

A note on the explicit constructions of tree codes over polylogarithmic-sized alphabet

Siddharth Bhandari* Prahladh Harsha*

February 19, 2020

Abstract

Recently, Cohen, Haeupler and Schulman gave an explicit construction of binary tree codes over polylogarithmic-sized output alphabet based on Pudlák’s construction of maximum-distance-separable (MDS) tree codes using totally-non-singular triangular matrices. In this short note, we give a unified and simpler presentation of Pudlák and Cohen-Haeupler-Schulman’s constructions.

1 Introduction

We begin by recalling the definition of tree codes as introduced by Schulman in 1993 [Sch96]

Definition 1.1 (tree codes). *Given two finite alphabets Σ and Γ and a parameter $\delta \in (0, 1)$, a tree code $\text{TC}: \Sigma^* \rightarrow \Gamma^*$ with distance δ is a function mapping n -length strings over the alphabet Σ to n -length strings over the alphabet Γ for every positive integer n satisfying the following two properties.*

Online encoding: *For every positive integer m , there exist functions $\varphi_m: \Sigma^m \rightarrow \Gamma$ such that for all $n \in \mathbb{N}$, $x \in \Sigma^n$ and $i \in [n]$, we have $\text{TC}(x)_i = \varphi_i(x_1, \dots, x_i)$. In other words, the i^{th} symbol of the output only depends on the first i symbols of the input and not on the latter symbols. The functions φ_i are referred to as the encoding functions.*

Distance: *The (relative) distance of the tree code TC , denoted by δ_{TC} , (defined as follows) is at least δ .*

$$\delta_{\text{TC}} := \inf_{n, x \neq x' \in \Sigma^n} \frac{\Delta(\text{TC}(x), \text{TC}(x'))}{n - \text{split}(x, x')},$$

where $\text{split}(x, x')$ is the largest integer s such that for all $i \leq s$, we have $x_i = x'_i$. In other words, for every positive integer n and any two distinct n -length strings $x, x' \in \Sigma^n$, we have that the Hamming distance between $y = \text{TC}(x)$ and $y' = \text{TC}(x')$ is at least $\delta(n - \text{split}(x, x'))$.

The string $y = \text{TC}(x)$ is said to be the tree code encoding of x . This encoding is said to be explicit if each of the encoding functions φ_i can be computed in time polynomial in i .

It is known via the probabilistic method that for every finite alphabet Σ and distance parameter $\delta \in (0, 1)$, there exists an alphabet $\Gamma = \Gamma(\Sigma, \delta)$ and a tree code $\text{TC}: \Sigma^* \rightarrow \Gamma^*$ with distance at least δ (see Theorem 1.3 for the precise statement when Σ is a finite field). A long standing open problem is to construct explicit tree codes matching the above construction (or even one for some distance parameter $\delta \in (0, 1)$). Since we do not yet know such explicit constructions, several intermediate variants and special cases of the above definition of tree codes have been studied, some of which are interesting of their own right. Below, we mention some of these variants.

*Tata Institute of Fundamental Research, INDIA. email: {siddharth.bhandari, prahladh}@tifr.res.in. Research of the authors supported by the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.01-0500. Research of the second author supported in part by the Swarnajayanti fellowship.

Truncated Tree Code: Instead of constructing a single tree code $\text{TC}: \Sigma^* \rightarrow \Gamma^*$ that works for all input lengths, sometimes it is easier to construct a family of tree codes (indexed by n) each of which works for all strings up to a fixed length n , i.e., $\text{TC}^{(n)}: \Sigma^{(\leq n)} \rightarrow \Gamma^{(\leq n)}$. We will refer to these tree-codes as *n-truncated* tree codes and the family $\{\text{TC}^{(n)}\}_{n \in \mathbb{Z}_{>0}}$ as truncated tree codes. While dealing with *n-truncated* tree codes, we will use the superscript n to denote the maximum input length. The notion of distance now becomes:

$$\delta_{\text{TC}^{(n)}} := \inf_{i \leq n, x \neq x' \in \Sigma^i} \frac{\Delta(\text{TC}(x), \text{TC}(x'))}{i - \text{split}(x, x')}$$

Note that for a truncated code $\{\text{TC}^{(n)}\}_n$, we have a family of encoding functions $\{\varphi_i^{(n)} : 1 \leq i \leq n\}$ for each n . In this case, we say that the truncated code is *explicit* if for each n and $i \leq n$, $\varphi_i^{(n)}$ is computable in time polynomial in n . Note that it is polynomial in n and not just i as in the case of (non-truncated) tree codes. Clearly, tree codes imply truncated tree codes but the other direction is not known.

Input Alphabet: The classical tree-code considers binary input alphabet, i.e., $\Sigma = \{0, 1\}$. Sometimes, it is convenient to have an algebraic structure on the input alphabet, say Σ is a finite field \mathbb{F} or even the set of non-negative integers $\mathbb{Z}_{\geq 0}$. When dealing with truncated tree-codes $\text{TC}^{(n)}$, one may also consider tree codes whose input and output alphabets grow with n (i.e., $\{\Sigma_n\}_n, \{\Gamma_n\}_n$ corresponding to $\{\text{TC}^{(n)}\}_n$).

Output Alphabet: The holy grail is to construct tree codes over the binary alphabet with constant output alphabet size. As an intermediate step, one typically constructs tree codes whose output alphabets grow with n . This is best formalized in the setting of truncated tree codes $\{\text{TC}^{(n)}: \Sigma^{(\leq n)} \rightarrow \Gamma_n^{(\leq n)}\}_{n \in \mathbb{Z}_{\geq 0}}$ where $\{\Gamma_n\}_n$ is a family of output alphabets whose size (may) grow with n . The formalism for the more standard (single) tree code is a little more involved. We let $\Gamma = \cup_n \Gamma_n$ and $\text{TC}: \Sigma^* \rightarrow \Gamma^*$ subject to the constraint that the encoding functions φ_i satisfy $\varphi_i: \Sigma^i \rightarrow \Gamma_i$. Note Γ is not necessarily a finite alphabet, though each Γ_n is.

We are now ready to state the remarkable recent result of Cohen, Haeupler and Schulman that constructs explicit binary tree codes over a polylogarithmic-sized output alphabet.¹

Theorem 1.2 (explicit tree codes with polylog alphabet [CHS18]). *For all $\eta \in [0, 1)$ there exists a family of alphabets $\{\Gamma_n\}_n$ such that $|\Gamma_n| = O_\eta(\text{polylog}(n))$ with the following property. Let $\Gamma = \cup_n \Gamma_n$. There exists a tree code $\text{TC}: \{0, 1\}^* \rightarrow \Gamma^*$ with distance η such that the encoding functions φ_i satisfy $\varphi_i: \{0, 1\}^i \rightarrow \Gamma_i$. Furthermore, the encoding functions φ_n are constructible in time polynomial in n .*

Cohen, Haeupler and Schulman construct such tree codes in two steps. In the first step, they construct tree codes over the integer alphabet using Newton polynomials. This construction is a special case of Pudlák's construction of "maximum-distance-separable (MDS)" tree codes using totally-non-singular triangular matrices. In the second step, they perform an alphabet reduction to reduce the alphabet from integers to a polylogarithmic-sized output alphabet.

