

Generalisation of Kraft inequality for source coding into permutations

Kristo Visk and Ago-Erik Riet

Institute of Mathematics and Statistics

Faculty of Science and Technology, University of Tartu, J. Liivi 2, Tartu 50409, Estonia

Abstract—We develop a general framework to prove Kraft-type inequalities for prefix-free permutation codes for source coding with various notions of permutation code and prefix. We also show that the McMillan-type converse theorem in most of these cases fails, and give a general form of a counterexample. Our approach is more general and works for other structures besides permutation codes. The classical Kraft inequality for prefix-free codes and results about permutation codes follow as corollaries of our main theorem and main counterexample.

Index Terms—Permutation codes, source coding, Kraft inequality, LYM inequality, prefix-free, Ulam distance, subsequence-free, permutation pattern.

I. INTRODUCTION

Permutation codes with different distance metrics are applicable to flash memories. Coding each block of multi-level memory cells into permutations and managing injections of electric charges, one can reduce drift and overshoot errors, using rank modulation [9]. Permutation codes for rank modulation [3], [6], [2], [15], [7], [8] provide further robustness. They were also proposed for powerline communications [18].

In this work, we study a question of unique decoding into permutation codes. This question is important for understanding theoretical limits for optimal (source) coding for storing information in flash memories. For other aspects of source coding with permutations see e.g. [20] and references therein.

To this end, we prove a generalisation (Theorem V.2) of Bollobás-Lubell-Yamamoto-Meshalkin (LYM) [4], [13], [17], [21] and Kraft [11] inequalities, which holds for certain graded posets. In Corollaries VI.3 and VI.4 we obtain a Kraft-type inequality for prefix-free (permutation) codes in different contexts, where we give several definitions of permutation codes and what it means to be a ‘prefix’. From our general counterexample (Theorem VI.2), in Corollaries VI.3 and VI.4 we obtain that a McMillan-type converse theorem fails in most of these cases, but not for the case of classical notion of prefix.

II. NOTATION AND DEFINITIONS

Denote $[k] := \{1, 2, \dots, k\}$, let $\mathbb{N} := \{1, 2, \dots\}$ and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. For a set S , its cardinality is denoted $\#S$, sometimes also written $|S|$.

An *alphabet* is a finite set $S := \{s_i \mid i \in [n]\}$ of cardinality $n \neq 0$. A *symbol* is an element $s \in S$, a finite sequence of symbols $s_1 \dots s_k$ is a *string*. The *length* of a string is the number of symbols it consists of. Let $S^l := \{s_1 \dots s_l \mid s_j \in S \text{ for all } j \in [l]\}$ denote the set of strings of length l . Define $S^0 := \{\varepsilon\}$ where ε is the unique string of length 0 called

the *empty string*. Let $S^* := \bigcup_{j \in \mathbb{N}_0} S^j = S^0 \cup S^1 \cup \dots$ be the set of all finite strings. For strings $t, u \in S^*$ where $t = s_{i_1} s_{i_2} \dots s_{i_k} \in S^k$ and $u = s_{j_1} s_{j_2} \dots s_{j_r} \in S^r$ their *concatenation* is $tu = s_{i_1} \dots s_{i_k} s_{j_1} \dots s_{j_r} \in S^{k+r}$. A string t is a *prefix* to the string u if $u = tw$ for some string w . If $w \neq \varepsilon$, then t is called a *proper prefix* of u .

A *permutation* of the set $[k]$ is a bijection $\sigma : [k] \rightarrow [k]$. To denote permutations $\sigma : i \mapsto b_i$ ($i \in [k]$) we use vectors (b_1, \dots, b_k) as well as strings $b_1 \dots b_k$. We also denote $\mathbb{S}_k := \{\sigma \mid \sigma : [k] \rightarrow [k] \text{ is a bijection}\}$ for the set (in fact, group) of permutations on $[k]$, called the *symmetric group of degree k* . For example $\sigma = \mathbf{2314} = (\mathbf{2}, \mathbf{3}, \mathbf{1}, \mathbf{4}) \in \mathbb{S}_4$ is the permutation $\sigma(\mathbf{1}) = \mathbf{2}$, $\sigma(\mathbf{2}) = \mathbf{3}$, $\sigma(\mathbf{3}) = \mathbf{1}$, $\sigma(\mathbf{4}) = \mathbf{4}$.

