

Optimal Quasi-Gray Codes: The Alphabet Matters*

Diptarka Chakraborty^{†1}, Debarati Das^{‡2}, Michal Koucký^{§3}, and Nitin Saurabh^{¶4}

^{1,2,3}Computer Science Institute of Charles University, Malostranské náměstí 25, 118 00
Praha 1, Czech Republic

⁴Max-Planck-Institut für Informatik, Saarbrücken, Germany

July 17, 2018

Abstract

A quasi-Gray code of dimension n and length ℓ over an alphabet Σ is a sequence of distinct words w_1, w_2, \dots, w_ℓ from Σ^n such that any two consecutive words differ in at most c coordinates, for some fixed constant $c > 0$. In this paper we are interested in the read and write complexity of quasi-Gray codes in the bit-probe model, where we measure the number of symbols read and written in order to transform any word w_i into its successor w_{i+1} .

We present construction of quasi-Gray codes of dimension n and length 3^n over the ternary alphabet $\{0, 1, 2\}$ with worst-case read complexity $O(\log n)$ and write complexity 2. This generalizes to arbitrary odd-size alphabets. For the binary alphabet, we present quasi-Gray codes of dimension n and length at least $2^n - 20n$ with worst-case read complexity $6 + \log n$ and write complexity 2. This complements a recent result by Raskin [Raskin '17] who shows that any quasi-Gray code over binary alphabet of length 2^n has read complexity $\Omega(n)$.

Our results significantly improve on previously known constructions and for the odd-size alphabets we break the $\Omega(n)$ worst-case barrier for space-optimal (non-redundant) quasi-Gray codes with constant number of writes. We obtain our results via a novel application of algebraic tools together with the principles of catalytic computation [Buhrman et al. '14, Ben-Or and Cleve '92, Barrington '89, Coppersmith and Grossman '75].

*The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 616787. The third author was also partially supported by the Center of Excellence CE-ITI under the grant P202/12/G061 of GA ČR.

[†]diptarka@iuuk.mff.cuni.cz

[‡]debaratix710@gmail.com

[§]koucky@iuuk.mff.cuni.cz

[¶]nsaurabh@mpi-inf.mpg.de

1 Introduction

One of the fundamental problems in the domain of algorithm design is to list down all the objects belonging to a certain combinatorial class. Researchers are interested in efficient generation of a list such that an element in the list can be obtained by a small amount of change to the element that precedes it. One of the classic examples is the binary *Gray code* introduced by Gray [22], initially used in pulse code communication. The original idea of a Gray code was to list down all binary strings of length n , i.e, all the elements of \mathbb{Z}_2^n , such that any two successive strings differ by exactly one bit. The idea was later generalized for other combinatorial classes (e.g. see [36, 27]). Gray codes have found applications in a wide variety of areas, such as information storage and retrieval [8], processor allocation [9], computing the permanent [36], circuit testing [40], data compression [39], graphics and image processing [1], signal encoding [31], modulation schemes for flash memories [25, 21, 47] and many more. Interested reader may refer to an excellent survey by Savage [41] for a comprehensive treatment on this subject.

In this paper we study the construction of Gray codes over \mathbb{Z}_m^n for any $m \in \mathbb{N}$. Originally, Gray codes were meant to list down all the elements from its domain but later studies (e.g. [19, 37, 5, 6]) focused on the generalization where we list ℓ distinct elements from the domain, each two consecutive elements differing in one position. We refer to such codes as Gray codes of length ℓ [19]. When the code lists all the elements from its domain it is referred to as *space-optimal*. It is often required that the last and the first strings appearing in the list also differ in one position. Such codes are called *cyclic Gray codes*. Throughout this paper we consider only cyclic Gray codes and we refer to them simply as Gray codes. Researchers also study codes where two successive strings differ in at most c positions, for some fixed constant $c > 0$, instead of differing in exactly one position. Such codes are called *quasi-Gray codes* [5]¹ or *c-Gray codes*.

We study the problem of constructing quasi-Gray codes over \mathbb{Z}_m^n in the cell probe model [46], where each cell stores an element from \mathbb{Z}_m . The efficiency of a construction is measured using three parameters. First, we want the length of a quasi-Gray code to be as large as possible. Ideally, we want space-optimal codes. Second, we want to minimize the number of coordinates of the input string the algorithm reads in order to generate the next (or, previous) string in the code. Finally, we also want the number of cells written in order to generate the successor (or, predecessor) string to be as small as possible. Since our focus is on quasi-Gray codes, the number of writes will always be bounded by a universal constant. We are interested in the worst-case behavior and we use *decision assignment trees* (DAT) of Fredman [19] to measure these complexities.

The second requirement of the above is motivated from the study of *loopless generation* of combinatorial objects. In the loopless generation we are required to generate the next string from the code in constant time. Different loopless algorithms to generate Gray codes are known in the literature [16, 4, 27]. However, those algorithms use extra memory cells in addition to the space required for the input string which makes it impossible to get a space-optimal code from them. More specifically, our goal is to design a decision assignment tree on n variables to generate a code over the domain \mathbb{Z}_m^n . If we allow extra memory cells (as in the case of loopless algorithms) then the corresponding DAT will be on $n + b$ variables, where b is the number of extra memory cells used.

Although there are known quasi-Gray codes with logarithmic read complexity and constant write complexity [19, 37, 5, 6], none of these constructions is space-optimal. The best result misses at least 2^{n-t} strings from the domain when having read complexity $t + O(\log n)$ [6]. Despite of an extensive research under many names, e.g., construction of Gray codes [19, 35, 15, 23],

¹Readers may note that the definition of quasi-Gray code given in [19] was different. The code referred as quasi-Gray code by Fredman [19] is called Gray code of length ℓ where $\ell < m^n$, in our notation.

dynamic language membership problem [18], efficient representation of integers [37, 6], so far we do not have any quasi-Gray code of length $2^n - 2^{\epsilon n}$, for some constant $\epsilon < 1$, with worst-case read complexity $o(n)$ and write complexity $o(n)$. The best worst-case read complexity for *space-optimal* Gray code is $n - 1$ [20]. Recently, Raskin [38] showed that any space-optimal quasi-Gray code over the domain \mathbb{Z}_2^n must have read complexity $\Omega(n)$. This lower bound is true even if we allow non-constant write complexity. It is worth noting that this result can be extended to the domain \mathbb{Z}_m^n when m is even.

In this paper we show that such lower bound does not hold for quasi-Gray codes over \mathbb{Z}_m^n , when m is odd. In particular, we construct space-optimal quasi-Gray codes over $\{0, 1, 2\}^n$ with read complexity $4 \log_3 n$ and write complexity 2. As a consequence we get an exponential separation between the read complexity of space-optimal quasi-Gray code over \mathbb{Z}_2^n and that over \mathbb{Z}_3^n .

Theorem 1.1. *Let $m \in \mathbb{N}$ be odd and $n \in \mathbb{N}$ be such that $n \geq 15$. Then, there is a space-optimal quasi-Gray code C over \mathbb{Z}_m^n for which, the two functions $\text{next}(C, w)$ and $\text{prev}(C, w)$ can be implemented by inspecting at most $4 \log_m n$ cells while writing only 2 cells.*

In the statement of the above theorem, $\text{next}(C, w)$ denotes the element appearing after w in the cyclic sequence of the code C , and analogously, $\text{prev}(C, w)$ denotes the preceding element. Using the argument as in [19, 35] it is easy to see a lower bound of $\Omega(\log_m n)$ on the read complexity when the domain is \mathbb{Z}_m^n . Hence our result is optimal up to some small constant factor.

Raskin shows $\Omega(n)$ lower bound on the read complexity of space-optimal binary quasi-Gray codes. The existence of binary quasi-Gray codes with sub-linear read complexity of length $2^n - 2^{\epsilon n}$, for some constant $\epsilon < 1$, was open. Using a different technique than that used in the proof of Theorem 1.1 we get a quasi-Gray code over the binary alphabet which enumerates all but $O(n)$ many strings. This result generalizes to the domain \mathbb{Z}_q^n for any prime power q .

Theorem 1.2. *Let $n \geq 15$ be any natural number. Then, there is a quasi-Gray code C of length at least $2^n - 20n$ over \mathbb{Z}_2^n , such that the two functions $\text{next}(C, w)$ and $\text{prev}(C, w)$ can be implemented by inspecting at most $6 + \log n$ cells while writing only 2 cells.*

We remark that the points that are missing from C in the above theorem are all of the form $\{0, 1\}^{O(\log n)} 0^{n-O(\log n)}$.

If we are allowed to read and write constant fraction of n bits then Theorem 1.2 can be adapted to get a quasi-Gray code of length $2^n - O(1)$ (see Section 5). In this way we get a trade-off between length of the quasi-Gray code and the number of bits read in the worst-case. All of our constructions can be made uniform (see the remark after Corollary 5.6).

Using the Chinese Remainder Theorem (cf. [13]), we also develop a technique that allows us to compose Gray codes over various domains. Hence, from quasi-Gray codes over domains $\mathbb{Z}_{m_1}^n, \mathbb{Z}_{m_2}^n, \dots, \mathbb{Z}_{m_k}^n$, where m_i 's are pairwise co-prime, we can construct quasi-Gray codes over \mathbb{Z}_m^n , where $m = m_1 \cdot m_2 \cdot \dots \cdot m_k$. Using this technique on our main results, we get a quasi-Gray code over \mathbb{Z}_m^n for any $m \in \mathbb{N}$ that misses only $O(n \cdot o^n)$ strings where $m = 2^\ell o$ for an odd o , while achieving the read complexity similar to that stated in Theorem 1.1. It is worth mentioning that if we get a space-optimal quasi-Gray code over the binary alphabet with non-trivial savings in read complexity, then we will have a space-optimal quasi-Gray code over the strings of alphabet \mathbb{Z}_m for any $m \in \mathbb{N}$ with similar savings.

The technique by which we construct our quasi-Gray codes relies heavily on simple algebra which is a substantial departure from previous mostly combinatorial constructions. We view Gray codes as permutations on \mathbb{Z}_m^n and we decompose them into k simpler permutations on \mathbb{Z}_m^n , each being computable with read complexity 3 and write complexity 1. Then we apply a

different composition theorem, than mentioned above, to obtain space-optimal quasi-Gray codes on $\mathbb{Z}_m^{n'}$, $n' = n + \log k$, with read complexity $O(1) + \log k$ and write complexity 2. The main issue is the decomposition of permutations into few simple permutations. This is achieved by techniques of catalytic computation [7] going back to the work of Coppersmith and Grossman [12, 2, 3].

It follows from the work of Coppersmith and Grossman [12] that our technique is incapable of designing a space-optimal quasi-Gray code on $\mathbb{Z}_2^{n'}$ as any such code represents an *odd* permutation. The tools we use give inherently only *even* permutations. However, we can construct quasi-Gray codes from cycles of length $2^n - 1$ on \mathbb{Z}_2^n as they are even permutations. Indeed, that is what we do for our Theorem 1.2. We note that any efficiently computable odd permutation on \mathbb{Z}_2^n , with say read complexity $(1 - \epsilon)n$ and write complexity $O(1)$, could be used together with our technique to construct a space-optimal quasi-Gray code on $\mathbb{Z}_2^{n'}$ with read complexity at most $(1 - \epsilon')n'$ and constant write complexity. This would represent a major progress on space-optimal Gray codes. (We would compose the odd permutation with some even permutation to obtain a full cycle on \mathbb{Z}_2^n . The size of the decomposition of the even permutation into simpler permutations would govern the read complexity of the resulting quasi-Gray code.)

Interestingly, Raskin's result relies on showing that a decision assignment tree of sub-linear read complexity must compute an even permutation.

1.1 Related works

The construction of Gray codes is central to the design of algorithms for many combinatorial problems [41]. Frank Gray [22] first came up with a construction of Gray code over binary strings of length n , where to generate the successor or predecessor strings one needs to read n bits in the worst-case. The type of code described in [22] is known as *binary reflected Gray code*. Later Bose *et al.* [5] provided a different type of Gray code construction, namely *recursive partition Gray code* which attains $O(\log n)$ average case read complexity while having the same worst-case read requirements. The read complexity we referred here is in the bit-probe model. It is easy to observe that any space-optimal binary Gray code must read $\log n + 1$ bits in the worst-case [19, 35, 20]. Recently, this lower bound was improved to $n/2$ in [38]. An upper bound of even $n - 1$ was not known until very recently [20]. This is also the best known so far.

