stream of finite-length codewords as well, thus demonstrating how to overcome the associated finite-length challenges. We use the generalized method to devise closed forms for the spectra of finite-length LOCO and A-LOCO codes from their transition diagrams. Our LOCO and A-LOCO spectral derivations can be performed for any code length and can be extended to other constrained codes. We

*These two authors contributed equally to this work.

J. Centers, X. Tan, A. Hareedy, and R. Calderbank are with Duke University, Durham, North Carolina, 27708, USA. J. Centers, A. Hareedy, and R. Calderbank are with the Electrical and Computer Engineering Department. X. Tan is with the Computer Science Department. (E-mail: jessica.bilskie@duke.edu, xinyu.tan@duke.edu, ahmed.hareedy@duke.edu, and robert.calderbank@duke.edu).

plot these power spectra, and discuss various important spectral properties for both LOCO and A-LOCO codes. We also briefly discuss an alternative method for deriving the power spectrum, and introduce an idea towards reaching the spectra of self-clocked codes.

Index Terms

Power spectra, constrained codes, finite-length, LOCO codes, data storage, data transmission.

I. INTRODUCTION

Shannon introduced the concept of coding for constrained channels in 1948 [1]. He represented a constrained sequence via a finite-state transition diagram (FSTD) and showed how the capacity can be derived. About 20 years later, researchers started to use constrained codes in magnetic recording (MR) devices adopting peak detection to increase their density [2], [3]. These codes were (d, k) binary run-length-limited (RLL) codes, which control the minimum (resp., maximum) separation between consecutive transitions via the parameter d (resp., k) to mitigate inter-symbol interference (ISI) (resp., enhance self-clocking) [4], [5]. RLL codes are associated with transition-based signaling, which is bipolar *non-return-to-zero inverted (NRZI) signaling* in the binary case. In NRZI signaling, a 0 is represented by no transition, while a 1 is represented by a transition. Constrained codes are still being used in modern MR systems adopting sequence detection [6], [7], which is demonstrated in [8] and [9].

Constrained codes find application in Flash memory devices to help increase their lifetime. A Flash cell is a MOSFET transistor with a floating gate, the amount of charge in which controls its operation. Consider three adjacent Flash cells, where the outer two are to be programmed to the highest charge level, while the inner one is to be programmed to a lower level. Charges propagate from the outer cells to the inner one, resulting in inter-cell interference (ICI) that can corrupt the stored data [10]. For Flash devices with 2 levels per cell, codes preventing the pattern 101 were presented in [11] and [12]. For Flash devices with more than 2 levels per cell, efficient constrained codes were introduced in recent literature [13], [14]. These codes are associated with level-based signaling, which is bipolar *non-return-to-zero (NRZ) signaling* in the binary case. In NRZ signaling, a 0 is represented by a level, while a 1 is represented by another level. In this paper, the two levels are always -1 for 0 and $+1$ for 1, respectively.

There are various other applications of constrained codes. Two-dimensional magnetic recording (TDMR) is a new technology that promises a boost in the density of MR devices [15], [16].

Constrained codes preventing certain, error-prone TD patterns increase the reliability of TDMR devices [17]. Constrained codes find application in optical recording devices [18]. Constrained codes are also used in various data transmission systems, as they can mitigate cross-talk between wires over which the data is transmitted [19], and they can achieve DC-balance, i.e., zero average power at frequency zero [20].

Even though RLL codes presented in [2] were based on lexicographic indexing, the coding community diverted to codes based on finite-state machines (FSMs) shortly after that [4], [21]. However, it is a complicated task to design FSM-based constrained codes that are capacity-approaching, and the task becomes even more complicated for high rate codes [14]. Constrained codes based on lexicographic indexing were presented in [22], which is guided by [23]. Recently, we returned to the fundamentals of constrained coding [1], [2], and designed binary symmetric lexicographically-ordered constrained (LOCO) codes primarily for magnetic recording applications to help increase density [9]. We then designed binary asymmetric LOCO (A-LOCO) codes for single-level cell (SLC) Flash devices, i.e., with 2 levels per cell, to help increase lifetime [12]. We also developed constrained codes for Flash devices with more than 2 levels per cell [14] and for TDMR devices [17]. All our codes are capacity-achieving, offer simple encoding-decoding, and reconfiguring them is as simple as reprogramming an adder.

Given the set of patterns to forbid, one can design an infinite sequence as well as a finite-length block code satisfying the constraint. In data storage and data transmission, deriving the power spectral density (PSD) of a random stream of data/codewords constrained by forbidding error-prone patterns after signaling is quite 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 [24], [25]. Consequently, the system designer identifies pivotal properties such as the average power of the stream at DC, the stream bandwidth, i.e., the frequency range in which most of the power is allocated, in addition to the discrete power components, if any, which result from the periodicity of the auto-correlation function [24].

There is a rich literature on studying power spectra of signals in various systems. In the context of data transmission, power spectra of coded streams generated by an FSM and transmitted over digital repeatered lines were discussed in [26] and [27]. In the context of wireless communications, numerical [28] and analytical [29] methods for obtaining the spectra of coded streams generated by an FSM were presented. The power spectrum of a block-coded modulated signal having a multi-dimensional constellation was discussed in [30]. In the context of data storage,

power spectra of certain FSM-based codes in MR systems were derived in [31], and spectral null analysis of (d, k) RLL codes in MR and magneto-optic systems was presented in [32]. The effect of asymmetry in coded data associated with NRZ signaling on the power spectrum was detailed in [33]. Finally, [25], which we shortly refer to again in this paper, introduced a method to find power spectra of RLL codes associated with NRZI signaling for storage and transmission, emphasizing more on infinite sequences.

In this paper, we start from the method in [25] and devise a method for obtaining the power spectrum of a binary constrained stream that is associated with NRZ signaling. Let y^r be a run of r consecutive y 's. Define $\mathcal{S}_x = \{010, 101, 0110, 1001, \dots, 01^x0, 10^x1\}$, which is symmetric, i.e., closed under taking pattern complements, and define $\mathcal{A}_x = \{101, 1001, \dots, 10^x1\}$, which is asymmetric. We use our method to obtain the power spectra of infinitely long \mathcal{S}_x -constrained and \mathcal{A}_x -constrained sequences. We then generalize this method for streams of finite-length constrained codewords, illustrating how to work with transition diagrams that depend on the code length and how to move from wide-sense-stationary into cyclostationary random processes. We apply the new generalized model to obtain the spectra of LOCO and A-LOCO codes, which are binary finite-length constrained codes. Our new generalized method can also be used for other constrained codes, and its complexity growth with the code length is limited. We use the plots of power spectra for various code parameters to understand pivotal properties about LOCO and A-LOCO coded streams such as the power at DC and the 3dB bandwidth. Finally, we verify our spectral derivations for \mathcal{S}_x -constrained and \mathcal{A}_x -constrained sequences via a different method, and take a step towards obtaining the spectra of self-clocked LOCO and A-LOCO codes.

The rest of the paper is organized as follows. In Section II, we introduce our method for deriving the spectrum of a constrained code with NRZ signaling. In Section III, we find closed-form expressions for the PSDs of infinitely long constrained codes. In Section IV, we show how to generate transition diagrams for finite-length codes. In Section V, we demonstrate how to deal with cyclostationary processes, and we find closed-form expressions for the PSDs of A-LOCO and LOCO codes. We discuss important properties of these power spectra. In Section VI, we conclude the paper. The appendices discuss an alternative method and self-clocked codes.

II. DERIVING THE PSD FOR A CONSTRAINED CODE WITH NRZ SIGNALING

Gallopoulos et al. presented a method to calculate the spectrum of an RLL code with NRZI signaling based on its finite-state transition diagram (FSTD) [25]. We modify their method

to address lexicographically-ordered (see Section IV) as well as FSM-based constrained codes associated with NRZ signaling. We note that the FSTDs we discuss in this paper represent each state as a fixed number of the most recently generated bits.

We first derive the one-step state transition matrix (OSTM) via the following steps:

- 1) Derive the FSTD that generates a valid code sequence $\{X_n\}$ of binary data satisfying the system constraints and keep track of the probabilities of each transition. Make sure that each edge generates a single symbol (either a 1 or 0) and that all the incoming edges for a given vertex are generated by the same symbol.
- 2) Derive the one-step state transition diagram (OSTD) of the same system 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 (or symbols) it takes to get from one state to the next according to the FSTD. It is necessary to maintain record of the probabilities associated with these edges. We define the sequence produced by the OSTD as $\{T_i\}$ (iid). In particular, T_i represents the number of consecutive 0's along with the ending 1 of the i th run in $\{X_n\}$.
- 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

$$\mathbf{G}_{ij}(D) = \sum_{t=1}^{\infty} p_{ij}(t) D^t, \quad (1)$$

where $p_{ij}(t)$ is the probability of transiting from state i to state j in the OSTD and t is the run length characterizing that transition.

We then derive the closed-form expression for the power spectral density (PSD) of a written or transmitted sequence using the OSTM. We adopt the following signal generation scheme:

$$\text{Code, } \{X_n\} \xrightarrow[\text{signaling}]{\text{NRZ}} \text{Modulation sequence, } \{Y_n\} \xrightarrow[\text{shaping}]{\text{Pulse}} \text{Write signal, } W(t). \quad (2)$$

The PSD function of a wide-sense-stationary discrete-time process $\{X_n\}$ is given by

$$S_X(D) = \sum_{k=-\infty}^{\infty} \mathbb{R}_X(k) D^k = \sum_{k=-\infty}^{\infty} \mathbb{E}[X_0 X_k] D^k, \quad (3)$$

where $\mathbb{R}_X(k)$ is the auto-correlation function with time lag k , and it is an even function. D is defined as the complex exponential $e^{i2\pi fT}$ for some frequency f and bit interval T . In the case that the process $\{X_n\}$ is cyclostationary (aside from the cyclic properties in $W(t)$ that arise

from the pulse shaping), the PSD is found by obtaining its continuous and discrete portions separately. We will provide discussion and details on this topic in Section V.

We first present a fundamental result that is analogous to [25, Theorem 1]. Vectors are by default row vectors unless otherwise stated.