Organization: The rest of this exposition is organized as follows. We give a short (but complete) exposition of MDS codes, à la Pudlák, in Section 2 followed by the alphabet reduction technique of Cohen-Haeupler-Schulman in Section 3.

The purpose of this note is only to give an exposition of the Cohen-Haeupler-Schulman construction. In particular, we place the Cohen-Haeupler-Schulman construction in the context of Pudlák's framework for constructing explicit tree codes. For a more comprehensive treatment of tree codes, we refer the interested reader to Gelles's survey on interactive communication [Gel17] and Pudlák's excellent paper [Pud16] discussing various potential approaches to obtain explicit tree codes.

¹The theorem as stated here is marginally stronger than the one proved by Cohen-Haeupler-Schulman, in the sense that it refers to explicit constructions of tree codes as opposed to truncated codes. See Remark 3.8 for more details.

1.1 Linear tree codes

A particularly appealing setting is when the input and output alphabet Σ, Γ are finite-dimensional vector spaces $\mathbb{F}^s, \mathbb{F}^r$ respectively over some finite field \mathbb{F} . In this setting, if the tree code $\text{TC}: (\mathbb{F}^s)^* \rightarrow (\mathbb{F}^r)^*$ further satisfies that $\text{TC}|_{(\mathbb{F}^s)^n}$ is a linear mapping for each n , then the tree code is said to be \mathbb{F} -linear.

For the case when $s = 1$, it is easy to check that the online encoding property of tree codes ensures that if $\text{TC}: \mathbb{F}^* \rightarrow (\mathbb{F}^r)^*$ is \mathbb{F} -linear, then there exist r -lower triangular matrices $A_0, \dots, A_{r-1} \in \mathbb{F}^{\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}}$ such that for any $x \in \mathbb{F}^n$, we have $\text{TC}(x)_i = ((A_0^{(n)}(x))_i, (A_1^{(n)}(x))_i, \dots, (A_{r-1}^{(n)}(x))_i) \in \mathbb{F}^r$ where $A^{(n)}$ refers to the $n \times n$ -matrix formed by the restricting A to the top n columns and rows. By a change of basis of the input strings, we may without loss of generality assume that $A_0 = I$, the identity matrix (we will not use this property, however our construction of distance-half linear tree codes will have this property).

As in linear codes, the distance of a linear tree code can be characterized by the minimum weight of non-zero codewords. More specifically, if $\text{TC}: \mathbb{F}^* \rightarrow (\mathbb{F}^r)^*$ is \mathbb{F} -linear then

$$\delta_{\text{TC}} = \inf_{n, x \in \mathbb{F}^n \setminus \{0^n\}} \frac{\text{wt}_{\mathbb{F}^r}(\text{TC}(x))}{n - \text{split}(x, 0^n)},$$

where $\text{wt}_{\Gamma}(x)$ refers to the Hamming weight of x with respect to the alphabet Γ . Since $\Gamma = \mathbb{F}^r$ and $\text{TC}(x) \in (\mathbb{F}^r)^n = \mathbb{F}^{rn}$, it is also natural to consider a variation of the above definition with Hamming weight with respect to the finite field alphabet \mathbb{F} instead.

$$\tilde{\delta}_{\text{TC}} := \inf_{n, x \in \mathbb{F}^n \setminus \{0^n\}} \frac{\text{wt}_{\mathbb{F}}(\text{TC}(x))}{r(n - \text{split}(x, 0^n))}.$$

Clearly, $\tilde{\delta}_{\text{TC}} \leq \delta_{\text{TC}}$.

Schulman proved the existence of tree codes via the probabilistic method. The following statement is due to Pudlák [Pud16, Theorem 2.1]².

Theorem 1.3 (probabilistic construction of linear tree codes [Sch96, Pud16]). *Let $q = |\mathbb{F}|$, $r = q^d = |\Sigma|$ and $0 < \delta < \frac{r-1}{r}$ such that*

$$\log_r(2q) + H_r(\delta) \leq 1$$

where H_r denotes the r -entropy function defined by

$$H_r(x) = x \log_r(r-1) - x \log_r x - (1-x) \log_r(1-x).$$

Then, there exists a linear tree code $\text{TC}: \mathbb{F}^* \rightarrow \Sigma^*$ with $\tilde{\delta}_{\text{TC}} > \delta$. Moreover, if q, r and δ are fixed, then the encoding functions φ_n can be constructed for every n in time $2^{O(n)}$.

Note that there exists $\delta > 0$ such that for every $q > 2$, there exist tree codes with rate $1/2$ (i.e., $d = 2$) and minimum relative distance $\geq \delta$. We do not know if binary (i.e., $q = 2$) tree codes with rate $1/2$ can have asymptotically positive minimum relative distance.

2 MDS tree codes, a la Pudlák

In this section, we present the treatment of linear tree codes following Pudlák [Pud16]. More precisely, we present the ‘‘Singleton bound’’ for linear codes, define maximum-distance-separable (MDS) tree codes and present a construction of MDS codes assuming the existence of totally-non-singular triangular matrices.

Proposition 2.1 (Singleton bound for tree codes [Pud16, Proposition 5.1]). *For every n -truncated code, $\text{TC}^{(n)}: \Sigma^{(\leq n)} \rightarrow \Gamma^{(\leq n)}$ we have $\delta_{\text{TC}^{(n)}} \leq \frac{1}{n} \lfloor n(1 - \log |\Sigma| / \log |\Gamma|) + 1 \rfloor$.*

²Pudlák provides two probabilistic constructions, one for truncated codes and another for untruncated codes. The theorem stated here refers to the latter construction using cyclic Toeplitz matrices.