We use notation

$$T_k^l := \{(n_1, \dots, n_l) \mid \forall j : n_j \in [k]; i \neq j \Rightarrow n_i \neq n_j\}$$

for the set of l -element *partial permutations* on the set $[k]$, i.e. the injective mappings $\tau : [l] \rightarrow [k]$. Let

$$T_k := T_k^1 \cup T_k^2 \cup \dots \cup T_k^k$$

and write

$$\begin{aligned} \mathbb{T}_k &:= \{(n_1, \dots, n_l) \mid \forall j : n_j \in [l], l \leq k; i \neq j \Rightarrow n_i \neq n_j\} \\ &= \mathbb{S}_1 \cup \mathbb{S}_2 \cup \dots \cup \mathbb{S}_k. \end{aligned}$$

We say $\sigma \in \mathbb{S}_l$ is the *pattern* of $\tau \in T_k^l$ if the relative ordering of symbols is the same, i.e. if, for all $i, j \in [l]$, we have $\sigma(i) < \sigma(j)$ if and only if $\tau(i) < \tau(j)$. A permutation $\sigma \in \mathbb{S}_l$ is a *pattern* in a partial permutation $\tau \in T_k^m$ if there are indices $1 \leq b_1 < b_2 < \dots < b_l \leq m$ such that, for all $i, j \in [l]$, $\sigma(i) < \sigma(j)$ if and only if $\tau(b_i) < \tau(b_j)$. We call $\sigma \in T_k^m$ a (not necessarily consecutive) *subsequence* of $\tau \in T_k^l$ if $m \leq l$ and there are m indices $1 \leq b_1 < b_2 < \dots < b_m \leq l$ with $\sigma(i) = \tau(b_i)$ for all $i \in [m]$. Note that σ is a pattern in $\tau \in T_k$ if and only if it is the pattern of a subsequence of τ .

For example $\mathbf{253} \in T_6^3$ is a subsequence in $\mathbf{2513} \in T_6^3$, the pattern of $\mathbf{253} \in T_6^3$ is $\mathbf{132} \in \mathbb{S}_3$ and thus $\mathbf{132} \in \mathbb{S}_3$ is a pattern in $\mathbf{2513} \in T_6^3$.

We call $\sigma \in T_k^m$ a (consecutive) *substring* of $\tau \in T_k^l$ if $m \leq l$ and there is $0 \leq n \leq l - m$ with $\sigma(i) = \tau(i + n)$ for all $i \in [m]$. If σ is a substring of τ then it is also a subsequence.

For example $\mathbf{51} \in T_6^2$ is a substring in $\mathbf{2513} \in T_6^4$ but the subsequence $\mathbf{253} \in T_6^3$ is a **not** a substring in $\mathbf{2513} \in T_6^4$.

We say $\sigma \in \mathbb{S}_m$ is a *substring pattern* of $\tau \in T_k$ if it is the pattern of a substring of τ . So $\mathbf{21} \in \mathbb{S}_2$ is a substring pattern in $\mathbf{2513} \in T_6^4$ since it is the pattern of the substring $\mathbf{51} \in T_6^2$.

A *partially ordered set* or a *poset* is a set P together with a binary relation \leq that is reflexive, transitive and antisymmetric, i.e. for all $a \in P$, $a \leq a$, for all $a, b, c \in P$, if $a \leq b$ and $b \leq c$ then $a \leq c$ and for all $a, b \in P$, if $a \leq b$ and $b \leq a$ then $a = b$. Write $a < b$ for $a \leq b$ and $a \neq b$. We say b *covers* a if $a < b$ and there is no $c \in P$ with $a < c$ and $c < b$. An element $a \in P$ is *minimal* if there is no $b \in P$ with $b < a$. A *graded poset* is a poset P with a *rank function* $\rho : P \rightarrow \mathbb{N}_0$ satisfying $\rho(c) = 0$ for all minimal $c \in P$, and, $\rho(b) = \rho(a) + 1$ if b covers a , and, if $a < b$ then $\rho(a) < \rho(b)$; say elements of the same rank are on the same *level*. An *antichain* is a subset $A \subseteq P$ whose elements are pairwise incomparable, i.e. $a \not\leq b$ for all $a, b \in A$.

III. CODES, THEIR PREFIX-FREENESS AND UNIQUE DECODABILITY

Let S_1 and S_2 be alphabets. A (non-singular, classical) *code* is an injective map $c : S_1 \rightarrow S_2^*$. An image $c(s)$ is called a *codeword* corresponding to symbol $s \in S_1$. The definition of a code extends to all strings as $c(s_1 \dots s_n) := c(s_1) \dots c(s_n)$. It is easily checked the new map $c : S_1^* \rightarrow S_2^*$, called the *extension* of the code c , is well-defined. A code $c : S_1 \rightarrow S_2^*$ is *uniquely decodable* if its extension is injective, i.e. if $c(s_1 \dots s_n) = c(s'_1 \dots s'_m)$ implies $m = n$ and $s_j = s'_j$ for all $j \in [n]$. A code $c : S_1 \rightarrow S_2^*$ is *prefix-free* if there do not exist $s_i, s_j \in S_1$ with $c(s_i)$ a proper prefix of $c(s_j)$.

