Convolutional Codes with Maximum Column Sum Rank for Network Streaming

Rafid Mahmood, *Student Member, IEEE*, Ahmed Badr, *Member, IEEE*, and Ashish Khisti *Member, IEEE*

***Abstract*—The column Hamming distance of a convolutional code determines the error correction capability when streaming over a class of packet erasure channels. We introduce a metric known as the column sum rank, that parallels column Hamming distance when streaming over a network with link failures. We prove rank analogues of several known column Hamming distance properties and introduce a new family of convolutional codes that maximize the column sum rank up to the code memory. Our construction involves finding a class of super-regular matrices that preserve this property after multiplication with non-singular block diagonal matrices in the ground field.**

***Index Terms*—Column distance, maximum rank distance (MRD) codes, network coding, super-regular matrices, maximum-distance profile (MDP) codes.**

I. Introduction

In streaming communication, source packets arrive sequentially at the transmitter and are only useful for playback by the receiver in the same order. Erased packets must be recovered within a given maximum delay or be considered permanently lost. Streaming codes recover packets within these decoding deadlines and have previously been studied for single-link communication [2]–[5]. The works referenced in [2]–[4] focused primarily on low-delay recovery against burst losses, which are the predominant erasure patterns in Internet streams [6]. Alternatively, [4], [5] considered coding for channels with arbitrary erasure patterns, restricting only the number of erasures in a window. It was shown that the column Hamming distance determines the maximum tolerable number of erasures that can occur in any window of the stream for decoding to remain successful. If there are fewer erasures than the distance in every sliding window, each source symbol is recovered within a given delay. A family of memory m convolutional codes, known as m-Maximum Distance Separable (m-MDS) codes, attain the maximum column Hamming distance up to the code memory. Furthermore, these are constructed from block Toeplitz super-regular matrices [5], [7]–[9]. These codes can also be used as constituent codes in the construction of optimal burst error correction codes for streaming systems [4].

Suppose that a transmitter sends packets to several users through a series of intermediate nodes. Using generation-based linear network codes, the problem of decoding is reduced to inverting the channel transfer matrix between the transmitted and received packets [10]–[12]. If links in the network fail to transmit packets for a given time instance, the rank of the channel matrix decreases, making packet recovery infeasible. One solution is to use end-to-end schemes to precode channel packets before transmission. Rank metric codes such as Gabidulin codes [13], [14] are capable of protecting packets in rank-deficient channels. The minimum rank distance of a code determines the maximum permissible rank loss in a channel matrix. This method can be considered for single-shot network coding [15], [16] or for multishot extensions [17], [18].

In this work we study a streaming setup that extends the single-link model of [2]–[4] to a network. We assume that the intermediate nodes implement linear network coding and produce a channel transfer matrix relating the transmitted and received packets. To combat link failures, the source stream is further precoded at the source using a *streaming code*. We define a new metric, the column sum rank and introduce a new family of convolutional codes that attain the maximum value for this metric. These are rank metric analogues of m-MDS codes [7]. Just as the column Hamming distance determines the maximum allowable number of erasures in single link streaming, we show that the column sum rank determines the maximum rank deficiency of the channel. Interestingly, there has been little prior work on rank metric convolutional codes. To our knowledge, the only previously studied construction appears in [18], where the authors consider the active column sum rank as the metric of importance. Consequently, their approach and results differ from the present work both in the code constructions and applications.

This paper is outlined as follows. The network streaming problem is introduced in Section II. We provide an overview of rank metric block codes and m-MDS codes in Section III. The column sum rank is defined in Section IV, where we derive several properties and establish their relevance in network streaming. Codes that maximize the column sum rank are referred to as Maximum Sum Rank (MSR) codes. We introduce a class of super-regular matrices in Section V that preserve super-regularity after multiplication with block diagonal matrices in the ground field and use these to construct an MSR code in Section VI. We conclude this paper with code examples and a discussion on the necessary field size.

II. Network Streaming Problem

The streaming problem is defined in three steps: the encoder, network model, and decoder. Encoding is performed in a causal fashion, as the incoming packets are not known until the time at which they must be transmitted. A linear network code has been applied to the network and each node receives and sends linear combinations of the symbols in the channel

R. Mahmood, A. Badr, and A. Khisti ({rmahmood, abadr, akhisti}@comm.utoronto.ca) are with the University of Toronto, Toronto, ON, Canada.

Part of this work was presented at the 2015 International Symposium on Information Theory (ISIT) in Hong Kong [1].

packet. Consequently the network is abstracted to a channel matrix, which is assumed to be known to the receiver [12]. The decoder observes linear combinations of the symbols in each transmitted packet and must recover the source within the imposed deadline.

A. Encoder

Let q be a prime power and $M \geq 0$. At each time instance $t \geq 0$, a source packet $\mathbf{s}_t \in \mathbb{F}_{q^M}^k$ arrives at the transmitter node. A channel packet $\mathbf{x}_t \in \mathbb{F}_{q^M}^n$ is constructed via a causal function of the previous source packets $\gamma_t(\mathbf{s}_0, \ldots, \mathbf{s}_t)$. We consider the class of linear time invariant encoders. A rate $R = \frac{k}{n}$ encoder with memory m generates the channel packet

$$\mathbf{x}_t = \sum_{i=0}^{m} \mathbf{s}_{t-i} \cdot \mathbf{G}_i, \tag{1}$$

using a set of generator matrices $\mathbf{G}_i \in \mathbb{F}_{q^M}^{n \times k}$ for $0 \leq i \leq m$.

B. Network Model

The transmitter node sends one channel packet through the network at each time instance. Although there is a natural delay in the end-to-end transmission due to link delays, we assume that such delays are sufficiently small so that one time instance contains the encoding, transmission, and decoding of a single channel packet[1]. The transmission of a single channel packet over one time instance is referred to as a *shot*. In each shot, the destination node observes $\mathbf{y}_t = \mathbf{x}_t \mathbf{A}_t$, where $\mathbf{A}_t \in \mathbb{F}_q^{n \times n}$ is the channel matrix at time t, and is known to the receiver. In practice, the coefficients for the linear transformations applied by each node can be encoded into header bits, which the receiver uses to reconstruct the channel matrix [12].

Each shot is independent of all others. Communication over a window $[t, t+W-1]$ of W shots is described using $\mathbf{y}_{[t,t+W-1]} = \mathbf{x}_{[t,t+W-1]} \mathbf{A}_{[t,t+W-1]}$, where $\mathbf{A}_{[t,t+W-1]} = \operatorname{diag}(\mathbf{A}_t, \ldots, \mathbf{A}_{t+W-1})$ is a block diagonal channel matrix [17], [19]. Let $\rho_t \triangleq \operatorname{rank}(\mathbf{A}_t)$ denote the rank of $\mathbf{A}_t$, for all $t \geq 0$. The sum of the ranks of the individual channel matrices is equal to the rank of the channel matrix in the window, i.e., $\sum_{i=t}^{t+W-1} \rho_i = \operatorname{rank}(\mathbf{A}_{[t,t+W-1]})$. Suppose that at any time instance, a link in the network may fail to transmit its intended symbol. Intermediate nodes that do not receive an intended symbol simply do not include that symbol in the linear combination they transmit. If all links are functional in the shot at any time t, then $\rho_t = n$, but failing links may result in a rank-deficient channel matrix at that time. One failing link can eliminate at most one of the min-cut paths connecting the transmitter and receiver. It follows that $\operatorname{rank}(\mathbf{A}_t)$ is reduced by at most 1 for every failing link [20]. We introduce a sliding window model to characterize rank deficiencies in the network.

Definition 1. Consider a network where for all $t \geq 0$, the receiver observes $\mathbf{y}_t = \mathbf{x}_t \mathbf{A}_t$, with $\rho_t \triangleq \operatorname{rank}(\mathbf{A}_t)$. The **Rank-Deficient Sliding Window Network** $\mathcal{CH}(S, W)$ has the property that in any sliding window of length W, the rank of the block diagonal channel decreases by no more than S, i.e., $\sum_{i=t}^{t+W-1} \rho_i \geq nW - S$ for each $t \geq 0$.

In analysis, we disregard the linearly dependent columns of the channel matrix and the associated received symbols. At each time instance, the receiver effectively observes $\mathbf{y}_t^* = \mathbf{x}_t \mathbf{A}_t^*$, where $\mathbf{A}_t^* \in \mathbb{F}_q^{n \times \rho_t}$ contains only the linearly independent columns of $\mathbf{A}_t$ and is referred to as the reduced channel matrix.

Remark 1. A sliding window model has been used in prior works on delay-constrained coding over single-link channels [4], [21]. In these works, the channel is adversarially permitted to erase symbols in each channel packet up to a maximum number of erasures within each sliding window. The Rank-Deficient Sliding Window Network can be viewed as an extension of this model, where the channel introduces rank deficiencies rather than erasures.

C. Decoder

Let T be the maximum delay permissible by the receiver node. A packet received at time t must be recovered by time $t+T$ using a delay-constrained decoder, i.e., $\hat{\mathbf{s}}_t = \eta_t(\mathbf{y}_0, \ldots, \mathbf{y}_{t+T})$ is the reconstructed packet. If the decoded source packet $\hat{\mathbf{s}}_t$ is equal to $\mathbf{s}_t$, then the source packet is perfectly recovered by the deadline; otherwise, it is declared lost. A linear code $\mathcal{C}$ over $\mathbb{F}_{q^M}$ is defined as *feasible* for $\mathcal{CH}(S, W)$ if the encoding and decoding functions for the code are capable of perfectly recovering every source packet transmitted over the channel with delay T.

In this paper we will assume that the window length satisfies $W = T + 1$. A source packet $\mathbf{s}_t$ must be decoded by time $t+T$ at the receiver. Thus its active duration spans the interval $[t, t+T]$, which maps to a window length of $W = T + 1$. Nevertheless, we will also discuss how our codes can handle the case when $W \neq T + 1$.

Our objective in this paper is to construct codes that guarantee recoverability under the worst channel conditions for a fixed delay and rate, i.e., identifying the largest rank deficiency S for which a code with a given rate is feasible. Towards this end, we introduce a new metric called the column sum rank distance of a convolutional code. We show that it is both a necessary and sufficient metric to determine the maximum rank deficiency from which the code guarantees perfect recovery. Thus, maximizing S reduces to finding codes with maximum column sum rank distance over the interval $[0, T]$.

Codes which achieve the maximum column sum rank distance will be referred to as Maximum Sum Rank (MSR) codes in this paper. Furthermore the column sum rank distance possess a profile property. Achieving the maximum distance at one point implies that it is also maximized at all points before it. Operationally we show that this property guarantees that the product of the generator matrix with elements of a specific set of channel matrices is full-rank. Finally we propose a family of super-regular matrices that permit this multiplication property

[1]For example in audio streaming, coded speech packets are generated once every 20 ms. When the propagation delays are much smaller than this value, they can be ignored.

of the generator matrix, which we then use to construct MSR codes. Our proposed family of codes uses the properties of rank metric block codes and m-MDS convolutional codes, which are introduced as preliminaries in the following section.

III. Background

A. Rank Metric Codes

Consider a vector $\mathbf{x} \in \mathbb{F}_{q^M}^n$ over the extension field. We refer to $\mathbf{x}$ as a channel packet. The vector $\mathbf{x}$ over the extension field is isomorphic to an $n \times M$ matrix over the ground field $\mathbb{F}_q$. Formally, a bijective mapping $\phi_n : \mathbb{F}_{q^M}^n \to \mathbb{F}_q^{n\times M}$ allows for the conversion of this vector to a matrix over the ground field. We more thoroughly detail this mapping in Appendix A. Noting that a normal basis can describe every element in the extension field, the elements of $\mathbb{F}_{q^M}$ are mapped to linearized polynomials evaluated at a normal element, whose coefficients form column vectors. A row vector in $\mathbb{F}_{q^M}^n$ then maps to a matrix whose columns are the coefficients of the corresponding linearized polynomials.