Fredman [19] extended the definition of Gray codes by considering codes that may not enumerate all the strings (though presented in a slightly different way in [19]) and also introduced the notion of *decision assignment tree* (DAT) to study the complexity of any code in the bit-probe model. He provided a construction that generates a Gray code of length $2^{c \cdot n}$ for some constant $c < 1$ while reducing the worst-case bit-read to $O(\log n)$. Using the idea of Lucal's modified reflected binary code [30], Munro and Rahman [37] got a code of length 2^{n-1} with worst-case read complexity only $4 + \log n$. However in their code two successive strings differ by 4 coordinates in the worst-case, instead of just one and we refer to such codes as quasi-Gray codes following the nomenclature used in [5]. Brodal *et al.* [6] extended the results of [37] by constructing a quasi-Gray code of length $2^n - 2^{n-t}$ for arbitrary $1 \leq t \leq n - \log n - 1$, that has $t + 3 + \log n$ bits ($t + 2 + \log n$ bits) worst-case read complexity and any two successive strings in the code differ by at most 2 bits (3 bits).

In contrast to the Gray codes over binary alphabets, Gray codes over non-binary alphabets received much less attention. The construction of binary reflected Gray code was generalized to the alphabet \mathbb{Z}_m for any $m \in \mathbb{N}$ in [17, 11, 26, 39, 27, 24]. However, each of those constructions reads n coordinates in the worst-case to generate the next element. As mentioned before, we measure the read complexity in the well studied cell probe model [46] where we assume that each cell stores an element of \mathbb{Z}_m . The argument of Fredman in [19] implies a lower bound of

Reference	Value of m	length	Worst-case cell read	Worst-case cell write
[22]	2	2^n	n	1
[19]	2	$2^{\Theta(n)}$	$O(\log n)$	$O(1)$
[18]	2	$\Theta(2^n/n)$	$\log n + 1$	$\log n + 1$
[37]	2	2^{n-1}	$\log n + 4$	4
[5]	2	$2^n - O(2^n/n^t)$	$O(t \log n)$	3
[6]	2	$2^n - 2^{n-t}$	$\log n + t + 3$	2
[6]	2	$2^n - 2^{n-t}$	$\log n + t + 2$	3
[20]	2	2^n	$n - 1$	1
Theorem 1.2	2	$2^n - O(n)$	$\log n + 4$	2
[11]	any m	m^n	n	1
Theorem 1.1	any odd m	m^n	$4 \log_m n + 3$	2

Table 1: Taxonomy of construction of Gray/quasi-Gray codes over \mathbb{Z}_m^n

$\Omega(\log_m n)$ on the read complexity of quasi-Gray code on \mathbb{Z}_m^n . To the best of our knowledge, for non-binary alphabets, there is nothing known similar to the results of Munro and Rahman or Brodal et al. [37, 6]. We summarize the previous results along with ours in Table 1.

Additionally, many variants of Gray codes have been studied in the literature. A particular one that has garnered a lot of attention in the past 30 years is the well-known *middle levels conjecture*. See [32, 33, 34, 23], and the references therein. It has been established only recently [32]. The conjecture says that there exists a Hamiltonian cycle in the graph induced by the vertices on levels n and $n + 1$ of the hypercube graph in $2n + 1$ dimensions. In other words, there exists a Gray code on the middle levels. Mütze et al. [33, 34] studied the question of efficiently enumerating such a Gray code in the word RAM model. They [34] gave an algorithm to enumerate a Gray code in the middle levels that requires $O(n)$ space and *on average* takes $O(1)$ time to generate the next vertex. In this paper we consider the bit-probe model, and Gray codes over the complete hypercube. It would be interesting to know whether our technique can be applied for the middle level Gray codes.

1.2 Our technique

Our construction of Gray codes relies heavily on the notion of s -functions defined by Copersmith and Grossman [12]. An s -function is a permutation τ on \mathbb{Z}_m^n defined by a function $f : \mathbb{Z}_m^s \rightarrow \mathbb{Z}_m$ and an $(s + 1)$ -tuple of indices $i_1, i_2, \dots, i_s, j \in [n]$ such that $\tau(\langle x_1, x_2, \dots, x_n \rangle) = (\langle x_1, x_2, \dots, x_{j-1}, x_j + f(x_{i_1}, \dots, x_{i_s}), x_{j+1}, \dots, x_n \rangle)$, where the addition is inside \mathbb{Z}_m . Each s -function can be computed by some decision assignment tree that given a vector $x = \langle x_1, x_2, \dots, x_n \rangle$, inspects $s + 1$ coordinates of x and then it writes into a single coordinate of x .

A counter C (quasi-Gray code) on \mathbb{Z}_m^n can be thought of as a permutation on \mathbb{Z}_m^n . Our goal is to construct some permutation α on \mathbb{Z}_m^n that can be written as a composition of 2-functions $\alpha_1, \dots, \alpha_k$, i.e., $\alpha = \alpha_k \circ \alpha_{k-1} \circ \dots \circ \alpha_1$.

Given such a decomposition, we can build another counter C' on \mathbb{Z}_m^{r+n} , where $r = \lceil \log_m k \rceil$, for which the function $next(C', x)$ operates as follows. The first r -coordinates of x serve as an *instruction pointer* $i \in [m^r]$ that determines which α_i should be executed on the remaining n coordinates of x . Hence, based on the current value i of the r coordinates, we perform α_i on the remaining coordinates and then we update the value of i to $i + 1$. (For $i > k$ we can execute the identity permutation which does nothing.)

We can use known Gray codes on \mathbb{Z}_m^r to represent the instruction pointer so that when incrementing i we only need to write into one of the coordinates. This gives a counter C' which

can be computed by a decision assignment tree that reads $r + 3$ coordinates and writes into 2 coordinates of x . (A similar composition technique is implicit in Brodal et al. [6].) If C is of length $\ell = m^n - t$, then C' is of length $m^{n+r} - tm^r$. In particular, if C is space-optimal then so is C' .

Hence, we reduce the problem of constructing 2-Gray codes to the problem of designing large cycles in \mathbb{Z}_m^n that can be decomposed into 2-functions. Coppersmith and Grossman [12] studied precisely the question of, which permutations on \mathbb{Z}_2^n can be written as a composition of 2-functions. They show that a permutation on \mathbb{Z}_2^n can be written as a composition of 2-functions if and only if the permutation is even. Since \mathbb{Z}_2^n is of even size, a cycle of length 2^n on \mathbb{Z}_2^n is an odd permutation and thus it cannot be represented as a composition of 2-functions. However, their result also implies that a cycle of length $2^n - 1$ on \mathbb{Z}_2^n can be decomposed into 2-functions.

We want to use the counter composition technique described above in connection with a cycle of length $2^n - 1$. To maximize the length of the cycle C' in \mathbb{Z}_2^{n+r} , we need to minimize k , the number of 2-functions in the decomposition. By a simple counting argument, most cycles of length $2^n - 1$ on \mathbb{Z}_2^n require k to be exponentially large in n . This is too large for our purposes. Luckily, there are cycles of length $2^n - 1$ on \mathbb{Z}_2^n that can be decomposed into polynomially many 2-functions, and we obtain such cycles from linear transformations.

There are linear transformations $\mathbb{Z}_2^n \rightarrow \mathbb{Z}_2^n$ which define a cycle on \mathbb{Z}_2^n of length $2^n - 1$. For example, the matrix corresponding to the multiplication by a fixed generator of the multiplicative group \mathbb{F}_2^* of the Galois field $GF[2^n]$ is such a matrix. Such matrices are full rank and they can be decomposed into $O(n^2)$ elementary matrices, each corresponding to a 2-function. Moreover, there are matrices derived from primitive polynomials that can be decomposed into at most $4n$ elementary matrices.² We use them to get a counter on $\mathbb{Z}_2^{n'}$ of length at least $2^{n'} - 20n'$ whose successor and predecessor functions are computable by decision assignment trees of read complexity $\leq 6 + \log n'$ and write complexity 2. Such counter represents 2-Gray code of the prescribed length. For any prime q , the same construction yields 2-Gray codes of length at least $q^{n'} - 5q^2n'$ with decision assignment trees of read complexity $\leq 6 + \log_q n'$ and write complexity 2.

The results of Coppersmith and Grossman [12] can be generalized to \mathbb{Z}_m^n as stated in Richard Cleve's thesis [10].³ For odd m , if a permutation on \mathbb{Z}_m^n is even then it can be decomposed into 2-functions. Since m^n is odd, a cycle of length m^n on \mathbb{Z}_m^n is an even permutation and so it can be decomposed into 2-functions. If the number k of those functions is small, so the $\log_m k$ is small, we get the sought after counter with small read complexity. However, for most cycles of length m^n on \mathbb{Z}_m^n , k is exponential in n .

We show though, that there is a cycle α of length m^n on \mathbb{Z}_m^n that can be decomposed into $O(n^3)$ 2-functions. This in turn gives space-optimal 2-Gray codes on $\mathbb{Z}_m^{n'}$ with decision assignment trees of read complexity $O(\log_m n')$ and write complexity 2.

We obtain the cycle α and its decomposition in two steps. First, for $i \in [n]$, we consider the permutation α_i on \mathbb{Z}_m^n which maps each element $0^{i-1}ay$ onto $0^{i-1}(a+1)y$, for $a \in \mathbb{Z}_m$ and $y \in \mathbb{Z}_m^{n-i}$, while other elements are mapped to themselves. Hence, α_i is a product of m^{n-i} disjoint cycles of length m . We show that $\alpha = \alpha_n \circ \alpha_{n-1} \circ \dots \circ \alpha_1$ is a cycle of length m^n . In the next step we decompose each α_i into $O(n^2)$ 2-functions.

For $i \leq n - 2$, we can decompose α_i using the technique of Ben-Or and Cleve [3] and its refinement in the form of catalytic computation of Buhrman et al. [7]. We can think of $x \in \mathbb{Z}_m^n$ as content of n memory registers, where x_1, \dots, x_{i-1} are the input registers, x_i is the output register, and x_{i+1}, \dots, x_n are the working registers. The catalytic computation technique gives a

²Primitive polynomials were previously also used in a similar problem, namely to construct shift-register sequences (see e.g. [27]).

³Unfortunately, there is no written record of the proof.

program consisting of $O(n^2)$ instructions, each being equivalent to a 2-function, which performs the desired adjustment of x_i based on the values of x_1, \dots, x_{i-1} without changing the ultimate values of the other registers. (We need to increment x_i iff x_1, \dots, x_{i-1} are all zero.) This program directly gives the desired decomposition of α_i , for $i \leq n - 2$. (Our proof in Section 6 uses the language of permutations.)

The technique of catalytic computation fails for α_{n-1} and α_n as the program needs at least two working registers to operate. Hence, for α_{n-1} and α_n we have to develop entirely different technique. This is not trivial and quite technical but it is nevertheless possible, thanks to the specific structure of α_{n-1} and α_n .

Organization of the paper

In Section 2 we define the notion of counter, Gray code and our computational model, namely decision assignment tree and also provide certain known results regarding the construction of Gray codes. Then in Section 3 we describe how to combine counters over smaller alphabets to get another counter over larger alphabet, by introducing the Chinese Remainder Theorem for counters. Next we provide some basic facts about the permutation group and the underlying structure behind all of our constructions of quasi-Gray codes. We devote Section 5 to construction of a quasi-Gray code over binary alphabet that misses only a few words, by using full rank linear transformation. In Section 6 we construct a space-optimal quasi-Gray code over any odd-size alphabet. Finally in Section 7 we rule out the existence of certain kind of space-optimal binary counters.

2 Preliminaries

In the rest of the paper we only present constructions of the successor function $next(C, w)$ for our codes. Since all the operations in those constructions are readily invertible, the same arguments also give the predecessor function $prev(C, w)$.

Notations: We use the standard notions of groups and fields, and mostly we will use only elementary facts about them (see [14, 29] for background.). By \mathbb{Z}_m we mean the set of integers modulo m , i.e., $\mathbb{Z}_m := \mathbb{Z}/m\mathbb{Z}$. Throughout this paper whenever we use addition and multiplication operation between two elements of \mathbb{Z}_m , then we mean the operations within \mathbb{Z}_m that is modulo m . For any $m \in \mathbb{N}$, we let $[m]$ denote the set $\{1, 2, \dots, m\}$. Unless stated otherwise explicitly, all the logarithms we consider throughout this paper are based 2.

Now we define the notion of counters used in this paper.

Definition 1 (Counter). *A counter of length ℓ over a domain \mathcal{D} is any cyclic sequence $C = (w_1, \dots, w_\ell)$ such that w_1, \dots, w_ℓ are distinct elements of \mathcal{D} . With the counter C we associate two functions $next(C, w)$ and $prev(C, w)$ that give the successor and predecessor element of w in C , that is for $i \in [\ell]$, $next(C, w_i) = w_j$ where $j - i = 1 \pmod{\ell}$, and $prev(C, w_i) = w_k$ where $i - k = 1 \pmod{\ell}$. If $\ell = |\mathcal{D}|$, we call the counter a space-optimal counter.*

Often elements in the underlying domain \mathcal{D} have some “structure” to them. In such cases, it is desirable to have a counter such that consecutive elements in the sequence differ by a “small” change in the “structure”. We make this concrete in the following definition.