It is easy to see that a prefix-free code is uniquely decodable: we can read symbols in an output string $c(s_1 \dots s_n) = c(s_1) \dots c(s_n)$ from left to right, and prefix-freeness guarantees that no proper prefix of $c(s_1)$ is a codeword and also that $c(s_1)$ is not a proper prefix to any codeword. Hence, encountering the substring $c(s_1)$ at the beginning of the output $c(s_1 \dots s_n)$, we are guaranteed that it arose by encoding the symbol s_1 . We then continue by decoding $c(s_2 \dots s_n)$ similarly. Because of this property, prefix-free codes are also called *instantaneous*, see for example [5] (Ch. 5).

Let $c : S_1 \rightarrow S_2^*$ be a code. Then the sequence (a_0, a_1, \dots) , where $a_j := \#\{s \in S_1 \mid c(s) \in S_2^j\}$, $j \in \mathbb{N}_0$, is the *parameter sequence* of the code c . The parameters count the codewords of each length.

We shall now state the **Kraft Inequality (Proposition III.1)** [11], [16], see also [5] (Ch. 5) from classical source coding, which holds for all uniquely decodable classical codes, in particular for all prefix-free classical codes.

Proposition III.1. *Let $c : S_1 \rightarrow S_2^*$ be a uniquely decodable classical code with parameter sequence (a_0, a_1, \dots) and $\#S_2 = r$. Then*

$$K_c := \sum_{i=0}^{\infty} \frac{a_i}{r^i} \leq 1. \quad (1)$$

The number K_c is the *Kraft number*, also known as the Kraft sum or the Kraft-McMillan number. Note that Proposition III.1 states that the sum of densities of used codewords of a fixed length, over all lengths, is at most 1.

McMillan [16] proved the following strong converse of Proposition III.1, known as **Converse McMillan Theorem (Proposition III.2)**:

Proposition III.2. *Let $r \in \mathbb{N}$. If non-negative integers $a_i \in \mathbb{N}_0$, $i \in \mathbb{N}_0$, are such that inequality (1) holds, then there exists a prefix-free code $c : S_1 \rightarrow S_2^*$ with $\#S_2 = r$ and parameter sequence (a_0, a_1, \dots) .*

We remark that its proof is the construction of a code by picking vertices in the r -ary code tree greedily in the quasi-lexicographic order, i.e. starting with vertices closer to the root and within a length going in the lexicographic order.

IV. PERMUTATION CODES

We shall define permutation codes in two ways, by restricting the output space of classical codes.

Definition IV.1. *Let S be an alphabet and $k \in \mathbb{N}$. We define a permutation code as an injection $c : S \rightarrow T_k$. Note that*

$$\#T_k = \sum_{l=1}^k \binom{k}{l} l!.$$

We have $\#S \leq \sum_{l=1}^k \binom{k}{l} l!$ because of the injectivity of c . The parameter sequence of the code is (a_0, a_1, \dots) , where $a_j := \#\{s \in S \mid c(s) \in T_k^j\}$, as for classical codes.

Definition IV.2. *A more restrictive definition of a permutation code is an injection $c : S \rightarrow \mathbb{T}_k$. Note that*

$$\#\mathbb{T}_k = \sum_{l=1}^k l!.$$

We have $\#S \leq \sum_{l=1}^k l!$ because of the injectivity of c . The parameter sequence of the code is then (a_0, a_1, \dots) where $a_j := \#\{s \in S \mid c(s) \in \mathbb{S}_j\}$.

A. Definitions of 'prefix-freeness' for permutation codes

Let us define the *permutation constant* of a permutation code c in these cases respectively as

$$P_c := \sum_{l=1}^k \frac{a_l}{\binom{k}{l} l!}, \quad \text{or,} \quad \mathbb{P}_c := \sum_{l=1}^k \frac{a_l}{l!}.$$

This is the analogue of the Kraft number from classical codes.

The extension, prefix-freeness and unique decodability of a permutation code $c : S \rightarrow T_k$ or $c : S \rightarrow \mathbb{T}_k$ is understood as that notion for the same code viewed as a code $c : S \rightarrow [k]^*$.