The rank of $\mathbf{x}$ is defined as the rank of its associated matrix $\phi_n(\mathbf{x})$. The *rank distance* between any two vectors $\mathbf{x}, \hat{\mathbf{x}} \in \mathbb{F}_{q^M}^n$ is defined as

$$d_R(\mathbf{x}, \hat{\mathbf{x}}) \triangleq \operatorname{rank}\left(\phi_n(\mathbf{x}) - \phi_n(\hat{\mathbf{x}})\right).$$

The rank distance is a metric and is upper bounded by the Hamming distance [13]. For any linear block code $\mathcal{C}[n,k]$ over $\mathbb{F}_{q^M}$, the *minimum rank distance* is defined as the smallest rank amongst all non-zero channel packets. Similar to the minimum Hamming distance, the minimum rank distance of a code must satisfy a Singleton-like bound, i.e., $d_R(\mathcal{C}) \leq \min\left\{1, \frac{M}{n}\right\}(n-k)+1$ [13]. We simplify the notation when $\mathcal{C}$ is obvious. It is assumed that $M \geq n$ from here on; d_R is then bounded exactly by the classic Singleton bound. Any code that meets this bound with equality is referred to as a Maximum Rank Distance (MRD) code. Such codes possess the following property.

Theorem 1 (Gabidulin, [13])**.** Let $\mathbf{G} \in \mathbb{F}_{q^M}^{k\times n}$ be the generator matrix of an MRD code. The product of $\mathbf{G}$ with any full-rank matrix $\mathbf{A} \in \mathbb{F}_q^{n\times k}$ satisfies $\operatorname{rank} \mathbf{GA} = k$.

A complementary theorem was proven in [13] for the parity-check matrix of an MRD code. We use the equivalent generator matrix property, which arises from the fact that the dual of an MRD code is also an MRD code [13].

Gabidulin codes are an important family of MRD codes. To construct a Gabidulin code, let $g_0, \ldots, g_{n-1} \in \mathbb{F}_{q^M}$ be a set of elements that are linearly independent over $\mathbb{F}_q$. The generator matrix for a Gabidulin code $\mathcal{C}[n,k]$ is given by

$$\mathbf{G} = \begin{pmatrix} g_0 & g_1 & \cdots & g_{n-1} \\ g_0^{[1]} & g_1^{[1]} & \cdots & g_{n-1}^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ g_0^{[k-1]} & g_1^{[k-1]} & \cdots & g_{n-1}^{[k-1]} \end{pmatrix}.$$

where we use the notation $g^{[j]} \triangleq g^{q^j}$ to denote the j-th Frobenius power of $g \in \mathbb{F}_{q^M}$ (see Appendix A). Gabidulin codes can be applied directly as end-to-end codes over networks where a channel matrix $\mathbf{A} \in \mathbb{F}_q^{n\times n}$ transforms symbols of the transmitted channel packet [16], [19]. For a source packet $\mathbf{s} \in \mathbb{F}_{q^M}^k$ encoded by a Gabidulin code, a receiver observes $\mathbf{y} = \mathbf{sGA}$. By Theorem 1, the product $\mathbf{GA}$ is an invertible matrix as long as $\operatorname{rank} \mathbf{A} \geq k$.

B. The Column Hamming Distance

Let $\mathcal{C}[n,k,m]$ be a linear time-invariant convolutional code, where m is the code memory. For a source packet sequence $\mathbf{s}_{[0,j]} = (\mathbf{s}_0, \ldots, \mathbf{s}_j) \in \mathbb{F}_{q^M}^{k(j+1)}$, the channel packet sequence[2] $\mathbf{x}_{[0,j]} = \mathbf{s}_{[0,j]}\mathbf{G}_j^{\mathrm{EX}}$ is determined using the extended form generator matrix

$$\mathbf{G}_j^{\mathrm{EX}} \triangleq \begin{pmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \cdots & \mathbf{G}_j \\ & \mathbf{G}_0 & \cdots & \mathbf{G}_{j-1} \\ & & \ddots & \vdots \\ & & & \mathbf{G}_0 \end{pmatrix}, \tag{2}$$

where $\mathbf{G}_j \in \mathbb{F}_{q^M}^{k\times n}$ and $\mathbf{G}_j = \mathbf{0}$ for $j > m$ [22, Chapter 1]. We assume from here on that $\mathbf{G}_0$ always has full row rank. This guarantees that $\mathbf{G}_j^{\mathrm{EX}}$ also possesses full row rank, which is a property used in subsequent results.

The Hamming weight of $\mathbf{x}_{[0,j]}$ is a sum of the Hamming weight of each channel packet $\mathbf{x}_t$ for $0 \leq t \leq j$. The j-th column Hamming distance of a convolutional code is defined

$$d_H(j) \triangleq \min_{\mathbf{x}_{[0,j]} \in \mathcal{C}, \mathbf{s}_0 \neq \mathbf{0}} \mathrm{wt}_H(\mathbf{x}_{[0,j]}),$$

as the minimum Hamming weight amongst all channel packet sequences for which the initial source packet $\mathbf{s}_0$ is non-zero [7], [23]. Note that because $\mathbf{G}_0$ is full-rank, $\mathbf{s}_0 \neq \mathbf{0}$ immediately implies that $\mathbf{x}_0 \neq \mathbf{0}$ as well.

Several properties pertaining to the column Hamming distance were treated in [5], [7]. We summarize two relevant ones below. In Section IV, we prove analogous properties to these for the rank metric.

Property 1 (Tomas et al., [5])**.** Consider an erasure channel being used for each $t \geq 0$, where the prior source sequence $\mathbf{s}_{[0,t-1]}$ is known to the decoder by time $t+j$. If there are at most $d_H(j)-1$ symbol erasures in the window $[t, t+j]$, then $\mathbf{s}_t$ is recoverable by time $t+j$. Conversely, there is at least one hypothetical channel window $[t, t+j]$ containing $d_H(j)$ erasures for which $\mathbf{s}_t$ is not recoverable by time $t+j$.

Property 1 states that for sliding window erasure channels featuring windows of length W, a convolutional code with column Hamming distance $d_H(W-1)$ can guarantee perfect decoding with delay $W-1$, provided that there are less than $d_H(W-1)$ erasures in the window [4].

Property 2 (Gluessing-Luerssen et al., [7])**.** The j-th column Hamming distance of a code is upper bounded by a Singleton-like bound, i.e., $d_H(j) \leq (n-k)(j+1)+1$. If $d_H(j)$ meets this bound with equality, then $d_H(i)$ meets its respective bound for all $i \leq j$.

[2]In network coding literature, each $\mathbf{x}_t \in \mathbb{F}_{q^M}^n$ is referred to as a generation of n channel packets. We denote $\mathbf{x}_t$ as a channel packet containing n symbols and $\mathbf{x}_{[t,t+j]}$ as a sequence of j packets. Similar notation is used for the source.

This property asserts a convolutional code extension of the Singleton bound. The desirability of achieving large column Hamming distances is given in the previous Property 1. In conjunction with this, a code capable of recovering from $d_H(j)-1$ erasures with delay j can be shown to be further capable of the respective maximum recovery for all $i \leq j$.

There exist several families of codes which achieve the maximum $d_H(j)$ for some given j [7], [8], [24]. One such class of codes are m-MDS codes. These achieve the upper bound up to the code memory i.e., $d_H(j) = (n-k)(j+1)+1$ for $0 \leq j \leq m$. The MSR codes, which we introduce in this work, are rank metric analogues of m-MDS codes. One approach to constructing the generator matrix of m-MDS codes is by taking a sub-matrix of $k(m+1)$ rows from a block Toeplitz super-regular matrix [7]. A prior construction of a block Toeplitz super-regular matrix was given in [9]. As this construction is modified for our purposes, we include a summary of this construction in Section V, as well as a review of super-regular matrices in Appendix A.

C. The Active Column Sum Rank Distance

Let $\mathcal{C}[n,k,m]$ be a linear time-invariant convolutional code over $\mathbb{F}_{q^M}$, whose codewords are generated using (2). The *active column sum rank distance* is a metric for convolutional codes that was proposed in a prior work [18]. This metric is defined using the state Trellis graph of the convolutional code. Let $\mathcal{C}_j^a$ be the set of all channel packet sequences $\mathbf{x}_{[0,j]}$ that are constructed by exiting the zero-state of the Trellis at time 0 and not re-entering it for $1 \leq t \leq j-1$. The j-th active column sum rank of a linear convolutional code $\mathcal{C}[n,k,m]$ is then defined as as the minimum sum rank of all channel packet sequences in $\mathcal{C}_j^a$, i.e.,

$$d_R^a(j) \triangleq \min_{\mathbf{x}_{[0,j]} \in \mathcal{C}_j^a} \sum_{t=0}^{j} \operatorname{rank}\left(\phi_n(\mathbf{x}_t)\right),$$

where $\phi_n(\cdot)$ is the previously introduced mapping from vectors in the extension field to matrices in the ground field in Section III-A.

Note that by restricting itself to only consider channel packet sequences in $\mathcal{C}_j^a$, the active column sum rank of a convolutional code does not impose any guarantees on the sum rank of any channel packets that enter the zero state before time j. An example of such a sequence is provided in Fig. 1. Consequently, the active column sum rank is not a sufficient metric to guarantee delay-constrained decoding over a sliding window channel, as we show in the next section.

IV. The Column Sum Rank Distance

Let $\mathcal{C}[n,k,m]$ be a linear time-invariant convolutional code over $\mathbb{F}_{q^M}$, constructed in the same manner as in the previous section. We introduce the j-th column sum rank distance of a code

$$d_R(j) \triangleq \min_{\mathbf{x}_{[0,j]} \in \mathcal{C}, \mathbf{s}_0 \neq 0} \sum_{t=0}^{j} \operatorname{rank}\left(\phi_n(\mathbf{x}_t)\right),$$

as an analogue of the column Hamming distance. Unlike the j-th active column sum rank distance, this metric permits returning to the zero-state before time j. For example, the channel packet sequence generated in Fig. 1 is valid for the column sum rank distance. As a result, this metric is stronger than the active version, i.e., $d_R(j) \leq d_R^a(j)$. In the following theorem, we show that the column sum rank distance of a convolutional code is both necessary and sufficient to guarantee low-delay decoding over $\mathcal{CH}(S,W)$.

Theorem 2. Let $\mathcal{C}[n,k,m]$ be a convolutional code used over the window $[0,W-1]$. For $0 \leq t \leq W-1$, let $\mathbf{A}_t^* \in \mathbb{F}_q^{n \times \rho_t}$ be full-rank matrices and $\mathbf{A}_{[0,W-1]}^* = \operatorname{diag}\left(\mathbf{A}_0^*, \ldots, \mathbf{A}_{W-1}^*\right)$ be a channel matrix. The following statements are true:

1) If $d_R(W-1) > nW - \sum_{t=0}^{W-1} \rho_t$, then $\mathbf{s}_0$ is always recoverable by time $W-1$.
2) If $d_R(W-1) \leq nW - \sum_{t=0}^{W-1} \rho_t$, then there exists at least one channel packet sequence and channel matrix for which $\mathbf{s}_0$ is not recoverable by time $W-1$.