Proof. This proof will be an adaptation of the standard proof of the Singleton bound for codes to the tree code setting. The proposition is trivially true when $|\Gamma| \geq |\Sigma^n|$ (in fact, in this case, there exist n -truncated tree codes with distance 1). So we might as well assume that $|\Gamma| < |\Sigma^n|$. Let ℓ be the smallest integer such that $\ell \geq n \cdot (\log |\Sigma| / \log |\Gamma|) - 1$, which is at least 1 since $n \cdot (\log |\Sigma| / \log |\Gamma|) > 1$. Consider the mapping $f_{n \rightarrow \ell}: \Sigma^n \rightarrow \Gamma^\ell$ defined as follows $f_{n \rightarrow \ell}(x) := \text{TC}^{(n)}(x)|_{[n-\ell+1, n]}$ (i.e., the tree code encoding restricted to the last ℓ symbols). Since $|\Sigma^n| > |\Gamma^\ell|$, there must exist two strings $x, x' \in \Sigma^n$ such that $f_{n \rightarrow \ell}(x) = f_{n \rightarrow \ell}(x')$ (i.e., the tree code encodings of x and x' agree on the last ℓ symbols). Let $i \geq 1$ be the first place where x and x' differ. Note that $i \leq n - \ell$. Due to the online nature of tree codes we know that $\text{TC}^{(n)}(x)$ and $\text{TC}^{(n)}(x')$ agree on the first $i - 1$ coordinates (in addition to the last ℓ coordinates). Therefore,

$$\delta_{\text{TC}^{(n)}} \leq \frac{n - (\ell + (i - 1))}{n - i + 1} = 1 - \frac{\ell}{n - i + 1} \leq 1 - \frac{\ell}{n} = 1 - \frac{1}{n} \left(\left\lceil n \cdot \frac{\log |\Sigma|}{\log |\Gamma|} \right\rceil - 1 \right) = \frac{1}{n} \left\lfloor n \left(1 - \frac{\log |\Sigma|}{\log |\Gamma|} \right) + 1 \right\rfloor.$$

□

Notice that the aforementioned proof would have also worked with the following relaxed notion of distance (that only considers strings of length n):

$$\delta_{\text{TC}^{(n)}}^{\text{relaxed}} := \inf_{x \neq x' \in \Sigma^n} \frac{\Delta(\text{TC}(x), \text{TC}(x'))}{n - \text{split}(x, x')}.$$

For any non-zero $y \in \Gamma^n$ and $\alpha \in (0, 1)$, the condition $\text{wt}(y) \geq \lfloor \alpha \cdot n + 1 \rfloor$ is equivalent to the condition that $\text{wt}(y) > \alpha \cdot n$. This combined with [Proposition 2.1](#) motivates the following definition.

Definition 2.2 (MDS tree codes). *A n -truncated tree code $\text{TC}^{(n)}: \Sigma^{(\leq n)} \rightarrow \Gamma^{(\leq n)}$ is said to be maximum-distance separable (MDS) if $\delta_{\text{TC}^{(n)}} > 1 - \log |\Sigma| / \log |\Gamma|$. An MDS tree code is defined similarly.*

Pudlák showed how the existence of certain type of matrices implies the existence of MDS codes. We first define these special type of matrices. In the following, for a $n \times n$ matrix A and subsets $I, J \subseteq [n]$ of the rows and columns, the matrix $A[I|J]$ refers to the submatrix of A restricted to the rows I and columns J .

Definition 2.3 (totally-non-singular triangular matrix). *A lower-triangular matrix $A \in \mathbb{F}^{n \times n}$ is said to be totally-non-singular if for every $1 \leq r \leq n$ and pair of sets $I, J \subseteq [n]$ such that $|I| = |J| = r$ and $I = \{1 \leq i_1 < \dots < i_r \leq n\}$, $J = \{1 \leq j_1 < \dots < j_r \leq n\}$ satisfying $i_s \geq j_s$ for all $s \in [r]$, we have that the submatrix $A[I|J]$ is non-singular.*

Given a lower-triangular matrix $A \in \mathbb{F}^{n \times n}$, consider the linear tree code $\text{TC}_A^{(n)}: \mathbb{F}^{(\leq n)} \rightarrow (\mathbb{F}^2)^{\leq n}$ defined by the pair of generator matrices ($A_0 = I_n, A_1 = A$). In other words, for any $x \in \mathbb{F}^k$ and $i \leq k$, we have

$$\left(\text{TC}_A^{(n)}(x) \right)_i := \left(x_i, \left(A^{(k)}(x) \right)_i \right).$$

Theorem 2.4 (MDS linear tree codes [[Pud16](#), Theorem 5.3]). *If $A \in \mathbb{F}^{n \times n}$ is a totally-non-singular lower-triangular matrix, then the n -truncated tree code $\text{TC}_A^{(n)}$ has distance greater than $1/2$ and is hence MDS.*

Proof. We will in fact show the stronger property that $\tilde{\delta} > 1/2$. More precisely, for any non-zero code-word $x \in \mathbb{F}^k$ with $\ell = \text{split}(0^k, x) \in [0, k - 1]$, we need to show that the \mathbb{F} -weight of $y = \text{TC}(x) \in (\mathbb{F}^2)^k$ when viewed as a string in \mathbb{F}^{2k} is greater than $(k - \ell)$. To this end, we define the following two sets of column and row indices.

$$\mathcal{C}_x := \{j \in [k]: x_j \neq 0\}, \quad \mathcal{R}_x := \{i \in [k]: i > \ell, (A^{(k)}(x))_i = 0\}.$$

The number of non-zero coordinates in y (when viewed a string over \mathbb{F}) is exactly $|\mathcal{C}_x| + (k - \ell - |\mathcal{R}_x|)$ which is at strictly greater than $(k - \ell)$ if $|\mathcal{C}_x| > |\mathcal{R}_x|$, which is proved in the claim below. □

Claim 2.5. $|\mathcal{C}_x| > |\mathcal{R}_x|$

Proof. Assume otherwise and suppose that $\mathcal{C}_x = \{j_1 < j_2 < \dots < j_r\}$ and $\mathcal{R}_x = \{i_1 < j_2 < \dots < i_s\}$ for some $s \geq r$. By definition of \mathcal{C}_x , we have $j_1 = \ell + 1$. Since $|\mathcal{R}_x| \geq |\mathcal{C}_x| \geq 1$, we have that $i_1 > \ell$ and hence $i_1 \geq j_1$. Let t be the maximum integer such that $i_1 \geq j_1, i_2 \geq j_2, \dots, i_t \geq j_t$. Let us observe the action of the matrix $A[\{i_1, \dots, i_t\} \mid \{j_1, \dots, j_t\}]$ on the vector $x|_{\{j_1, \dots, j_t\}}$. By maximality of t , we get that either $r = t$ or $i_{t+1} < j_{t+1}$. In either case, it is easy to see that $A[\{i_1, \dots, i_t\} \mid \{j_1, \dots, j_t\}](x|_{\{j_1, \dots, j_t\}}) = \mathbf{0}$. However, this is contradiction to the total non-singularity of A . Therefore, $|\mathcal{C}_x| > |\mathcal{R}_x|$. \square