Definition 2 (Gray Code). *Let $\mathcal{D}_1, \dots, \mathcal{D}_n$ be finite sets. A Gray code of length ℓ over the domain $\mathcal{D} = \mathcal{D}_1 \times \dots \times \mathcal{D}_n$ is a counter (w_1, \dots, w_ℓ) of length ℓ over \mathcal{D} such that any two consecutive strings w_i and w_j , $j - i = 1 \pmod{\ell}$, differ in exactly one coordinate when viewed as*

an n -tuple. More generally, if for some constant $c \geq 1$, any two consecutive strings w_i and w_j , $j - i = 1 \pmod{\ell}$, differ in at most c coordinates such a counter is called a c -Gray Code.

By a quasi-Gray code we mean c -Gray code for some unspecified fixed $c > 0$. In the literature sometimes people do not place any restriction on the relationship between w_ℓ and w_1 and they refer to such a sequence a (quasi)-Gray code. In their terms, our codes would be *cyclic* (quasi)-Gray codes. If $\ell = |\mathcal{D}|$, we call the codes *space-optimal* (quasi)-Gray codes.

Decision Assignment Tree: The computational model we consider in this paper is called *Decision Assignment Tree* (DAT). The definition we provide below is a generalization of that given in [19]. It is intended to capture random access machines with small word size.

Let us fix an underlying domain \mathcal{D}^n whose elements we wish to enumerate. In the following, we will denote an element in \mathcal{D}^n by $\langle x_1, x_2, \dots, x_n \rangle$. A decision assignment tree is a $|\mathcal{D}|$ -ary tree such that each internal node is labeled by one of the variables x_1, x_2, \dots, x_n . Furthermore, each outgoing edge of an internal node is labeled with a distinct element of \mathcal{D} . Each leaf node of the tree is labeled by a set of assignment instructions that set new (fixed) values to chosen variables. The variables which are not mentioned in the assignment instructions remain unchanged.

The execution on a decision assignment tree on a particular input vector $\langle x_1, x_2, \dots, x_n \rangle \in \mathcal{D}^n$ starts from the root of the tree and continues in the following way: at a non-leaf node labeled with a variable x_i , the execution queries x_i and depending on the value of x_i the control passes to the node following the outgoing edge labeled with the value of x_i . Upon reaching a leaf, the corresponding set of assignment statements is used to modify the vector $\langle x_1, x_2, \dots, x_n \rangle$ and the execution terminates. The modified vector is the output of the execution.

Thus, each decision assignment tree computes a mapping from \mathcal{D}^n into \mathcal{D}^n . We are interested in decision assignment trees computing the mapping $next(C, \langle x_1, x_2, \dots, x_n \rangle)$ for some counter C . When C is space-optimal we can assume, without loss of generality, that each leaf assigns values only to the variables that it reads on the path from the root to the leaf. (Otherwise, the decision assignment tree does not compute a bijection.) We define the *read complexity* of a decision assignment tree T , denoted by $READ(T)$, as the maximum number of non-leaf nodes along any path from the root to a leaf. Observe that any mapping from \mathcal{D}^n into \mathcal{D}^n can be implemented by a decision assignment tree with read complexity n . We also define the *write complexity* of a decision assignment tree T , denoted by $WRITE(T)$, as the maximum number of assignment instructions in any leaf.

Instead of the domain \mathcal{D}^n , we will sometimes also use domains that are a cartesian product of different domains. The definition of a decision assignment tree naturally extends to this case of different variables having different domains.

For any counter $C = (w_1, \dots, w_\ell)$, we say that C is computed by a decision assignment tree T if and only if for $i \in [\ell]$, $next(C, w_i) = T(w_i)$, where $T(w_i)$ denotes the output string obtained after an execution of T on w_i . Note that any two consecutive strings in the cyclic sequence of C differ by at most $WRITE(T)$ many coordinates.

For a small constant $c \geq 1$, some domain \mathcal{D} , and all large enough n , we will be interested in constructing cyclic counters on \mathcal{D}^n that are computed by decision assignment trees of write complexity c and read complexity $O(\log n)$. By the definition such cyclic counters will necessarily be c -Gray codes.

2.1 Construction of Gray codes

For our construction of quasi-Gray codes on a domain \mathcal{D}^n with decision assignment trees of small read and write complexity we will need ordinary Gray codes on a domain $\mathcal{D}^{O(\log n)}$. Several constructions of space-optimal binary Gray codes are known where the oldest one is the binary

reflected Gray code [22]. This can be generalized to space-optimal (cyclic) Gray codes over non-binary alphabets (see e.g. [11, 27]).

Theorem 2.1 ([11, 27]). *For any $m, n \in \mathbb{N}$, there is a space-optimal (cyclic) Gray code over \mathbb{Z}_m^n .*

3 Chinese Remainder Theorem for Counters

In this paper we consider quasi-Gray codes over \mathbb{Z}_m^n for $m \in \mathbb{N}$. Below we describe how to compose decision assignment trees over different domains to get a decision assignment tree for a larger mixed domain.

Theorem 3.1 (Chinese Remainder Theorem for Counters). *Let $r, n_1, \dots, n_r \in \mathbb{N}$ be integers, and let $\mathcal{D}_{1,1}, \dots, \mathcal{D}_{1,n_1}, \mathcal{D}_{2,1}, \dots, \mathcal{D}_{r,n_r}$ be some finite sets of size at least two. Let $\ell_1 \geq r - 1$ be an integer, and ℓ_2, \dots, ℓ_r be pairwise co-prime integers. For $i \in [r]$, let C_i be a counter of length ℓ_i over $\mathcal{D}_i = \mathcal{D}_{i,1} \times \dots \times \mathcal{D}_{i,n_i}$ computed by a decision assignment tree T_i over n_i variables. Then, there exists a decision assignment tree T over $\sum_{i=1}^r n_i$ variables that implements a counter C of length $\prod_{i=1}^r \ell_i$ over $\mathcal{D}_1 \times \dots \times \mathcal{D}_r$. Furthermore, $\text{READ}(T) = n_1 + \max\{\text{READ}(T_i)\}_{i=2}^r$, and $\text{WRITE}(T) = \text{WRITE}(T_1) + \max\{\text{WRITE}(T_i)\}_{i=2}^r$.*

Proof. For any $i \in [r]$, let the counter $C_i = (w_{i,1}, \dots, w_{i,\ell_i})$. Let x_1, \dots, x_r be variables taking values in $\mathcal{D}_1, \dots, \mathcal{D}_r$, respectively. The following procedure, applied repeatedly, defines the counter C :

If $x_1 = w_{1,i}$ for some $i \in [r - 1]$ **then**
 $x_{i+1} \leftarrow \text{next}(C_{i+1}, x_{i+1})$
 $x_1 \leftarrow \text{next}(C_1, x_1)$
else
 $x_1 \leftarrow \text{next}(C_1, x_1)$.

It is easily seen that the above procedure defines a valid cyclic sequence when starting at $w_{1,i_1}, \dots, w_{r,i_r}$ for any $\langle i_1, i_2, \dots, i_r \rangle \in [\ell_1] \times \dots \times [\ell_r]$. That is, every element has a unique predecessor and a unique successor, and that the sequence is cyclic. It can easily be implemented by a decision assignment tree, say T . First it reads the value of x_1 . Since $x_1 \in \mathcal{D}_1 = \mathcal{D}_{1,1} \times \dots \times \mathcal{D}_{1,n_1}$, it queries n_1 components. Then, depending on the value of x_1 , it reads and updates another component, say x_j . This can be accomplished using the decision assignment tree T_j . We also update the value of x_1 , and to that end we use the appropriate assignments from decision assignment tree T_1 . Observe that irrespective of how efficient T_1 is, we read x_1 completely to determine which of the remaining $r - 1$ counters to update. Hence, $\text{READ}(T) = n_1 + \max\{\text{READ}(T_i)\}_{i=2}^r$, and $\text{WRITE}(T) = \text{WRITE}(T_1) + \max\{\text{WRITE}(T_i)\}_{i=2}^r$.

Now it only remains to show that the counter described above is indeed of length $\prod_{i=1}^r \ell_i$. Thus, it suffices to establish that starting with the string $\langle w_{1,1}, \dots, w_{r,1} \rangle$, we can generate the string $\langle w_{1,i_1}, \dots, w_{r,i_r} \rangle$ for any $\langle i_1, \dots, i_r \rangle \in [\ell_1] \times \dots \times [\ell_r]$. Let us assume $i_1 = 1$. At the end of the proof we will remove this assumption. Suppose the string $\langle w_{1,1}, w_{2,i_2}, \dots, w_{r,i_r} \rangle$ is reachable from $\langle w_{1,1}, w_{2,1}, \dots, w_{r,1} \rangle$ in t steps. As our procedure always increments x_1 , t must be divisible by ℓ_1 . Let $d = t/\ell_1$. Furthermore, the procedure increments a variable x_i , $i \neq 1$, exactly after ℓ_1

steps. Thus, $\langle w_{1,1}, w_{2,i_2}, \dots, w_{r,i_r} \rangle$ is reachable if and only if d satisfies the following equations:

$$\begin{aligned} d &\equiv i_2 - 1 \pmod{\ell_2} \\ d &\equiv i_3 - 1 \pmod{\ell_3} \\ &\vdots \\ d &\equiv i_r - 1 \pmod{\ell_r}. \end{aligned}$$

Since ℓ_2, \dots, ℓ_r are pairwise co-prime, Chinese Remainder Theorem (for a reference, see [13]) guarantees the existence of a unique integral solution d such that $0 \leq d < \prod_{i=2}^r \ell_i$. Hence, $\langle w_{1,1}, w_{2,i_2}, \dots, w_{r,i_r} \rangle$ is reachable from $\langle w_{1,1}, w_{2,1}, \dots, w_{r,1} \rangle$ in at most $\prod_{i=1}^r \ell_i$ steps.

Now we remove the assumption $i_1 = 1$, i.e., $w_{1,i_1} \neq w_{1,1}$. Consider the string $\langle w_{1,1}, w_{2,i'_2}, \dots, w_{r,i'_r} \rangle$ where $w_{j,i'_j} = w_{j,i_j-1}$ for $2 \leq j \leq \min\{i_1, r\}$, and $w_{j,i'_j} = w_{j,i_j}$ for $j > \min\{i_1, r\}$. From the arguments in the previous paragraph, we know that this tuple is reachable. We now observe that the next $i_1 - 1$ steps increment $w_{1,1}$ to w_{1,i_1} and w_{j,i'_j} to w_{j,i_j} for $2 \leq j \leq \min\{i_1, r\}$, thus, reaching the desired string $\langle w_{1,i_1}, \dots, w_{r,i_r} \rangle$. \square

Remark. We remark that if C_i 's are space-optimal in Theorem 3.1, then so is C .

In the above proof, we constructed a special type of a counter where we always read the first coordinate, incremented it, and further depending on its value, we may update the value of another coordinate. From now on we refer to such type of counters as *hierarchical counters*. In Section 7 we will show that for such type of a counter the co-primality condition is necessary at least for $\ell_1 = 2, 3$. One can further note that the above theorem is similar to the well known Chinese Remainder Theorem and has similar type of application for constructing of space-optimal quasi-Gray codes over \mathbb{Z}_m^n for arbitrary $m \in \mathbb{N}$.

Lemma 3.2. Let $n, m \in \mathbb{N}$ be such that $m = 2^k o$, where o is odd and $k \geq 0$. Given decision assignment trees T_1 and T_2 computing space-optimal (quasi-)Gray codes over $(\mathbb{Z}_{2^k})^{n-1}$ and \mathbb{Z}_o^{n-1} , respectively, there exists a decision assignment tree T implementing a space-optimal quasi-Gray code over \mathbb{Z}_m^n such that $\text{READ}(T) = 1 + \max\{\text{READ}(T_1), \text{READ}(T_2)\}$, and $\text{WRITE}(T) = 1 + \max\{\text{WRITE}(T_1), \text{WRITE}(T_2)\}$.

Proof. We will view \mathbb{Z}_m^n as $\mathbb{Z}_m \times (\mathbb{Z}_{2^k})^{n-1} \times (\mathbb{Z}_o)^{n-1}$ and simulate a decision assignment tree operating on $\mathbb{Z}_m \times (\mathbb{Z}_{2^k})^{n-1} \times (\mathbb{Z}_o)^{n-1}$ on \mathbb{Z}_m^n . From the Chinese Remainder Theorem (see [13]), we know that there exists a bijection (in fact, an isomorphism) $f: \mathbb{Z}_m \rightarrow \mathbb{Z}_{2^k} \times \mathbb{Z}_o$. We denote the tuple $f(z)$ by $\langle f_1(z), f_2(z) \rangle$. From Theorem 3.1 we know that there exists a decision assignment tree T' over $\mathbb{Z}_m \times (\mathbb{Z}_{2^k})^{n-1} \times (\mathbb{Z}_o)^{n-1}$ computing a space-optimal quasi-Gray code such that $\text{READ}(T') = 1 + \max\{\text{READ}(T_1), \text{READ}(T_2)\}$, and $\text{WRITE}(T') = 1 + \max\{\text{WRITE}(T_1), \text{WRITE}(T_2)\}$.