Proof: Due to code linearity, we only need to show that all output channel packet sequences are distinguishable from the all-zero sequence. We prove this by contradiction. Consider a source packet sequence $\mathbf{s}_{[0,W-1]} = (\mathbf{s}_0, \ldots, \mathbf{s}_{W-1})$, where $\mathbf{s}_0 \neq \mathbf{0}$. Suppose that this sequence generates the channel packet sequence $\mathbf{x}_{[0,W-1]}$, for which $\mathbf{x}_{[0,W-1]}\mathbf{A}_{[0,W-1]}^* = \mathbf{0}$. This implies that $\operatorname{rank}\left(\phi_n(\mathbf{x}_t)\right) \leq n - \operatorname{rank}\left(\mathbf{A}_t^*\right)$ for $0 \leq t \leq W-1$. By summing each of the inequalities, we arrive at the following contradiction on the sum rank of the channel packet sequence:

$$\begin{aligned}\sum_{t=0}^{W-1} \operatorname{rank}\left(\phi_n(\mathbf{x}_t)\right) &\leq nW - \sum_{t=0}^{W-1} \rho_t \\ &< d_R(W-1).\end{aligned}$$

For the converse, let $\mathbf{s}_{[0,W-1]} = (\mathbf{s}_0, \ldots, \mathbf{s}_{W-1})$, with $\mathbf{s}_0 \neq \mathbf{0}$ be a source packet sequence that maps to $\mathbf{x}_{[0,W-1]}$, for which $\sum_{t=0}^{W-1} \operatorname{rank}\left(\phi_n(\mathbf{x}_t)\right) = d_R(W-1)$. For $0 \leq t \leq W-1$, let $\rho_t = n - \operatorname{rank}\left(\phi_n(\mathbf{x}_t)\right)$. There exist matrices $\mathbf{A}_t^* \in \mathbb{F}_q^{n \times \rho_t}$ such that each $\mathbf{x}_t\mathbf{A}_t^* = \mathbf{0}$. We let $\mathbf{A}_{[0,W-1]}^* = \operatorname{diag}\left(\mathbf{A}_0^*, \ldots, \mathbf{A}_{W-1}^*\right)$ be the channel matrix. Summing all of the ρ_t reveals $\operatorname{rank}\left(\mathbf{A}_{[0,W-1]}^*\right) = nW - d_R(W-1)$. Furthermore, $\mathbf{s}_{[0,W-1]}$ is indistinguishable from the all-zero source packet sequence over this channel. ∎

Remark 2. The constraint that $\mathbf{s}_0 \neq \mathbf{0}$ is necessary in order to differentiate the first source packet from the all-zero source packet. Note however, that there is no necessary constraint on the state Trellis transitions; any non-zero source packet sequence should be differentiable from the all-zero sequence. Using the active column sum rank in place of the column sum rank in the above Theorem then leads to a partial guarantee. If $\mathbf{x}_{[0,W-1]} \in \mathcal{C}_{W-1}^a$, then the active column sum rank determines the maximum rank deficiency in the channel from which the first source packet is recoverable. However, if $\mathbf{x}_{[0,W-1]} \notin \mathcal{C}_{W-1}^a$, then the $W-1$-th active column sum rank does not provide any guarantees on recoverability. Consequently, we view the active column sum rank as an over-estimate of the recovery capability of the code.

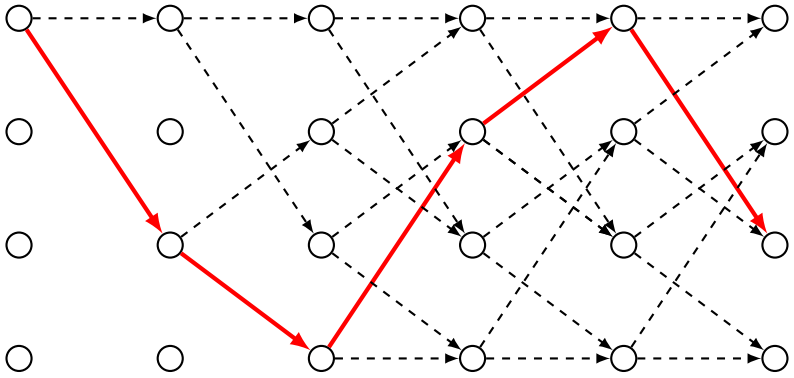


Fig. 1: A hypothetical transition along the state Trellis graph is highlighted. The active column sum rank distance $d_R^a(5)$ does not guarantee the sum rank of the channel packet sequence generated by this path, i.e., $\mathbf{x}_{[0,5]} \notin \mathcal{C}_5^a$. In this work, we introduce the column sum rank, which does consider this channel packet sequence.

For time-invariant encoders, Theorem 2 can be used to guarantee that all source packets are recovered with delay at most $W-1$ over a sliding window channel. Assuming all prior packets have been decoded, we recover each $\mathbf{s}_t$ using the window $[t, t+W-1]$. The contributions of $\mathbf{s}_{[0,t-1]}$ can be negated from the received packet sequence used for decoding. Theorem 2 is then effectively a rank metric analogue to Property 1 from Section III-B, which describes how column Hamming distance bounds the number of tolerable erasures in single-link streaming [5].

Remark 3. Aside from rank-deficient channel matrices, adversarial errors can also be considered using rank metric codes. Consider a single-link single-shot system where the receiver observes $\mathbf{y} = \mathbf{x} + \mathbf{e}$, with $\mathbf{e} \in \mathbb{F}_{q^M}^n$ being an additive error vector. If $\operatorname{rank}(\phi_n(\mathbf{e})) \leq \frac{d_R-1}{2}$, the decoder for an MRD code can recover the source [13]. MRD codes reflect a rank analogue of the error correcting capability of MDS codes. It can easily be shown that the column sum rank can ensure $\mathbf{x}_0$ is recoverable by time j in the channel $\mathbf{y}_{[0,j]} = \mathbf{x}_{[0,j]} + \mathbf{e}_{[0,j]}$, if the sum rank of $\mathbf{e}_{[0,j]} \in \mathbb{F}_{q^M}^{n(j+1)}$ is constrained to be at most $\frac{d_R(j)-1}{2}$. The proof follows similarly to that for Theorem 2.

We next propose an analogue to Property 2 from Section III-B. First, we bound the column sum rank by $d_R(j) \leq (n-k)(j+1)+1$. The sum rank of a channel packet cannot exceed its Hamming weight, meaning that the upper bound on column Hamming distance is inherited by the rank metric analogue. Furthermore, we show that if the j-th column sum rank achieves its upper bound, all prior column sum ranks do so as well for their respective bounds.

Lemma 1. If $d_R(j) = (n-k)(j+1)+1$, then $d_R(i) = (n-k)(i+1)+1$ for all $i \leq j$.

Proof: It suffices to prove for $i = j-1$. Let $\mathcal{C}[n,k,m]$ be a code for which $d_R(j-1)$ does not attain the upper bound i.e., $d_R(j-1) \leq (n-k)j$, but $d_R(j)$ achieves the maximum i.e., $d_R(j) = (n-k)(j+1)+1$. We will argue by contradiction that such a code cannot exist.

Consider a source packet sequence $\mathbf{s}_{[0,j-1]}$ that generates $\mathbf{x}_{[0,j-1]}$ whose sum rank equal to $d_R(j-1)$ i.e., $\sum_{t=0}^{j-1} \operatorname{rank}(\phi_n(\mathbf{x}_t)) = d_R(j-1)$ holds. We argue that this sequence can be augmented to include $\mathbf{x}_j$, such that $\sum_{t=0}^{j} \operatorname{rank}(\phi_n(\mathbf{x}_t)) \leq (n-k)(j+1) < d_R(j)$ holds. This will complete the contradiction.

To exhibit such a choice of $\mathbf{x}_j$ recall that $\mathbf{x}_j = \sum_{t=0}^{j-1} \mathbf{s}_t \mathbf{G}_{j-t} + \mathbf{s}_j \mathbf{G}_0$. The summation up to $j-1$ produces a vector whose Hamming weight is at most n. Because $\operatorname{rank}(\mathbf{G}_0) = k$, the source packet $\mathbf{s}_j$ can be selected specifically in order to negate up to k non-zero entries of the first summation. This implies that $\operatorname{wt}_H(\mathbf{x}_j) \leq n-k$ and consequently, $\operatorname{rank}(\phi_n(\mathbf{x}_j)) \leq n-k$. Therefore, we bound the sum rank of $\mathbf{x}_{[0,j]}$

$$\begin{aligned}\sum_{t=0}^{j} \operatorname{rank}(\phi_n(\mathbf{x}_t)) &= d_R(j-1) + \operatorname{rank}(\phi_n(\mathbf{x}_j)) \\ &\leq d_R(j-1) + n - k \\ &\leq (n-k)(j+1),\end{aligned}$$

as required. ∎

Codes achieving the Singleton bound for $d_R(m)$ are referred to as MSR codes. They directly parallel m-MDS codes, which maximize the m-th column Hamming distance [7]. In fact, since $d_R(j) \leq d_H(j)$, MSR codes automatically maximize the column Hamming distance and can be seen as a special case of m-MDS codes.

By Theorem 2, an MSR code with memory $T = W-1$ recovers source packets with delay T, when the rank of the channel matrix is at least $k(T+1)$ in each sliding window of length W. We prove the existence of MSR codes in the next section, but first discuss a matrix multiplication property for the generator matrix. The following theorem serves as an extension of Theorem 1 to convolutional codes transmitted over independent network uses.

Theorem 3. For $0 \leq t \leq j$, let $0 \leq \rho_t \leq n$ satisfy

$$\sum_{i=0}^{t} \rho_i \leq k(t+1) \tag{3}$$

for all $t \leq j$ and with equality for $t = j$. The following are equivalent for any convolutional code:

1) $d_R(j) = (n-k)(j+1)+1$

2) For all full-rank $\mathbf{A}^*_{[0,j]} = \text{diag}\left(\mathbf{A}^*_0, \ldots, \mathbf{A}^*_j\right)$ constructed from full-rank blocks $\mathbf{A}^*_t \in \mathbb{F}_q^{n\times\rho_t}$ and ρ_t that satisfy (3), the product $\mathbf{G}^{\text{EX}}_j \mathbf{A}^*_{[0,j]}$ is non-singular.

Proof: We first prove $1 \Rightarrow 2$. Suppose there exists an $\mathbf{A}^*_{[0,j]}$ whose blocks satisfy (3), for which $\mathbf{G}^{\text{EX}}_j \mathbf{A}^*_{[0,j]}$ is singular. Then there exists a channel packet sequence $\mathbf{x}_{[0,j]}$, where $\mathbf{x}_{[0,j]}\mathbf{A}^*_{[0,j]} = \mathbf{0}$. We show that this leads to a contradiction of 1. The contradiction is immediate if $\mathbf{x}_0 \neq \mathbf{0}$. In this case the sum rank of $\mathbf{x}_{[0,j]}$ is at least $d_R(j)$, i.e., $\sum_{t=0}^{j} \text{rank}\left(\phi_n(\mathbf{x}_t)\right) \geq d_R(j)$ must hold. Note however that:

$$\begin{aligned}\sum_{t=0}^{j} \text{rank}\left(\phi_n(\mathbf{x}_t)\right) &\leq n(j+1) - \sum_{t=0}^{j} \rho_t \\ &= (n-k)(j+1)\end{aligned}$$

contradicts $d_R(j) = (n-k)(j+1)+1$. Note that we use (3) with equality at $t = j$ in the second step.

If $\mathbf{x}_0 = \mathbf{0}$, then the sum rank of $\mathbf{x}_{[0,j]}$ is not constrained by $d_R(j)$. Let $l = \arg\min_t \mathbf{x}_t \neq \mathbf{0}$ be the smallest index for which $\mathbf{x}_l$ is non-zero and consider the channel packet sequence $\mathbf{x}_{[l,j]}$, whose sum rank is at least $d_R(j-l)$. Because $\mathbf{x}_t\mathbf{A}^*_t = \mathbf{0}$ for $t = l, \ldots, j$, we bound $\text{rank}\left(\phi_n(\mathbf{x}_t)\right) \leq n - \rho_t$ in this window. The sum rank of $\mathbf{x}_{[l,j]}$ is bounded:

$$\begin{aligned}\sum_{t=l}^{j} \text{rank}\left(\phi_n(\mathbf{x}_t)\right) &\leq n(j-l+1) - \sum_{t=l}^{j} \rho_t \\ &\leq (n-k)(j-l+1).\end{aligned}$$

The second line follows from $\sum_{t=l}^{j} \rho_t \geq k(j-l+1)$, which can be derived when (3) is met with equality for $t = j$. Due to Lemma 1, the column sum rank achieves $d_R(j-l) = (n-k)(j-l+1)+1$. The sum rank of $\mathbf{x}_{[l,j]}$ is less than $d_R(j-l)$, which is a contradiction.

We prove $2 \Rightarrow 1$ by using a code with $d_R(j) \leq (n-k)(j+1)$ and constructing a full-rank $\mathbf{A}^*_{[0,j]}$ for which $\mathbf{G}^{\text{EX}}_j \mathbf{A}^*_{[0,j]}$ is singular. Let $m = \arg\min_i d_R(i) \leq (n-k)(i+1)$ be the first instance where the column sum rank fails to attain its upper bound and consider the sequence $\mathbf{x}_{[0,m]}$ with the minimum column sum rank. We show that there exist full-rank matrices $\mathbf{A}^*_t \in \mathbb{F}_q^{n\times\rho_t}$ satisfying both (3) and $\mathbf{x}_t\mathbf{A}^*_t = \mathbf{0}$ for $0 \leq t \leq m$. In addition, we aim to have equality in (3) at $t = m$ and thus $\mathbf{A}^*_{[0,m]}$ will be of dimension $(m+1)n \times (m+1)k$. This is relevant later in the proof.