Theorem 1. *Assume that the Markov chain describing the code sequence generation is at equilibrium. The power spectrum $S_X(D)$ of the process $\{X_n\}$ is given by*

$$S_X(D) = p(1)\boldsymbol{\pi} [(\mathbf{I} - \mathbf{G}(D))^{-1} + (\mathbf{I} - \mathbf{G}(D^{-1}))^{-1} - \mathbf{I}] \mathbf{u}^\top, \quad (4)$$

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

Proof: The stationary distribution $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi}\mathbf{u}^\top = 1$ and $\boldsymbol{\pi}\mathbf{G}(1) = \boldsymbol{\pi}$, where $\mathbf{G}(1)$ is the ordinary OSTM of the Markov chain. We connect $\mathbf{G}(D)$ with $S_X(D)$. Define

$$\Phi_l(D) = \sum_{k=0}^{\infty} D^k \cdot \mathbb{P}[\text{length of } l \text{ consecutive } T_i \text{ is } k | \text{run started}]. \quad (5)$$

It follows from (1) and the definition of $\{T_i\}$ that (see also [25])

$$\Phi_l(D) = \boldsymbol{\pi}[\mathbf{G}(D)]^l \mathbf{u}^\top \text{ and } \mathbb{E}[T_i] = \frac{1}{p(1)} = \boldsymbol{\pi}\mathbf{G}'(1)\mathbf{u}^\top, \quad (6)$$

where $\mathbf{G}'(1)$ is the derivative of $\mathbf{G}(D)$ with respect to D when $D = 1$. Then,

$$\begin{aligned} \sum_{l=0}^{\infty} \Phi_l(D) &= \sum_{k=0}^{\infty} D^k \left(\sum_{l=0}^{\infty} \mathbb{P}[\text{length of } l \text{ consecutive } T_i \text{ is } k | \text{run started}] \right) \\ &= 1 + \sum_{k=1}^{\infty} D^k \cdot \mathbb{P}[X_k = X_0 = 1 | X_0 = 1] \\ &= 1 + \sum_{k=1}^{\infty} D^k \cdot \mathbb{E}[X_0 X_k | X_0 = 1]. \end{aligned} \quad (7)$$

Therefore, the power spectral density $S_X(D)$ of $\{X_n\}$ is

$$S_X(D) = \sum_{k=-\infty}^{\infty} D^k \cdot \mathbb{E}[X_0 X_k] = p(1) \left(\sum_{l=0}^{\infty} \Phi_l(D) + \sum_{l=0}^{\infty} \Phi_l(D^{-1}) - 1 \right). \quad (8)$$

It follows from (5) that $\Phi_l(D) = [\Phi_1(D)]^l = [\Phi(D)]^l$. Thus, using the power series

$$\sum_{l=0}^{\infty} x^l = \frac{1}{1-x}, \text{ for } |x| < 1,$$

along with (6), we get

$$\begin{aligned} S_X(D) &= p(1) \left[\frac{1}{1 - \Phi(D)} + \frac{1}{1 - \Phi(D^{-1})} - 1 \right] \\ &= p(1) \boldsymbol{\pi} [(\mathbf{I} - \mathbf{G}(D))^{-1} + (\mathbf{I} - \mathbf{G}(D^{-1}))^{-1} - \mathbf{I}] \mathbf{u}^T, \end{aligned} \quad (9)$$

which completes the proof. \blacksquare

After NRZ signaling, we have the modulation sequence $\{Y_n\}$ with $Y_n = 2X_n - 1$. Thus, for any $k \in \mathbb{Z}$,

$$\mathbb{R}_Y(k) = \mathbb{E}[Y_0 Y_k] = \mathbb{E}[(2X_0 - 1)(2X_k - 1)] = 4\mathbb{E}[X_0 X_k] - 4p(1) + 1. \quad (10)$$

Therefore, the power spectral density $S_Y(D)$ of $\{Y_n\}$ is

$$\begin{aligned} S_Y(D) &= \sum_{k=-\infty}^{\infty} D^k \cdot \mathbb{E}[Y_0 Y_k] \\ &= 4S_X(D) + [1 - 4p(1)] \sum_{k=-\infty}^{\infty} D^k. \end{aligned}$$

As a result, we get

$$S_Y(D) = \begin{cases} 4S_X(D), & D \neq 1, \\ 4S_X(D) + 1 - 4p(1), & D = 1. \end{cases} \quad (11)$$

The write signal $W(t)$ is given by

$$W(t) = \sum_{n=-\infty}^{\infty} Y_n P_T(t - nT), \quad (12)$$

where $P_T(t)$ is the modulation pulse function (chosen as a rectangular pulse in the context of this work) with bit interval T , i.e.,

$$P_T(t) = \begin{cases} 1, & 0 \leq t < T, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Therefore, the power spectral density $S_W(f)$ of the write signal $W(t)$ is given by

$$S_W(f) = \text{sinc}^2(\pi f T) T^2 S_Y(e^{i2\pi f T}). \quad (14)$$

For $f = 0$, which is equivalent to $D = 1$, (14) becomes

$$S_W(0) = T^2 S_Y(1) = T^2 (4S_X(1) + 1 - 4p(1)). \quad (15)$$

In the following, we consider f to be the normalized frequency, i.e., T is set to 1.

In this paper, for any coding schemes except LOCO codes, we obtain the PSD through (4), (11), (14), and (15). Given that LOCO codes require a three-level waveform because of their bridging [9], we modify the general procedure in Subsection V-C.

III. CLOSED-FORM PSD OF INFINITE-LENGTH CONSTRAINED CODES

Consider the following two infinite-length codes constrained by the separation of bit transitions: \mathcal{A}_x -constrained codes and \mathcal{S}_x -constrained codes. Though technically the transition separation is constrained only when the first transition occurs from a 1 to a 0 for \mathcal{A}_x -constrained codes, we still categorize them into the family of codes constrained by transition separation. Code sequences here represent wide-sense stationary processes.

A. PSD of Infinite-Length \mathcal{A}_x -Constrained Codes

Definition 1. Let $\mathcal{A}_x = \{101, 1001, \dots, 10^x 1\}$ with $|\mathcal{A}_x| = x$. Define a binary \mathcal{A}_x -constrained code which forbids any pattern in \mathcal{A}_x from appearing in the coded sequences.

\mathcal{A}_x -constrained codes are used in Flash memory systems to eradicate data patterns, such as 101, that cause parasitic capacitances to result in maximum charge propagation across cells during the programming phase (ICI). These codes are typically associated with NRZ signaling. Thus, Theorem 1 is applicable when calculating the PSD for this class of codes.

Example 1. We first provide an example of deriving the OSTM for infinite-length \mathcal{A}_1 -constrained code, which forbids 101 in the code sequence. Using mathematical induction, the interested reader can finish the derivation of the OSTM for any \mathcal{A}_x -constrained code (see Result 1).

Fig. 1a shows the FSTD of \mathcal{A}_1 -constrained code, where each edge is labeled by its single binary symbol transition. Fig. 1b records the probability for each corresponding transition, and it also labels in gray only the states resulting from a 1 transition.

We obtain Fig. 1c by using only the labeled states and drawing all the possible transitions from one state to another. The edges are labeled by run-length(s) t_i , indicating a transition corresponding to a sequence $0^{t_i-1}1$. We also keep track of the probabilities associated with the edge labels in the OSTD in Fig. 1d. It then follows from (1) that the OSTM of infinite-length \mathcal{A}_1 -constrained code is given by

$$\mathbf{G}(D) = \begin{bmatrix} \alpha & \frac{1}{2}D \\ \alpha & \frac{1}{2}D \end{bmatrix}, \quad (16)$$

where

$$\alpha = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{2+k} D^{3+k} = \frac{D^3}{2(2-D)}. \quad (17)$$

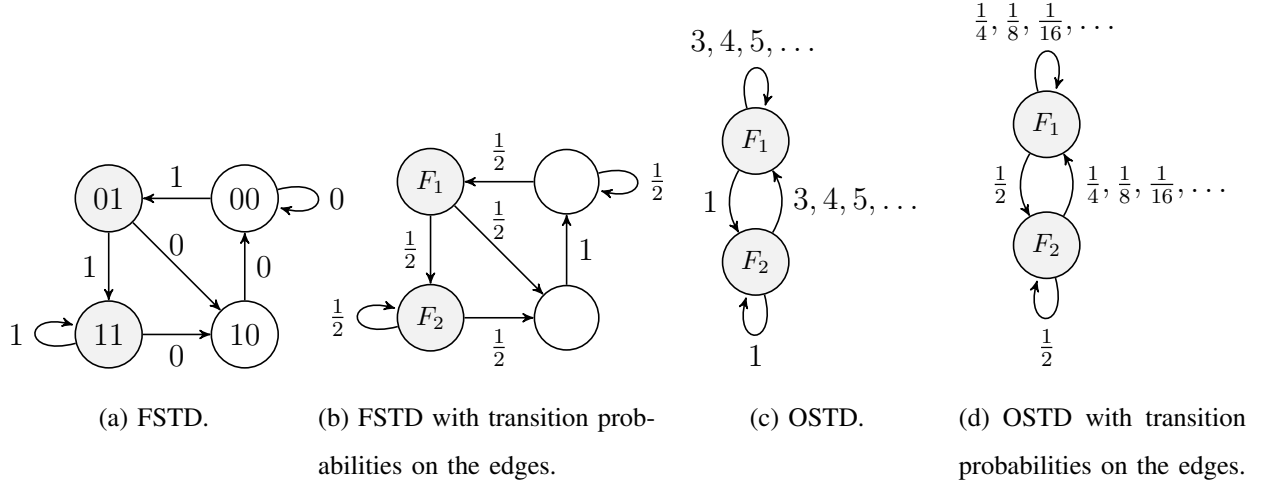


Fig. 1: FSTD and OSTD for infinite-length \mathcal{A}_1 -constrained sequence.

Result 1. According to our derivation by induction, the OSTM for any \mathcal{A}_x -constrained code is then

$$\mathbf{G}(D) = \begin{bmatrix} \alpha & & & & \\ \vdots & \frac{1}{2}D\mathbf{I}_x & & & \\ \alpha & 0 & \cdots & 0 & \frac{1}{2}D \end{bmatrix}, \quad (18)$$

where

$$\alpha = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{2+k} D^{2+x+k} = \frac{D^{2+x}}{2(2-D)} \quad (19)$$

and \mathbf{I}_x is the identity matrix of size $x \times x$.

We then generate the PSD for an infinite-length \mathcal{A}_x -constrained code following (4), (11), (14), and (15). Fig. 2 shows the PSDs for these codes with $x \in \{1, 2, \dots, 5\}$.

B. PSD of Infinite-Length \mathcal{S}_x -Constrained Codes

Definition 2. Let $\mathcal{S}_x = \{010, 101, 0110, 1001, \dots, 01^x0, 10^x1\}$ with $|\mathcal{S}_x| = 2x$. Define a binary \mathcal{S}_x -constrained code which forbids any pattern in \mathcal{S}_x from appearing in the coded sequences.

\mathcal{S}_x -constrained codes forbid data patterns that contribute the most to ISI in magnetic recording systems, i.e., the most error-prone data patterns in these systems. They find another application in some standards for data transmission. These codes are also typically associated with NRZ signaling, thus justifying the use of Theorem 1 to calculate their PSDs.