Claim 2.5 lends itself to the following alternate construction of a tree code which also has distance greater than $1/2$. For $v \in \mathbb{F}$ define $v' = (v, 0) \in \mathbb{F}^2$ as the vector obtained by appending a zero to v . Similarly, for a vector $x \in \mathbb{F}^k$ let $x' \in (\mathbb{F}^2)^k$ be $(x'_1, x'_2, \dots, x'_k)$. Given a totally-non-singular lower-triangular matrix $A \in \mathbb{F}^{2n \times 2n}$, consider the n -truncated tree code $\text{TC}_{A,(1,1)}^{(n)} : \mathbb{F}^{(\leq n)} \rightarrow (\mathbb{F}^2)^{(\leq n)}$ that maps $x \in \mathbb{F}^k$ to $A^{(2k)}(x') \in \mathbb{F}^{2k}$. For a non-zero $x \in \mathbb{F}^k$, let $\ell = \text{split}(0^k, x)$. Let $x' \in \mathbb{F}^{2k}$ be the corresponding non-zero vector and let $\ell' = \text{split}(0^{2k}, x')$. Clearly, $\ell' = 2\ell$. Let $y' = \text{TC}_{A,(1,1)}^{(n)}(x')$. Define sets $\mathcal{C}_{x'}$ and $\mathcal{R}_{x'}$ as before. By definition, $|\mathcal{C}_{x'}| \leq 1/2 \cdot (2k - \ell') = k - \ell$ and $\text{wt}_{\mathbb{F}}(y') = 2k - \ell' - |\mathcal{R}_{x'}| > 2k - \ell' - |\mathcal{C}_{x'}|$ (by **Claim 2.5**). Hence, $\text{wt}_{\mathbb{F}}(y') > 2k - \ell' - (k - \ell) = k - \ell = 1/2(2k - \ell')$. Hence, $\tilde{\delta}$ distance of $\text{TC}_{A,(1,1)}^{(n)}$ is greater than $1/2$.

This alternate construction has the advantage that it can be generalized to obtain MDS tree codes with distance $\delta_{\text{TC}} > r/r+s$ for integers $r, s \geq 1$. Let $A \in \mathbb{F}^{(r+s)n \times (r+s)n}$ be a totally-non-singular lower-triangular matrix. For $v \in \mathbb{F}^s$ define $v' \in \mathbb{F}^{r+s}$ as the vector obtained by appending r zeros to v . Similarly, for a vector $x \in (\mathbb{F}^s)^k$ let $x' \in \mathbb{F}^{(r+s)k}$ be $(x'_1, x'_2, \dots, x'_k)$. For all $x \in (\mathbb{F}^s)^k$ and $k \leq n$ define $\text{TC}_{A,(s,r)}^{(n)} : (\mathbb{F}^s)^{(\leq n)} \rightarrow (\mathbb{F}^{r+s})^{(\leq n)}$ as

$$x \in (\mathbb{F}^s)^k \mapsto A^{(r+s)k} x' \in (\mathbb{F}^{(r+s)})^k.$$

An argument identical to the one above proves that this is an MDS code.

Theorem 2.6. $\text{TC}_{A,(s,r)}^{(n)}$ is a n -truncated MDS tree code with distance $r/r+s$.

2.1 Instantiating the MDS tree codes

Pudlák suggested various examples of triangular totally-nonsingular matrices. One of these examples is the the Pascal Matrix defined as follows:

Definition 2.7 (Pascal matrix). The Pascal matrix $P \in \mathbb{Z}_{>0}^{\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}}$ is the matrix whose (i, j) -entry for $i, j \in \mathbb{Z}_{>0}$ is the binomial coefficient $\binom{i}{j}$. For any $n \in \mathbb{Z}_{>0}$, we let P_n be the restriction of the (infinite) matrix P to the first $(n + 1)$ rows and $(n + 1)$ columns. We will use the fact that for any n , the maximum value of any entry in P_n is at most 2^n .

Gessel and Viennot [GV85], using Lindström's combinatorial interpretation of certain determinants in terms of disjoint path systems [Lin73], proved the following remarkable theorem which immediately yields the total-non-singularity of the Pascal matrix.

Lemma 2.8 ([GV85], Corollary 2). Let $0 \leq a_1 < a_2 < a_3 < \dots < a_n$ and $0 \leq b_1 < b_2 < \dots < b_n$ be two sequences of n integers each. Define the $n \times n$ matrix M by $M_{i,j} = \binom{a_i}{b_j}$. If $a_i \geq b_i$ for each $i \in [n]$ then $\det(M) > 0$.

Combining the above theorem with **Theorem 2.4**, we obtain the following result.

Theorem 2.9 (tree code over integers [CHS18, Theorem 1.3]). ³ There exists an explicit tree code over the integers $\text{TC}_{\mathbb{Z}} : \mathbb{Z}_{\geq 0}^{\mathbb{Z}_{\geq 0}} \rightarrow (\mathbb{Z}_{\geq 0}^{\mathbb{Z}_{\geq 0}})^{\mathbb{Z}_{\geq 0}}$ with distance at least $1/2$ satisfying the property that for any $n \in \mathbb{Z}_{>0}$ and $i \in \{0, 1, \dots, n\}$, we have $(\overline{(\text{TC}_{\mathbb{Z}}(a_0, \dots, a_n))}_i)_1 = a_i$ and $(\overline{(\text{TC}_{\mathbb{Z}}(a_0, \dots, a_n))}_i)_2 \leq 2^n \max_j |a_j|$.

³Cohen-Haeupler-Schulman obtained this theorem by using the inverse of the Pascal matrix instead. If a lower-triangular is totally-non-singular, then so is its inverse.

3 Alphabet Reduction

In this section, we show how to convert the tree code over the integers constructed in [Theorem 2.9](#) to one over the binary alphabet. By merely restricting the input alphabet of $\text{TC}_{\mathbb{Z}}$ in [Theorem 2.9](#) to integers within the range $\{0, 1, \dots, 2^{s-1}\}$ and setting $n = s$, we obtain the following corollary.

Corollary 3.1 (tree code over large alphabet). *For every $s \in \mathbb{Z}_{>0}$, there exist an explicit tree code over the input alphabet $\{0, 1\}^s$, $\text{TC}^{(s)}: \{\{0, 1\}^s\}^{(\leq s)} \rightarrow \{\{0, 1\}^{3s}\}^{(\leq s)}$ with distance at least $1/2$.*

The tree code constructed in the above corollary has excellent distance albeit over a large super-constant alphabet (both with respect to input and output). In the rest of this section, we will show how to transform this tree code into one over the binary alphabet and polylogarithmic output alphabet.