When $m = 0$, the column rank of $\mathbf{x}_0$ cannot exceed $n-k$. For every $\rho_0 \leq n - \text{rank}\left(\phi_n(\mathbf{x}_0)\right)$, there exists an $\mathbf{A}^*_0$ for which $\mathbf{x}_0\mathbf{A}^*_0 = \mathbf{0}$. Clearly we can always choose an $\mathbf{A}^*_0$ with rank $\rho_0 = k$.

When $m > 0$, the sum rank of $\mathbf{x}_{[0,t]}$ satisfies

$$\sum_{i=0}^{t} \text{rank}(\phi_n(\mathbf{x}_i)) \geq (n-k)(t+1)+1 \tag{4}$$

for $0 \leq t \leq m-1$, and

$$\sum_{i=0}^{m} \text{rank}(\phi_n(\mathbf{x}_i)) \leq (n-k)(m+1). \tag{5}$$

Let $\rho_t = n - \text{rank}\left(\phi_n(\mathbf{x}_t)\right)$ for $0 \leq t \leq m-1$ and choose the appropriate $\mathbf{A}^*_t$ for which $\mathbf{x}_t\mathbf{A}^*_t = \mathbf{0}$. For all $0 \leq t \leq m-1$, we have that:

$$\begin{aligned}\sum_{i=0}^{t} \rho_i &= n(t+1) - \sum_{i=0}^{t} \text{rank}\left(\phi_n(\mathbf{x}_i)\right) &(6)\\ &\leq k(t+1) - 1, &(7)\end{aligned}$$

confirming that (3) is satisfied for $t \leq m-1$. Note that in (7) we apply the inequality in (4). We next specify an appropriate choice for ρ_m. We will select:

$$\rho_m = (m+1)k - \sum_{t=0}^{m-1} \rho_t, \tag{8}$$

and show that there exists an associated $\mathbf{A}^*_m \in \mathbb{F}_q^{n\times\rho_m}$ that will satisfy $\mathbf{x}_m\mathbf{A}^*_m = \mathbf{0}$. It thus suffices to show that ρ_m also satisfies $\rho_m \leq n - \text{rank}(\phi_n(\mathbf{x}_m))$. Note that

$$\begin{aligned}n - \rho_m &= n - (m+1)k + \sum_{i=0}^{m-1} \rho_i &(9)\\ &= (n-k)(m+1) - \sum_{i=0}^{m-1} \text{rank}\left(\phi_n(\mathbf{x}_i)\right) &(10)\\ &= (n-k)(m+1) - \sum_{i=0}^{m} \text{rank}\left(\phi_n(\mathbf{x}_i)\right) \\ &\quad + \text{rank}\left(\phi_n(\mathbf{x}_m)\right) &(11)\\ &\geq \text{rank}\left(\phi_n(\mathbf{x}_m)\right), &(12)\end{aligned}$$

where (10) follows by using (6) with $t = m-1$ and (12) follows via the inequality in (5). Finally, our choice (8) also guarantees that $\mathbf{A}^*_{[0,m]}$ has dimension $(m+1)n \times (m+1)k$ as claimed.

The remaining $\mathbf{A}^*_{m+1}, \ldots, \mathbf{A}^*_j$ can be any full-rank $n \times k$ matrices, thus satisfying (3) for all $t \leq j$. The product $\mathbf{G}^{\text{EX}}_j \mathbf{A}^*_{[0,j]}$ can be written as

$$\begin{aligned}\mathbf{G}^{\text{EX}}_j \mathbf{A}^*_{[0,j]} &= \begin{pmatrix}\mathbf{G}^{\text{EX}}_m & \mathbf{X} \\ & \mathbf{Y}\end{pmatrix}\begin{pmatrix}\mathbf{A}^*_{[0,m]} & \\ & \mathbf{A}^*_{[m+1,j]}\end{pmatrix} \\ &= \begin{pmatrix}\mathbf{G}^{\text{EX}}_m \mathbf{A}^*_{[0,m]} & \mathbf{X}\mathbf{A}^*_{[m+1,j]} \\ & \mathbf{Y}\mathbf{A}^*_{[m+1,j]}\end{pmatrix}\end{aligned}$$

where $\mathbf{X}$ and $\mathbf{Y}$ denote the remaining blocks that comprise $\mathbf{G}^{\text{EX}}_j$. The block $\mathbf{G}^{\text{EX}}_m \mathbf{A}^*_{[0,m]}$ is a square matrix with a zero determinant. Therefore, $\det \mathbf{G}^{\text{EX}}_j \mathbf{A}^*_{[0,j]}$ is also zero. ∎

Although in this work, we do not propose a general decoding algorithm for MSR codes, we remark that decoding in the network streaming problem can be reduced to matrix inversion. Consider a scenario where an MSR code with memory $T = W-1$ is used and that all source packets before time t have been recovered. To ensure that $\mathbf{s}_t$ is recoverable within its deadline of time $t+T$, we consider increasingly larger windows $[t, t+j]$ for $0 \leq j \leq T$. Theorem 2 states that if $\sum_{i=t}^{t+j} \rho_i < k(t+j+1)$, then the decoder cannot guarantee recovery by time $t+j$. The window length j is incremented at each time instance up to the first point where $\sum_{i=t}^{t+j} \rho_i \geq k(t+j+1)$ is achieved. By Theorem 3, the rank conditions in (3) are satisfied and $\mathbf{G}^{\text{EX}}_j \mathbf{A}^*_{[t,t+j]}$

is non-singular. We therefore invert this matrix and solve for $\mathbf{s}_{[t,t+j]}$, recovering all packets in the window simultaneously. Consequently, packets encoded by an MSR code can be recovered for the network streaming problem with complexity $\mathcal{O}((jk)^3)$.

In the next section, we construct an extended generator matrix. Theorem 3 is then useful afterwards to verify that the generator matrix does in fact define an MSR code.

V. Preservation of Super-regularity

In [7], the authors provided a construction for a Toeplitz super-regular matrix that exists in $\mathbb{F}_q$ for a prime q. From (2), the generator matrix of a convolutional code is block Toeplitz and we focus on super-regular matrices with similar structure. For simplicity, we consider block Hankel matrices and later convert the structure to block Toeplitz before code construction.

A. Block Hankel Super-regular Matrices

A block Hankel super-regular matrix construction, which we outline below, was proposed in [9] for $\mathbb{F}_{q^M}$ where q is a prime power and M is sufficiently large.

Theorem 4 (Almeida et al., [9])**.** For $n, m \in \mathbb{N}$, let $M = q^{n(m+2)-1}$. Let $\alpha \in \mathbb{F}_{q^M}$ be a primitive element and a root of the minimal polynomial $p_\alpha(X)$. For $0 \leq j \leq m$, let the blocks $\mathbf{T}_j \in \mathbb{F}_{q^M}^{n\times n}$ be defined by

$$\mathbf{T}_j = \begin{pmatrix} \alpha^{[nj]} & \alpha^{[nj+1]} & \dots & \alpha^{[n(j+1)-1]} \\ \alpha^{[nj+1]} & \alpha^{[nj+2]} & \dots & \alpha^{[n(j+1)]} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{[n(j+1)-1]} & \alpha^{[n(j+1)]} & \dots & \alpha^{[n(j+2)-2]} \end{pmatrix}. \quad (13)$$

Then the following block Hankel matrix

$$\mathbf{T} = \begin{pmatrix} & & & \mathbf{T}_0 \\ & & \mathbf{T}_0 & \mathbf{T}_1 \\ & \iddots & \vdots & \vdots \\ \mathbf{T}_0 & \dots & \mathbf{T}_{m-1} & \mathbf{T}_m \end{pmatrix} \quad (14)$$

is super-regular.

We refer the reader to [9] for the full proof. In this subsection however, several properties of $\mathbf{T}$ that lead to its super-regularity are highlighted. They will be useful in the subsequent section.

Let $T_j(r, s)$ be the element in the r-th row and s-th column of $\mathbf{T}_j$, where $0 \leq r, s \leq n-1$ and $0 \leq j \leq m$. Each entry of the matrix is a linearized monomial evaluated at $X = \alpha$, in particular [3],

$$T_j(r, s) = \alpha^{[nj+r+s]}. \quad (15)$$

The q-degrees of these monomials increase as one moves further down and to the right inside any $\mathbf{T}_j$. This can be described with the following inequalities:

$$\deg_{\mathrm{q}} T_j(r', s) > \deg_{\mathrm{q}} T_j(r, s), \quad 0 \leq r < r' \leq n-1, \quad (16a)$$
$$\deg_{\mathrm{q}} T_j(r, s') > \deg_{\mathrm{q}} T_j(r, s), \quad 0 \leq s < s' \leq n-1, \quad (16b)$$

for all $0 \leq j \leq m$.

Combining all of the $\mathbf{T}_j$ to construct $\mathbf{T}$ in (14) preserves the property of q-degrees strictly increasing downwards (along a single column) or rightwards (along a single row) now for the block Hankel matrix, i.e.,

$$\deg_{\mathrm{q}} T_{j'}(r', s) > \deg_{\mathrm{q}} T_j(r, s), \quad 0 \leq j < j' \leq m, \quad (17a)$$
$$\deg_{\mathrm{q}} T_{j'}(r, s') > \deg_{\mathrm{q}} T_j(r, s), \quad 0 \leq j < j' \leq m, \quad (17b)$$

for all $0 \leq r, r' \leq n-1$ and $0 \leq s, s' \leq n-1$.

To show that $\mathbf{T}$ is super-regular, let $\mathbf{D}$ be any $l \times l$ sub-matrix of $\mathbf{T}$. All sub-matrices of $\mathbf{T}$ preserve the degree inequalities in (16) and (17). $\mathbf{T}$ being super-regular is equivalent to $\det \mathbf{D} \neq 0$ for every $\mathbf{D}$ with a non-trivial determinant.

The determinant is evaluated using the Leibniz formula as a polynomial $D(\alpha) = \sum_{\sigma \in S_r} \operatorname{sgn}(\sigma) D_\sigma$. Each non-zero term in the formula, D_σ, is a product of linearized monomials

$$D_\sigma = \prod_{i=0}^{l-1} D(i, \sigma(i)).$$

Note that the determinant polynomial itself is not linearized. Using (16) and (17), the degree of this polynomial can be bounded.

Lemma 2 (Almeida et al., [9])**.** For $\mathbf{T}$ defined in (14), let $\mathbf{D}$ be any square sub-matrix with a non-trivial determinant (see Appendix A-B). Let $D(X)$ be the polynomial, such that $D(\alpha) = \det \mathbf{D}$. The degree of $D(X)$ is bounded by

$$1 \leq \deg D(X) < q^{n(m+2)-1}.$$

In [9], the matrices defined by (13) contained entries with Frobenius powers with fixed $q = 2$, rather than for a general q. A direct extension reveals that the bounds in Lemma 2 still holds when using Frobenius powers with arbitrary q in (13) provided that $M \geq q^{n(m+2)-1}$. We refer the reader to [9], [25] for the full proof.

An arbitrary q and appropriate choice of α permits the super-regular blocks $\mathbf{T}_j$ to resemble Gabidulin code generator matrices. We require this slight generalization since our channel in Section II can operate over any field $\mathbb{F}_q$.

The upper bound gives $\deg D(X) < \deg p_\alpha(X)$, which implies that, $p_\alpha(X) \nmid D(X)$. Consequently, α cannot be a root of $D(X)$ (see Lemma 4 of Appendix A).

The lower bound in our extension does not change from the original. It implies that $D(X)$ is not the zero polynomial and can be derived by algorithmically finding a unique permutation $\bar{\sigma} = \arg\max_\sigma \deg D_\sigma(X)$, which generates the highest degree monomial term in the Leibniz formula. Because $\bar{\sigma}$ is unique, there is no other term which can negate this monomial. Then, $\deg D(X) = \deg D_{\bar{\sigma}}(X)$, meaning $D(X)$ is not the zero polynomial.