Now we give some notions analogous to prefix-freeness. A permutation code $c : S \rightarrow T_k$ or $c : S \rightarrow \mathbb{T}_k$ is *{subsequence-, substring-, pattern- or substring-pattern-}free* if there are no two different codewords $c(s_1) \neq c(s_2)$ such that $c(s_1)$ is respectively a *{subsequence, substring, pattern or substring pattern}* in $c(s_2)$. These notions can also be defined for classical codes $c : S_1 \rightarrow S_2^*$, with the subsequence- and substring-freeness being perhaps the more natural notions.

We shall see that often for prefix-free, subsequence-free, substring-free, pattern-free or substring-pattern-free permutation codes, $P_c \leq 1$, or, $\mathbb{P}_c \leq 1$, so we have an analogue of

Kraft Inequality (Proposition III.1): the sum of densities of used codewords of fixed length, over all lengths, is at most 1.

However, we shall also see that $P_c \leq 1$, or $\mathbb{P}_c \leq 1$ for given code parameters does not in general imply that a subsequence-free, substring-free, pattern-free or substring-pattern-free permutation code with these parameters exists. That is, in some of these cases there is no analogue of Converse McMillan Theorem (Proposition III.2).

V. A GENERALISATION OF THE LYM AND KRAFT INEQUALITIES

We shall state the Bollobás-Lubell-Yamamoto-Meshalkin inequality, also known as the **LYM inequality** [4], [13], [17], [21], see also [10] (Ch. 8). Consider the set of all subsets of a finite set $[n]$, denoted by $\mathcal{P}([n])$, partially ordered by the subset relation \subseteq . Denote $[n]^{(k)} := \{A \subseteq [n] \mid \#A = k\}$.

Proposition V.1. *Suppose $\mathcal{A} \subseteq \mathcal{P}([n])$ is an antichain, i.e. $A, B \in \mathcal{A}$ and $A \subseteq B$ implies $A = B$. Then*

$$\sum_{i=0}^n \frac{\#(\mathcal{A} \cap [n]^{(i)})}{\binom{n}{i}} \leq 1.$$

That is, the sum of densities of \mathcal{A} in each *level*, i.e. of each cardinality, summed over all levels, is at most 1.

Now we shall prove a common generalisation of Kraft Inequality (Proposition III.1) for prefix-free codes and LYM Inequality (Proposition V.1). As consequences we obtain some analogues of Kraft Inequality (Proposition III.1), namely $P_c \leq 1$, and, $\mathbb{P}_c \leq 1$, for some of prefix-free, subsequence-free, substring-free, pattern-free and substring-pattern-free permutation codes. We remark that our theorem follows from the so-called AZ identity for general finite posets [1] but our proof here is self-contained and useful to obtain our general counterexample to the converse. We also remark that our proof follows closely the known proof of the Proposition V.1 via the Local LYM inequality, and our framework of level-regular graded posets, to be defined, is chosen so that this proof still works, while being general enough for our corollaries.

We shall also investigate when a converse statement akin to Converse McMillan Theorem (Proposition III.2) can hold. For example for LYM Inequality (Proposition V.1) such a converse statement does not hold in general.

A. Level-regular graded posets

A *directed (multi)graph* $G = (V, E)$ is a set V , called its *vertex set*, together with a multiset $E \subseteq V \times V$ of ordered pairs of vertices, called its *edge set* (there may be multiple edges (u, v) for fixed $u, v \in V$). An element $v \in V$ is a *vertex* and an element $e \in E$ is an *edge*. An edge $e = (u, v)$ is *directed from u to v* or goes from u to v . A vertex v is a *neighbour* of a vertex u if there is an edge (u, v) or (v, u) . A directed graph is *weakly connected*, if its underlying undirected graph is connected, i.e., without regard to directions of edges, one can walk from any vertex to any other vertex along edges (we can walk from a vertex to any of its neighbours). The *up-degree* of a vertex u is $\#\{e \in E \mid \exists v \in V : e = (u, v)\}$,

i.e. the number of edges directed from u , and the *down-degree* of a vertex v is $\#\{e \in E \mid \exists u \in V : e = (u, v)\}$, i.e. the number of edges directed to v . Sometimes the up- or down-degree is just called degree. For a graph $G = (V, E)$ and a subset $V' \subseteq V$, the graph $G' = (V', E \cap (V' \times V'))$ is called an *induced subgraph* of G , i.e. we keep all edges with both endpoints in V' and only them; then G' is *induced by V'* .

The *Hasse diagram* of a graded poset P is a directed graph with vertex set P , and an edge from a to b if and only if b covers a ; it is drawn with elements of the same rank on the same horizontal level and elements of higher ranks higher.