We follow, roughly, the treatment of Cohen-Haeupler-Schulman [[CHS18](#)]. (Also, see Pudlák [[Pud16](#), Proposition 3.1] for the standard well-known alphabet reduction procedure from a polynomial-sized output alphabet to constant-sized output alphabet. In some sense, the Cohen-Haeupler-Schulman reduction is an extension of this procedure to super-polynomial-sized alphabets.)

To this end, we define the notion of a lagged tree code.

Definition 3.2 (tree codes with lag). *Given positive integers $\ell \leq L$, a tree code $\text{TC}: \Sigma^* \rightarrow \Gamma^*$ is said to have ℓ -lagged distance δ if the (relative) ℓ -lagged distance $\delta_{\text{TC}}^{\text{Lag}^\ell}$ (defined below) is at least δ . It is said to have (ℓ, L) -lagged distance δ if the (relative) (ℓ, L) -lagged distance $\delta_{\text{TC}}^{\text{Lag}^{(\ell, L)}}$ (defined below) is at least δ .*

$$\delta_{\text{TC}}^{\text{Lag}^\ell} := \inf_{n, x \neq x' \in \Sigma^n, n - \text{split}(x, x') \geq \ell} \frac{\Delta(\text{TC}(x), \text{TC}(x'))}{n - \text{split}(x, x')},$$

$$\delta_{\text{TC}}^{\text{Lag}^{(\ell, L)}} := \inf_{n, x \neq x' \in \Sigma^n, \ell \leq n - \text{split}(x, x') \leq L} \frac{\Delta(\text{TC}(x), \text{TC}(x'))}{n - \text{split}(x, x')}.$$

The truncated versions of these lagged codes are defined appropriately.

In other words, lagged distance is a relaxation of the standard notion of distance when we consider only pairs of strings x, x' whose $\text{split}(x, x')$ lies in a particular interval.

The rest of the construction proceeds as follows. In step 1, we show that composing the integer tree code in [Corollary 3.1](#) with a standard error-correcting code yields a truncated tree code with constant ℓ -lagged distance for some $\ell = \ell(s)$. In step 2, we show that we can then obtain an untruncated code with constant (ℓ, L) -lagged distance by placing interleaving copies of the truncated code obtained in step 1 along the infinite tree for some fixed function $L = L(\ell)$. Finally, in step 3, we show that a superimposition of several (ℓ, L) -lagged distance codes for increasing values of ℓ and L yields our final code. The size of the final alphabet is determined by how fast the function $L = L(\ell)$ grows which is in turn determined by the function $\ell = \ell(s)$.

We begin by stating some basic error-correcting code that we need for step 1.

Proposition 3.3. *For all $\delta \in [0, 1)$ there are constants s_δ and c_δ such that for all $s \geq s_\delta$ there is an explicit code $\mathcal{C}: \{0, 1\}^{3s} \rightarrow (\{0, 1\}^{c_\delta})^s$ with (relative) distance at least δ .*

Proof. Using standard error-correcting codes, we know that for every $\delta \in [0, 1)$, there are constants $n_\delta \in \mathbb{Z}_{>0}$, $\alpha_\delta \in (0, 1)$ and alphabet Σ_δ such that there exist explicit codes $\tilde{\mathcal{C}}: \{0, 1\}^n \rightarrow \Sigma_\delta^{n/\alpha_\delta}$ for all $n \geq n_\delta$ with distance δ over the alphabet Σ_δ ⁴.

Set $s_\delta := n_\delta/3$ and for $s \geq s_\delta$, let $\tilde{\mathcal{C}}: \{0, 1\}^{3s} \rightarrow \Sigma_\delta^{3s/\alpha_\delta}$ be the above code for $n = 3s$. Choose c_δ such that $2^{c_\delta} \geq |\Sigma_\delta|^{3/\alpha_\delta}$ so that there exists an injective mapping $\Sigma_\delta^{3/\alpha_\delta} \hookrightarrow \{0, 1\}^{c_\delta}$. Under this mapping, the code $\tilde{\mathcal{C}}: \{0, 1\}^{3s} \rightarrow \Sigma_\delta^{3s/\alpha_\delta}$ can be viewed as $\mathcal{C}: \{0, 1\}^{3s} \rightarrow (\{0, 1\}^{c_\delta})^s$. Clearly, the distance of \mathcal{C} is at least that of $\tilde{\mathcal{C}}$. \square

⁴There are several ways to obtain such codes. Take a constant-rate-constant-distance code and distance amplify it using the ABNNR distance amplification technique. Or take Algebraic Geometric codes.

Step 1: truncated code with constant ℓ -lagged distance: We describe the truncated-lagged tree code

$$\text{TC}_{\text{Lag } \ell}^{(s)}: \{0, 1\}^{(\leq s^2)} \rightarrow \{\{0, 1\}^{c_\delta}\}^{(\leq s^2)}$$

that results by composing the tree code, $\text{TC}^{(s)}: (\{0, 1\}^s)^{(\leq s)} \rightarrow (\{0, 1\}^{3s})^{(\leq s)}$, in [Corollary 3.1](#), which is a distance half tree code, with the code \mathcal{C} from [Proposition 3.3](#).

Formally, let \mathcal{C} be the code from [Proposition 3.3](#) with output alphabet $\{0, 1\}^{c_\delta}$ and distance δ and $a \in \mathbb{N}$ a large constant to be determined later. For every $s \geq s_\delta$ and $\ell \geq as^5$ we describe an explicit truncated- ℓ -lagged tree code over the binary input alphabet, $\text{TC}_{\text{Lag } \ell}^{(s)}: \{0, 1\}^{(\leq s^2)} \rightarrow \{\{0, 1\}^{c_\delta}\}^{(\leq s^2)}$ with ℓ -lagged distance at least $\delta(1/2 - 3/2a)$.