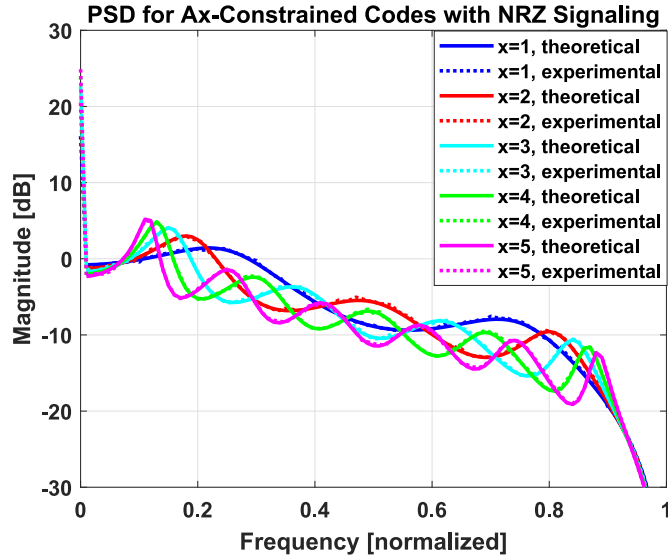


Fig. 2: PSDs for infinite-length \mathcal{A}_x -constrained codes with NRZ signaling. The spike at DC indicates the expected discrete DC power component due to asymmetry.

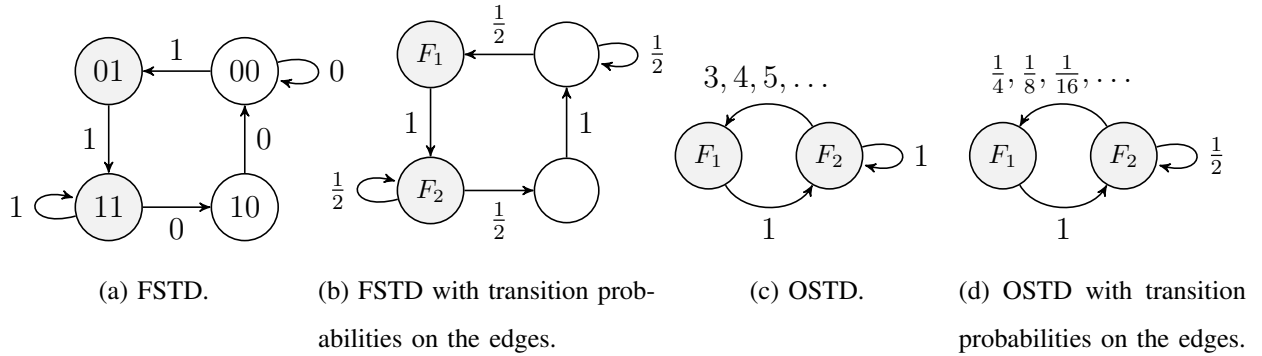


Fig. 3: FSTD and OSTD for infinite-length \mathcal{S}_1 -constrained sequence.

Example 2. In a way similar to Example 1, we will derive the OSTM for infinite-length \mathcal{S}_1 -constrained code, which forbids 010 and 101 in the code sequence. The interested reader can again finish the derivation of the general OSTM using mathematical induction (see Result 2).

Fig. 3a shows the FSTD of \mathcal{S}_1 -constrained code where each edge is labeled by its single binary symbol transition. Fig. 3b records the probability for each corresponding transition, and it also labels in gray only the states resulting from a 1 transition.

We again obtain Fig. 3c by using only the labeled states and drawing all the possible transitions from one state to another. The edges are labeled by run-length(s) t_i , indicating a transition corresponding to a sequence $0^{t_i-1}1$. We continue to keep track of the probabilities associated with the edge labels in the OSTD in Fig. 3d. It then follows from (1) that the OSTM of infinite-

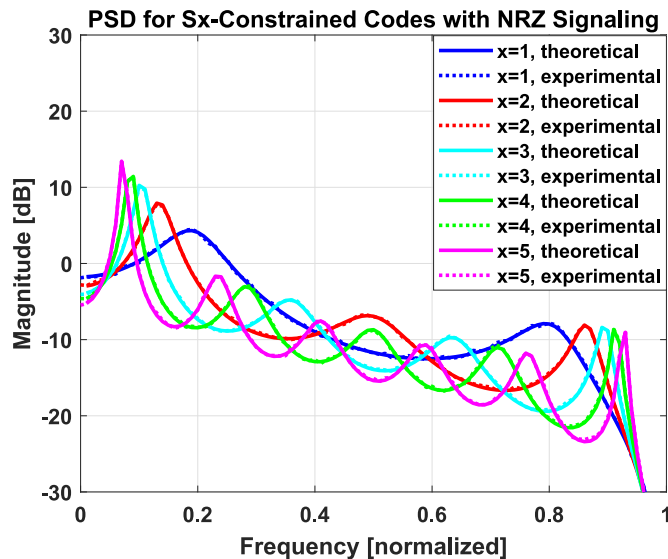


Fig. 4: PSDs for infinite-length \mathcal{S}_x -constrained codes with NRZ signaling.

length \mathcal{S}_1 -constrained code is given by

$$\mathbf{G}(D) = \begin{bmatrix} 0 & D \\ \alpha & \frac{1}{2}D \end{bmatrix}, \quad (20)$$

where

$$\alpha = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{2+k} D^{3+k} = \frac{D^3}{2(2-D)}. \quad (21)$$

Result 2. According to our derivation by induction, the OSTM for any \mathcal{S}_x -constrained code is then

$$\mathbf{G}(D) = \begin{bmatrix} 0 & & & & \\ \vdots & & D\mathbf{I}_x & & \\ 0 & & & & \\ \alpha & 0 & \dots & 0 & \frac{1}{2}D \end{bmatrix}, \quad (22)$$

where

$$\alpha = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{2+k} D^{2+x+k} = \frac{D^{2+x}}{2(2-D)}. \quad (23)$$

We then generate the PSD for an infinite-length \mathcal{S}_x -constrained code following (4), (11), (14), and (15). Fig. 4 shows the PSDs for these codes with $x \in \{1, 2, \dots, 5\}$.

It is clear from both Fig. 2 and Fig. 4 that the theoretical PSD plots match perfectly the experimental (Monte-Carlo) ones.

It should be noted that it is possible to use the exact method described in [25] without the modifications described in this paper to find the PSD of these infinite-length constrained codes. For further details, see Appendix A.

IV. OSTM GENERATION METHOD FOR FINITE-LENGTH CODES

When the constraints adopted in \mathcal{A}_x -constrained and \mathcal{S}_x -constrained codes are imposed on a finite-length codebook, the FSTD and thus the OSTM become more complicated due to the possible transitions being dependent on the bit location within the codeword and its bridging pattern. In these finite-length constrained codes, preventing patterns constrains the beginning bits and the ending bits of a codeword in a way that differs from how it constrains the middle bits. Observe that the length of the shortest forbidden pattern is 3. Another characteristic of finite-length codes is that they require bridging patterns. Given any two consecutive codewords, bridging patterns are bits/symbols appearing between the codewords to prevent forbidden patterns from occurring on the transition from any codeword to the following codeword.

The method for generating the FSTD/OSTM that is based on [25] and described in Section II provides the backbone to a generalized method for finding the FSTD/OSTM for finite-length codes constrained by the separation of transitions such as LOCO codes and A-LOCO codes described in detail later. The steps of our generalized method are:

- 1) Derive FSTD states by considering 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, since there are $m + x$ bit positions and 2^{x+1} possible $x + 1$ binary sequences. However, we will see that forbidden patterns result in some states being unneeded. Note that x and m are parameters of LOCO and A-LOCO codes. Here, x defines the set of forbidden patterns as seen in \mathcal{S}_x -constrained and \mathcal{A}_x -constrained codes, while m is the length of each codeword without its bridging.
- 2) Connect states in each column to states in the next column with reference to allowed patterns according to the codebook constraint. These are always single bit transitions.
- 3) Note the probabilities of each transition in the FSTD using the codebook. For convenience, all codewords in the codebook are assumed to be available for usage.
- 4) Discard states that are never achieved, i.e., never connected to other states.
- 5) Label states that result from the most recent bit being a 1.

- 6) If desirable, combine states that have the same exiting paths and probabilities.
- 7) Develop the OSTM $G(D)$ associated with this final FSTD by using the edge probabilities and run lengths, as done in Section II.
- 8) If desirable for a class of codes, look for a pattern in the OSTMs of codes with specific parameters that leads to a general solution for the OSTM of that class of codes.
- 9) Follow Theorem 1 for obtaining the final PSD of the code.

If examined closely, this method provides key insights regarding how the OSTM of these codes grows with respect to x and m . The maximum number of states generated, assuming that they all can be achieved and therefore not discarded, is $(m+x)2^{x+1}$. The maximum number of labeled states is $\frac{(m+x)2^{x+1}}{2} = (m+x)2^x$, since only states resulting from a 1 are labeled. Because the OSTM is a square matrix whose size is defined by the maximum number of labeled states, the OSTM is at maximum size an $(m+x)2^x \times (m+x)2^x$ matrix. This means that as the length of the codewords m grows, the matrix dimension grows linearly. However, as the transition-separation constraint variable x grows, the matrix dimension grows exponentially. In practice, x tends to stay small, within $\{1, 2, 3\}$ typically, while m can be large depending on the specific rate requirement [9], [12], so these trends are favorable.

As discussed later in this paper, some codes do not result in FSTDs with this bound on the maximum number of labeled states, specifically clocked LOCO and clocked A-LOCO codes [9], [14]. Self-clocking is an important property for various line codes, as it enables the receiver to obtain the clock of the transmitter from the signal. Thus, in Appendix B, we give the definitions of self-clocked A-LOCO and LOCO codes and, as an alternative, describe a general but slower algorithmic approach to construct the OSTM using breadth first search (BFS). Additionally, it is often the case that as the OSTM size grows, the OSTM becomes more sparse. This is particularly true in the upcoming examples of LOCO and A-LOCO codes.

We apply this method to study the spectra of A-LOCO and LOCO codes in Section V. We provide the general OSTM solution for any A-LOCO and LOCO code with standard, optimal bridging, and we also show how to address other finite-length challenges.