We can simulate actions of T' on an input \mathbb{Z}_m^n to obtain the desired decision assignment tree T . Indeed, whenever T' queries x_1 , T queries the first coordinate of its input. Whenever T' queries the i -th coordinate of $(\mathbb{Z}_{2^k})^{n-1}$, T queries the $(i+1)$ -th coordinate of its input and makes its decision based on the $f_1(\cdot)$ value of that coordinate. Similarly, whenever T' queries the j -th coordinate of $(\mathbb{Z}_o)^{n-1}$, T queries the $(j+1)$ -th coordinate and makes its decision based on the $f_2(\cdot)$ value of that coordinate. Assignments by T are handled in similar fashion by updating only the appropriate part of $\langle f_1(z), f_2(z) \rangle$. (Notice, queries made by T might reveal more information than queries made by T' .) \square

Before proceeding further, we would also like to point out that to get a space-optimal decision assignment tree over \mathbb{Z}_{2^k} , it suffices to get space-optimal decision assignment trees over \mathbb{Z}_2

for arbitrary dimensions. Thus, to get a decision assignment tree implementing space-optimal quasi-Gray codes over \mathbb{Z}_m , we only need decision assignment trees implementing space-optimal quasi-Gray codes over \mathbb{Z}_2 and \mathbb{Z}_o . This also justifies our sole focus on construction of space-optimal decision assignment trees over \mathbb{Z}_2 and \mathbb{Z}_o in the later sections.

Lemma 3.3. *If, for all $n \in \mathbb{N}$, there exists a decision assignment tree T implementing a space-optimal (quasi-)Gray code over \mathbb{Z}_2^n , then for any k and $n \in \mathbb{N}$, there exists a decision assignment tree T' implementing a space-optimal (quasi-)Gray code over $(\mathbb{Z}_{2^k})^n$ such that the read and write complexity remain the same.*

Proof. Consider any bijective map $f : \mathbb{Z}_{2^k} \rightarrow \mathbb{Z}_2^k$. For example, one can take standard binary encoding of integers ranging from 0 to $2^k - 1$ as the bijective map f . Next, define another map $g : (\mathbb{Z}_{2^k})^n \rightarrow \mathbb{Z}_2^{kn}$ as follows: $g(x_1, \dots, x_n) = \langle f(x_1), \dots, f(x_n) \rangle$. Now consider T that implements a space-optimal (quasi-)Gray code over \mathbb{Z}_2^{kn} . We fix a partition of the variables $\{1, \dots, k\} \uplus \dots \uplus \{(n-1)k+1, \dots, nk\}$ into n blocks of k variables each.

We now construct a decision assignment tree T' over $(\mathbb{Z}_{2^k})^n$ using T and the map f . As in the proof of Lemma 3.2, our T' follows T in the decision making. That is, if T queries a variable, then T' queries the block in the partition where the variable lies. (Again, as noted before, T' may get more information than required by T .) Upon reaching a leaf, using f , T' updates the blocks depending on T 's updates to the variables. □

We devote the rest of the paper to the construction of counters over \mathbb{Z}_2^n , and \mathbb{Z}_m^n for any odd m .

4 Permutation Group and Decomposition of Counters

We start this section with some basic notation and facts about the permutation group which we will use heavily in the rest of the paper. The set of all permutations over a domain \mathcal{D} forms a group under the composition operation, denoted by \circ , which is defined as follows: for any two permutations σ and α , $\sigma \circ \alpha(x) = \sigma(\alpha(x))$, where $x \in \mathcal{D}$. The corresponding group, denoted \mathcal{S}_N , is the *symmetric group* of order $N = |\mathcal{D}|$. We say, a permutation $\sigma \in \mathcal{S}_N$ is a *cycle* of length ℓ if there are distinct elements $a_1, \dots, a_\ell \in [N]$ such that for $i \in [\ell - 1]$, $a_{i+1} = \sigma(a_i)$, $a_1 = \sigma(a_\ell)$, and for all $a \in [N] \setminus \{a_1, a_2, \dots, a_\ell\}$, $\sigma(a) = a$. We denote such a cycle by $(a_1, a_2, \dots, a_\ell)$. Below we state few simple facts about composition of cycles.

Proposition 4.1. *Consider two cycles $C_1 = (t, a_1, \dots, a_{\ell_1})$ and $C_2 = (t, b_1, \dots, b_{\ell_2})$ where for any $i \in [\ell_1]$ and $j \in [\ell_2]$, $a_i \neq b_j$. Then, $C = C_2 \circ C_1$ is the cycle $(t, a_1, \dots, a_{\ell_1}, b_1, \dots, b_{\ell_2})$ of length $\ell_1 + \ell_2 + 1$.*

Proposition 4.2. *If $\sigma \in \mathcal{S}_N$ is a cycle of length ℓ , then for any $\alpha \in \mathcal{S}_N$, $\alpha \circ \sigma \circ \alpha^{-1}$ is also a cycle of length ℓ . Moreover, if $\sigma = (a_1, a_2, \dots, a_\ell)$, then $\alpha \circ \sigma \circ \alpha^{-1} = (\alpha(a_1), \alpha(a_2), \dots, \alpha(a_\ell))$.*

The permutation $\alpha \circ \sigma \circ \alpha^{-1}$ is called the *conjugate* of σ with respect to α . The above proposition is a special case of a well known fact about the cycle structure of conjugates of any permutation and can be found in any standard text book on Group Theory (e.g., Proposition 10 in Chapter 4.3 of [14].).

Roughly speaking, a counter of length ℓ over \mathcal{D} , in the language of permutations, is nothing but a cycle of the same length in $\mathcal{S}_{|\mathcal{D}|}$. We now make this correspondence precise and give a construction of a decision assignment tree that implements such a counter.

Lemma 4.3. Let $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_r$ be a domain. Suppose $\sigma_1, \dots, \sigma_k \in \mathcal{S}_{|\mathcal{D}|}$ are such that $\sigma = \sigma_k \circ \sigma_{k-1} \circ \cdots \circ \sigma_1$ is a cycle of length ℓ . Let T_1, \dots, T_k be decision assignment trees that implement $\sigma_1, \dots, \sigma_k$ respectively. Let $\mathcal{D}' = \mathcal{D}'_1 \times \cdots \times \mathcal{D}'_{r'}$ be a domain such that $|\mathcal{D}'| \geq k$, and let T' be a decision assignment tree that implements a counter C' of length k' over \mathcal{D}' where $k' \geq k$.

Then, there exists a decision assignment tree T that implements a counter of length $k'\ell$ over $\mathcal{D}' \times \mathcal{D}$ such that $\text{READ}(T) = r' + \max\{\text{READ}(T_i)\}_{i=1}^k$, and $\text{WRITE}(T) = \text{WRITE}(T') + \max\{\text{WRITE}(T_i)\}_{i=1}^k$.

Proof. Suppose $C' = (a_1, \dots, a_{k'})$. Now let us consider the following procedure P : on any input $\langle x_1, x_2 \rangle \in \mathcal{D}' \times \mathcal{D}$,

If $x_1 = a_j$ for some $j \in [k]$ **then**
 $x_2 \leftarrow \sigma_j(x_2)$
 $x_1 \leftarrow \text{next}(C', x_1)$
else
 $x_1 \leftarrow \text{next}(C', x_1)$.

Now using a similar argument as in the proof of Theorem 3.1, the above procedure is easily seen to be implementable using a decision assignment tree T of the prescribed complexity. Each time we check the value of $x_1 \in \mathcal{D}' = \mathcal{D}'_1 \times \cdots \times \mathcal{D}'_{r'}$. Thus, we need to read r' components. Depending on the value of x_1 , we may apply σ_j on x_2 using the decision assignment tree T_j . Then we update the value of x_1 . Hence, $\text{READ}(T) = r' + \max\{\text{READ}(T_i)\}_{i=1}^k$, and $\text{WRITE}(T) = \text{WRITE}(T') + \max\{\text{WRITE}(T_i)\}_{i=1}^k$.

Let $(w_1, w_2, \dots, w_\ell)$ be the cycle of length ℓ given by σ . We now argue that the procedure P generates a counter of length $k'\ell$ over $\mathcal{D}' \times \mathcal{D}$ starting at $\langle a_1, w_1 \rangle$. Without loss of generality, let us assume that $\sigma = \sigma_{k'} \circ \cdots \circ \sigma_{k+1} \circ \sigma_k \circ \sigma_{k-1} \circ \cdots \circ \sigma_1$, where for $j \geq k+1$, σ_j is the identity map. Fix $j \in [k']$. Define $\alpha_j = \sigma_{j-1} \circ \cdots \circ \sigma_1$, and $\tau_j = \alpha_j \circ \sigma \circ \alpha_j^{-1} = \sigma_{j-1} \circ \cdots \circ \sigma_1 \circ \sigma_{k'} \circ \cdots \circ \sigma_j$. For $i = 0, 1, \dots, \ell$, let $\langle g_i, e_i \rangle = P^{ik'}(\langle a_j, \alpha_j(w_1) \rangle)$ where $P^{ik'}$ denotes ik' invocations of P . Since P increments x_1 in every invocation, for $i = 1, 2, \dots, \ell$, $g_i = a_j$ and $e_i = \tau_j(e_{i-1})$.

By Proposition 4.2, τ_j is a cycle $(\alpha_j(w_1)\alpha_j(w_2)\cdots\alpha_j(w_\ell))$ of length ℓ . Hence, e_1, \dots, e_ℓ are all distinct and $e_\ell = e_0$.

As a consequence we conclude that for any $x \in \mathcal{D}' \times \mathcal{D}$ and $1 \leq j_1 \neq j_2 \leq k'\ell$, $P^{j_1}(x) \neq P^{j_2}(x)$ and $P^{k'\ell}(x) = x$. This completes the proof. \square

In the next two sections we describe the construction of $\sigma_1, \dots, \sigma_k \in \mathcal{S}_N$ where $N = m^n$ for some $m, n \in \mathbb{N}$ and how the value of k depends on the length of the cycle $\sigma = \sigma_k \circ \sigma_{k-1} \circ \cdots \circ \sigma_1$.

5 Counters via Linear Transformation

The construction in this section is based on linear transformations. Consider the vector space \mathbb{F}_q^n , and let $L : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^n$ be a linear transformation. A basic fact in linear algebra says that if L has *full* rank, then the mapping given by L is a bijection. Thus, when L is full rank, the mapping can also be thought of as a permutation over \mathbb{F}_q^n . Throughout this section we use many basic terms related to linear transformation without defining them, for the details of which we refer the reader to any standard text book on linear algebra (e.g. [28]).

A natural way to build counter out of a full rank linear transformation is to fix a starting element, and repeatedly apply the linear transformation to obtain the next element. Clearly this only list out elements in the cycle containing the starting element. Therefore, we would like to

reduce A^{-1} to a permutation matrix. A permutation matrix is a $\{0, 1\}$ -matrix that has exactly one 1 in each row and column. Now we need at most $(n - 1)$ row swaps to further reduce the obtained permutation matrix to the identity matrix. We claim that a row swap can be implemented by left multiplication with at most 4 elementary matrices. Indeed, to swap row i and row j , the following sequence of operation suffices: (i) add j -th row to i -th row, (ii) subtract i -th row from j -th row, (iii) add j -th row to i -th row, and (iv) multiply j -th row with -1 . (The last operation is not required if the characteristic of the underlying field is 2.)

Hence, the inverse of A^{-1} which is our original matrix A is the product of k elementary matrices. \square

5.1 Construction of the counter

Let A be a full rank linear transformation from \mathbb{F}_q^n to \mathbb{F}_q^n such that when viewed as permutation it is a single cycle of length $q^n - 1$. More specifically, A is an invertible matrix in $\mathbb{F}_q^{n \times n}$ such that for any $x \in \mathbb{F}_q^n$ where $x \neq (0, \dots, 0)$, $Ax, A^2x, \dots, A^{(q^n-1)}x$ are distinct. Such a matrix exists, for example, take A to be the matrix of a linear transformation that corresponds to multiplication from left by a fixed generator of the multiplicative group of \mathbb{F}_{q^n} under the standard vector representation of elements of \mathbb{F}_{q^n} . Let $A = E_k E_{k-1} \cdots E_1$ where E_i 's are elementary matrices.