The encoding process is as follows. Consider any $x \in \{0, 1\}^k$ and let $x' = (x'_1, \dots, x'_{\lfloor k/s \rfloor}) \in (\{0, 1\}^s)^{\lfloor k/s \rfloor}$ obtained by grouping together blocks of s symbols and ignoring any leftover symbols. Let $y' = (y'_1, \dots, y'_{\lfloor k/s \rfloor}) = \text{TC}^{(s)}(x')$, so $y' \in (\{0, 1\}^{3s})^{\lfloor k/s \rfloor}$ and $\mathcal{C}(y') = (\mathcal{C}(y'_1), \dots, \mathcal{C}(y'_{\lfloor k/s \rfloor})) \in (\{0, 1\}^{c_\delta})^{\lfloor k/s \rfloor}$. We construct $\text{TC}_{\text{Lag } \ell}^{(s)}(x)$ by letting for all $i \in \{1, \dots, s\}$ and for all $1 \leq j \leq \lfloor k/s \rfloor$ such that $js + i - 1 \leq k$:

$$\text{TC}_{\text{Lag } \ell}^{(s)}(x)_{js+i-1} = (\mathcal{C}(y'_j))_i.$$

In other words, we read groups of s symbols from $x \in \{0, 1\}^k$, say $x|_{[js+1, (j+1)s]}$ and interpret each block of s symbols as a step in $\text{TC}^{(s)}$. Thus, we obtain a symbol in $\{0, 1\}^{3s}$ which we then encode to a symbol in $\{0, 1\}^{c_\delta}$ using \mathcal{C} . Then we "write" this s -character long string on the appropriate s -length segment of $\text{TC}_{\text{Lag } \ell}^{(s)}(x)$, i.e., from $(j+1)s$ to $(j+2)s - 1$.

Proposition 3.4. *For all $x \in \{0, 1\}^k$ the encoding procedure of $\text{TC}_{\text{Lag } \ell}^{(s)}(x)$ is online and $\text{TC}_{\text{Lag } \ell}^{(s)}(x)_{[1, s-1]} = \varepsilon$ where ε is the empty string.*

Proof. The online nature of the encoding is clear by construction. For the latter part notice that we don't begin "writing" the encoding of $\text{TC}_{\text{Lag } \ell}^{(s)}(x)$ until we see the first block of s symbols and hence the encoding starts only from $\text{TC}_{\text{Lag } \ell}^{(s)}(x)_s$. \square

Lemma 3.5. $\text{TC}_{\text{Lag } \ell}^{(s)}: \{0, 1\}^{(\leq s^2)} \rightarrow \{\{0, 1\}^{c_\delta}\}^{(\leq s^2)}$ has ℓ -lagged distance at least $\delta(1/2 - 3/2a)$.

Proof. We need to show that for all $u, v \in \{0, 1\}^k$ such that $b = k - \text{split}(u, v) \geq \ell$ we have

$$\frac{\Delta(\text{TC}_{\text{Lag } \ell}^{(s)}(u), \text{TC}_{\text{Lag } \ell}^{(s)}(v))}{b} \geq \delta(1/2 - 3/2a).$$

Notice that $|u'| - \text{split}(u', v') \geq \lfloor b/s \rfloor$ where u' and v' are analogues of x' for u and v respectively. By distance of tree code $\text{TC}^{(s)}$, we have $\Delta(\text{TC}^{(s)}(u'), \text{TC}^{(s)}(v')) \geq \frac{\lfloor b/s \rfloor}{2}$. By composition with \mathcal{C} we obtain that $\Delta(\text{TC}_{\text{Lag } \ell}^{(s)}(u), \text{TC}_{\text{Lag } \ell}^{(s)}(v)) \geq \delta s(\lfloor b/s \rfloor \cdot 1/2 - 1)$. The $\lfloor b/s \rfloor - 1$ term arises as we may not have enough length to encode $\mathcal{C}(u')|_{\lfloor k/s \rfloor}$, i.e., it may be that $k < \lfloor k/s \rfloor s + s - 1$.

$$\text{Thus, } \frac{\Delta(\text{TC}_{\text{Lag } \ell}^{(s)}(u), \text{TC}_{\text{Lag } \ell}^{(s)}(v))}{b} \geq \frac{\delta s(\lfloor b/s \rfloor \cdot 1/2 - 1)}{b} \geq \delta(1/2 - 3/2a) \text{ since } b \geq \ell \geq as. \quad \square$$

Step 2: untruncated code with constant $(\ell, O(\ell^2))$ -lagged distance: We use interleaving copies of the above truncated- ℓ -lagged distance $\delta(1/2 - 3/2a)$ tree code to construct a un-truncated- $(\ell, O(\ell^2))$ -lagged distance $\delta(1/2 - 3/2a)$ tree code. Formally, let \mathcal{C} be the code from [Proposition 3.3](#) with output alphabet $\{0, 1\}^{c_\delta}$ and distance δ . For all $\ell \geq as$ (where $s \geq s_\delta$ as in [Proposition 3.3](#)), we describe an explicit- $(\ell, O(\ell^2))$ -lagged tree code over the binary input alphabet and output alphabet $\{0, 1\}^{2c_\delta}$,

$$\text{TC}_{\text{Lag } \ell}: \{0, 1\}^* \rightarrow (\{0, 1\}^{2c_\delta})^*.$$

⁵We think of a as a constant and of ℓ and s as growing.

Consider any $x \in \{0,1\}^k$. We describe the encoding of x , i.e., $\text{TC}_{\text{Lag } \ell}(x)$ below. Let $i \leq k$ be $j(s^2/2) + r$ where $r < s^2/2$.⁶ Then,

$$\text{TC}_{\text{Lag } \ell}(x)_i = (\text{TC}_{\text{Lag } \ell}^{(s)}(x|_{[(j-1)s^2/2, i]}), \text{TC}_{\text{Lag } \ell}^{(s)}(x|_{[(j)s^2/2, i]})).$$

Notice that if $j = 0$ then $x|_{[(j-1)s^2/2, i]}$ is the empty string.

In other words, we place a ℓ -lagged tree code $\text{TC}_{\text{Lag } \ell}^{(s)}: \{0,1\}^{s^2} \rightarrow \{\{0,1\}^{c_\delta}\}^{s^2}$ at intervals of length $s^2/2$. So, for every $x \in \{0,1\}^k$ and $i > s^2/2$ there are two ℓ -lagged tree codes whose encodings are "written" on $\text{TC}_{\text{Lag } \ell}(x)_i$.

Proposition 3.6. *For all $x \in \{0,1\}^k$ the encoding procedure of $\text{TC}_{\text{Lag } \ell}(x)$ is online and $\text{TC}_{\text{Lag } \ell}(x)|_{[1, s-1]} = \varepsilon$ where ε is the empty string.*

Proof. The online nature of the encoding is by construction.

$\text{TC}_{\text{Lag } \ell}(x)|_{[1, s-1]} = \varepsilon$ follows from $\text{TC}_{\text{Lag } \ell}^{(s)}(x)|_{[1, s-1]} = \varepsilon$ as shown in [Proposition 3.4](#). \square

Lemma 3.7. $\text{TC}_{\text{Lag } \ell}: \{0,1\}^* \rightarrow (\{0,1\}^{2c_\delta})^*$ has $(\ell, O(\ell^2))$ -lagged distance at least $\delta(1/2 - 3/2a)$.