V. CLOSED-FORM PSD OF FINITE-LENGTH CONSTRAINED CODES

Before calculating the PSD for a finite-length constrained code, we begin by referencing Fig. 2, the PSDs of infinite-length \mathcal{A}_x -constrained codes. A spike (delta function) at frequency 0 (DC) is seen. This implies that there is some discrete amount of power at DC, which intuitively makes

sense because $p(1)$ is not $\frac{1}{2}$ as in \mathcal{S}_x -constrained codes. Instead, $p(1) = \frac{2}{x+4}$ for \mathcal{A}_x -constrained codes. This implies that the level-based sequence $\{Y_n\}$ has some DC offset. Mathematically, this means that there is a periodic component in the auto-correlation function associated with $\{Y_n\}$. In this specific case of infinite-length \mathcal{A}_x -constrained codes, this component is a constant $= (\mathbb{E}[Y_n])^2 = \frac{x^2}{(x+4)^2}$, which results in a delta function at DC when calculating the spectrum.

Besides for the case of having discrete power at DC, it is possible for a code to have discrete power at other frequencies. This occurs when the code sequence process is cyclostationary, and the expected value at any stream position is not fixed at 0 after signaling. In this case, the auto-correlation function has a periodic component that is not a constant.

A. Discussion on the Cyclostationarity of Finite-Length Codes

Time-synchronous pulse amplitude modulated signals are known to be cyclostationary with a period equal to the channel bit interval. A stream of m -length codewords exhibits higher level cyclostationarity with a period equal to $m + x$ channel bit intervals when bridging patterns are designed to be x -bits long. To account for these cyclic statistics in the PSD, phase-averaging is performed on the auto-correlation function over $m + x$ bits (the higher level cyclostationarity). As mentioned, it is cyclostationarity along with non-zero averages for $\{Y_n\}$ that cause the PSD of a finite-length code to have a discrete component and a continuous component.

The continuous component and stationary statistics are obtained by considering the aperiodic component of the phase-averaged auto-correlation function. This is equivalent to considering the auto-covariance function rather than the auto-correlation function. The periodic component of the auto-correlation function then results in a discrete component in the PSD. This is seen as concentrated power at integer multiples of the codeword normalized frequency, $\frac{1}{m+x}$, including power at DC. In terms of data transmission, the discrete component is typically considered a somewhat undesirable use of power. However, these spectral lines can be used to extract timing or position information from the sequence in some applications [24].

A general expression for the (average) PSD function of a cyclostationary code after NRZ signaling, i.e., of its write process $W(t)$ obtained from $\{Y_n\}$, is

$$S_W(f) = S_W^c(f) + S_W^d(f), \quad (24)$$

where $S_W^c(f)$ and $S_W^d(f)$ are the continuous and discrete PSD components, respectively. For a sequence of bridged codewords $\{X_n\}$, the phase-averaged auto-correlation function can be written as

$$\mathbb{R}_X(k) = \overline{\mathbb{E}}[X_\ell X_{\ell+k}] = \frac{1}{m+x} \sum_{\ell=0}^{m+x-1} \mathbb{E}[X_\ell X_{\ell+k}]. \quad (25)$$

Averaging ℓ over the $m+x$ period is what makes the auto-correlation function phase-averaged.

For a code with codeword length m and bridging length x , the purely aperiodic component of the auto-correlation function used to calculate the continuous component of the PSD only exists for $k = 0, \pm 1, \pm 2, \dots, \pm(m+x-1)$. Since the periodic component of the auto-correlation function is cyclic with intervals equal to integer multiples of $m+x$, the coefficients of the second period in the auto-correlation function can be subtracted from the overall auto-correlation coefficients within the first period to determine $\mathbb{R}_X^{\text{aperiodic}}(k)$:

$$\mathbb{R}_X^{\text{aperiodic}}(k) = \begin{cases} \mathbb{R}_X(k) - \mathbb{R}_X(k+m+x), & 0 \leq k < m+x, \\ \mathbb{R}_X(k) - \mathbb{R}_X(k-m-x), & -m-x < k < 0, \\ 0, & \text{otherwise.} \end{cases} \quad (26)$$

It then follows that the periodic component of the auto-correlation function is

$$\mathbb{R}_X^{\text{periodic}}(k) = \mathbb{R}_X(k) - \mathbb{R}_X^{\text{aperiodic}}(k). \quad (27)$$

All the characteristics regarding periodicity/aperiodicity for the process $\{Y_n\}$ are inherited from the process $\{X_n\}$. Thus, the analysis also extends to $\{Y_n\}$.

It is important to understand that Theorem 1 and the discussion following it in Section II can be used to theoretically generate the continuous component of the PSD of a cyclostationary sequence. First, the process of using the OSTM $\mathbf{G}(D)$ to derive the continuous component of the PSD $S_X(D)$ inherently takes care of the phase-averaging for the code binary sequence $\{X_n\}$. Second, recall (10), and now average the expectation over the period covering a codeword plus its bridging pattern in order to phase-average for $\mathbb{R}_Y(k)$ as shown in (25):

$$\begin{aligned} \mathbb{R}_Y(k) &= \overline{\mathbb{E}}[Y_\ell Y_{\ell+k}] = \frac{1}{m+x} \sum_{\ell=0}^{m+x-1} \mathbb{E}[Y_\ell Y_{\ell+k}] \\ &= \frac{1}{m+x} \sum_{\ell=0}^{m+x-1} \mathbb{E}[(2X_\ell - 1)(2X_{\ell+k} - 1)] \\ &= \frac{4}{m+x} \sum_{\ell=0}^{m+x-1} \mathbb{E}[X_\ell X_{\ell+k}] - \frac{2}{m+x} \sum_{\ell=0}^{m+x-1} \mathbb{E}[X_\ell] - \frac{2}{m+x} \sum_{\ell=0}^{m+x-1} \mathbb{E}[X_{\ell+k}] + 1 \\ &= 4\overline{\mathbb{E}}[X_\ell X_{\ell+k}] - 2\overline{\mathbb{E}}[X_\ell] - 2\overline{\mathbb{E}}[X_{\ell+k}] + 1 \\ &= 4\overline{\mathbb{E}}[X_\ell X_{\ell+k}] - 4\overline{p}(1) + 1 = 4\mathbb{R}_X(k) - 4\overline{p}(1) + 1. \end{aligned} \quad (28)$$

The overline notation here refers to averaging over a period of length $m + x$ via the variable ℓ . Since $\overline{\mathbb{E}[X_\ell]}$ and $\overline{\mathbb{E}[X_{\ell+k}]}$ are averaged over the period of cyclostationarity, they are exactly the same. Their value is $\bar{p}(1)$, which is computed as shown in (6). More specifically, we see that they are the equilibrium probability. Using the overline notation leads to the same result as seen in (10), therefore justifying the use of Theorem 1 and its following discussion (the equations from (10) to (15)) for cyclostationary finite-length constrained codes.

Assuming NRZ signaling and a rectangular pulse shape, the continuous component of the PSD of the code, i.e., of $W(t)$, can then be calculated as

$$S_W^c(f) = \text{sinc}^2(\pi f) S_Y^c(e^{i2\pi f}) = \text{sinc}^2(\pi f) \left[\sum_{k=-m-x+1}^{m+x-1} \mathbb{R}_Y^{\text{aperiodic}}(k) D^k \right]_{D=e^{i2\pi f}}, \quad (29)$$

where $S_Y^c(\cdot)$ is derived from $S_X^c(\cdot)$ (both are the continuous PSD components) the same way described in (11) and further justified by (28).

Using Fourier analysis, the discrete component of the PSD then follows as

$$S_W^d(f) = \text{sinc}^2(\pi f) \left(\frac{a_0}{2} \delta(f) + \sum_{n=1}^{\infty} \frac{a_n}{2} \left(\delta \left(f - \frac{n}{m+x} \right) + \delta \left(f + \frac{n}{m+x} \right) \right) \right), \quad (30)$$

$$\text{where } a_0 = \frac{2}{m+x} \sum_{k=0}^{m+x-1} \mathbb{R}_Y^{\text{periodic}}(k)$$

$$\text{and } a_n = \frac{2}{m+x} \sum_{k=0}^{m+x-1} \mathbb{R}_Y^{\text{periodic}}(k) \cos \left(2\pi \frac{n}{m+x} k \right).$$

The $\delta(\cdot)$ is the dirac delta function, and it occurs at integer multiples of the sequence frequency, 0 and $\pm \frac{n}{m+x}$, where $n \in \mathbb{N}^+$. Further details on calculating these components can be found in [24]. By calculating the phase-averaged auto-correlation function using all possible codewords in the codebook or using many random streams of the desired code, an experimental solution for the PSD of the cyclostationary process was generated. This Monte-Carlo type solution is what we used to verify the theoretical results.

At frequencies where a discrete component is present, the OSTM used in the theoretical calculation becomes singular, causing inaccuracies in PSD values when used in (4) of Theorem 1. These points on the PSD can be calculated separately. We first note that

$$\mathbb{R}_X^{\text{periodic}}(k) = \frac{1}{m+x} \sum_{\ell=0}^{m+x-1} \mathbb{E}[X_\ell] \mathbb{E}[X_{\ell+\tilde{k}}], \text{ where } \tilde{k} = \text{mod}(|k|, m+x) + m+x, \quad (31)$$

which requires only $m+x$ calculations. The same applies for $\mathbb{R}_Y^{\text{periodic}}(k)$. Then, we use (30).

B. General Solution for the PSD of any A-LOCO Code

We first consider A-LOCO codes, adopting their definition from [12].

Definition 3. An A-LOCO code $\mathcal{AC}_{m,x}$, with parameters $m \geq 1$ and $x \geq 1$, is defined via the following properties:

- 1) Codewords in $\mathcal{AC}_{m,x}$ are binary and of length m .
- 2) Codewords in $\mathcal{AC}_{m,x}$ are ordered lexicographically.
- 3) Any pattern in the asymmetric set \mathcal{A}_x (see Definition 1) does not appear in any codeword \mathbf{c} in $\mathcal{AC}_{m,x}$.
- 4) The code $\mathcal{AC}_{m,x}$ contains all the codewords satisfying the previous three properties.

Lexicographic ordering of codewords means that they are ordered in an ascending manner following the rule $0 < 1$ for any bit, and the bit significance reduces from left to right.

As was done with infinite-length \mathcal{A}_x -constrained codes, we have categorized A-LOCO codes into the family of codes constrained by transition separation. We adopt a simple bridging pattern of 0^x unless the ending bit of a codeword and the beginning bit of the next codeword are both 1's, at which point the bridging pattern is 1^x .

Example 3. We illustrate our method for obtaining the PSD of A-LOCO codes in this example. Consider a code whose set of forbidden patterns is $\{101\}$ and all codewords are of length 4, i.e., $x = 1$ and $m = 4$. We adopt the aforementioned bridging. This is equivalently $\mathcal{AC}_{4,1}$ with its standard method of bridging which is optimal in terms of bit protection, easy to implement, and requires the minimum number of added bits [12].