Theorem 5.2. *Let q , A , and k be as defined above. Let $r \geq \log_q k$. There exists a quasi-Gray code on the domain $(\mathbb{F}_q)^{n+r}$ of length $q^{n+r} - q^r$ that can be implemented using a decision assignment tree T such that $\text{READ}(T) \leq r + 2$ and $\text{WRITE}(T) \leq 2$.*

Proof. The proof follows readily from Lemma 4.3, where E_i 's play the role of σ_i 's, and noting that the permutation given by any elementary matrix can be implemented using a decision assignment tree that reads at most two coordinates and writes at most one. For the counter C' on $(\mathbb{F}_q)^r$ we chose a Gray code of trivial read complexity r and write complexity 1. \square

Thus, we obtain a counter on a domain of size roughly kq^n that misses at most qk elements. Clearly, we would like to minimize k . A trivial bound on k is $O(n^2)$ that follows from Proposition 5.1. We now discuss the choice of A so that k becomes $O(n)$. We start with recalling a notion of primitive polynomials over finite fields.

Definition 5 (Primitive polynomials). *A primitive polynomial $p(z) \in \mathbb{F}_q[z]$ of degree n is a monic irreducible polynomial over \mathbb{F}_q such that any root of it in \mathbb{F}_{q^n} generates the multiplicative group of \mathbb{F}_{q^n} .*

Theorem 5.3 ([29]). *The number of primitive polynomials of degree n over \mathbb{F}_q equals $\phi(q^n - 1)/n$, where $\phi(\cdot)$ is the Euler ϕ -function.*

Let $p(z)$ be a primitive polynomial of degree n over \mathbb{F}_q . The elements of \mathbb{F}_{q^n} can be uniquely expressed as a polynomial in z over \mathbb{F}_q of degree at most $n - 1$. In particular, we can identify an element of \mathbb{F}_{q^n} with a vector in \mathbb{F}_q^n that is given by the coefficient vector of the unique polynomial expression of degree at most $n - 1$. But, since $p(z)$ is primitive, we also know that $\mathbb{F}_{q^n} = \{0, 1, z, z^2, \dots, z^{q^n-2}\}$. This suggests a particularly nice linear transformation to consider: *left multiplication by z* . This is so because the matrix A of the linear transformation computing the multiplication by z is very *sparse*. In particular, if $p(z) = z^n + c_{n-1}z^{n-1} + c_{n-2}z^{n-2} + \cdots + c_1z + c_0$,

then A looks as follows:

$$\begin{pmatrix} -c_{n-1} & 1 & 0 & \cdots & 0 & 0 & 0 \\ -c_{n-2} & 0 & 1 & \cdots & 0 & 0 & 0 \\ -c_{n-3} & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -c_2 & 0 & 0 & \cdots & 0 & 1 & 0 \\ -c_1 & 0 & 0 & \cdots & 0 & 0 & 1 \\ -c_0 & 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}.$$

Thus, from the proof of Proposition 5.1, it follows that A can be written as a product of at most $n + 4(n - 1)$ elementary matrices. (When q is a power of 2, then the number of elementary matrices in the product is at most $n + 3(n - 1)$.) Hence, from the discussion above and using Theorem 5.2, we obtain the following corollaries. Setting $r = \lceil \log(4n - 3) \rceil$ in Theorem 5.2 gives:

Corollary 5.4. *For any $n' \geq 2$, and $n = n' + \lceil \log(4n' - 3) \rceil$, there exists a counter on $(\mathbb{Z}_2)^n$ that misses at most $8n$ strings and can be implemented by a decision assignment tree that reads at most $4 + \log n$ bits and writes at most 2 bits.*

By doubling the number of missed strings and increasing the number of read bits by one we can construct given counters for any \mathbb{Z}_2^n , where $n \geq 15$. In the above corollary the number of missed strings grows linearly with n . One might wonder if it is possible to keep the number of missing strings $o(n)$, while keeping the read complexity essentially the same. The next corollary shows that this is indeed possible, but at the cost of increasing the write complexity.

Corollary 5.5. *For $n \geq 2$, there exists a counter on $(\mathbb{Z}_2)^{n+O(\log n)}$ that misses out at most $O(n/\lceil \log n \rceil)$ strings. Furthermore, it can be implemented by a decision assignment tree that reads and writes at most $O(\log n)$ bits.*

Proof. The idea is simply to increase the underlying alphabet size. In particular, let $q = 2^{\lceil \log n \rceil}$ in Theorem 5.2. □

We also remark that by taking q to be $2^{\frac{n}{C}}$, where $C > 1$ is a universal constant, one would get a counter on $(\mathbb{Z}_2)^{n+O(1)}$ that misses only $O(C)$ strings (i.e. independent of n). However, the read and write complexity gets worse. They are at most $2 \binom{n}{C} + O(1)$.

For the general case, when q is a prime power, we obtain the following corollary by setting r to $\lceil \log_q(5n - 4) \rceil$ or $1 + \lceil \log_q(5n - 4) \rceil$ in Theorem 5.2.

Corollary 5.6 (Generalization of Theorem 1.2). *Let q be any prime power. For $n \geq 15$, there exists a counter on \mathbb{Z}_q^n that misses at most $5q^2n$ strings and that is computed by a decision assignment tree with read complexity at most $6 + \log_q n$ and write complexity 2.*

Remark. *We remark that the algorithm given in this section can be made uniform. To achieve that we need to obtain a primitive polynomial of degree n over \mathbb{F}_q uniformly. To this end, we can use a number of algorithms (deterministic or probabilistic); for example, [42, 45, 43]. For a thorough treatment, we refer to Chapter 2 in [44].*

6 Space-optimal Counters over \mathbb{Z}_m^n for any Odd m

This section deals with space-optimal counter over an odd-size alphabet. We start by recalling Theorem 1.1 in terms of decision assignment tree complexity.

Theorem 6.1 (Restatement of Theorem 1.1). *For any odd $m \in \mathbb{N}$ and any positive integer $n \geq 15$, there is a space-optimal 2-Gray code over \mathbb{Z}_m^n that can be computed by a decision assignment tree T such that $\text{READ}(T) \leq 4 \log_m n$.*

Before providing our construction of the quasi-Gray code, we give a short overview of how the construction will proceed. First we set $n' = n - c \cdot \log n$ for some constant $c > 0$ that will be fixed later. Then we define suitable permutations $\alpha_1, \dots, \alpha_{n'} \in \mathcal{S}_N$ where $N = m^{n'}$ such that their composition $\alpha_{n'} \circ \dots \circ \alpha_1$ is a cycle of length $m^{n'}$. Next we show that each α_i can be further decomposed into $\alpha_{i,1}, \dots, \alpha_{i,j} \in \mathcal{S}_N$ for some j , such that each $\alpha_{i,r}$ for $r \in [j]$ can be computed using a decision assignment tree with read complexity 3 and write complexity 1. Finally to complete the construction we use Lemma 4.3 with $\alpha_{i,r}$'s playing the role of $\sigma_1, \dots, \sigma_k$ in the lemma.

We recall the notion of r -functions over \mathbb{Z}_m^n that was introduced by Coppersmith and Grossman [12] for $m = 2$. Below we generalize that definition for any $m \in \mathbb{N}$.

Definition 6. *For any $r \in [n - 1]$, an r -function on \mathbb{Z}_m^n is a permutation τ over \mathbb{Z}_m^n identified by a subset $\{i_1, \dots, i_r, j\} \subseteq [n]$ of size $r + 1$ and a function $f : \mathbb{Z}_m^r \rightarrow \mathbb{Z}_m$ such that for any $\langle a_1, \dots, a_n \rangle \in \mathbb{Z}_m^n$,*

$$\tau(\langle a_1, \dots, a_n \rangle) = \langle a_1, \dots, a_{j-1}, a_j + f(a_{i_1}, \dots, a_{i_r}), a_{j+1}, \dots, a_n \rangle.$$

Observe that any r -function can be implemented using a decision assignment tree T that queries x_{i_1}, \dots, x_{i_r} and x_j at internal nodes, and at leaves it assigns value only to the variable x_j . Thus, $\text{READ}(T) = r + 1$ and $\text{WRITE}(T) = 1$.

Claim 6.2. *Any r -function on \mathbb{Z}_m^n can be implemented using a decision assignment tree T with n variables such that $\text{READ}(T) = r + 1$ and $\text{WRITE}(T) = 1$.*

We are now ready to provide details of our construction of a space-optimal quasi-Gray code over \mathbb{Z}_m^n for any odd m . Define $n' := n - c \cdot \log_m n$ for some constant $c > 0$ that will be fixed later.

Step 1: Construction of $\alpha_1, \dots, \alpha_{n'}$.

We consider specific permutations $\alpha_1, \dots, \alpha_{n'}$ over $\mathbb{Z}_m^{n'}$ such that $\alpha = \alpha_{n'} \circ \dots \circ \alpha_1$ is a cycle of length $m^{n'}$. We define them below.

Definition 7. *Let m and n' be natural numbers. For $i \in [n']$, we define α_i to be the permutation given by the following map: for any $\langle x_1, \dots, x_i, \dots, x_{n'} \rangle \in \mathbb{Z}_m^{n'}$,*

$$\alpha_i(\langle x_1, \dots, x_i, \dots, x_{n'} \rangle) = \begin{cases} \langle x_1, \dots, x_{i-1}, x_i + 1, x_{i+1}, \dots, x_{n'} \rangle & \text{if } x_j = 0 \text{ for all } j \in [i - 1], \\ \langle x_1, \dots, x_i, \dots, x_{n'} \rangle & \text{otherwise.} \end{cases}$$

The addition operation in the mapping $x_i \leftarrow x_i + 1$ is within \mathbb{Z}_m .

The following observation is easily seen from the definitions of r -functions and α_i .

Claim 6.3. *For any $i \in [n']$, α_i is an $(i - 1)$ -function on $\mathbb{Z}_m^{n'}$. Furthermore, each α_i is composed of disjoint cycles of length m over $\mathbb{Z}_m^{n'}$.*

We now establish a crucial property of the α_i 's, i.e., their composition is a *full* length cycle.

Claim 6.4. *$\alpha = \alpha_{n'} \circ \dots \circ \alpha_1$ is a cycle of length $m^{n'}$.*

Proof. Consider the sequence of permutations $\tau_1, \dots, \tau_{n'}$ such that $\tau_i = \alpha_i \circ \alpha_{i-1} \circ \dots \circ \alpha_1$ for $i \in [n']$. Clearly, $\tau_{n'} = \alpha$. We now prove the claim. In fact, we establish the following stronger claim: for $i \in [n']$, τ_i is a permutation composed of $m^{n'-i}$ disjoint cycles, each of the cycles being of length m^i . Furthermore, for every $\langle a_{i+1}, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'-i}$ there is a cycle that involves all tuples of the form $\langle x_1, \dots, x_i, a_{i+1}, \dots, a_{n'} \rangle$. The claim, α is a cycle of length $m^{n'}$, follows as a consequence. We prove the stronger claim by induction on i .

Base case: $\tau_1 = \alpha_1$. From the definition of α_1 , it follows that there is a cycle of length m of the form $(\langle 0, a_2, \dots, a_{n'} \rangle, \langle 1, a_2, \dots, a_{n'} \rangle, \dots, \langle m-1, a_2, \dots, a_{n'} \rangle)$ for each $\langle a_2, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'-1}$. Hence our induction hypothesis clearly holds for the base case.

Induction step: Suppose our induction hypothesis holds until some $i \in [n']$ and we would like to establish it for $i+1$. Let us consider the permutation α_{i+1} . We know that it is composed of $m^{n'-(i+1)}$ disjoint cycles of length m . Indeed, for each $\langle a_{i+2}, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'-(i+1)}$, α_{i+1} contains a cycle that involves all m tuples where the first i coordinates are all 0 and the last $n' - (i+1)$ coordinates are set to $\langle a_{i+2}, \dots, a_{n'} \rangle$. From the cycle decomposition of α_{i+1} and τ_i into disjoint cycles, it is clear that for any cycle say C in α_{i+1} there are m disjoint cycles C_1, \dots, C_m in τ_i , each of them intersecting C in exactly one element. Consider $C' = C \circ C_1 \circ \dots \circ C_m$. By repeated application of Proposition 4.1, we conclude that C' is a cycle of length $\sum_{i=1}^m |C_i| = m^{i+1}$. (Here by $|C_i|$ we mean the length of the cycle C_i .) Also C' involves tuples where the last $n' - (i+1)$ coordinates are set to some fixed $\langle a_{i+2}, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'-(i+1)}$. Thus, C' is a cycle of length m^{i+1} containing all tuples of the form $\langle x_1, \dots, x_{i+1}, a_{i+2}, \dots, a_{n'} \rangle$. Since $\tau_{i+1} = \alpha_{i+1} \circ \tau_i$ and α_{i+1} contains $m^{n'-(i+1)}$ disjoint cycles, we conclude that τ_{i+1} consists of exactly $m^{n'-(i+1)}$ disjoint cycles, each of length m^{i+1} and containing tuples of the required form.