Example V.1. See Figure 1 for the Hasse diagram of the poset of subsets of $\{1, 2\}$ with the subset relation \subseteq .

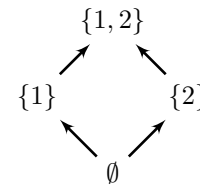


Fig. 1. The Hasse diagram of the poset of subsets of $\{1, 2\}$.

We shall consider a special kind of graded posets.

Let us say a graded poset P is *level-regular* if the bipartite (multi)graph induced by any two consecutive levels of its Hasse diagram is biregular — that is, all elements on the same level, i.e. with the same rank, are covered by an equal number of elements, and also cover an equal (perhaps different) number of elements (with multiplicity, i.e. we count edges).

Let $\mathcal{A} \subseteq P$ be any set of elements of the same rank, i.e. $\rho(a) = \rho(b)$ for all $a, b \in \mathcal{A}$. Then its *upper shadow* $\delta^+(\mathcal{A}) := \{b \mid \exists a \in \mathcal{A} : b \text{ covers } a\}$ is the set of all elements covering some element of \mathcal{A} , and its *lower shadow* $\delta^-(\mathcal{A}) := \{b \mid \exists a \in \mathcal{A} : a \text{ covers } b\}$ is defined analogously.

B. A common generalisation of the LYM and Kraft inequalities

Our main Theorem is the following generalisation of LYM (Proposition V.1) and Kraft (Proposition III.1) Inequalities.

Theorem V.2. *Let P be a level-regular graded poset and let $\mathcal{A} \subseteq P$ be an antichain. Denote $P^{(i)} := \{p \in P \mid \rho(p) = i\}$ and $\mathcal{A}^{(i)} := \{a \in \mathcal{A} \mid \rho(a) = i\}$. Assume all levels are finite, i.e. $\#P^{(i)} < \infty$ for each $i \in \mathbb{N}_0$. Define the LYM number of the antichain as*

$$L_{\mathcal{A}} := \sum_{i=0}^{\infty} \frac{\#\mathcal{A}^{(i)}}{\#P^{(i)}}.$$

Then

$$L_{\mathcal{A}} \leq 1.$$

Proof. First let us assume that the poset P is finite. Then the LYM number is a finite sum. We shall proceed by induction on $k := \max\{\rho(a) \mid a \in \mathcal{A}\}$. If $k = 0$ then $\#\mathcal{A}^{(i)} = 0$ for all $i > 0$ and

$$\sum_{i=0}^{\infty} \frac{\#\mathcal{A}^{(i)}}{\#P^{(i)}} = \frac{\#\mathcal{A}^{(0)}}{\#P^{(0)}} \leq 1,$$

as needed. Let us define $A' := \{a \in A \mid \rho(a) < k\} \cup \{p \in P \mid \rho(p) = k - 1, p \in \delta^-(A)\}$, i.e. replace the k -th level by its lower shadow. We shall prove that, on replacing A by A' , its LYM number does not decrease, i.e. $L_A \leq L_{A'}$, and that A' is still an antichain. The claim now follows by induction, since $L_{A'} \leq 1$ by the induction hypothesis.

Let $a, b \in A'$. We shall prove that $a \not\prec b$. If $a, b \in A$ then $a \not\prec b$ as A is an antichain. If $a, b \in P$ with $\rho(a) = \rho(b)$ then also $a \not\prec b$ by the definition of a graded poset. Suppose for a contradiction that $a < b$. The only way it might happen is with $a \in A$ and $b \in \delta^-(A)$ with $\rho(b) = k - 1$. But since b is in the lower shadow of $\{a \in A \mid \rho(a) = k\}$, there exists $c \in A$ with $b < c$. By transitivity, $a < c$ with $a, c \in A$ — a contradiction with A being an antichain. Hence A' is an antichain.

Note that $\delta^-(A^{(k)}) \cap A^{(k-1)} = \emptyset$, as A is an antichain. Hence, to prove that $L_{A'} \geq L_A$, it is enough to show that

Lemma V.3.

$$\frac{\#\delta^-(A^{(k)})}{\#P^{(k-1)}} \geq \frac{\#A^{(k)}}{\#P^{(k)}}.$$

This is the analogue of what is known as the Local LYM inequality, which reads that shadows have greater density.