The upper and lower bounds cover both potential cases to ensure $D(\alpha) \neq 0$. Therefore, any $\mathbf{D}$ with a non-trivial determinant is also non-singular and consequently, $\mathbf{T}$ is a

[3]We use the variable X when discussing properties of linearized polynomials and evaluate the polynomials at $X = \alpha$ specifically when calculating the determinant of a matrix. However, we suppress the dependence on X whenever it can be determined from the context.

super-regular matrix.

A block Toeplitz or Hankel super-regular matrix can be used to construct the extended generator matrix of an m-MDS code [8]. We will use a similar technique to construct MSR codes. Super-regularity alone however, is insufficient for the network streaming problem. We show that the super-regular matrix construction technique above can be modified to produce the desired property for the network streaming problem.

B. Preservation of Super-regularity after Multiplication

In this section, we connect the matrix multiplication property in Theorem 3 to super-regularity. The block Hankel matrix $\mathbf{T}$ from (14) is constructed using an α that is both a primitive and a normal element of the field $\mathbb{F}_{q^M}$. We show in this case that the super-regularity of $\mathbf{T}$ is preserved after multiplication with any block diagonal matrix in the ground field.

Theorem 5. For $0 \leq t \leq m$, let $\mathbf{A}_t \in \mathbb{F}_q^{n\times n}$ be any non-singular matrices. We then construct $\mathbf{A}_{[0,m]} = \operatorname{diag}(\mathbf{A}_0, \ldots, \mathbf{A}_m)$. Let $\mathbf{T} \in \mathbb{F}_{q^M}^{n(m+1)\times n(m+1)}$ be the super-regular matrix in (14). If $M \geq q^{n(m+2)-1}$ and α is a primitive normal element of $\mathbb{F}_{q^M}$, then $\mathbf{F} = \mathbf{T}\mathbf{A}_{[0,m]}$ is super-regular.

Several properties of $\mathbf{F}$ are introduced before proving Theorem 5. To simplify the notation, we assume that $\mathbf{A}_0 = \mathbf{A}_1 = \cdots = \mathbf{A}_m = \mathbf{A}$. This allows free use of the polynomial structures and bounds from the previous subsection. We will then explain why the proof extends immediately to the general case. Hence, $\mathbf{A}_{[0,m]} = \operatorname{diag}(\mathbf{A}, \ldots, \mathbf{A})$, where $\mathbf{A} \in \mathbb{F}_q^{n\times n}$ is a non-singular matrix in the ground field and we consider properties of the product

$$\mathbf{F} = \mathbf{T}\mathbf{A}_{[0,m]}.$$

Property 3. The product $\mathbf{F} = \mathbf{T}\mathbf{A}_{[0,m]}$ is a block Hankel matrix, whose blocks each possess a Moore structure.

The product can be written as

$$\mathbf{F} = \mathbf{T}\mathbf{A}_{[0,m]} = \begin{pmatrix} & & & \mathbf{T}_0\mathbf{A} \\ & & \mathbf{T}_0\mathbf{A} & \mathbf{T}_1\mathbf{A} \\ & \iddots & \vdots & \vdots \\ \mathbf{T}_0\mathbf{A} & \cdots & \mathbf{T}_{m-1}\mathbf{A} & \mathbf{T}_m\mathbf{A} \end{pmatrix}.$$

Let $\mathbf{F}_j = \mathbf{T}_j\mathbf{A}$ for $j = 0, \ldots, m$. To show that $\mathbf{F}_j$ has a Moore structure, let $(f_0, \ldots, f_{n-1})$ be the first row of $\mathbf{F}_0$, i.e.,

$$(f_0, \ldots, f_{n-1}) = (\alpha^{[0]}, \ldots, \alpha^{[n-1]})\mathbf{A}.$$

Each element f_i is a linearized polynomial $f_i(X)$ evaluated at $X = \alpha$, with coefficients from $\mathbf{A} = [A_{l,i}]$, i.e.,

$$f_i = \sum_{l=0}^{n-1} A_{l,i}\alpha^{[l]}. \tag{18}$$

Since $\alpha \in \mathbb{F}_{q^M}$ and $A_{l,i} \in \mathbb{F}_q$, invoking Freshman's rule

$$f_i(X^{[s]}) = f_i^{[s]}(X), \quad s \geq 0, \tag{19}$$

results in the following Moore structure,

$$\mathbf{F}_j = \begin{pmatrix} f_0^{[nj]} & f_1^{[nj]} & \cdots & f_{n-1}^{[nj]} \\ f_0^{[nj+1]} & f_1^{[nj+1]} & \cdots & f_{n-1}^{[nj+1]} \\ \vdots & \vdots & \ddots & \vdots \\ f_0^{[n(j+1)-1]} & f_1^{[n(j+1)-1]} & \cdots & f_{n-1}^{[n(j+1)-1]} \end{pmatrix} \tag{20}$$

for $0 \leq j \leq m$.

Remark 4. Since α is normal over $\mathbb{F}_{q^M}$ and $\mathbf{A}$ is full rank, $(f_0, \ldots, f_{n-1})$ are linearly independent over $\mathbb{F}_q$ (see [13]).

Property 4. The q-degrees of the polynomial elements of $\mathbf{F}_j$ strictly increase downwards on any fixed column. They are not necessarily monotonic across a fixed row.

From (18) and (19), the q-degree of each linearized polynomial is bounded by

$$s \leq \deg_{\mathrm{q}} f_i^{[s]}(X) \leq n - 1 + s \tag{21}$$

for $0 \leq s \leq n(m+1) - 1$. Note that both the upper and lower bounds in (21) do not depend on the column index i. Furthermore all elements on any fixed row in $\mathbf{F}_j$ share the same degree (c.f. (20)). Consequently, the polynomial entries on any fixed row of the block $\mathbf{F}_j$ share the same bound and are not necessarily in an increasing order. Thus we do not have a counterpart of (16b) that guarantees that the q-degrees of elements are monotonically increasing across each row in $\mathbf{T}_j$.

The counterpart of (16a) is satisfied however. Let $F_j(r, s)$ be the element in r-th row and s-th column of $\mathbf{F}_j$, i.e.,

$$F_j(r, s) = f_s^{[nj+r]}, \tag{22}$$

where $0 \leq r, s \leq n-1$ and $0 \leq j \leq m$. The Moore structure of $\mathbf{F}_j$ implies that the q-degrees of the polynomial entries strictly increase from top to bottom along any fixed column of $\mathbf{F}_j$, i.e.,

$$\deg_{\mathrm{q}} F_j(r', s) > \deg_{\mathrm{q}} F_j(r, s), \quad 0 \leq r < r' \leq n-1, \tag{23}$$

for all $0 \leq j \leq m$.

The following lemma establishes the relations between q-degrees of the polynomial entries inside $\mathbf{F}$.

Property 5. The following inequalities

$$\deg_{\mathrm{q}} F_{j'}(r', s) > \deg_{\mathrm{q}} F_j(r, s), \quad 0 \leq j < j' \leq m, \tag{24a}$$
$$\deg_{\mathrm{q}} F_{j'}(r, s') > \deg_{\mathrm{q}} F_j(r, s), \quad 0 \leq j < j' \leq m, \tag{24b}$$

hold for all $0 \leq r, r' \leq n-1$ and $0 \leq s, s' \leq n-1$.

Proof: The inequality in (24a) is established by

$$\begin{aligned} &\deg_{\mathrm{q}} F_{j'}(r', s) - \deg_{\mathrm{q}} F_j(r, s) \\ &\quad = \deg_{\mathrm{q}} f_s^{[nj'+r']} - \deg_{\mathrm{q}} f_s^{[nj+r]} \\ &\quad = (nj' + r') + \deg_{\mathrm{q}} f_s - (nj + r) - \deg_{\mathrm{q}} f_s \\ &\quad \geq n + r' - r \\ &\quad > 0, \end{aligned} \tag{25}$$

where the first inequality follows from $j' \geq j + 1$, and the second due to $r' - r > -n$.

Using (21), we prove (24b),

$$\begin{aligned}\deg_{\mathrm{q}} F_{j'}(r,s') - \deg_{\mathrm{q}} F_j(r,s) &= \deg_{\mathrm{q}} f_s^{[nj'+r]} - \deg_{\mathrm{q}} f_s^{[nj+r]} \\ &\geq nj' + r - (n-1+nj+r) \\ &> 0. \end{aligned} \tag{26}$$

Here, the first inequality follows from (21), and the second due to $j' \geq j+1$. ∎

We note that (23), (24a) and (24b) parallel the inequalities (16a), (17a) and (17b) for entries of $\mathbf{T}$. It remains to find an equivalent relationship of (16b) for the entries of $\mathbf{F}$.

To state the next property we let $\mathbf{C}_i \in \mathbb{F}_{q^M}^{n(m+1)\times n}$ and $\mathbf{R}_i \in \mathbb{F}_{q^M}^{n\times n(m+1)}$ be the i-th column and row blocks of $\mathbf{F}$, i.e.,

$$\mathbf{F} = \big(\mathbf{C}_0, \mathbf{C}_1, \cdots, \mathbf{C}_m\big) = \begin{pmatrix} \mathbf{R}_0 \\ \mathbf{R}_1 \\ \vdots \\ \mathbf{R}_m \end{pmatrix}, \tag{27}$$

where

$$\mathbf{C}_i \triangleq \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{F}_0 \\ \mathbf{F}_1 \\ \vdots \\ \mathbf{F}_i \end{pmatrix} \qquad \mathbf{R}_i \triangleq \big(\mathbf{0}, \cdots, \mathbf{0}, \mathbf{F}_0, \mathbf{F}_1, \cdots, \mathbf{F}_i\big). \tag{28}$$

The authors in [9] showed the following property.

Property 6. Let $\mathbf{D}$ be an $l \times l$ sub-matrix of $\mathbf{F}$ possessing a non-trivial determinant. The matrix $\mathbf{D}$ can be written as

$$\mathbf{D} = \left(\begin{array}{c|c|c} \mathbf{O}_h & & \\ \hline \mathbf{O}_{h-1} & & \\ \hline \vdots & \mathbf{D}_{h-1} & \mathbf{D}_h \\ \hline \mathbf{O}_0 & & \\ \hline \mathbf{D}_0 & \cdots & \end{array}\right), \tag{29}$$

where $0 \leq h \leq m$. Each $\mathbf{O}_i$ is a zero matrix drawn from a single row block of $\mathbf{F}$, whereas each $\mathbf{D}_i$ is a matrix containing non-zero polynomials drawn from a single column block of $\mathbf{F}$. Let $l_i \in \{1, \ldots, n\}$ be the number of columns in each $\mathbf{D}_i$ for $i \in \{0, \ldots, h\}$. Then, $\sum_{i=0}^{h} l_i = l$.

Remark 5. In the special case of $h = m$, $\mathbf{D}_i$ consists of columns drawn from the non-zero elements of $\mathbf{C}_i$, whereas $\mathbf{O}_i$ consists of rows drawn from the zero elements of $\mathbf{R}_{m-i}$. When $h < m$, it can be shown that $\mathbf{D}$ preserves the structure in (29). If the j-th column block $\mathbf{C}_j$ of $\mathbf{F}$ was skipped when generating the submatrix $\mathbf{D}$, one has to skip the $(m-j)$-th row block $\mathbf{R}_{m-j}$ to avoid generating a trivial sub-matrix $\mathbf{D}$. We refer the reader to [9] for the complete proof.

Remark 6. The degrees of the polynomials entries of $\mathbf{D}$ satisfy (23), (24a) and (24b).

Lemma 3. For $1 \leq l \leq n$, let $i_1, i_2, \ldots, i_l$ be l distinct elements of the set $\{0, \ldots, n-1\}$. There exists a full-rank matrix $\mathbf{M} \in \mathbb{F}_q^{l\times l}$, such that the q-degrees of

$$(\hat{f}_{i_1}, \hat{f}_{i_2}, \ldots, \hat{f}_{i_l}) = (f_{i_1}, f_{i_2}, \ldots, f_{i_l})\mathbf{M}$$

are monotonically increasing. Moreover, the q-degrees of the polynomial entries of the product $(f_{i_1}^{[s]}, f_{i_2}^{[s]}, \ldots, f_{i_l}^{[s]})\mathbf{M}$ are also sorted in the same order, for all $0 \leq s \leq n(m+1)-1$.