This finishes the proof. \square

Readers may note that this step does not use the fact that m is odd and, thus, it is true for any $m \in \mathbb{N}$. If we were to directly implement α_i by a decision assignment tree, its read complexity would be i . Hence, we would not get any savings in Lemma 4.3. So we need to further decompose α_i into permutations of small read complexity.

Step 2: Further decomposition of α_i 's.

Our goal is to describe α_i as a composition of 2-functions. Recall, Claim 6.3, each α_i is an $(i-1)$ -function on $\mathbb{Z}_m^{n'}$. Suppose, for $i \in [n']$, there exists a set of 2-functions $\alpha_{i,1}, \dots, \alpha_{i,k_i}$ such that $\alpha_i = \alpha_{i,k_i} \circ \dots \circ \alpha_{i,1}$. Then using Lemma 4.3, where $\alpha_{i,k}$'s play the role of σ_j 's, we obtain a decision assignment tree implementing a 2-Gray code with potentially low read complexity. Indeed, each $\alpha_{i,k}$ has low read complexity by Claim 6.2, hence the read complexity essentially depends on how large is $\sum_i k_i$. In the following we will argue that α_i 's can be decomposed into a *small* set of 2-functions, thus keeping the maximum k_i small. As a result, the read complexity bound in Theorem 6.1 will follow.

Note α_1, α_2 and α_3 are already 2-functions. In the case of α_i , $4 \leq i \leq n' - 2$, we can directly adopt the technique from [3, 7] to generate the desired set of 2-functions. However, as discussed in Section 1.2, that technique falls short when $i > n' - 2$. (It needs two free registers to operate.) For $i = n' - 1$, it is possible to generalize the proof technique of [12] to decompose $\alpha_{n'-1}$. Unfortunately all the previously known techniques fail to decompose $\alpha_{n'}$ and we have to develop a new technique.

First we provide the adaptation of [3, 7], and then develop a new technique that allows us to express both $\alpha_{n'-1}$ and $\alpha_{n'}$ as a composition of small number of 2-functions, thus overcoming the challenge described above.

Lemma 6.5. *For any $4 \leq i \leq n' - 2$, let α_i be the permutation given by Definition 7. Then there exists a set of 2-functions $\alpha_{i,1}, \dots, \alpha_{i,k_i}$ such that $\alpha_i = \alpha_{i,k_i} \circ \dots \circ \alpha_{i,1}$, and $k_i \leq 4(i-1)^2 - 3$.*

It is worth noting that, although in this section we consider m to be odd, the above lemma holds for any $m \in \mathbb{N}$. In [3], computation uses only addition and multiplication from the ring, whereas we can use any function $g : \mathbb{Z}_m \rightarrow \mathbb{Z}_m$. This subtle difference makes the lemma to be true for any $m \in \mathbb{N}$ instead of being true only for prime powers.

Proof. Pick $i \leq n' - 2$. Let us represent α_i as an $(i-1)$ -function. From the definition we have,

$$\alpha_i(\langle a_1, \dots, a_i, \dots, a_{n'} \rangle) = \langle a_1, \dots, a_{i-1}, a_i + f(a_1, \dots, a_{i-1}), a_{i+1}, \dots, a_{n'} \rangle,$$

where the map $f : \mathbb{Z}_m^{i-1} \rightarrow \mathbb{Z}_m$ is defined as follows:

$$f(a_1, \dots, a_{i-1}) = \begin{cases} 1 & \text{if } (a_1, \dots, a_{i-1}) = (0, \dots, 0), \\ 0 & \text{otherwise.} \end{cases}$$

Observe that f is an indicator function of a tuple; in particular, of the all-zeroes tuple. To verify the lemma, we would prove a stronger claim than the statement of the lemma. Consider the set \mathcal{S} of r -functions, $1 \leq r \leq n' - 3$, such that the function f used to define them is the indicator function of the all-zeroes tuple. That is, $\tau \in \mathcal{S}$ if and only if there exists a set $\{i_1, i_2, \dots, i_r\} \subseteq [n']$ of size r and a $j \in [n'] \setminus \{i_1, i_2, \dots, i_r\}$ such that

$$\tau(\langle a_1, \dots, a_{n'} \rangle) = \langle a_1, \dots, a_{j-1}, a_j + f(a_{i_1}, \dots, a_{i_r}), a_{j+1}, \dots, a_{n'} \rangle,$$

where $f(x_1, \dots, x_r) = 1$ if $(x_1, \dots, x_r) = (0, \dots, 0)$, and 0 otherwise. Observe that $\alpha_i \in \mathcal{S}$ for $4 \leq i \leq n' - 2$. We establish the following stronger claim.

Claim 6.6. *For an r -function $\tau \in \mathcal{S}$, there exist 2-functions $\tau_1, \dots, \tau_{k_r}$ such that $\tau = \tau_{k_r} \circ \dots \circ \tau_1$ and $k_r \leq 4r^2 - 3$. (We stress that $\tau_1, \dots, \tau_{k_r}$ need not belong to \mathcal{S} .)*

Clearly, the claim implies the lemma. We now prove the claim by induction on r . The base case is $r \leq 2$, in which case the claim trivially holds. Suppose the claim holds for all $(r-1)$ -functions in \mathcal{S} . Let $\tau \in \mathcal{S}$ be an r -function identified by the set $S := \{i_1, i_2, \dots, i_r\} \subseteq [n']$ and $j \in [n'] \setminus S$. Since f is an indicator function, it can be expressed as a product of indicator functions. That is, $f(a_{i_1}, \dots, a_{i_r}) = \prod_{s \in S} g(a_s)$ where $g : \mathbb{Z}_m \rightarrow \mathbb{Z}_m$ is the following $\{0, 1\}$ -map:

$$g(y) = \begin{cases} 1 & \text{if } y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Consider a partition of S into two sets A and B of sizes $\lfloor r/2 \rfloor$ and $\lceil r/2 \rceil$, respectively. Let j_1 and j_2 be two distinct integers in $[n'] \setminus S \cup \{j\}$. The existence of such integers is guaranteed by the bound on r . We now express τ as a composition of $(r/2)$ -functions and 2-functions, and then use induction hypothesis to complete the proof. The decomposition of τ is motivated by the following identity:

$$\begin{aligned} a_j + \prod_{s \in S} g(a_s) &= a_j + \left(a_{j_1} + \prod_{s \in A} g(a_s) - a_{j_1} \right) \left(a_{j_2} + \prod_{s \in B} g(a_s) - a_{j_2} \right) \\ &= a_j + \left(a_{j_1} + \prod_{s \in A} g(a_s) \right) \left(a_{j_2} + \prod_{s \in B} g(a_s) \right) \\ &\quad - \left(a_{j_1} + \prod_{s \in A} g(a_s) \right) a_{j_2} - a_{j_1} \left(a_{j_2} + \prod_{s \in B} g(a_s) \right) + a_{j_1} a_{j_2}. \end{aligned}$$

Therefore, we consider three permutations γ , τ_A and τ_B such that for any $\langle a_1, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'}$ their maps are given as follows:

$$\begin{aligned}\gamma(\langle a_1, \dots, a_n \rangle) &= \langle a_1, \dots, a_{j-1}, a_j + a_{j_1} a_{j_2}, a_{j+1}, \dots, a_{n'} \rangle, \\ \tau_A(\langle a_1, \dots, a_{n'} \rangle) &= \langle a_1, \dots, a_{j_1-1}, a_{j_1} + \prod_{s \in A} g(a_s), a_{j_1+1}, \dots, a_{n'} \rangle, \text{ and} \\ \tau_B(\langle a_1, \dots, a_{n'} \rangle) &= \langle a_1, \dots, a_{j_2-1}, a_{j_2} + \prod_{s \in B} g(a_s), a_{j_2+1}, \dots, a_{n'} \rangle,\end{aligned}$$

where both the multiplications and additions are in \mathbb{Z}_m . Using the identity it is easy to verify the following decomposition of τ :

$$\tau = \tau_B^{-1} \circ \gamma^{-1} \circ \tau_A^{-1} \circ \gamma \circ \tau_B \circ \gamma^{-1} \circ \tau_A \circ \gamma.$$

Clearly, γ is a 2-function, while τ_A and τ_B are $\lfloor r/2 \rfloor$ -function and $\lceil r/2 \rceil$ -function, respectively, and belong to \mathcal{S} . By induction hypothesis τ_A and τ_B can be expressed as a composition of 2-functions. Thus their inverses too. Hence we obtain a decomposition of τ in terms of 2-functions. The bound on k_r , the length of the decomposition, follows from the following recurrence:

$$T(r) \leq 2T(\lfloor r/2 \rfloor) + 2T(\lceil r/2 \rceil) + 4.$$

□

We would like to mention that another decomposition of τ in terms of 2-functions can be obtained by following the proof of [12], albeit with a much worse bound on the value of k_r . Further, by strengthening the induction hypothesis, it is easily seen that the above proof can be generalized to hold for certain special type of r -functions. Let β be an r -function, $r \leq n' - 3$, such that for any $\langle a_1, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'}$,

$$\beta(\langle a_1, \dots, a_{n'} \rangle) = \langle a_1, \dots, a_{i-1}, a_i + f_e(a_{i_1}, \dots, a_{i_r}), a_{i+1}, \dots, a_{n'} \rangle,$$

where the function $f_e : \mathbb{Z}_m^r \rightarrow \mathbb{Z}_m$ is defined by:

$$f_e(x) = \begin{cases} b & \text{if } x = e, \\ 0 & \text{otherwise,} \end{cases}$$

for some $b \in \mathbb{Z}_m$ and $e \in \mathbb{Z}_m^r$, i.e., f_e is some constant multiple of the characteristic function of the tuple e . A crucial step in the proof is to express f_e as a product of indicator functions. In this case we consider the following functions $g_{i_j} : \mathbb{Z}_m \rightarrow \mathbb{Z}_m$ for $1 \leq j \leq r$. Define $g_{i_1} : \mathbb{Z}_m \rightarrow \mathbb{Z}_m$ such that for any $y \in \mathbb{Z}_m$, $g_{i_1}(y) = b$ if $y = e_1$, and 0 otherwise. For any $2 \leq j \leq r$, define $g_{i_j} : \mathbb{Z}_m \rightarrow \mathbb{Z}_m$ as $g_{i_j}(y) = 1$ if $y = e_i$, and 0 otherwise. By definition, we have $f_e(x_1, \dots, x_r) = g_{i_1}(x_1)g_{i_2}(x_2) \cdots g_{i_r}(x_r)$. Thus we get the following generalization of Lemma 6.5.

Lemma 6.7. *For any $m \in \mathbb{N}$ and $1 \leq r \leq n' - 3$, let τ be an r -function such that for any $\langle a_1, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'}$,*

$$\tau(\langle a_1, \dots, a_{n'} \rangle) = \langle a_1, \dots, a_{j-1}, a_j + f_e(a_{i_1}, \dots, a_{i_r}), a_{j+1}, \dots, a_{n'} \rangle$$

where the function $f_e : \mathbb{Z}_m^r \rightarrow \mathbb{Z}_m$ is defined by: $f_e(x) = b$ if $x = e$; and 0 otherwise. Then there exists a set of 2-functions $\tau_1, \dots, \tau_{k_r}$ such that $\tau = \tau_{k_r} \circ \cdots \circ \tau_1$, and $k_r \leq 4r^2 - 3$.

Comment on decomposition of general r -functions for $3 \leq r \leq n' - 3$: Any function $f : \mathbb{Z}_m^r \rightarrow \mathbb{Z}_m$ can be expressed as $f = \sum_e c_e \cdot \chi_e$, where χ_e is the characteristic function of the tuple $e \in \mathbb{Z}_m^r$, and $c_e \in \mathbb{Z}_m$. Thus Lemma 6.7 suffices to argue that any r -function can be decomposed into a set of 2-functions. However, the implied bound on k_r may not be small. In particular, the number of tuples where f takes non-zero value might be large.

It remains to decompose $\alpha_{n'-1}$ and $\alpha_{n'}$. The following lemma about cycles that intersect at exactly one point serves as a key tool in our decomposition.