In accordance with our method, the initial FSTD is given in Fig. 5a. Then, if unused states are deleted, only states resulting from the most recent bit being a 1 are labeled in gray, and states are combined when desirable, the final FSTD becomes Fig. 5b.

The resulting OSTM is then

$$\mathbf{G}(D) = \begin{bmatrix} \beta_{1,5} & \frac{3}{5}D + \beta_{\frac{3}{5},6} & \beta_{\frac{2}{5},7} & \frac{1}{5}D^3 + \beta_{\frac{1}{5},8} & 0 \\ \beta_{\frac{5}{3},4} & \beta_{1,5} & \frac{2}{3}D + \beta_{\frac{2}{3},6} & \beta_{\frac{1}{3},7} & 0 \\ \beta_{\frac{5}{2},3} & \beta_{\frac{3}{2},4} & \beta_{1,5} & \frac{1}{2}D + \beta_{\frac{1}{2},6} & 0 \\ \beta_{\frac{5}{12},7} & \beta_{3,3} & \beta_{2,4} & \beta_{1,5} & \frac{5}{12}D \\ D & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (32)$$

where $\beta_{a,b} = \sum_{k=1}^{\infty} a \left(\frac{1}{12}\right)^k D^{b+5(k-1)} = \frac{aD^b}{12-D^5}$.

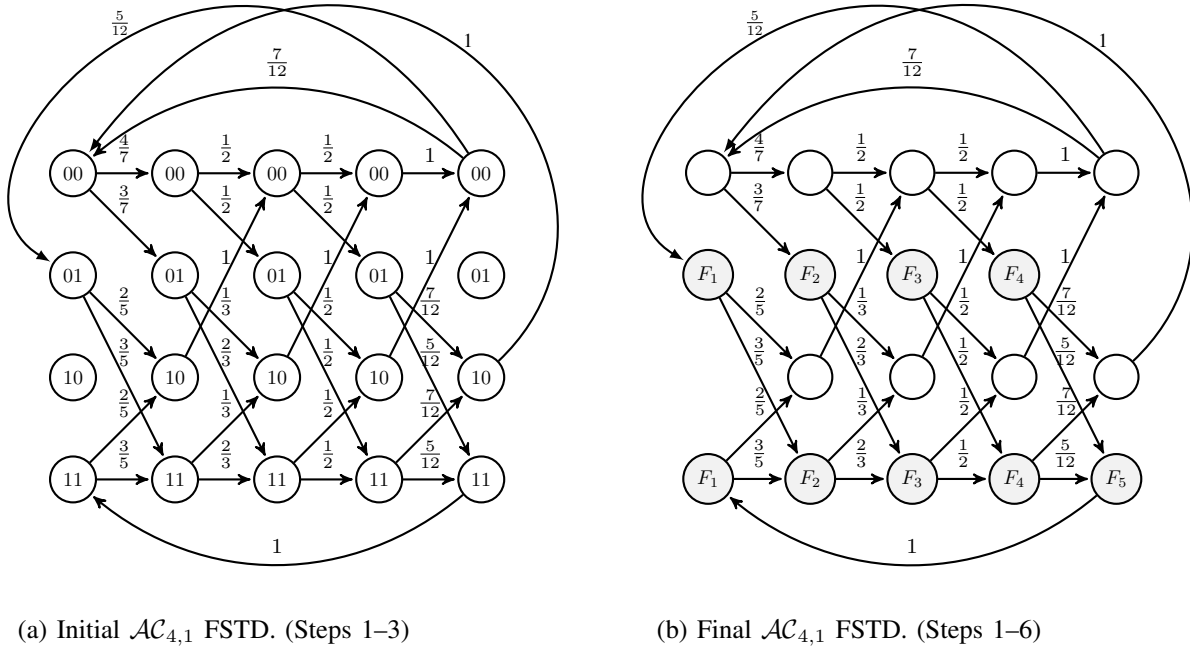


Fig. 5: FSTD used for OSTM generation of an $\mathcal{AC}_{4,1}$ sequence.

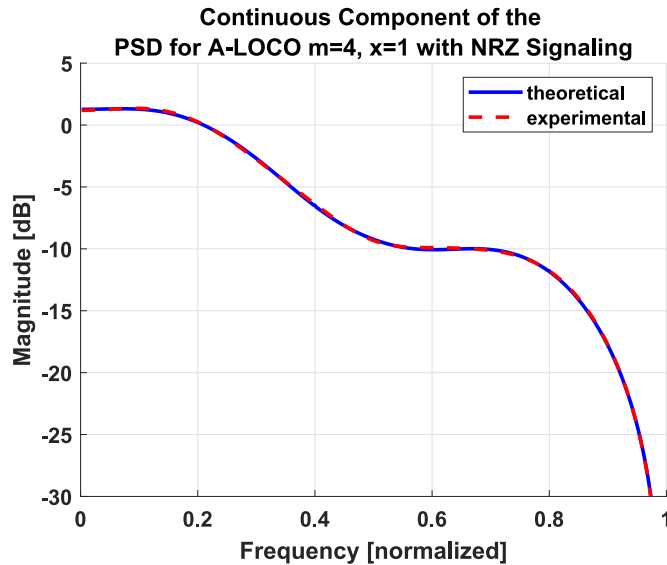


Fig. 6: The theoretical (OSTM method) versus experimental (Monte-Carlo) continuous component of the PSD for A-LOCO code $\mathcal{AC}_{4,1}$ with NRZ signaling.

Once the OSTM is found, we use Theorem 1 and its following discussion to calculate the continuous component of the PSD for $\mathcal{AC}_{4,1}$ theoretically, see Fig. 6.

The discrete component is then calculated using (30) and is shown in Fig. 7. For this A-LOCO code, the amplitudes in one period of the periodic component of the auto-correlation function

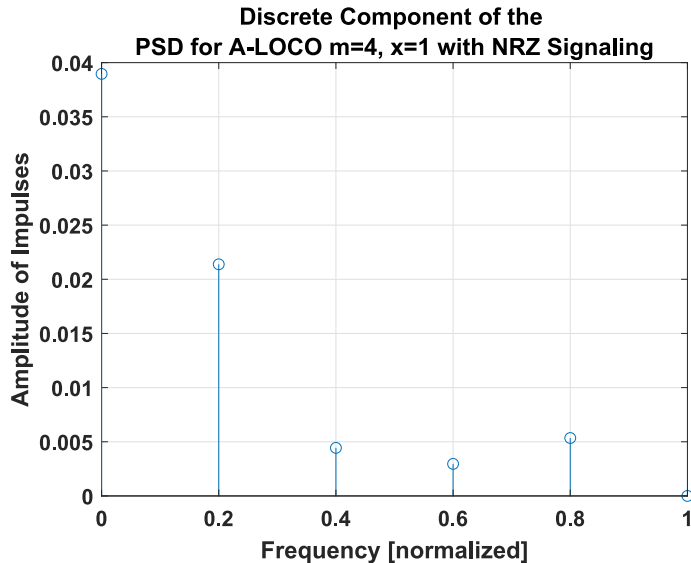


Fig. 7: The calculated discrete component of the PSD for A-LOCO code $\mathcal{AC}_{4,1}$ with NRZ signaling. Since the discrete component consists of delta functions, the result is displayed as the amplitude of the impulses (describing the area under the impulses) at the frequencies indicated.

were found to be $\mathbb{R}_Y^{\text{periodic}}(0) = 0.0964$, $\mathbb{R}_Y^{\text{periodic}}(1) = 0.0436$, $\mathbb{R}_Y^{\text{periodic}}(2) = 0.0056$, $\mathbb{R}_Y^{\text{periodic}}(3) = 0.0056$, and $\mathbb{R}_Y^{\text{periodic}}(4) = 0.0436$.

Result 3. The general OSTM solution for any A-LOCO code $\mathcal{AC}_{m,x}$ defined by the parameters m and x , and using x -bit bridging is the $m + x$ square matrix

$$\mathbf{G}(D) = \begin{bmatrix} & & & 0 & \mathbf{0} \\ & & & \vdots & \vdots \\ & \mathbf{A}_{m \times m} & & 0 & \vdots \\ & & & \zeta D & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & D\mathbf{I}_{x-1} \\ D & 0 & \dots & \dots & \mathbf{0} \end{bmatrix}, \quad (33)$$

where $\zeta = \frac{N_2(m,x) + N_3(m,x)}{N(m,x)}$. 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 $\mathcal{AC}_{m,x}$ (the number of codewords starting with 11 from the left), and the cardinality of Group 3 in $\mathcal{AC}_{m,x}$ (the number of codewords starting with 10^{x+1} from the left), respectively, adopted from [12].

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 the following rules:

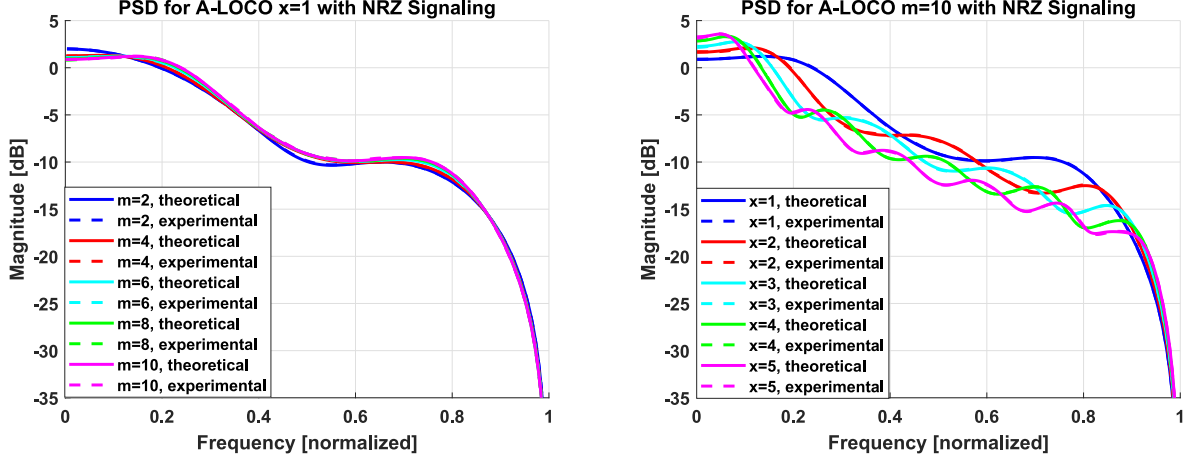


Fig. 8: Comparisons of the continuous component of various $\mathcal{AC}_{m,x}$ spectra.

- 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 $\mathbf{A}_{i,j} = \beta_{\zeta, m+2x+1}$.