Proof: As discussed in Remark 4, because α is a normal element of $\mathbb{F}_{q^M}$ and $\mathbf{A}$ is full rank, the polynomials $(f_{i_1}, f_{i_2}, \ldots, f_{i_l})$ are linearly independent over $\mathbb{F}_q$. As a result, the isomorphic matrix

$$\big(\phi_n(f_{i_1}), \phi_n(f_{i_2}), \ldots, \phi_n(f_{i_l})\big)$$

is full column rank and can be transformed to reduced column echelon form $\big(\phi(\hat{f}_{i_1}), \phi(\hat{f}_{i_2}), \ldots, \phi(\hat{f}_{i_l})\big)$ through elementary column operations, i.e.,

$$\begin{aligned}&\big(\phi_n(\hat{f}_{i_1}), \phi_n(\hat{f}_{i_2}), \ldots, \phi_n(\hat{f}_{i_l})\big) \\ &\qquad = \big(\phi_n(f_{i_1}), \phi_n(f_{i_2}), \ldots, \phi_n(f_{i_l})\big)\,\mathbf{M},\end{aligned}$$

where $\mathbf{M} \in \mathbb{F}_q^{l\times l}$ is full rank matrix in the ground field. This is equivalent to saying that the degrees of the polynomials $(\hat{f}_{i_1}, \ldots, \hat{f}_{i_l})$ are strictly increasing.

The second part immediately follows, as $(f_{i_1}^{[s]}, f_{i_2}^{[s]}, \ldots, f_{i_l}^{[s]})\mathbf{M} = (\hat{f}_{i_1}^{[s]}, \hat{f}_{i_2}^{[s]}, \ldots, \hat{f}_{i_l}^{[s]})$ by using Freshman's rule. ∎

C. Proof of Theorem 5

Having established the various properties in the previous section we are now in a position to complete the proof of Theorem 5.

Proof of Theorem 5: Consider any nontrivial minor $\mathbf{D}$ of $\mathbf{F}$. We will show that $\det(\mathbf{D}) \neq 0$. First note from Property 4 and 5 the elements of $\mathbf{D}$ satisfy (23), (24a) and (24b), which are the counterparts of (16a), (17a) and (17b) respectively. We will show that $\mathbf{D}$ can be transformed into a matrix $\hat{\mathbf{D}}$ that satisfies the counterpart of (16b).

Using Property 6 we have that the structure of $\mathbf{D}$ should satisfy (29). Consider the sub-matrices $\mathbf{D}_0, \ldots, \mathbf{D}_h$ associated with $\mathbf{D}$. Each $\mathbf{D}_j$ is contained entirely within a column block of $\mathbf{F}$. Thus it follows from (20) that any row of $\mathbf{D}_j$ can be expressed in the form $(f_{i_1}^{[s]}, f_{i_2}^{[s]}, \ldots, f_{i_{l_j}}^{[s]})$ for some $0 \leq s \leq n(m+1)-1$ and $0 \leq i_1 < i_2 < \cdots < i_{l_j} \leq n-1$. In particular the indices $i_1, \ldots, i_{l_j}$ are common across all rows in $\mathbf{D}_j$ and only the q-degree, denoted by s, varies across the rows. Using Lemma 3, we construct a full-rank matrix $\mathbf{M}_j \in \mathbb{F}_q^{l_j \times l_j}$, such that the degrees of the polynomials of $\hat{\mathbf{D}}_j = \mathbf{D}_j \cdot \mathbf{M}_j$ will increase monotonically across the columns within each fixed row. Performing this for each $\mathbf{D}_j$, we construct

$$\hat{\mathbf{D}} = \mathbf{D} \cdot \mathbf{M}, \tag{30}$$

where $\mathbf{M} = \operatorname{diag}(\mathbf{M}_h, \ldots, \mathbf{M}_0)$ is full-rank with elements in $\mathbb{F}_q$. It follows that the transformed matrix $\hat{\mathbf{D}}$ satisfies the counterpart of (16b).

We argue that the remaining conditions are also preserved under the transformation of $\mathbf{M}$. Note that the elements on any

given row of $\hat{\mathbf{D}}_j$ are obtained through linear combinations (over $\mathbb{F}_q$) of the elements in the corresponding row of $\mathbf{D}_j$. Since both (21) and (22) are satisfied by all elements belonging to a fixed row of $\mathbf{D}_j$, they must also be satisfied by $\hat{\mathbf{D}}_j$. Then the proof of (24b), (23) and (24a) can be carried out for elements in $\hat{\mathbf{D}}_j$.

Consequently, $\hat{\mathbf{D}}$ completely satisfies (16) and (17), in contrast to $\mathbf{D}$ which only satisfies (23) and (24). The proof showing that $\hat{\mathbf{D}}$ is non-singular then follows verbatim to the proof showing all sub-matrices of $\mathbf{T}$ are non-singular. Then, $\det \hat{\mathbf{D}} = \det \mathbf{D} \det \mathbf{M}$ implies that $\mathbf{D}$ is also non-singular and therefore, $\mathbf{F}$ is super-regular.

Recall for this proof, we fixed $\mathbf{A}_0 = \mathbf{A}_1 = \cdots = \mathbf{A}_m$. Without the assumption, the product $\mathbf{F}$ would be comprised of a different set of $f_i(\alpha)$ for each column block. Because each $\mathbf{D}_i$ is a sub-matrix from a single column block and column operations are performed for each $\mathbf{D}_i$ independently, the polynomial degrees can always be transformed in order to satisfy (16) and (17). Consequently, the proof is identical when considering a $\mathbf{A}_{[0,m]}$ constructed from different blocks. ∎

We summarize the main steps in the proof of Theorem 5 as follows. A comparison of our proof and the original proof from [9] is provided in Fig. 2.

1) We start by constructing the matrices $\mathbf{T}_j$ in (13) for $0 \leq j \leq m$ with α being a primitive normal over $\mathbb{F}_{q^M}$.
2) We consider the product $\mathbf{F} = \mathbf{T}\mathbf{A}_{[0,m]}$, where $\mathbf{T}$ is the block Hankel matrix in (14) and $\mathbf{A}_{[0,m]}$ is a block diagonal matrix in the ground field.
3) We next consider non-trivial sub-matrices $\mathbf{D}$ of $\mathbf{F}$ and study their structure and ordering of the q-degrees (Properties 4, 5, 6).
4) In Lemma 3 we show that the matrix $\mathbf{D}$ can be transformed into the matrix $\hat{\mathbf{D}}$, where the q-degrees are sorted as required by Lemma 2.
5) Applying Lemma 2 it follows that $\hat{\mathbf{D}}$ is non-singular. In turn this also implies that $\mathbf{D}$ is non-singular.
6) This is equivalent to showing that $\mathbf{F}$ is super-regular and the Theorem follows.

D. Example Illustrating the Proof of Theorem 5

In this subsection, we provide an example of the preservation of super-regularity, detailing the properties of $\mathbf{F}$ and the key concepts in the proof of Theorem 5.

Example 1. *Let $q = 2, n = 4, m = 1$ and α be a primitive normal element of $\mathbb{F}_{2^M}$, where $M = q^{n(m+2)-1} = 2048$ satisfies Theorem 5. Then,*

$$\mathbf{T} = \begin{pmatrix} \mathbf{0} & \mathbf{T}_0 \\ \mathbf{T}_0 & \mathbf{T}_1 \end{pmatrix} = \begin{pmatrix} & & & & \alpha^{[0]} & \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} \\ & & & & \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} \\ & & & & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} \\ & & & & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} \\ \alpha^{[0]} & \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} \\ \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} & \alpha^{[8]} \\ \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} & \alpha^{[8]} & \alpha^{[9]} \\ \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} & \alpha^{[8]} & \alpha^{[9]} & \alpha^{[10]} \end{pmatrix}$$

is a super-regular matrix.

Let $\mathbf{A}_0$ and $\mathbf{A}_1$ be the following two non-singular square matrices:

$$\mathbf{A}_0 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

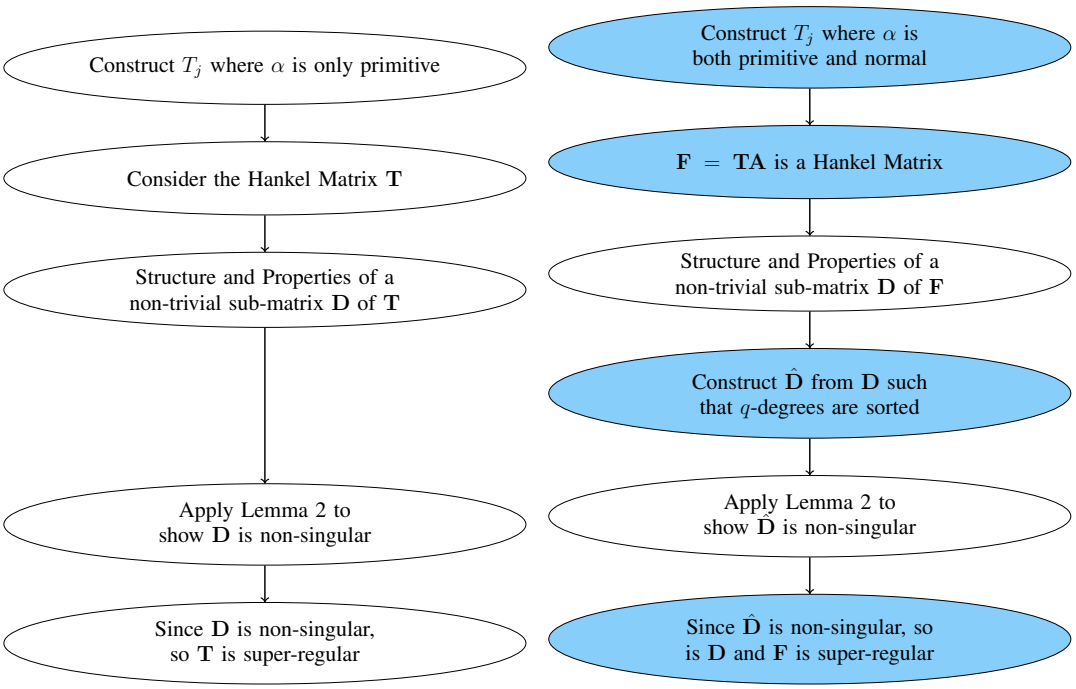
We use $\mathbf{A}_{[0,1]} = \operatorname{diag}(\mathbf{A}_0, \mathbf{A}_1)$ as the block-diagonal matrix and generate the product $\mathbf{F}$ given in (31)*.*

Because $\mathbf{A}_0 \neq \mathbf{A}_1$, the matrix $\mathbf{F}$ is not a Hankel matrix. However, note that every block matrix posses a Moore structure. The elements in every row of a given block are Frobenius powers of the corresponding elements in the first row. Since α is a normal element in $\mathbb{F}_{2^{2048}}$ and $\mathbf{A}_0$ and $\mathbf{A}_1$ are full rank, the isomorphism $\phi_n(\cdot)$ in Appendix A-A implies that the entries generating the left column block $\alpha^{[1]} + \alpha^{[2]}, \alpha^{[0]}, \alpha^{[0]} + \alpha^{[2]}, \alpha^{[3]}$ are linearly independent polynomials. Similarly, the entries generating the right column block $\alpha^{[0]} + \alpha^{[1]}, \alpha^{[1]} + \alpha^{[2]} + \alpha^{[3]}, \alpha^{[1]}, \alpha^{[0]} + \alpha^{[2]}$ are also respectively linearly independent amongst each other as stated in Remark 4. Note that the q-degree of these polynomials are upper and lower bounded between 3 *and* 0 *which satisfies* (21)*.*

Furthermore, the q-degrees are sorted according to Property 5. On a given column of $\mathbf{F}$, the q-degrees are increasing by 1 *from top to bottom. Moreover, for any fixed row, any polynomial in the left block has a lower degree than any polynomial in the right block.*