Lemma 6.8. *Suppose there are two cycles, $\sigma = (t, a_1, \dots, a_{\ell-1})$ and $\tau = (t, b_1, \dots, b_{\ell-1})$, of length $\ell \geq 2$ such that $a_i \neq b_j$ for every $i, j \in [\ell - 1]$. Then, $(\sigma \circ \tau)^\ell \circ (\tau \circ \sigma)^\ell = \sigma^2$.*

Proof. By Proposition 4.1, we have

$$\begin{aligned}\beta &:= \tau \circ \sigma = (t, a_1, \dots, a_{\ell-1}, b_1, \dots, b_{\ell-1}), \text{ and} \\ \gamma &:= \sigma \circ \tau = (t, b_1, \dots, b_{\ell-1}, a_1, \dots, a_{\ell-1}).\end{aligned}$$

Both β and γ are cycles of length $2\ell - 1$. Also note that $2\ell - 1$ is co-prime with ℓ . Thus both β^ℓ and γ^ℓ are also cycles of length $2\ell - 1$ given as follow:

$$\begin{aligned}\beta^\ell &= (t, b_1, a_1, b_2, a_2, \dots, b_{\ell-1}, a_{\ell-1}), \text{ and} \\ \gamma^\ell &= (t, a_1, b_1, a_2, b_2, \dots, a_{\ell-1}, b_{\ell-1}).\end{aligned}$$

Now by Proposition 4.2,

$$\begin{aligned}\sigma \circ \beta^\ell \circ \sigma^{-1} &= (\sigma(t), \sigma(b_1), \sigma(a_1), \dots, \sigma(a_{\ell-2}), \sigma(b_{\ell-1}), \sigma(a_{\ell-1})) \\ &= (a_1, b_1, a_2, b_2, \dots, a_{\ell-1}, b_{\ell-1}, t) = \gamma^\ell.\end{aligned}$$

Therefore,

$$\begin{aligned}\beta^\ell \circ \sigma^{-1} &= (t, a_{\ell-1}, a_{\ell-2}, \dots, a_1) \circ (t, a_1, b_1, a_2, b_2, \dots, a_{\ell-1}, b_{\ell-1}) \\ &= (a_1, b_1)(a_2, b_2) \cdots (a_{\ell-1}, b_{\ell-1}).\end{aligned}$$

It is thus evident that $(\beta^\ell \circ \sigma^{-1})^2$ is the identity permutation. Hence,

$$\begin{aligned}\gamma^\ell \circ \beta^\ell &= \sigma \circ \beta^\ell \circ \sigma^{-1} \circ \beta^\ell \\ &= \sigma \circ \beta^\ell \circ \sigma^{-1} \circ \beta^\ell \circ \sigma^{-1} \circ \sigma \\ &= \sigma^2.\end{aligned}$$

□

Before going into the detailed description of the decomposition procedure, let us briefly discuss the main idea. Here we first consider $\alpha_{n'}$. The case of $\alpha_{n'-1}$ will be analogous. Recall that $\alpha_{n'} = (\langle 00 \cdots 00 \rangle, \langle 00 \cdots 01 \rangle, \langle 00 \cdots 02 \rangle, \dots, \langle 00 \cdots 0(m-1) \rangle)$ is a cycle of length m . For $a = (m+1)/2$, we define $\sigma = (\langle 00 \cdots 0(0 \cdot a) \rangle, \langle 00 \cdots 0(1 \cdot a) \rangle, \langle 00 \cdots 0(2 \cdot a) \rangle, \dots, \langle 00 \cdots 0((m-1) \cdot a) \rangle)$, and $\tau = (\langle (0 \cdot a)00 \cdots 0 \rangle, \langle (1 \cdot a)00 \cdots 0 \rangle, \langle (2 \cdot a)00 \cdots 0 \rangle, \dots, \langle ((m-1) \cdot a)00 \cdots 0 \rangle)$, where the multiplication is in \mathbb{Z}_m . Since m is co-prime with $(m+1)/2$, σ and τ are cycles of length m . (Here we use the fact that m is odd.) Observe that $\sigma^2 = \alpha_{n'}$, so by applying Lemma 6.8 to σ and τ we get $\alpha_{n'}$. It might seem we didn't make much progress towards decomposition, as now instead of one $(n' - 1)$ -function $\alpha_{n'}$ we have to decompose two $(n' - 1)$ -functions σ and τ . However, we will not decompose σ and τ directly, but rather we obtain a decomposition for $(\sigma \circ \tau)^m$ and $(\tau \circ \sigma)^m$. Surprisingly this can be done using Lemma 6.7 although indirectly.

We consider an $(n' - 3)$ -function σ' whose cycle decomposition contains σ as one of its cycles. Similarly we consider a 3-function τ' whose cycle decomposition contains τ as one of its cycles. We carefully choose these σ' and τ' such that $(\sigma' \circ \tau')^m = (\sigma \circ \tau)^m$ and $(\tau' \circ \sigma')^m = (\tau \circ \sigma)^m$. We will use Lemma 6.7 to directly decompose σ' and τ' to get the desired decomposition.

Lemma 6.9. *For any $n' - 1 \leq i \leq n'$, let α_i be the permutation over $\mathbb{Z}_m^{n'}$ given by Definition 7 where m is odd. Then, there exists a set of 2-functions $\alpha_{i,1}, \dots, \alpha_{i,k_i}$ such that $\alpha_i = \alpha_{i,k_i} \circ \dots \circ \alpha_{i,1}$, and $k_i = O(m \cdot (i - 1)^2)$.*

Proof. For the sake of brevity, we will only describe the procedure to decompose $\alpha_{n'}$ into a set of 2-functions. The decomposition of $\alpha_{n'-1}$ is completely analogous. (We comment on this more at the end of the proof.)

Let σ be the following permutation: for any $\langle a_1, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'}$,

$$\sigma(\langle a_1, \dots, a_{n'} \rangle) = \langle a_1, \dots, a_{n'-1}, a_{n'} + f(a_1, \dots, a_{n'-1}) \rangle$$

where the function $f : \mathbb{Z}_m^{n'-1} \rightarrow \mathbb{Z}_m$ is defined as follows:

$$f(x) = \begin{cases} (m+1)/2 & \text{if } x = \langle 0, \dots, 0 \rangle, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $(m+1)/2$ is well defined because m is odd. Further, since m and $(m+1)/2$ are co-prime, σ is a m length cycle. Moreover, $\sigma^2 = \alpha_{n'}$. The description of σ uses crucially that m is odd. If m were not odd, then the description fails. Indeed finding a substitute for σ is the main hurdle that needs to be addressed to handle the case when m is even.

We also consider another permutation τ such that for any $\langle a_1, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'}$,

$$\tau(\langle a_1, \dots, a_{n'} \rangle) = \langle a_1 + f(a_2, \dots, a_{n'}), a_2, \dots, a_{n'} \rangle$$

where the function f is the same as in the definition of σ . So τ is also a cycle of length m . Let $t = (m+1)/2$. Then the cycle decomposition of σ and τ are,

$$\begin{aligned} \sigma &= (\langle 0, 0, \dots, 0, 0 \cdot t \rangle, \langle 0, 0, \dots, 0, 1 \cdot t \rangle, \langle 0, 0, \dots, 0, 2 \cdot t \rangle, \dots, \langle 0, 0, \dots, 0, (m-1) \cdot t \rangle), \text{ and} \\ \tau &= (\langle 0 \cdot t, 0, 0, \dots, 0 \rangle, \langle 1 \cdot t, 0, 0, \dots, 0 \rangle, \langle 2 \cdot t, 0, 0, \dots, 0 \rangle, \dots, \langle (m-1) \cdot t, 0, 0, \dots, 0 \rangle) \end{aligned}$$

where the multiplication is in \mathbb{Z}_m . Observe that $\langle 0, \dots, 0 \rangle$ is the only common element involved in both cycles σ and τ . Therefore, by Lemma 6.8,

$$(\sigma \circ \tau)^m \circ (\tau \circ \sigma)^m = \sigma^2 = \alpha_{n'}.$$

We note that both σ and τ are $(n' - 1)$ -functions. Thus so far it is not clear how the above identity helps us to decompose $\alpha_{n'}$. We now define two more permutations σ' and τ' such that they are themselves decomposable into 2-functions and, moreover, $(\sigma' \circ \tau')^m = (\sigma \circ \tau)^m$ and $(\tau' \circ \sigma')^m = (\tau \circ \sigma)^m$.

The permutations σ' and τ' are defined as follows: for any $\langle a_1, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'}$,

$$\begin{aligned} \sigma'(\langle a_1, \dots, a_{n'} \rangle) &= \langle a_1, \dots, a_{n'-1}, a_{n'} + g(a_1, \dots, a_{n'-3}) \rangle, \text{ and} \\ \tau'(\langle a_1, \dots, a_{n'} \rangle) &= \langle a_1 + h(a_{n'-2}, a_{n'-1}, a_{n'}), a_2, \dots, a_{n'} \rangle \end{aligned}$$

where $g : \mathbb{Z}_m^{n'-3} \rightarrow \mathbb{Z}_m$ is the following map:

$$g(x) = \begin{cases} (m+1)/2 & \text{if } x = \langle 0, \dots, 0 \rangle, \\ 0 & \text{otherwise,} \end{cases}$$

and $h : \mathbb{Z}_m^3 \rightarrow \mathbb{Z}_m$ is similarly defined but on lower dimension:

$$h(x) = \begin{cases} (m+1)/2 & \text{if } x = \langle 0, 0, 0 \rangle, \\ 0 & \text{otherwise.} \end{cases}$$

Since m and $(m+1)/2$ are co-prime, σ' is composed of m^2 disjoint cycles, each of length m . Similarly, τ' is composed of $m^{(n'-4)}$ disjoint cycles, each of length m . Moreover, σ is one of the cycles among m^2 disjoint cycles of σ' and τ is one of the cycles among $m^{(n'-4)}$ disjoint cycles of τ' . So, we can write

$$\begin{aligned} \sigma' &= \sigma \circ C_1 \circ \cdots \circ C_{m^2-1}, \text{ and} \\ \tau' &= \tau \circ C'_1 \circ \cdots \circ C'_{m^{(n'-4)}-1} \end{aligned}$$

where each of C_i 's and C'_j 's is a m length cycle. An important fact regarding σ' and τ' is that the *only* element moved by both is the all-zeros tuple $\langle 0, \dots, 0 \rangle$. This is easily seen from their definitions. Recall we had observed that the all-zeroes is, in fact, moved by σ and τ . In other words, cycles in the set $\{C_1, \dots, C_{m^2-1}, C'_1, \dots, C'_{m^{(n'-4)}-1}\}$ are mutually disjoint, as well as they are disjoint from σ and τ . Thus, using the fact that C_i 's and C'_j 's are m length cycles, we have

$$\begin{aligned} (\tau' \circ \sigma')^m &= (\tau \circ \sigma)^m, \text{ and} \\ (\sigma' \circ \tau')^m &= (\sigma \circ \tau)^m. \end{aligned}$$

Therefore, we can express $\alpha_{n'}$ in terms of σ' and τ' ,

$$\alpha_{n'} = \sigma^2 = (\sigma \circ \tau)^m \circ (\tau \circ \sigma)^m = (\sigma' \circ \tau')^m \circ (\tau' \circ \sigma')^m.$$

But, by definition, σ' and τ' are an $(n'-3)$ -function and a 3-function, respectively. Furthermore, they satisfy the requirement of Lemma 6.7. Hence both σ' and τ' can be decomposed into a set of 2-functions. As a result we obtain a decomposition of $\alpha_{n'}$ into a set of $k_{n'}$ many 2-functions, where $k_{n'} \leq 2m(4(n'-3)^2 - 3 + 4 \cdot 3^2 - 3) = m(8(n'-3)^2 + 60) \leq 60 \cdot m(n'-3)^2$.