In those rules,

$$\beta_{a,b} = \sum_{k=1}^{\infty} a \left(\frac{1}{N(m,x)} \right)^k D^{b+(m+x)(k-1)} = \frac{aD^b}{N(m,x) - D^{m+x}}, \quad (34)$$

$$\text{and } \lambda_{d,g} = \prod_{k=0}^g (1 - \alpha_{d-k}), \quad (35)$$

$$\text{where } \alpha_c = \frac{N_3(c,x)}{N_2(c,x) + N_3(c,x)}. \quad (36)$$

Since A-LOCO codes are asymmetric, there is an expected amount of discrete power at frequency 0 (and other frequencies as well) since 0's are more likely to occur than 1's. Moreover, as x becomes smaller compared with m , the continuous PSD at frequency 0 decreases. These and other characteristics of A-LOCO spectra, such as the number of spectral peaks increasing as x increases, are demonstrated in Fig. 8.

One standard performance measurement of a signal is the 3dB bandwidth defined as twice the distance from frequency 0 to the point at which the magnitude becomes 3 dB lower than

TABLE I: 3dB Bandwidth for Various A-LOCO Codes

A-LOCO 3dB Bandwidth Measurements									
x [bits]	1	1	1	1	1	2	3	4	5
m [bits]	2	4	6	8	10	10	10	10	10
3dB BW [normalized]	0.480	0.542	0.577	0.591	0.596	0.431	0.334	0.273	0.231

the magnitude at frequency 0. Due to the spike at frequency 0 in asymmetric codes such as A-LOCO codes, the magnitude at frequency 0 here is considered to be the limit of the magnitude as frequency approaches 0, which is the magnitude of the continuous PSD at frequency 0. For the examples in Fig. 8, the 3dB bandwidths are presented in Table I. Here, the 3dB bandwidth decreases for the same m as x increases since fast transitions in $W(t)$ become more restricted.

C. General Solution for the PSD of any LOCO Code

We next consider LOCO codes, adopting their definition from [12].

Definition 4. A LOCO code $\mathcal{C}_{m,x}$, with parameters $m \geq 1$ and $x \geq 1$, is defined via the following properties:

- 1) Codewords in $\mathcal{C}_{m,x}$ are binary and of length m .
- 2) Codewords in $\mathcal{C}_{m,x}$ are ordered lexicographically.
- 3) Any pattern in the symmetric set \mathcal{S}_x (see Definition 2) does not appear in any codeword \mathbf{c} in $\mathcal{C}_{m,x}$.
- 4) The code $\mathcal{C}_{m,x}$ contains all the codewords satisfying the previous three properties.

Lexicographic ordering of codewords means that they are ordered in an ascending manner following the rule $0 < 1$ for any bit, and the bit significance reduces from left to right.

We define the symbol z as the no writing (no transmission) symbol. A run of x consecutive z symbols is denoted by \mathbf{z}^x . To prevent forbidden patterns from appearing across two consecutive codewords, we simply separate any two consecutive LOCO codewords in a stream by \mathbf{z}^x .

Definition 5. Given a LOCO code sequence $\{X_n\}$, where binary LOCO codewords are separated by \mathbf{z}^x , we obtain $\{Y_n\}$ via NRZ signaling. We define three signals $\{A_n\}$, $\{B_n\}$, and $\{C_n\}$ by:

- If $X_i = 0$, it is translated to $Y_i = -1$, $A_i = 0$, $B_i = 1$, $C_i = 1$.
- If $X_i = 1$, it is translated to $Y_i = 1$, $A_i = 1$, $B_i = 0$, $C_i = 1$.

- If $X_i = z$, it is translated to $Y_i = 0$, $A_i = 0$, $B_i = 0$, $C_i = 0$.

Theorem 2. *The power spectral density of the modulation sequence $\{Y_n\}$ of a LOCO code $\mathcal{C}_{m,x}$ with z^x bridging pattern and NRZ signaling is given by*

$$S_Y(D) = 4S_A(D) - S_C(D). \quad (37)$$

Proof: It follows from Definition 5 that

$$\begin{aligned} S_C(D) &= \sum_{j=-\infty}^{\infty} \overline{\mathbb{E}}[C_\ell C_{\ell+j}] \cdot D^j = \sum_{j=-\infty}^{\infty} \overline{\mathbb{P}}[C_\ell = C_{\ell+j} = 1] \cdot D^j \\ &= \sum_{j=-\infty}^{\infty} (\overline{\mathbb{P}}[Y_\ell = Y_{\ell+j} = 1] + \overline{\mathbb{P}}[Y_\ell = Y_{\ell+j} = -1] \\ &\quad + \overline{\mathbb{P}}[Y_\ell = 1, Y_{\ell+j} = -1] + \overline{\mathbb{P}}[Y_\ell = -1, Y_{\ell+j} = 1]) \cdot D^j. \end{aligned} \quad (38)$$

Similarly,

$$S_A(D) = \sum_{j=-\infty}^{\infty} \overline{\mathbb{P}}[Y_\ell = Y_{\ell+j} = 1] \cdot D^j, \text{ and } S_B(D) = \sum_{j=-\infty}^{\infty} \overline{\mathbb{P}}[Y_\ell = Y_{\ell+j} = -1] \cdot D^j. \quad (39)$$

Consequently,

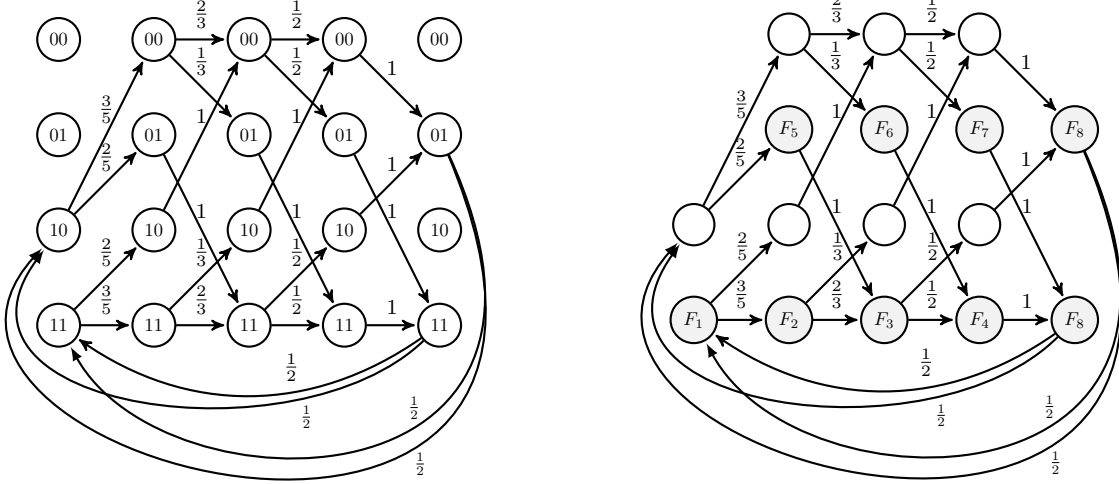
$$\begin{aligned} S_Y(D) &= \sum_{j=-\infty}^{\infty} \overline{\mathbb{E}}[Y_\ell Y_{\ell+j}] \cdot D^j \\ &= \sum_{j=-\infty}^{\infty} (\overline{\mathbb{P}}[Y_\ell = Y_{\ell+j} = 1] + \overline{\mathbb{P}}[Y_\ell = Y_{\ell+j} = -1] \\ &\quad - \overline{\mathbb{P}}[Y_\ell = 1, Y_{\ell+j} = -1] - \overline{\mathbb{P}}[Y_\ell = -1, Y_{\ell+j} = 1]) \cdot D^j \\ &= 2[S_A(D) + S_B(D)] - S_C(D). \end{aligned} \quad (40)$$

Because of the inherent symmetry of LOCO codes, we have $S_A(D) = S_B(D)$. Therefore, (40) can be simplified as (37). ■

Observe that the OSTM $\mathbf{G}_C(D)$ of $\{C_n\}$ is simply given by

$$\mathbf{G}_C(D) = \begin{bmatrix} \mathbf{0} & D \cdot \mathbf{I}_{m-1} \\ D^{x+1} & \mathbf{0} \end{bmatrix}. \quad (41)$$

Example 4. *We illustrate our method for obtaining the PSD of LOCO codes in this example. Consider a code whose set of forbidden patterns is $\{010, 101\}$ and all codewords are of length 4, i.e., $x = 1$ and $m = 4$. We adopt the bridging pattern described for signal A_n in Definition 5. In practice, we flip all bits of A_n to eliminate infinite sums in the general solution for the*

(a) Initial $\mathcal{C}_{4,1}$ FSTD. (Steps 1–3)(b) Final $\mathcal{C}_{4,1}$ FSTD. (Steps 1–6)Fig. 9: FSTD used for OSTM generation of the A sub-signal of a $\mathcal{C}_{4,1}$ sequence.

OSTM of $S_A(D)$. Because LOCO codes are symmetric and level-based signaling is being used, the continuous spectrum resulting from this OSTM will still be $S_A(D)$.

In accordance with our method, the initial FSTD is given in Fig. 9a. Then, if unused states are deleted, only states resulting from the most recent bit being a 1 are labeled in gray, and states are combined when desirable, the final FSTD becomes Fig. 9b.

The resulting OSTM is then

$$\mathbf{G}_A(D) = \begin{bmatrix} 0 & \frac{3}{5}D & 0 & 0 & 0 & 0 & \frac{1}{5}D^3 & \frac{1}{5}D^4 \\ 0 & 0 & \frac{2}{3}D & 0 & 0 & 0 & 0 & \frac{1}{3}D^3 \\ 0 & 0 & 0 & \frac{1}{2}D & 0 & 0 & 0 & \frac{1}{2}D^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & D \\ 0 & 0 & D & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & D & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & D \\ \frac{1}{2}D & 0 & 0 & 0 & \frac{1}{5}D^2 & \frac{1}{10}D^3 & \frac{1}{10}D^4 & \frac{1}{10}D^5 \end{bmatrix}. \quad (42)$$

Once the OSTM is found, we use Theorem 1 to calculate the sub-signal PSD $S_A(D)$, and we then use Theorem 2 for $S_Y(D)$ and (14) to calculate the continuous component of the PSD of $\mathcal{C}_{4,1}$ theoretically, see Fig. 10. A LOCO code does not have a discrete PSD component, and this will be discussed shortly.