Proof. We shall prove the Lemma by degree considerations of the Hasse diagram of levels k and $k - 1$. Let the down-degree, i.e. the number of elements it covers, of each $v \in P^{(k)}$ be d , and the up-degree, i.e. the number of elements covering it, of each $w \in P^{(k-1)}$ be u . The number of edges between the sets $P^{(k)}$ and $P^{(k-1)}$ is $d \cdot \#P^{(k)} = u \cdot \#P^{(k-1)}$. The number of edges in the subgraph induced by $A^{(k)}$ and $\delta^-(A^{(k)})$ is equal to $d \cdot \#A^{(k)}$ on the one hand and at most $u \cdot \#\delta^-(A^{(k)})$ on the other hand — a vertex of the original has d neighbours in the lower shadow and a vertex of the lower shadow has at most u neighbours in the original. Hence

$$\frac{\#P^{(k)}}{\#P^{(k-1)}} = \frac{u}{d} \geq \frac{\#A^{(k)}}{\#\delta^-(A^{(k)})},$$

and rearranging proves the Lemma. \square

This proves the Theorem if P is finite. To prove the infinite case, restrict the poset to levels up to N , for every $N \in \mathbb{N}$, and then by the finite case we have

$$L_A = \sum_{i=1}^{\infty} \frac{\#A^{(i)}}{\#P^{(i)}} = \lim_{N \rightarrow \infty} \sum_{i=1}^N \frac{\#A^{(i)}}{\#P^{(i)}} \leq 1.$$

\square

VI. FAILURE OF THE CONVERSE MCMILLAN THEOREM

We may ask about the general analogue of Proposition III.2:

Question VI.1. *Let P be an (infinite or finite) level-regular graded poset. Assume that all levels are finite, i.e. $\#P^{(i)} < \infty$ for each $i \in \mathbb{N}_0$. Let $a_i \in \mathbb{N}_0$ for each $i \in \mathbb{N}_0$. Assume*

$$\sum_{i=1}^{\infty} \frac{a_i}{\#P^{(i)}} \leq 1.$$

Is it true that then there exists an antichain $A \subseteq P$ with $\#A^{(i)} = a_i$ for each i ?

The answer is in general “No” but “Yes” for example if the Hasse diagram is a tree. The next Theorem gives the general form of our counterexample to Question VI.1. We can hope for a counterexample only when Lemma V.3 is not tight, i.e. when $\#\delta^-(A^{(k)})/\#P^{(k-1)} > \#A^{(k)}/\#P^{(k)}$.

Theorem VI.2. *Consider two consecutive levels $P^{(i)}$ and $P^{(i+1)}$ of a level-regular graded poset P . Suppose that the equal up-degrees from level i to level $i + 1$ are $u > 1$ and the equal down-degrees from level $i + 1$ to level i are $d > 1$, i.e. each element of $P^{(i)}$ is covered by u elements of $P^{(i+1)}$ and that each element of $P^{(i+1)}$ covers d elements of $P^{(i)}$. Assume that the graph, induced by levels i and $i + 1$ of the Hasse diagram, is weakly connected. Further assume that the greatest common divisor $\gcd(\#P^{(i)}, \#P^{(i+1)}) =: g > 1$.*

Define $a_i := \frac{g-1}{g}\#P^{(i)}$; $a_{i+1} := \frac{1}{g}\#P^{(i+1)}$; $a_j := 0$ for $j \neq i, i + 1$, and note that they are integers. Then

$$\sum_{i=1}^{\infty} \frac{a_i}{\#P^{(i)}} \leq 1$$

but there is no antichain A with $\#A^{(j)} = a_j$ for all $j \in \mathbb{N}_0$.

Proof. Fix any subset $A^{(i+1)} \subseteq P^{(i+1)}$ with $\#A^{(i+1)} = a_{i+1} = \frac{1}{g}\#P^{(i+1)}$. Consider the graph $G_i = (V_i, E_i)$ induced by $V_i := A^{(i+1)} \cup \delta^-(A^{(i+1)})$. It is a bipartite graph with parts $A^{(i+1)}$ and $\delta^-(A^{(i+1)})$, i.e. all edges $e \in E_i$ are of the form (v, w) with $v \in \delta^-(A^{(i+1)})$ and $w \in A^{(i+1)}$. The average of down-degrees of vertices in $A^{(i+1)}$ is equal to d since all neighbours of vertices of $A^{(i+1)}$ lying in the i -level of the Hasse diagram of P are contained in the graph G_i . The average of up-degrees of vertices in $\delta^-(A^{(i+1)})$, however, is strictly less than u . To see this, note that all degrees are at most u . But there is a vertex of degree less than u since not all neighbours of all vertices of $\delta^-(A^{(i+1)})$ in the graph induced by $P^{(i)} \cup P^{(i+1)}$ (where all these degrees are u) lie in the set $A^{(i+1)}$. Indeed, otherwise the graph induced by $P^{(i)} \cup P^{(i+1)}$ would not be weakly connected, with $\delta^-(A^{(i+1)}) \cup A^{(i+1)}$ being disconnected from the rest of the graph. Hence,

$$\frac{\#\delta^-(A^{(i+1)})}{\#P^{(i)}} > \frac{\#A^{(i+1)}}{\#P^{(i+1)}}.$$

Thus $a_i > \#(P^{(i)} \setminus \delta^-(A^{(i+1)}))$ and we cannot pick a_i elements to our antichain A on level $P^{(i)}$. Since the choice of a_{i+1} elements comprising $A^{(i+1)}$ was arbitrary, there is no antichain A with the required properties. \square