Now consider the sub-matrix formed from rows $\mathbf{r}_i, i \in$

$$\mathbf{F} = \mathbf{T}\mathbf{A}_{[0,1]} = \begin{pmatrix} & & & & \alpha^{[0]} + \alpha^{[1]} & \alpha^{[1]} + \alpha^{[2]} + \alpha^{[3]} & \alpha^{[1]} & \alpha^{[0]} + \alpha^{[2]} \\ & & & & \alpha^{[1]} + \alpha^{[2]} & \alpha^{[2]} + \alpha^{[3]} + \alpha^{[4]} & \alpha^{[2]} & \alpha^{[1]} + \alpha^{[3]} \\ & & & & \alpha^{[2]} + \alpha^{[3]} & \alpha^{[3]} + \alpha^{[4]} + \alpha^{[5]} & \alpha^{[3]} & \alpha^{[2]} + \alpha^{[4]} \\ & & & & \alpha^{[3]} + \alpha^{[4]} & \alpha^{[4]} + \alpha^{[5]} + \alpha^{[6]} & \alpha^{[4]} & \alpha^{[3]} + \alpha^{[5]} \\ \alpha^{[1]} + \alpha^{[2]} & \alpha^{[0]} & \alpha^{[0]} + \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} + \alpha^{[5]} & \alpha^{[5]} + \alpha^{[6]} + \alpha^{[7]} & \alpha^{[5]} & \alpha^{[4]} + \alpha^{[6]} \\ \alpha^{[2]} + \alpha^{[3]} & \alpha^{[1]} & \alpha^{[1]} + \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} + \alpha^{[6]} & \alpha^{[6]} + \alpha^{[7]} + \alpha^{[8]} & \alpha^{[6]} & \alpha^{[5]} + \alpha^{[7]} \\ \alpha^{[3]} + \alpha^{[4]} & \alpha^{[2]} & \alpha^{[2]} + \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} + \alpha^{[7]} & \alpha^{[7]} + \alpha^{[8]} + \alpha^{[9]} & \alpha^{[7]} & \alpha^{[6]} + \alpha^{[8]} \\ \alpha^{[4]} + \alpha^{[5]} & \alpha^{[3]} & \alpha^{[3]} + \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} + \alpha^{[8]} & \alpha^{[8]} + \alpha^{[9]} + \alpha^{[10]} & \alpha^{[8]} & \alpha^{[7]} + \alpha^{[9]} \end{pmatrix} \tag{31}$$



(a) Almeida et. al. [9] Proof Summary (b) MSR Proof Summary

Fig. 2: A summary of the proof of Theorem 5, compared with the original proof given in [9]. The highlighted sections denote the key differences.

$\{1,2,4,5\}$ *and columns* $\mathbf{c}_j, j \in \{3,4,5,6\}$. *This matrix*

$$\mathbf{D} = \begin{pmatrix} & \alpha^{[1]}+\alpha^{[2]} & \alpha^{[2]}+\alpha^{[3]}+\alpha^{[4]} & \alpha^{[2]} \\ & \alpha^{[2]}+\alpha^{[3]} & \alpha^{[3]}+\alpha^{[4]}+\alpha^{[5]} & \alpha^{[3]} \\ \alpha^{[3]} & \alpha^{[4]}+\alpha^{[5]} & \alpha^{[5]}+\alpha^{[6]}+\alpha^{[7]} & \alpha^{[5]} \\ \alpha^{[4]} & \alpha^{[5]}+\alpha^{[6]} & \alpha^{[6]}+\alpha^{[7]}+\alpha^{[8]} & \alpha^{[6]} \end{pmatrix}$$

does not have increasing q*-degrees along rows, i.e.,* (16b) *is not satisfied. The degree of the polynomials in the second and fourth columns are equal and lower than the degree of the polynomials in the third column. The degrees can be sorted by subtracting the fourth column from the second and then swapping the third and the fourth columns. The transformed matrix*

$$\hat{\mathbf{D}} = \begin{pmatrix} & \alpha^{[1]} & \alpha^{[2]} & \alpha^{[2]}+\alpha^{[3]}+\alpha^{[4]} \\ & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[3]}+\alpha^{[4]}+\alpha^{[5]} \\ \alpha^{[0]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[5]}+\alpha^{[6]}+\alpha^{[7]} \\ \alpha^{[1]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[6]}+\alpha^{[7]}+\alpha^{[8]} \end{pmatrix}$$

completely satisfies (16) *and* (17). *The above column operations are equivalent to right multiplication with the matrix,*

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix},$$

which is full-rank. It follows that $\hat{\mathbf{D}}$ *is non-singular and therefore,* $\mathbf{D}$ *as well.*

VI. Code Construction

The rows of $\mathbf{T}$ are permuted to form the block Toeplitz structure of an extended generator matrix

$$\bar{\mathbf{T}} = \begin{pmatrix} \mathbf{T}_0 & \mathbf{T}_1 & \dots & \mathbf{T}_m \\ & \mathbf{T}_0 & \dots & \mathbf{T}_{m-1} \\ & & \ddots & \vdots \\ & & & \mathbf{T}_0 \end{pmatrix}. \tag{32}$$

Since every sub-matrix of $\mathbf{T}$ has a counterpart in $\bar{\mathbf{T}}$ identical up to row permutations, this block Toeplitz matrix is also super-regular. $\mathbf{G}_m^{\mathrm{EX}}$ is then constructed as a sub-matrix of $k(m+1)$ rows from $\mathbf{T}$. This process parallels the construction of m-MDS generator matrices [7].

Theorem 6. Let $\bar{\mathbf{T}}$ be the super-regular matrix in (32) generated using a primitive normal $\alpha \in \mathbb{F}_{q^M}$, where $M = q^{n(m+2)-1}$. Let $0 \le i_1 < \cdots < i_k < n$ and construct a $k(m+1)\times n(m+1)$ sub-matrix $\mathbf{G}_m^{\mathrm{EX}}$ of $\bar{\mathbf{T}}$ from rows indexed $jn+i_1, \dots, jn+i_k$ for $0 \le j \le m$. This matrix is the extended generator of an MSR convolutional code $\mathcal{C}[n,k,m]$.

Proof: We show that $\mathbf{G}_m^{\mathrm{EX}}$ satisfies Theorem 3. Assume without loss of generality that $i_1 = 0, \dots, i_k = k-1$. Each $\mathbf{T}_i$ is divided into $\begin{pmatrix} \mathbf{G}_i \\ \mathbf{T}'_i \end{pmatrix}$, where $\mathbf{G}_i \in \mathbb{F}_{q^M}^{k\times n}$ are the blocks of the extended generator matrix. For $0 \le t \le m$, let $\mathbf{A}_t \in \mathbb{F}_q^{n\times n}$ be non-singular matrices. We similarly divide $\mathbf{A}_t = \begin{pmatrix} \mathbf{A}_t^* & \mathbf{A}'_t \end{pmatrix}$, where the two blocks $\mathbf{A}_t^* \in \mathbb{F}_q^{n\times \rho_t}$ and $\mathbf{A}'_t \in \mathbb{F}_q^{n\times(n-\rho_t)}$ represent the reduced channel matrix and some remaining matrix respectively. Let $\mathbf{A}_{[0,m]} = \mathrm{diag}(\mathbf{A}_0, \dots, \mathbf{A}_m)$. The

product can be written as

$$\bar{\mathbf{T}}\mathbf{A}_{[0,m]} = \begin{pmatrix} \mathbf{T}_0\mathbf{A}_0 & \mathbf{T}_1\mathbf{A}_1 & \dots & \mathbf{T}_m\mathbf{A}_m \\ & \mathbf{T}_0\mathbf{A}_1 & \dots & \mathbf{T}_{m-1}\mathbf{A}_m \\ & & \ddots & \vdots \\ & & & \mathbf{T}_0\mathbf{A}_m \end{pmatrix},$$

where

$$\mathbf{T}_i\mathbf{A}_t = \begin{pmatrix} \mathbf{G}_i\mathbf{A}_t^* & \mathbf{G}_i\mathbf{A}_t' \\ \mathbf{T}_i'\mathbf{A}_t^* & \mathbf{T}_i'\mathbf{A}_t' \end{pmatrix}.$$

The sub-matrix of $\bar{\mathbf{T}}\mathbf{A}_{[0,m]}$ containing only the rows and columns involving $\mathbf{G}_i\mathbf{A}_t^*$ is equal to the product $\mathbf{G}_m^{\mathrm{EX}}\mathbf{A}_{[0,m]}^*$. Because $\bar{\mathbf{T}}\mathbf{A}_{[0,m]}$ is super-regular, the determinant of $\mathbf{G}_m^{\mathrm{EX}}\mathbf{A}_{[0,m]}^*$ is either trivially zero or non-trivial and therefore non-zero.

It can be shown that for all $\mathbf{A}_{[0,m]}^*$ whose blocks satisfy (3), the product

$$\mathbf{G}_m^{\mathrm{EX}}\mathbf{A}_{[0,m]}^* = \begin{pmatrix} \mathbf{G}_0\mathbf{A}_0^* & \mathbf{G}_1\mathbf{A}_1^* & \dots & \mathbf{G}_m\mathbf{A}_m^* \\ & \mathbf{G}_0\mathbf{A}_1^* & \dots & \mathbf{G}_{m-1}\mathbf{A}_m^* \\ & & \ddots & \vdots \\ & & & \mathbf{G}_0\mathbf{A}_m^* \end{pmatrix}$$

has a non-trivial determinant. Note that the structure is reversed from that of $\mathbf{D}$ in (29). In [9], the authors showed that for $\mathbf{D}$ to have a non-trivial determinant, the number of rows of each $\mathbf{D}_j$ block cannot be less than the number of columns of each $\mathbf{O}_j$ block. In this case each $\mathbf{D}_j$ block of $\mathbf{G}_m^{\mathrm{EX}}\mathbf{A}_{[0,m]}^*$ contains $k(j+1)$ rows and each $\mathbf{O}_j$ block contains $\sum_{t=0}^{j}\rho_t$ columns. Consequently if the condition in (3) is satisfied, then $k(j+1) \geq \sum_{t=0}^{j}\rho_t$ for $j \leq m$ implies that $\mathbf{G}_m^{\mathrm{EX}}\mathbf{A}_{[0,m]}^*$ has a non-trivial determinant and is therefore non-singular. Thus, $\mathbf{G}_m^{\mathrm{EX}}$ satisfies Theorem 3 and $\mathcal{C}[n,k,m]$ achieves $d_R(m) = (n-k)(m+1)+1$. ∎

Remark 7. Our construction thus far is feasible over any sliding window channel $\mathcal{CH}(S,W)$ with delay $T = W-1$ and $S < d_R(W-1)$. If instead $T \geq W$ holds, the same code $\mathcal{C}[n,k,T]$ is still feasible with the same threshold on S i.e., $S < d_R(W-1)$. However if $T < W-1$, then the code is feasible provided that $S < d_R(T)$.

A. Numerical Results

The bound on the field size given in Theorem 5 is only a sufficiency constraint required for the proof. For small code parameters, it is possible to numerically verify whether Theorem 3 holds for a given extended generator matrix. An example is provided below of an MSR code over a small field. For the results in this section, we represent the primitive normal elements in the field as elements within the polynomial ring.

Example 2. *Let $\alpha = X+1$ be a primitive normal element in $\mathbb{F}_{2^{11}} = \mathbb{F}_2/\langle X^{11}+X^2+1\rangle$. We construct the following extended generator matrix*

$$\mathbf{G}_1^{\mathrm{EX}} = \begin{pmatrix} \alpha^{[0]} & \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} \\ \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} & \alpha^{[8]} \\ & & & & \alpha^{[0]} & \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} \\ & & & & \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} \end{pmatrix}.$$

This matrix satisfies Theorem 3, making it the generator for an MSR code $\mathcal{C}[4,2,1]$. Theorem 6 guarantees the construction if $M = 2^{11}$, i.e., α is a primitive normal element of $\mathbb{F}_{2^{2048}}$.

Example 3. *Let $\alpha = X+1$ be a primitive normal element in $\mathbb{F}_{2^{11}} = \mathbb{F}_2/\langle X^{11}+X^2+1\rangle$. The extended generator $\mathbf{G}_2^{\mathrm{EX}}$ is given by*

$$\begin{pmatrix} \alpha & \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} & \alpha^{[8]} \\ \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} & \alpha^{[8]} & \alpha^{[9]} & \alpha^{[10]} \\ & & & \alpha & \alpha^{[1]} & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} \\ & & & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} & \alpha^{[5]} & \alpha^{[6]} & \alpha^{[7]} \\ & & & & & & \alpha & \alpha^{[1]} & \alpha^{[2]} \\ & & & & & & \alpha^{[2]} & \alpha^{[3]} & \alpha^{[4]} \end{pmatrix}.$$

This matrix satisfies Theorem 3, making it the generator for an MSR code $\mathcal{C}[3,2,2]$. Theorem 6 guarantees the construction if $M = 2^{11}$, i.e., α is a primitive normal element of $\mathbb{F}_{2^{2048}}$.