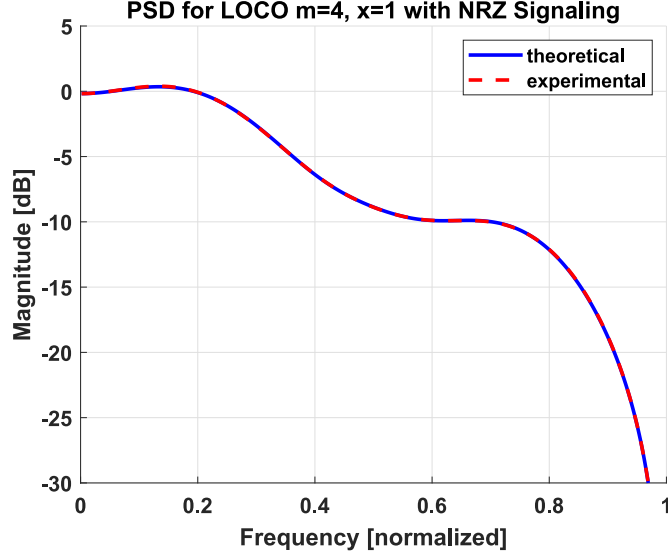


Fig. 10: The theoretical (OSTM method) versus experimental (Monte-Carlo) continuous component of the PSD for LOCO code $\mathcal{C}_{4,1}$ with NRZ signaling.

Result 4. *The general OSTM solution for the A sub-signal of any LOCO code $\mathcal{C}_{m,x}$ defined by the parameters m and x , and using \mathbf{z}^x bridging is the $m + xm + x - x^2$ square matrix $\mathbf{G}_A(D)$, which is sparse. The non-zero entries in $\mathbf{G}_A(D)$ can be found using the following rules:*

- *If $i = m + xm + m - x^2$ and $j = 1$, then $\mathbf{G}_A(D)_{i,j} = \frac{1}{2}D$.*
- *If $i = j - 1$ and $i < m$, then $\mathbf{G}_A(D)_{i,j} = \lambda_{m-i+1}D$.*
- *If $xm + x - x^2 - 1 < j < m + xm + x - x^2$ and $i < m + 1$, then consider the sub-matrix $\mathbf{A}_{m \times m}$ with non-zero entries $\mathbf{A}_{k,l}$, $1 \leq k, l \leq m$, set as follows.*

$$\diamond \text{ If } l = m \text{ or } l - k > x, \mathbf{A}_{k,l} = (1 - \lambda_{m-k+1})(1 - \lambda_{m-l+1})D^{l-k+1} \underbrace{\left(\prod_{g=m-l+2}^{m-x-1} \lambda_g \right)}_{\text{if } l > x+2}.$$

- *If $m < i < m + xm - x - x^2 + 1$ and $x + 1 < j < xm - x^2 + 2$, then $\mathbf{G}_A(D)_{i,j} = D$.*
- *If $m + xm - x - x^2 < i < m + xm + x - x^2$ and $m + xm - x - x^2 + 1 < j < m + xm + x - x^2 + 1$, then $\mathbf{G}_A(D)_{i,j} = D$.*
- *If $i = m + xm + m - x^2$ and $xm + x - x^2 - 1 < j < m + xm + x - x^2$, then consider the sub-vector $\mathbf{a}_{1 \times m}$ with entries \mathbf{a}_l , $1 \leq l \leq m$, set as follows.*

$$\diamond \mathbf{a}_l = \frac{1}{2}(1 - \lambda_{m-l+1}) \prod_{g=m-l+2}^m \lambda_g.$$

In those rules,

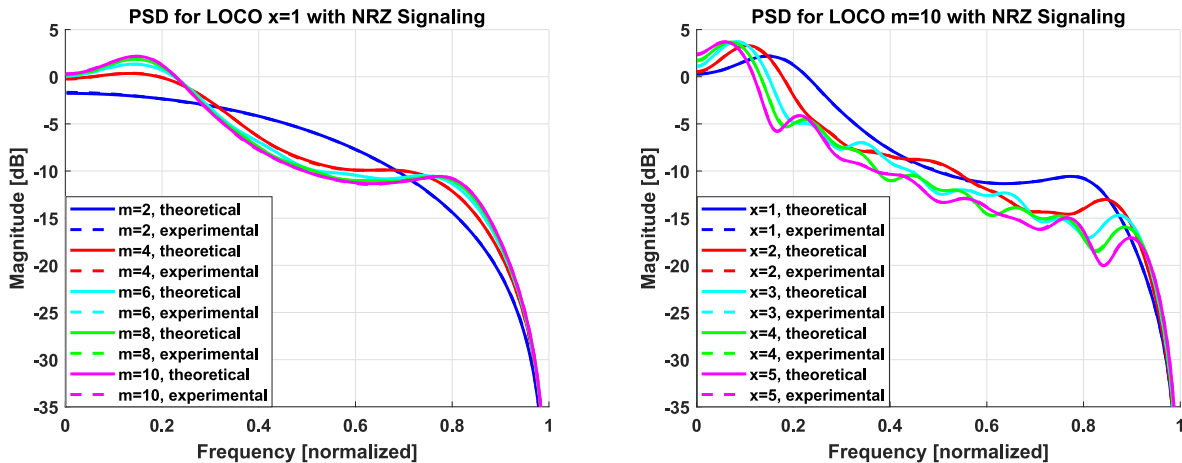


Fig. 11: Comparisons of the continuous component of various $\mathcal{C}_{m,x}$ spectra.

TABLE II: 3dB Bandwidth for Various LOCO Codes

LOCO 3dB Bandwidth Measurements									
x [bits]	1	1	1	1	1	2	3	4	5
m [bits]	2	4	6	8	10	10	10	10	10
3dB BW [normalized]	0.868	0.644	0.582	0.568	0.558	0.412	0.327	0.283	0.246

$$\lambda_a = \frac{N_1(a, x)}{\frac{1}{2}N(a, x)}. \quad (43)$$

The notation $N(m, x)$ and $N_1(m, x)$ refers to the cardinality of $\mathcal{C}_{m,x}$ and the cardinality of Group 1 in $\mathcal{C}_{m,x}$ (the number of codewords starting with 00 from the left), respectively, adopted from [9].

Since LOCO codes are symmetric, there is no discrete power at frequency 0 (DC) when NRZ signaling is adopted since 0's (-1 's) and 1's ($+1$'s) are equally likely to occur, i.e., $\mathbb{E}[Y_n] = 0$. Symmetric coding with NRZ signaling leads to the case where the auto-correlation function is equal to the auto-covariance function. Consequently, there is no discrete component at all in the PSD for all frequencies. Once again, the continuous PSD at frequency 0 increases and the number of spectral peaks increases for the same m as x increases. These spectral characteristics along with others are demonstrated in Fig. 11.

For the examples in Fig. 11, the 3dB bandwidths are presented in Table II. Here, the 3dB bandwidth decreases for the same m as x increases since the minimum and average time separations between consecutive transitions in $W(t)$ become bigger.

VI. CONCLUSION

We introduced a method for deriving the power spectra of constrained codes associated with level-based signaling. We applied this method to infinite-length \mathcal{A}_x -constrained and \mathcal{S}_x -constrained codes. We then generalized this method to address finite-length challenges in deriving power spectra, particularly how to find the OSTM, how to address cyclostationarity, and how to handle the periodic component of the auto-correlation function. We applied the generalized method to finite-length A-LOCO and LOCO codes. We discussed important spectral properties of these classes of codes based on our derivations. Our theoretical PSD results perfectly match the experimental PSD results for all codes. We present another method for deriving the spectra along with an algorithmic approach for self-clocked codes in the appendices. We suggest that our methods and ideas can provide helpful insights to data storage and data transmission engineers regarding different classes of constrained codes they plan to employ.

ACKNOWLEDGMENT

This research was supported in part by NSF under Grant CCF 1717602 and in part by AFOSR under Grant FA 9550-17-1-0291.

APPENDIX A

ALTERNATE METHOD FOR OBTAINING THE PSD OF \mathcal{A}_x - AND \mathcal{S}_x -CONSTRAINED CODES

Instead of altering the NRZI-focused method seen in [25] for a code with NRZ signaling as we did in this paper, the relationship between NRZI signals and their base binary sequence can be used to modify the constraints of the base sequence. NRZI signaling is transition-based, with each transition (from -1 to $+1$ or $+1$ to -1) being caused by a 1 in the base sequence. Therefore, in order to use the unaltered method, all we need to do is to generate a constrained sequence whose NRZI signal is equivalent to the NRZ signal of the desired sequence.

We focus here on infinite-length codes. We start from the desired binary sequence, and perform “*inverse NRZI*” operation on it. In particular, we take the bit-wise difference between each two adjacent bits in the desired sequence, resulting in a new binary sequence whose NRZI signal is the NRZ signal we are after, assuming the initial signal level is the same. What remains is that we identify the new constraint in this bit-wise difference sequence in order to build its FSTD and OSTM, and thus derive its PSD. This is done as follows for the codes we discuss here. Suppose the initial level is -1 .

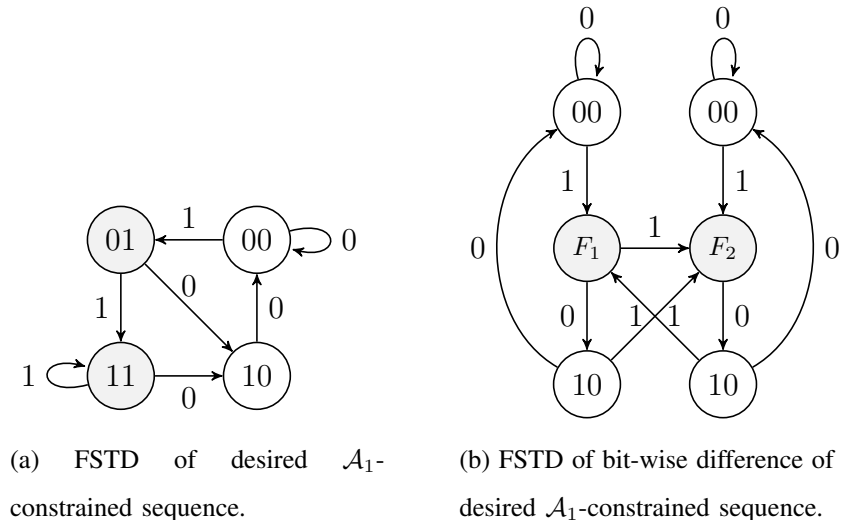


Fig. 12: FSTDs considered for using the unaltered method [25] for an \mathcal{A}_1 -constrained sequence.

- For \mathcal{A}_x -constrained codes, the new constraint is that separation between 1's indexed by $2j$ and $2j + 1$ (resp., $2j + 1$ and $2j + 2$) is unrestricted (resp., at least x 0's), where $j \in \mathbb{N}$.
- For \mathcal{S}_x -constrained codes, the new constraint is that separation between 1's indexed by j and $j + 1$ is at least x 0's, where $j \in \mathbb{N}$. This is exactly the d constraint in an RLL code for $d = x$.