From Theorem V.2 follows Kraft inequality in the classical case (Proposition III.1) and for permutation codes (Corollaries VI.3 and VI.4), from Theorem VI.2 follow counterexamples to Converse McMillan Theorem III.2 for some of our relations other than prefix relation (Corollaries VI.3 and VI.4).

Corollary VI.3. *The Kraft inequality (1), $K_c \leq 1$, holds for a classical code $c : S_1 \rightarrow S_2^*$, $\#S_2 = r$, if it is prefix-free, subsequence-free or substring-free. For $r \geq 2$ there exist*

parameter sequences (a_0, a_1, \dots) for which $K_c \leq 1$ but there is no subsequence-free or substring-free code $c : S_1 \rightarrow S_2^*$.

Proof. Consider the poset of all possible codewords, with the prefix, subsequence or substring relation. In this poset, an antichain is precisely a prefix-, subsequence-, resp. substring-free code. It is a graded poset: the rank of a codeword is its length. It is level-regular: the number of codewords $c \in S_2^{l+1}$ covering a codeword of length l is $r = \#S_2$ for the prefix, $(l+1)r$ for the subsequence and $2r$ for the substring relation, and a codeword of length l covers 1 codeword of S_2^{l-1} for the prefix, l codewords for the subsequence and 2 codewords for the substring relation (with multiplicity: the addition or removal of a different symbol or in a different position may produce equal outputs). Hence Theorem V.2 proves Kraft Inequality III.1 for these codes.

Conversely, for the subsequence and substring relations, note that level 1 of the poset has r elements, level 2 has r^2 elements, with $\gcd(r, r^2) = r > 1$. The graph induced by levels 1 and 2 of the Hasse diagram is weakly connected: any string can be transformed into any other string of the same length by alternate adding and removing of symbols at the beginning or end. Hence Theorem VI.2 shows the analogue of Converse McMillan Theorem III.2 fails in these cases. \square

Corollary VI.4. *We have $P_c \leq 1$ for {prefix-, subsequence- or substring-}free permutation codes $c : S \rightarrow T_k$ and $\mathbb{P}_c \leq 1$ for {prefix-, subsequence-, substring-, pattern- or substring-pattern-}free permutation codes $c : S \rightarrow \mathbb{T}_k$. For all $k \geq 3$, there are parameter sequences (a_0, a_1, \dots) with $P_c \leq 1$ but no subsequence-free or substring-free permutation codes $c : S \rightarrow T_k$, and, parameter sequences (a_0, a_1, \dots) with $\mathbb{P}_c \leq 1$ but no pattern-free or substring-pattern-free codes $c : S \rightarrow \mathbb{T}_k$.*

Proof. An antichain in a poset with some relation R is precisely an R -free permutation code; rank is number of symbols in the (partial) permutation. Level-regularity for prefix, subsequence and substring relations follows as the number of ways to add or remove a symbol only depends on the number of symbols in the (partial) permutation. For level-regularity for pattern and substring-pattern relations define edge multiplicities via a new poset, inserting a new level between every pair of consecutive levels, i.e. look at the pattern relation in two steps: first pick a subsequence or substring in T_k and then consider its pattern — draw an edge for each combination of these two steps; the pattern of a partial permutation is unique, while a permutation in \mathbb{S}_l is the pattern of $\binom{k}{l}$ partial permutations in T_k^l . For the converse, note that 2 divides $\gcd(3!, 2!)$ and $\gcd(\binom{k}{3}3!, \binom{k}{2}2!)$, for $k \geq 3$, so none of these is 1; weak connectivity is seen by alternate adding and removing of symbols at the beginning and end for substring or subsequence level pairs. So Theorems V.2 and VI.2 apply. \square

The subsequence relation has relevance to Ulam codes [19], [7], [8]. A permutation code $c : S \rightarrow T_k$ has minimum Ulam distance d if and only if every string $s \in [k]^{k-d+1}$ is a subsequence in at most one codeword. For fixed-length codes this was explored for example in [12], [14], [7], [8].