Proof. We show that for all $u, v \in \{0,1\}^k$ such that $b = k - \text{split}(u, v) \geq \ell$ and $b \leq s^2/2$ we have $\frac{\delta(\text{TC}_{\text{Lag } \ell}(u), \text{TC}_{\text{Lag } \ell}(v))}{b} \geq \delta(1/2 - 3/2a)$. By construction we know that there is some j such that $\text{split}(u, v) - js^2/2 \leq s^2/2$ and $k - js^2/2 \leq s^2$. Thus the encoding of u and v which include $\text{TC}_{\text{Lag } \ell}^{(s)}(u|_{[js^2/2, k]})$ and $\text{TC}_{\text{Lag } \ell}^{(s)}(v|_{[js^2/2, k]})$ differ on at least $\delta(1/2 - 3/2a)b$ locations by virtue of $\text{TC}_{\text{Lag } \ell}^{(s)}$. Thus, $\text{TC}_{\text{Lag } \ell}$ is a $(\ell, s^2/2)$ -lagged tree code with distance $\delta(1/2 - 3/2a)$. \square

Step 3: tree code with polylogarithmic output alphabet Finally, we superimpose several copies of the above $(\ell, O(\ell^2))$ -lagged tree code for varying ℓ to get an explicit family of constant distance n -truncated tree codes, $\{\text{TC}^{(n)}\}_n$ with binary input alphabet and output alphabet of size $O(\text{polylog}(n))$. Formally, there exists a family of alphabets $\{\Gamma_n\}_n$ such that $|\Gamma_n| = \text{polylog}(n)$ with the following property. There exists a constant-distance tree code $\text{TC}^{(n)}: \{0,1\}^{(\leq n)} \rightarrow \Gamma_n^{(\leq n)}$ such that the encoding functions are constructible in time polynomial in n . $\text{TC}^{(n)}$ is a superimposition of $\text{polylog}(n)$ many copies of the $(\ell, O(\ell^2))$ -lagged tree code for varying ℓ such that for any two strings u, v , it is always the case that $\text{split}(u, v)$ lies in the stipulated range for one of these codes which witnesses the distance. We describe $\text{TC}^{(n)}$ formally below.

For concreteness let us take $\delta = 1/4$, $a = 6$ and $\ell = 6s$ in [Lemma 3.7](#). Then for all $s \geq s_\delta$ we have a $(6s, s^2/2)$ -lagged tree code with distance $1/16$. In other words, for $\ell \geq 6s_{1/4}$ we have a distance- $1/16 - (\ell, \ell^2/72)$ -lagged tree code. To construct $\text{TC}^{(n)}$ we will superimpose the encodings of $\text{TC}_{\text{Lag } \ell_i}$ for all $i \in \{1, \dots, j\}$ where $\ell_1 = 6s_{1/4}$ and $\ell_{i+1} = \ell_i^2/72$ and $\ell_j^2/72 = n$, thus $j = O(\log \log(n))$.⁷ More precisely, for all $x \in \{0,1\}^k$ let

$$\text{TC}^{(n)}(x)_i = ((x|_{[i-6s_{1/4}, i]}), (\text{TC}_{\text{Lag } \ell_1}(x))_i, \dots, (\text{TC}_{\text{Lag } \ell_j}(x))_i).$$

Hence, $\text{TC}^{(n)}(x) \in \Gamma_n^k$ where $|\Gamma_n| = O(\text{polylog}(n))$ and the encoding functions take time polynomial in n since each coordinate is computable in time polynomial in n .

Remark 3.8. The above construction yields a n -truncated tree code. To obtain a honest-to-god untruncated tree code, we observe that the Cohen-Haeupler-Schulman construction can be done incrementally in the following sense. [Propositions 3.4, 3.6](#) and [3.9](#) state that the initial part of the tree code encodings are empty. Hence, we might begin with a constant-sized alphabet (without printing the ε) and grow the alphabet as and when needed

⁶We assume that s is even for simplicity.

⁷We ignore divisibility issues for the sake of simplicity.

Proposition 3.9. For all $x \in \{0, 1\}^n$, $i \in [n]$ and g such that $\ell_g > 6(i + 1)$ we have $\text{TC}_{\text{Lag } \ell_g}(x)_i = \varepsilon$. Hence, $\text{TC}^{(n)}(x)_i \in \Gamma_i$ where $|\Gamma_i| = O(\text{polylog}(i))$.

Proof. From [Proposition 3.6](#) (with $a = 6$) we know that $\text{TC}_{\text{Lag } \ell_i}(x)_i = \varepsilon$. In particular, we have that

$$\text{TC}^{(n)}(x)_i = ((x|_{[i-6s_{1/4}, i]}), (\text{TC}_{\text{Lag } \ell_1}(x))_i, \dots, (\text{TC}_{\text{Lag } \ell_g}(x))_i, \varepsilon, \dots, \varepsilon).$$

□

Putting the above together we get that there exists a family of alphabets $\{\Gamma_n\}_n$ such that $|\Gamma_n| = O_\eta(\text{polylog}(n))$ with the following property. Let $\Gamma = \cup_n \Gamma_n$. There exists a tree code $\text{TC}: \{0, 1\}^* \rightarrow \Gamma^*$ such that the encoding functions φ_i satisfy $\varphi_i: \{0, 1\}^i \rightarrow \Gamma_i$. Furthermore, the encoding functions φ_n are constructible in time polynomial in n .

Theorem 3.10. There exists a family of alphabets $\{\Gamma_n\}_n$ such that $|\Gamma_n| = O_\eta(\text{polylog}(n))$ with the following property. Let $\Gamma = \cup_n \Gamma_n$. There exists a tree code $\text{TC}: \{0, 1\}^* \rightarrow \Gamma^*$ with distance $1/16$ such that the encoding functions φ_i satisfy $\varphi_i: \{0, 1\}^i \rightarrow \Gamma_i$. Furthermore, the encoding functions φ_n are constructible in time polynomial in n .

Proof. The tree code we consider is the tree code TC as mentioned above the theorem statement. We have already shown that $\text{TC}: \{0, 1\}^* \rightarrow \Gamma^*$ is such that the encoding functions φ_i satisfy $\varphi_i: \{0, 1\}^i \rightarrow \Gamma_i$. Furthermore, the encoding functions φ_n are constructible in time polynomial in n .