The unaltered method in [25] is then used to derive the PSD via the OSTM built using the new constraint. Observe that we are discussing here the continuous component of the PSD. The delta function at DC in the case of asymmetric codes is computed separately as discussed in Section V. Observe that for an asymmetric sequence, the initial level is quite important for generating the desired NRZ signal. An example of the FSTDs for an \mathcal{A}_1 -constrained sequence and its bit-wise difference sequence is provided in Fig. 12.

Result 5. *By exploring FSTDs and OSTDs of the bit-wise difference of various \mathcal{A}_x -constrained sequences, it can be found that the alternative OSTM for any \mathcal{A}_x -constrained sequence is*

$$\mathbf{G}(D) = \begin{bmatrix} 0 & \sum_{k=1}^{\infty} (\frac{1}{2})^k D^k \\ \sum_{k=1}^{\infty} (\frac{1}{2})^k D^{x+k} & 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{D}{2-D} \\ \frac{D^{1+x}}{2-D} & 0 \end{bmatrix}. \quad (44)$$

Similarly, the alternative OSTM for any \mathcal{S}_x -constrained sequence is

$$\mathbf{G}(D) = \left[\sum_{k=1}^{\infty} (\frac{1}{2})^k D^{x+k} \right] = \left[\frac{D^{1+x}}{2-D} \right]. \quad (45)$$

APPENDIX B

GENERATING THE OSTM OF CLOCKED CONSTRAINED CODES

Definition 6. [9] [12] A self-clocked LOCO (C-LOCO) code $\mathcal{C}_{m,x}^{\text{clk}}$ is defined by

$$\mathcal{C}_{m,x}^{\text{clk}} = \mathcal{C}_{m,x} \setminus \{\mathbf{0}^m, \mathbf{1}^m\}, \text{ where } m \geq 2. \quad (46)$$

Furthermore, a self-clocked A-LOCO (CA-LOCO) code $\mathcal{AC}_{m,x}^{\text{clk}}$ is defined by

$$\mathcal{AC}_{m,x}^{\text{clk}} = \mathcal{AC}_{m,x} \setminus \{\mathbf{0}^m, \mathbf{1}^m\}, \text{ where } m \geq 2. \quad (47)$$

Self-clocked codes impose another constraint on the stream generation, causing the method used in Section IV for finding the FSTD and OSTM of the code to require more than the maximum $(m+x)2^{x+1}$ states introduced. One idea for handling this additional complexity is to utilize the breadth first search (BFS) algorithm to properly construct the code OSTD and OSTM. Details on the BFS algorithm are not discussed in this paper for brevity.

Algorithm 1 Deriving OSTD of Self-Clocked Codes

Inputs: \mathbf{P} , \mathbf{V} , target, and $k_{\text{eff}}^{\text{clk}}$.

for $d' = 1 : (k_{\text{eff}}^{\text{clk}} + 1)$ **do**

$\mathbf{V}_1 = \mathbf{P} \times \mathbf{V}$. (matrix multiplication)

$\mathbf{V}_2 = \text{target} \cdot \mathbf{V}_1$. (element-wise multiplication)

Record d' and all the non-zero entries in \mathbf{V}_2 .

$\mathbf{V} = \mathbf{V}_1 \cdot (\sim \mathbf{V}_2)$. ($\sim \mathbf{V}_2$ is the bit flipping matrix of the binary version of \mathbf{V}_2)

end for

Recover the recorded numbers of steps (values of d' for \mathbf{V}_2) as labels on one graph.

Recover the recorded probabilities (non-zero entries in \mathbf{V}_2) as labels on another graph.

Output: OSTD of the code with number of steps and corresponding probability for each edge.

To make use of the BFS algorithm, first note that for every codeword in $\mathcal{C}_{m,x}^{\text{clk}}$ or $\mathcal{AC}_{m,x}^{\text{clk}}$, there exists at least one transition for self-clocking. Thus, we can define the maximum number of successive symbols between two consecutive 1's as $k_{\text{eff}}^{\text{clk}} = 2(m-1) + x$, see [9] and [12]. We define the following variables as inputs to the BFS algorithm:

- Derive the stochastic probability $t \times t$ matrix \mathbf{P} , where t is the number of states in the FSTD and \mathbf{P}_{ij} is the probability of transiting from state j to state i in 1 step.

- Define a column vector \mathbf{w} of length t , where $w_i = 1$ if and only if all the incoming edges of vertex i in the FSTD have label 1, otherwise $w_i = 0$.
- Define a $t \times t$ matrix \mathbf{V} , where $V_{ij} = 1$ if and only if $i = j$ and $w_i = 1$, otherwise $V_{ij} = 0$. Each column of \mathbf{V} has at most one non-zero entry. We start our BFS with \mathbf{V} and update its value after we walk (transit) in the FSTD.
- Define a $t \times t$ matrix $\text{target} = \mathbf{w}\mathbf{w}^\top$, i.e., $\text{target}_{ij} = 1$ if and only if $w_j = 1$ and $w_i = 1$, otherwise $\text{target}_{ij} = 0$. In the algorithm, when we hit a non-zero entry in target , we find a possible $00 \dots 01$ sequence and its associated probability.

The BFS algorithm is Algorithm 1.

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] P. A. Franaszek, "Sequence-state methods for run-length-limited coding," *IBM J. Res. Dev.*, vol. 14, no. 4, pp. 376–383, Jul. 1970.
- [4] P. Siegel, "Recording codes for digital magnetic storage," *IEEE Trans. Magn.*, vol. 21, no. 5, pp. 1344–1349, Sep. 1985.
- [5] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for digital recorders," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2260–2299, Oct. 1998.
- [6] B. Vasic and E. Kurtas, *Coding and Signal Processing for Magnetic Recording Systems*. CRC Press, 2005.
- [7] A. Hareedy, R. Wu, and L. Dolecek, "A channel-aware combinatorial approach to design high performance spatially-coupled codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4834–4852, Aug. 2020.
- [8] R. Karabed and P. H. Siegel, "Coding for higher-order partial-response channels," in *Proc. SPIE Int. Symp. Voice, Video, and Data Commun.*, M. R. Raghuveer, S. A. Dianat, S. W. McLaughlin, and M. Hassner, Eds., Philadelphia, PA, Oct. 1995, vol. 2605, pp. 115–126.
- [9] A. Hareedy and R. Calderbank, "LOCO codes: Lexicographically-ordered constrained codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3572–3589, Jun. 2020.
- [10] J.-D. Lee, S.-H. Hur, and J.-D. Choi, "Effects of floating-gate interference on NAND flash memory cell operation," *IEEE Electron Device Lett.*, vol. 23, no. 5, pp. 264–266, May 2002.
- [11] M. Qin, E. Yaakobi, and P. H. Siegel, "Constrained codes that mitigate inter-cell interference in read/write cycles for flash memories," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 836–846, Apr. 2014.
- [12] 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.
- [13] V. Taranalli, H. Uchikawa, and P. H. Siegel, "Error analysis and inter-cell interference mitigation in multi-level cell flash memories," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, UK, Jun. 2015, pp. 271–276.
- [14] A. Hareedy, B. Dabak, and R. Calderbank, "Managing device lifecycle: Reconfigurable constrained codes for M/T/Q/P-LC Flash memories," Jan. 2020. [Online]. Available: <https://arxiv.org/abs/2001.02325>
- [15] R. Wood, M. Williams, A. Kavcic, and J. Miles, "The feasibility of magnetic recording at 10 terabits per square inch on conventional media," *IEEE Trans. Magn.*, vol. 45, no. 2, pp. 917–923, Feb. 2009.

- [16] K. S. Chan and M. R. Elidrissi, "A system level study of two-dimensional magnetic recording (TDMR)," *IEEE Trans. Magn.*, vol. 49, no. 6, pp. 2812–2817, Jun. 2013.
- [17] B. Dabak, A. Hareedy, and R. Calderbank, "Non-binary constrained codes for two-dimensional magnetic recording," *IEEE Trans. Magn.*, to be published, doi: 10.1109/TMAG.2020.3017511.
- [18] K. A. S. Immink, "Modulation systems for digital audio discs with optical readout," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Atlanta, Georgia, USA, Mar.–Apr. 1981, pp. 587–589.
- [19] S. Sridhara and N.R. Shanbhag, "Coding for reliable on-chip buses: A class of fundamental bounds and practical codes," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 26, no. 5, pp. 977–982, May 2007.
- [20] J. Saadé, A. Goulahsen, A. Picco, J. Huloux, and F. Pétrot, "Low overhead, DC-balanced and run length limited line coding," in *Proc. IEEE 19th Workshop on Signal and Power Integrity (SPI)*, Berlin, Germany, May 2015, pp. 1–4.
- [21] R. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes—An application of symbolic dynamics to information theory," *IEEE Trans. Inf. Theory*, vol. 29, no. 1, pp. 5–22, Jan. 1983.
- [22] K. A. S. Immink, "A practical method for approaching the channel capacity of constrained channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1389–1399, Sep. 1997.
- [23] T. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. 19, no. 1, pp. 73–77, Jan. 1973.
- [24] K. A. S. Immink, *Codes for Mass Data Storage Systems*. Shannon Foundation Publishers, 2004.
- [25] 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.
- [26] B. S. Bosik, "The spectral density of a coded digital signal," *Bell Syst. Tech. J.*, vol. 51, pp. 921–932, Apr. 1972.
- [27] G. Cariolaro and G. Tronca, "Spectra of block coded digital signals," *IEEE Trans. Commun.*, vol. 22, no. 10, pp. 1555–1564, Oct. 1974.
- [28] J. Justesen, "Calculation of power spectra for block coded signals," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 389–392, Mar. 2001.
- [29] C. Pimentel and V. C. da Rocha, "On the power spectral density of constrained sequences," *IEEE Trans. Commun.*, vol. 55, no. 3, pp. 409–416, Mar. 2007.
- [30] E. Biglieri and G. Caire, "Power spectrum of block-coded modulation," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1580–1585, Feb./Mar./Apr. 1994.
- [31] D. Lindholm, "Power spectra of channel codes for digital magnetic recording," *IEEE Trans. Magn.*, vol. 14, no. 5, pp. 321–323, Sep. 1978.
- [32] C. Menyennett and H. C. Ferreira, "Spectral properties of (d, k) codes with multiple spacing," *IEEE Trans. Commun.*, vol. 30, no. 6, pp. 4215–4217, Nov. 1994.
- [33] D. F. Bishop, S. Million, T. M. Nguyen and M. K. Simon, "Power spectrum of unbalanced NRZ and biphase signals in the presence of data asymmetry," *IEEE Trans. on Electromagn. Compat.*, vol. 40, no. 1, pp. 55–61, Feb. 1998.