The construction in Theorem 6 can generate MSR codes over much smaller field sizes than than those given by the bound of $M \geq q^{n(m+2)-1}$. Table I provides a list of code parameters and the field sizes on which they satisfy the matrix multiplication property of Theorem 3.

VII. Conclusion

We introduce a new distance metric for convolutional codes called the column sum rank distance. We prove several properties analogous to the column Hamming distance. A new family of codes — MSR Codes — that achieve the maximum distance up to the code memory is proposed. This is the rank metric counterpart to the m-MDS convolutional code. Our construction is based on matrices over extension fields that preserve super-regularity after multiplication with block diagonal matrices in the ground field.

The proof requires large field sizes but we numerically show that MSR codes do exist over smaller fields. Future work involves pursuing a more detailed study on field size requirements. Moreover, we have only considered a specific class of rank-deficient sliding window channels. In single-link streaming over burst erasure or mixed erasure channels, structured constructions using m-MDS codes as constituents have been revealed as more powerful alternatives [4]. A similar study pertaining MSR codes remains an interesting direction of further study.

Appendix A
Mathematical Preliminaries

A. Finite Fields

For $M \geq 0$ and a prime power q, let $\mathbb{F}_q$ be the finite field with q elements and $\mathbb{F}_{q^M}$ be an extension field of $\mathbb{F}_q$. A *primitive* element $\alpha \in \mathbb{F}_{q^M}$ is one whose consecutive powers can generate all non-zero elements of that field, i.e., $\mathbb{F}_{q^M} =$

Parameters $[n,k,m]$	Field Bound	Achievable Field	α
$[4,2,1]$	$\mathbb{F}_{2^{2048}}$	$\mathbb{F}_{2^{11}} = \mathbb{F}_2/\langle X^{11}+X^2+1\rangle$	$X+1$
$[3,2,2]$	$\mathbb{F}_{2^{2048}}$	$\mathbb{F}_{2^{11}} = \mathbb{F}_2/\langle X^{11}+X^2+1\rangle$	$X+1$
$[3,1,2]$	$\mathbb{F}_{2^{2048}}$	$\mathbb{F}_{2^{11}} = \mathbb{F}_2/\langle X^{11}+X^2+1\rangle$	$X+1$
$[2,1,2]$	$\mathbb{F}_{2^{128}}$	$\mathbb{F}_{2^7} = \mathbb{F}_2/\langle X^7+X^3+1\rangle$	X^3+1
$[2,1,1]$	$\mathbb{F}_{2^{64}}$	$\mathbb{F}_{2^5} = \mathbb{F}_2/\langle X^5+X^2+1\rangle$	$X+1$

TABLE I: Achievable field sizes under which codes constructed using Theorem 6. The bound required by the theorem for each set of code parameters is provided in the middle column.

$\{0, \alpha, \alpha^2, \ldots, \alpha^{q^M-1}\}$. Let $\mathbb{F}_q[X]$ be a polynomial ring of the ground field. The minimal polynomial of a primitive element α is the lowest degree monic polynomial $p_\alpha(X) \in \mathbb{F}_q[X]$ for which α is a root. The minimal polynomial is irreducible and the degree of $p_\alpha(X)$ is equal to M.

Lemma 4. If $f(\alpha) = 0$ for any $f(X) \in \mathbb{F}_q[X]$, then $p_\alpha(X) \mid f(X)$.

Proof: The proof can be found in [26, Chapter 4]. ∎

The extension field $\mathbb{F}_{q^M}$ is isomorphic to the M-dimensional vector space $\mathbb{F}_q^M$ over the ground field. Let $\alpha_0, \ldots, \alpha_{M-1} \in \mathbb{F}_{q^M}$ map to a basis for the vector space. A basis is defined as being *normal* when for $0 \le i \le m-1$, each $\alpha_i = \alpha^{q^i}$ for some $\alpha \in \mathbb{F}_{q^M}$. The generating element α is referred to as a normal element. The notation $\alpha^{[i]} = \alpha^{q^i}$ is used to describe the i-th Frobenius power of α. Every element $f \in \mathbb{F}_{q^M}$ can be written as a linear combination in $\mathbb{F}_q$ of the basis elements. Using the normal basis, f resembles a linearized polynomial $f(X) = \sum_{i=0}^{M-1} f_i X^{[i]} \in \mathbb{F}_q[X]$ evaluated at the normal element $X = \alpha$. The coefficients of this polynomial can be mapped

$$f(\alpha) = \sum_{i=0}^{M-1} f_i \alpha^{[i]} \mapsto \mathbf{f} = (f_0, \ldots, f_{M-1})^{\mathsf{T}} \tag{33}$$

to the entries of a unique vector $\mathbf{f} \in \mathbb{F}_q^{M\times 1}$. This mapping can be extended to vector spaces over the extension field and matrix spaces over the ground field. Using (33), we define $\phi_n : \mathbb{F}_{q^M}^n \to \mathbb{F}_q^{M\times n}$ as a bijection transforming a vector of linearized polynomial entries to a matrix whose columns are the coefficients of the polynomials.

For every finite extension of a finite field, there exists at least one element that is both a normal and a primitive element [27]. Such an element is referred to as being *primitive normal*. The properties of both normal and primitive elements are inherited in a primitive normal and will be useful in our code construction.

A linearized polynomial is defined by the property that every monomial term must have a Frobenius power. A linearized polynomial possesses a q-degree, denoted as $\deg_{\mathrm{q}} f(X)$, which gives the largest Frobenius power of the polynomial.

B. Super-regular Matrices

For $b \in \mathbb{N}$, let σ be a permutation of the set $\{0, \ldots, b-1\}$. A permutation is comprised of a series of transpositions, which are defined as two entries of the set switching positions. The sign function of a permutation measures its parity, i.e., $\mathrm{sgn}(\sigma)$ is equal to 1 when σ is constructed from an even number of transpositions, and equal to -1 otherwise. Let S_b denote the set of all possible permutations. The determinant of a $b \times b$ matrix $\mathbf{D}$ can be calculated by summing over all permutations in S_b in the Leibniz formula

$$\det \mathbf{D} \triangleq \sum_{\sigma \in S_b} \mathrm{sgn}(\sigma) \prod_{i=0}^{b-1} D_{i,\sigma(i)}. \tag{34}$$

Each product $\prod_{i=0}^{b-1} D_{i,\sigma(i)}$ is referred to as a term in the summation. When every term is equal to 0, the matrix is said to have a trivial determinant. Using the Leibniz formula, a *super-regular matrix* is a matrix for which every square submatrix with a non-trivial determinant is non-singular.

REFERENCES

[1] R. Mahmood, A. Badr, and A. Khisti, "Convolutional codes in rank metric for network streaming," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2015, to appear in 2015.

[2] E. Martinian and C.-E. W. Sundberg, "Burst erasure correction codes with low decoding delay," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2494–2502, 2004.

[3] D. Leong, A. Qureshi, and T. Ho, "On coding for real-time streaming under packet erasures," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2013, pp. 1012–1016.

[4] A. Badr, P. Patil, A. Khisti, W. Tan, and J. Apostolopoulos, "Layered constructions for low-delay streaming codes," *IEEE Trans. Inf. Theory*, to appear in 2015.

[5] V. Tomas, J. Rosenthal, and R. Smarandache, "Decoding of convolutional codes over the erasure channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 90–108, 2012.

[6] M. Ellis, D. P. Pezaros, and C. Perkins, "Performance analysis of AL-FEC for RTP-based streaming video traffic to residential users," in *IEEE Packet Video Workshop*, 2012.

[7] H. Gluesing-Luerssen, J. Rosenthal, and R. Smarandache, "Strongly-MDS convolutional codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 584–598, 2006.

[8] R. Hutchinson, R. Smarandache, and J. Trumpf, "Superregular matrices and the construction of convolutional codes having a maximum distance profile," *Lin. Algebra and App.*, vol. 428, pp. 2585–2596, 2008.

[9] P. Almeida, D. Napp, and R. Pinto, "A new class of super regular matrices and MDP convolutional codes," *Lin. Algebra and App.*, vol. 439, pp. 2145–2157, 2013.

[10] R. Ahlswede, N. Cai, S. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[11] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Allerton Conf. on Comm., Control, Comp.*, 2003.

[12] T. Ho, M. Médard, R. Kötter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 413–430, 2006.

[13] E. M. Gabidulin, "Theory of codes with maximum rank distance," *Probl. Inf. Transm.*, vol. 21, no. 1, pp. 1–12, 1985.

[14] R. M. Roth, "Maximum-rank array codes and their application to crisscross error correction," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 328–336, 1991.

[15] R. Kötter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3579–3591, 2008.

[16] D. Silva, F. R. Kschischang, and R. Kötter, "A rank-metric approach to error control in random network coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3951–3967, 2008.
[17] R. W. Nóbrega and B. F. Uchôa-Filho, "Multishot codes for network coding using rank-metric codes," in *IEEE Wireless Netw. Coding Conf. (WiNC)*, 2010, pp. 1–6.
[18] A. Wachter-Zeh, M. Stinner, and V. Sidorenko, "Convolutional codes in rank metric with application to random network coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3199–3213, 2015.
[19] A. Khisti, D. Silva, and F. R. Kschischang, "Secure-broadcast codes over linear-deterministic channels," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2010, pp. 555–559.
[20] S. Jalali, M. Effros, and T. Ho, "On the impact of a single edge on the network coding capacity," in *Inf. Theory and App. (ITA)*, 2011, pp. 1–5.
[21] A. Badr, D. Lui, and A. Khisti, "Streaming codes for multicast over burst erasure channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 4181–4208, 2015.
[22] R. Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*. IEEE Press, 1999.
[23] S. Lin and D. J. Costello, *Error Control Coding*. Pearson Education Inc., 2004.
[24] E. M. Gabidulin, "Convolutional codes over large alphabets," in *Proc. Int. Workshop on Algebraic Combinatorial and Coding Theory (ACCT)*, Varna, Bulgaria, 1988, pp. 80–84.
[25] R. Mahmood, "Rank metric convolutional codes with applications in network streaming," Master's thesis, University of Toronto, 2015.
[26] F. J. MacWilliams and N. J. A. Sloane, *Theory of Error-Correcting Codes*. North Holland Mathematical Library, 1977.
[27] H. W. Lenstra and R. J. Schoof, "Primitive normal bases for finite fields," *Math. Comp.*, vol. 48, no. 177, pp. 217–231, 1987.

Rafid Mahmood Rafid Mahmood received his B.A.Sc. and M.A.Sc. degrees in Electrical and Computer Engineering from the University of Toronto in Toronto, Ontario, Canada in 2013 and 2015 respectively. During this time, he foucsed his research on coding theory and information theory. He is pursuing his Ph.D. degree in Mechanical & Industrial Engineering at the University of Toronto.

Ahmed Badr Ahmed Badr received the B.Sc., M.Sc. and Ph.D. degrees in Electrical & Computer Engineering from Cairo University, Egypt in 2007, Nile University, Egypt in 2009 and University of Toronto, Canada in 2014. From September 2007 to August 2009, he was a Research Assistant in the Wireless Intelligent Networks Center (WINC), Nile University. In September 2009, he was a Research Assistant at the Signals Multimedia and Security Laboratory in University of Toronto. In 2014, he assumed his current position as a Postdoctoral Fellow in University of Toronto. His research interests include information theory, coding theory and real-time communication.

Ashish Khisti Ashish Khisti is an assistant professor and a Canada Research Chair in the Electrical and Computer Engineering (ECE) department at the University of Toronto, Toronto, Ontario Canada. He received his BASc degree in Engineering Sciences from University of Toronto and his S.M and Ph.D. Degrees from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. His research interests span the areas of information theory, wireless physical layer security and streaming in multimedia communication systems. At the University of Toronto, he heads the signals, multimedia and security laboratory. He is a recipient of the Ontario Early Researcher Award from the Province of Ontario as well as a Hewlett-Packard Innovation Research Award.