Next, we show that the distance of $\text{TC}^{(n)}$ is at least $1/16$. Consider $u, v \in \{0, 1\}^k$ and let $b = k - \text{split}(u, v)$. If $b \leq 6s_{1/4}$ then for all $i \in \{\text{split}(u, v) + 1, k\}$ we have $(\text{TC}^{(n)}(u))_i \neq (\text{TC}^{(n)}(v))_i$ as they differ in the first coordinate. Now, suppose $\ell_i \leq b \leq \ell_i^2/72$ for some $1 \leq i \leq j$. Then, by virtue of $\text{TC}_{\text{Lag } \ell_i}$ we know that $\frac{\Delta(\text{TC}^{(n)}(u), \text{TC}^{(n)}(v))}{b} \geq 1/16$. In any case, we have $\delta_{\text{TC}^{(n)}} \geq 1/16$. □

Notice that if we use the boosting trick mentioned in [Theorem 2.6](#) we obtain a distance $r/(r+s)$ tree code over the integers for any $r, s \in \mathbb{N}$. If we use this tree code in the procedure detailed in the proof of [Theorem 3.10](#) we obtain $\text{TC}^{(n)}$ with distance $(1 - O(\frac{r}{r+s} - O(\frac{1}{a})))\delta$ where δ is the distance of \mathcal{C} in [Proposition 3.3](#) and $a = \ell/s$ as describe in the proof of [Lemma 3.5](#). Hence, by choosing $r/(r+s)$ and δ close to 1 and a large enough we have established [Theorem 1.2](#), which is restated below for convenience.

Theorem 1.2 (explicit tree codes with polylog alphabet [[CHS18](#)]). For all $\eta \in [0, 1)$ there exists a family of alphabets $\{\Gamma_n\}_n$ such that $|\Gamma_n| = O_\eta(\text{polylog}(n))$ with the following property. Let $\Gamma = \cup_n \Gamma_n$. There exists a tree code $\text{TC}: \{0, 1\}^* \rightarrow \Gamma^*$ with distance η such that the encoding functions φ_i satisfy $\varphi_i: \{0, 1\}^i \rightarrow \Gamma_i$. Furthermore, the encoding functions φ_n are constructible in time polynomial in n .

4 Open Questions

Recall that the $\text{TC}_{\mathbb{Z}}$ of [Theorem 2.9](#) used in [Corollary 3.1](#) was an explicit tree code over the integers $\text{TC}_{\mathbb{Z}}: \mathbb{Z}_{\geq 0}^{\mathbb{Z}_{\geq 0}} \rightarrow (\mathbb{Z}_{\geq 0}^2)^{\mathbb{Z}_{\geq 0}}$ with distance at least $1/2$ satisfying the property that for any $n \in \mathbb{Z}_{>0}$ and $i \in \{0, 1, \dots, n\}$, we have $|(\text{TC}_{\mathbb{Z}}(a_0, \dots, a_n))_i| \leq 2^n \max_j |a_j|^2$. This in turn was used in step 1 of the reduction process to small alphabets. If instead our $\text{TC}_{\mathbb{Z}}$ had the property that $|(\text{TC}_{\mathbb{Z}}(a_0, \dots, a_n))_i| \leq 2^{\text{polylog}(n)} \max_j |a_j|^2$, then the above reduction process to small alphabets would actually yield constant-sized alphabets.⁸ Since, $\text{TC}_{\mathbb{Z}}$ of [Theorem 2.9](#) was obtained using a totally-non-singular lower-triangular matrix (specifically the Pascal Matrix), we are led to the following question about the existence of such $n \times n$ matrices with entries only $2^{\text{polylog}(n)}$ which if true would imply tree codes with constant distance and constant-sized alphabets. Furthermore, if the matrices are explicit then we would get explicit tree codes with constant distance and constant-sized alphabets.

⁸Strictly speaking, $|(\text{TC}_{\mathbb{Z}}(a_0, \dots, a_n))_i| \leq 2^{f(n)} \max_j |a_j|^2$ yields a tree code whose (relative) $(2f(n), nf(n))$ -lagged distance is constant. Thus, any function $f(n)$ would suffice for constant-sized alphabets as long as we can show that there exists a constant c such that $\underbrace{g \circ g \circ \dots \circ g}_{c\text{-times}}(n) = O(\log \log(n))$ where $g(nf(n)) = 2f(n)$.

Question 1 (Existence of totally-non-singular lower-triangular matrices over small alphabets). *Are there $n \times n$ totally-non-singular lower-triangular matrices as defined in [Definition 2.3](#) with entries of size $O(2^{\text{polylog}(n)})$. Further, are there such matrices which are explicit? (Pudlák [[Pud16](#), Problem 1] states this question asking whether the entries of such a matrix can be $\text{poly}(n)$.)*

The next question asks about the decoding process to accompany the constant-distance tree codes of [Theorem 1.2](#).

Question 2 (Efficient decoding of tree codes). *Is there an efficient decoding procedure for the tree codes of [Theorem 1.2](#) which corrects a constant fraction of errors?*

There has been recent work on this question by Narayanan and Weidner [[NW20](#)]. The authors use a randomized polynomial time decoding algorithm to correct errors which scale roughly as the block length to the three-fourths power, falling short of the constant fraction error correction guaranteed by the constant distance.

Acknowledgements

We thank Madhu Sudan for various discussions while preparing this exposition, in particular, we thank him for suggesting the alternate construction of distance $1/2$ -tree codes over the integers mentioned after [Claim 2.5](#).

References

- [CHS18] GIL COHEN, BERNHARD HAEUPLER, and LEONARD J. SCHULMAN. *Explicit binary tree codes with poly-logarithmic size alphabet*. In *Proc. 50th ACM Symp. on Theory of Computing (STOC)*, pages 535–544. 2018. [eccc:2018/TR18-032](#). 2, 5, 6, 9
- [Gel17] RAN GELLES. *Coding for interactive communication: A survey*. *Foundations and Trends in Theoretical Computer Science*, 13(1-2):1–157, 2017. [doi:10.1561/04000000079](#). 2
- [GV85] IRA GESSEL and GÉRARD VIENNOT. *Binomial determinants, paths, and hook length formulae*. *Adv. Math.*, 58(3):300 – 321, 1985. 5
- [Lin73] BERNT LINDSTRÖM. *On the vector representations of induced matroids*. *Bull. Lond. Math. Soc.*, 5(1):85–90, 1973. 5
- [NW20] ANAND KUMAR NARAYANAN and MATTHEW WEIDNER. *On decoding Cohen-Haeupler-Schulman tree codes*. In *Proc. 31st Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 1337–1356. 2020. [arXiv:1909.07413](#). 10
- [Pud16] PAVEL PUDLÁK. *Linear tree codes and the problem of explicit constructions*. *Linear Algebra and its Applications*, 490:124–144, 2016. [arXiv:1310.5684](#). 2, 3, 4, 6, 10
- [Sch96] LEONARD J. SCHULMAN. *Coding for interactive communication*. *IEEE Trans. Inform. Theory*, 42(6):1745–1756, 1996. (Preliminary version in 33rd FOCS, 1992). 1, 3