VII. OPEN PROBLEMS

Does the analogue of Kraft Inequality III.1 hold with definitions of ‘unique decodability’, expanding the freeness definitions, for classical or permutation codes? Does there exist an analogue of Huffman coding for these types of codes?

VIII. ACKNOWLEDGEMENTS

This work was supported in part by the Estonian Research Council through the research grants PUT405, PUT620 and IUT20-57. The authors wish to thank Vitaly Skachek for helpful discussions.

REFERENCES

- [1] R. Ahlswede, and Z. Zhang, “An identity in combinatorial extremal theory,” *Advances in Mathematics*, vol. 80, no. 2, pp. 137–151, 1990.
- [2] A. Barg and A. Mazumdar, “Codes in permutations and error correction for rank modulation,” *IEEE Trans. on Inform. Theory*, vol. 56, no. 7, pp. 3158–3165, Jul. 2010.
- [3] I.F. Blake, G. Cohen, and M. Deza, “Coding with permutations,” *Information and Control*, vol. 43, no. 1, pp. 1–19, 1979.
- [4] B. Bollobás, “On generalized graphs,” *Acta Math. Acad. Sci. Hung.*, vol. 16, no. 3–4, pp. 447–452, 1965.
- [5] T. Cover, and J.A. Thomas, “Elements of information theory,” *Wiley-Interscience New York, NY, USA*, 1991.
- [6] E. En Gad, M. Langberg, M. Schwartz, and J. Bruck, “Constant-weight Gray codes for local rank modulation,” *IEEE Trans. on Inform. Theory*, vol. 57, no. 11, pp. 7431–7442, Nov. 2011.
- [7] F. Farnoud, V. Skachek, and O. Milenkovic, “Error-correction in flash memories via codes in the Ulam metric,” *IEEE Trans. on Inform. Theory*, vol. 59, no. 5, pp. 3003–3020, May 2013.
- [8] F. Göloğlu, J. Lember, A.-E. Riet, and V. Skachek, “New bounds for permutation codes in Ulam metric,” *Proc. IEEE Intern. Symp. on Inform. Theory*, pp. 1726–1730, Jun. 2015.
- [9] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck, “Rank modulation for flash memories,” in *IEEE Trans. on Inform. Theory*, vol. 55, no. 6, pp. 2659–2673, June 2009.
- [10] S. Jukna, “Extremal combinatorics: with applications in computer science,” *Springer Publishing Company, Inc.*, 1st ed., 2010.
- [11] L. G. Kraft, “A device for quantizing, grouping, and coding amplitude modulated pulses,” *Cambridge, MA: MS Thesis, Electrical Engineering Department, Massachusetts Institute of Technology*, 1949.
- [12] V.I. Levenshtein, “On perfect codes in deletion and insertion metric,” *Discrete Math. Appl.*, vol. 2, no. 3, pp. 241–258, 1992.
- [13] D. Lubell, “A short proof of Sperner’s lemma,” *J. Combin. Theory*, vol. 1, no. 2, pp. 299–299, 1966.
- [14] R. Matheron and T. van Trung, “Directed t -packings and directed t -Steiner systems,” *Designs, Codes and Cryptography*, vol. 18, no. 1-3, pp. 187–198, 1999.
- [15] A. Mazumdar, A. Barg, and G. Zemor, “Constructions of rank modulation codes,” in *Proc. IEEE Intern. Symp. on Inform. Theory*, pp. 869–873, Jul./Aug. 2011.
- [16] B. McMillan, “Two inequalities implied by unique decipherability,” *IEEE Trans. on Inform. Theory* vol. 2, no. 4, pp.115–116, 1956.
- [17] L. D. Meshalkin, “Generalization of Sperner’s theorem on the number of subsets of a finite set,” *Theory Probab. Appl.* vol. 8, no. 2, pp. 203–204, 1963.
- [18] D. Slepian, “Permutation modulation,” *Proc. IEEE*, vol. 53, no. 3, pp. 228–236, Mar. 1965.
- [19] S. Ulam, “Monte-Carlo calculations in problems of mathematical physics,” *Modern Mathematics for the Engineer, Second Series*, (E. Beckenbach, ed.), pp. 261–281, 1961.
- [20] D. Wang, A. Mazumdar, and G. Wornell, “Compression in the space of permutations,” *IEEE Trans. on Inform. Theory*, vol. 61, no. 12, pp. 6417–6431, Dec. 2015.
- [21] K. Yamamoto, “Logarithmic order of free distributive lattice,” *J. Math. Soc. Japan* vol. 6, pp. 343–353, 1954.