The permutation $\alpha_{n'-1}$ can be decomposed in a similar fashion. Note that $\alpha_{n'-1}$ is composed of m disjoint m length cycles. Each of the m disjoint cycles can be decomposed using the procedure described above. If we do so, we get the length $k_{n'-1} = O(m^2 \cdot (n'-3)^2)$, which would suffice for our purpose. However, we can improve this bound to $O(m \cdot (n'-2)^2)$ by a slight modification of σ, τ, σ' and τ' . Below we define these permutations and leave the rest of the proof to the reader as it is analogous to the proof above. The argument should be carried with σ, τ, σ' and τ' defined as follows: for any $\langle a_1, \dots, a_{n'} \rangle \in \mathbb{Z}_m^{n'}$,

$$\begin{aligned} \sigma(\langle a_1, \dots, a_{n'} \rangle) &= \langle a_1, \dots, a_{n'-2}, a_{n'-1} + f(a_1, \dots, a_{n'-2}), a_{n'} \rangle, \\ \tau(\langle a_1, \dots, a_{n'} \rangle) &= \langle a_1 + f(a_2, \dots, a_{n'-1}), \dots, a_{n'-1}, a_{n'} \rangle, \\ \sigma'(\langle a_1, \dots, a_{n'} \rangle) &= \langle a_1, \dots, a_{n'-2}, a_{n'-1} + g(a_1, \dots, a_{n'-3}), a_{n'} \rangle, \text{ and} \\ \tau'(\langle a_1, \dots, a_{n'} \rangle) &= \langle a_1 + h(a_{n'-2}, a_{n'-1}), a_2, \dots, a_{n'-1}, a_{n'} \rangle, \end{aligned}$$

where the function $f : \mathbb{Z}_m^{n'-2} \rightarrow \mathbb{Z}_m$ is defined as: $f(x) = (m+1)/2$ if $x = \langle 0, \dots, 0 \rangle$; otherwise $f(x) = 0$, the function $g : \mathbb{Z}_m^{n'-3} \rightarrow \mathbb{Z}_m$ is defined as: $g(x) = (m+1)/2$ if $x = \langle 0, \dots, 0 \rangle$; otherwise $g(x) = 0$ and the function $h : \mathbb{Z}_m^2 \rightarrow \mathbb{Z}_m$ is defined as: $h(x) = (m+1)/2$ if $x = \langle 0, 0 \rangle$; otherwise $h(x) = 0$. The only difference is now that both σ and τ are composed of m disjoint cycles of length m , instead of just one cycle as in case of $\alpha_{n'}$. However every cycle of σ has non-empty intersection with exactly one cycle of τ , and furthermore, the intersection is singleton. Hence we can still apply Lemma 6.8 with m pairs of m length cycles, where each pair consists of one cycle from σ and another from τ such that they have non-empty intersection. \square

Step 3: Invocation of Lemma 4.3

To finish the construction we just replace α_i 's in $\alpha = \alpha_{n'} \circ \dots \circ \alpha_1$ by $\alpha_{i,k_i} \circ \dots \circ \alpha_{i,1}$. Take the resulting sequence as $\sigma = \sigma_k \circ \dots \circ \sigma_1$, where $k = \sum_{i=1}^{n'} k_i \leq c_3 \cdot m \cdot n'^3$ for some constant $c_3 > 0$. Now we apply Lemma 4.3 to get a space-optimal counter over \mathbb{Z}_m^n . In Lemma 4.3, we set $k' = m^{r'}$ such that r' is the smallest integer for which $m^{r'} \geq k$ and set $\mathcal{D}' = \mathbb{Z}_m^{r'}$. Hence $r' \leq \log_m k + 1 \leq 3 \log_m n + c$, for some constant $c > 1$. Since each of σ_i 's is a 2-function, by Claim 6.2 it can be implemented using a decision assignment tree T_i such that $\text{READ}(T_i) = 3$ and $\text{WRITE}(T_i) = 1$. In Lemma 4.3, we use the standard space-optimal Gray code over $\mathcal{D}' = \mathbb{Z}_m^{r'}$ as C' . The code C' can be implemented using a decision assignment tree T' with $\text{READ}(T') = r'$ and $\text{WRITE}(T') = 1$ (Theorem 2.1). Hence the final counter implied from Lemma 4.3 can be computed using a decision assignment tree T with $\text{READ}(T) \leq 4 \log_m n$ and $\text{WRITE}(T) = 2$.

Odd permutations: the real bottleneck We saw our decomposition procedure over \mathbb{Z}_m^n into 2-functions (Lemma 6.9) requires that m be odd. It is natural to wonder whether such a decomposition exists irrespective of the parity of m . That is, we would like to design a set of 2-functions that generates a cycle of length m^n where m is even. Unfortunately, the answer to this question is, unequivocally, **NO**. A cycle of length m^n is an *odd permutation* when m is even. However, Coppersmith and Grossman [12] showed that 2-functions are *even permutations* for $m = 2$. Their proof is easily seen to be extended to hold for *all* m . Hence, 2-functions can only generate even permutations. Thus, ruling out the possibility of decomposing a full cycle into 2-functions when m is even. In fact, Coppersmith and Grossman [12] showed that $(n - 2)$ -functions are all even permutations. Therefore, the decomposition remains impossible even if $(n - 2)$ -functions are allowed.

We would like to give further evidence that odd permutations are indeed the *bottleneck* towards obtaining an efficient decision assignment tree implementing a space-optimal counter over \mathbb{Z}_2^n . Suppose T is a decision assignment tree that implements an odd permutation σ over \mathbb{Z}_2^n such that $\text{READ}(T) = r$ and $\text{WRITE}(T) = w$. From the basic group theory we know that there exists an even permutation τ such that $\tau \circ \sigma$ is a cycle of length 2^n . Following the argument in [12], but using an efficient decomposition of $(n - 3)$ -functions into 2-functions (in particular, Lemma 6.7), we see that any even permutation can be decomposed into a set of 2-functions of size at most $O(2^{n-1}n^2)$. Indeed, any even permutation can be expressed as a composition of at most 2^{n-1} 3-cycles, while any 3-cycle can be decomposed into at most $O(n^2)$ 2-functions. We implement $\tau \circ \sigma$ using Lemma 4.3 to obtain a space-optimal counter over $\mathbb{Z}_2^{2n+O(\log n)}$ that is computed by a decision assignment tree with read complexity $n + O(\log n) + r$ and write complexity $w + 1$. Thus, savings made on implementing an odd permutation translate directly to the savings possible in the implementation of some space-optimal counter. By the argument in [38] we know that $r \in \Omega(n)$.

Recall the main idea behind 2-Gray codes in Section 5, when $m = 2$, is to generate the cycle of length $2^n - 1$ efficiently. Since a cycle of length $2^n - 1$ is an even permutation, we could decompose it into 2-functions. In particular, we could choose an appropriate 3-cycle as α_n in Claim 6.4 such that α becomes a cycle of length $2^n - 1$. However, the counter thus obtained misses out on $O(n^3)$ strings, instead of $O(n)$.

6.1 Getting counters for even m

We can combine the results from Theorem 1.2 and Theorem 1.1 to get a counter over \mathbb{Z}_m^n for any even m . We have already mentioned in Section 3 that if we have space-optimal quasi-Gray codes over the alphabet \mathbb{Z}_2 and \mathbb{Z}_o with o being odd then we can get a space-optimal

quasi-Gray code over the alphabet \mathbb{Z}_m for any even m . Unfortunately in Section 5, instead of space-optimal counters we were only able to generate a counter over binary alphabet that misses $O(n)$ many strings. As a consequence we cannot directly apply Theorem 3.1. The problem is following. Suppose $m = 2^\ell o$ for some $\ell > 0$ and odd o . By the argument used in the proof of Lemma 3.3 we know that there is a counter over $(\mathbb{Z}_{2^\ell})^{n-1}$ of the same length as that over $(\mathbb{Z}_2)^{\ell(n-1)}$. Furthermore the length of the counter is of the form $2^{O(\log n)}(2^{n'} - 1)$, for some n' that depends on the value of ℓn (see the construction in Section 5). Now to apply Theorem 3.1 as in the proof of Lemma 3.2, $2^{n'} - 1$ must be co-prime with o . However that may not always be the case. Nevertheless, we can choose the parameters in the construction given in Section 5 to make n' such that $2^{n'} - 1$ is co-prime with o . This is always possible because of the following simple algebraic fact. In the following proposition we use the notation \mathbb{Z}_o^* to denote the multiplicative group modulo o and $\text{ord}_o(e)$ to denote the order of any element $e \in \mathbb{Z}_o^*$.

Proposition 6.10. *For any $n \in \mathbb{N}$ and odd $o \in \mathbb{N}$, consider the set $\mathcal{S} = \{n, n+1, \dots, n + \text{ord}_o(2) - 1\}$. Then there always exists an element $n' \in \mathcal{S}$ such that $2^{n'} - 1$ is co-prime to o .*

Proof. Inside \mathbb{Z}_o^* , consider the cyclic subgroup G generated by 2, i.e., $G = \{1, 2, \dots, 2^{\text{ord}_o(2)}\}$. Clearly, $\{2^s \pmod{o} \mid s \in \mathcal{S}\} = G$. Hence there exists an element $n' \in \mathcal{S}$ such that $2^{n'} - 1 \pmod{o} = 1$. It is clear that if $\gcd(2^{n'} - 1, o) = 1$ then we are done. Now the proposition follows from the following easily verifiable fact: for any $a, b, c \in \mathbb{N}$, if $a \equiv b \pmod{c}$ then $\gcd(a, c) = \gcd(b, c)$. \square

So for any $n \in \mathbb{N}$, in the proof of Lemma 3.2 we take the first coordinate to be \mathbb{Z}_m^i for some suitably chosen $i \geq 1$ instead of just \mathbb{Z}_m . The choice of i will be such that the length of the counter over $(\mathbb{Z}_{2^\ell})^{n-i}$ will become co-prime with o . The above proposition guarantees the existence of such an $i \in [\text{ord}_o(2)]$. Hence we can conclude the following.

Theorem 6.11. *For any even $m \in \mathbb{N}$ so that $m = 2^\ell o$ where o is odd, there is a quasi-Gray code C over \mathbb{Z}_m^n of length at least $m^n - O(no^n)$, that can be implemented using a decision assignment tree which reads at most $O(\log_m n + \text{ord}_o(2))$ coordinates and writes at most 3 coordinates.*

Remark. *So far we have only talked about implementing $\text{next}(\cdot, \cdot)$ function for any counter. However we can similarly define complexity of computing $\text{prev}(\cdot, \cdot)$ function in the decision assignment tree model. We would like to emphasize that all our techniques can also be carried over to do that. To extend the result of Section 5, we take the inverse of the linear transformation matrix A and follow exactly the same proof to get the implementation of $\text{prev}(\cdot, \cdot)$. As a consequence we achieve exactly the same bound on read and write complexity. Now for the quasi-Gray code over \mathbb{Z}_m for any odd m , instead of α in the Step 1 (in Section 6), we simply consider the α^{-1} which is equal to $\alpha_1^{-1} \circ \alpha_2^{-1} \circ \dots \circ \alpha_n^{-1}$. Then we follow an analogous technique to decompose α_i^{-1} 's and finally invoke Lemma 4.3. Thus we get the same bound on read and write complexity.*

7 Lower Bound Results for Hierarchical Counter

We have mentioned that the best known lower bound on the depth of a decision assignment tree implementing any space-optimal quasi-Gray code is $\Omega(\log_m n)$. Suppose we are interested in getting a space-optimal quasi-Gray code over the domain $\mathbb{Z}_{m_1} \times \dots \times \mathbb{Z}_{m_n}$. A natural approach to achieve the $O(\log n)$ bound is to build a decision assignment tree with an additional restriction that a variable labels *exactly* one internal node. We call any counter associated with such a special type of decision assignment tree by *hierarchical counter*. We note that the aforementioned restriction doesn't guarantee an $O(\log n)$ bound on the depth; for example, a decision assignment tree implementing a counter that increments by 1 to obtain the next element, reads a variable

exactly once. Nevertheless, informally, one can argue that if a hierarchical counter is lopsided it must mimic the decision assignment tree in the previous example. We show that a space-optimal hierarchical counter over strings of length 3 exists if and only if m_2 and m_3 are co-prime.

Theorem 7.1. *Let $m_1 \in \{2, 3\}$ and $m_2, m_3 \geq 2$. There is no space-optimal hierarchical counter over $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \mathbb{Z}_{m_3}$ unless m_2 and m_3 are co-prime.*

In fact, we conjecture that the above theorem holds for all $m_1 \geq 2$.

Conjecture 1. *There is no space-optimal hierarchical counter over $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \mathbb{Z}_{m_3}$ unless m_2 and m_3 are co-prime.*

It follows already from Theorem 3.1 that a space-optimal hierarchical counter exists when m_2 and m_3 are co-prime. Hence Theorem 7.1 establishes the necessity of co-primality in Theorem 3.1. It also provides a theoretical explanation behind the already known fact, derived from an exhaustive search in [6], that there exists no space-optimal counter over $(\mathbb{Z}_2)^3$ that reads at most 2 bits to generate the next element.

Proof of Theorem 7.1. We first establish the proof when $m_1 = 2$, and then reduce the case when $m_1 = 3$ to $m_1 = 2$.

Following the definition of hierarchical counter, without loss of generality, we assume that it reads the second coordinate if the value of the first coordinate is 0; otherwise it reads the third coordinate. Let $C = (\langle a_0, b_0, c_0 \rangle, \dots, \langle a_{k-1}, b_{k-1}, c_{k-1} \rangle)$ be the cyclic sequence associated with a space-optimal hierarchical counter over $\mathbb{Z}_2 \times \mathbb{Z}_{m_2} \times \mathbb{Z}_{m_3}$. We claim that for any two consecutive elements in the sequence, $\langle a_i, b_i, c_i \rangle$ and $\langle a_{(i+1) \bmod k}, b_{(i+1) \bmod k}, c_{(i+1) \bmod k} \rangle$, the following must hold:

1. If $a_i = 0$, then $b_{i+1} \neq b_i$. For the sake of contradiction suppose not, i.e. $b_{i+1} = b_i$. Since $a_i = 0$ we do not read the third coordinate, thus, implying $c_{i+1} = c_i$. Hence, $a_i \neq a_{i+1} = 1$. Therefore, we observe that the second coordinate never changes henceforth. Thence it must be that $b_1 = b_2 = \dots = b_k$. We thus contradict space-optimality. Similarly, we can argue if $a_i = 1$, then $c_{i+1} \neq c_i$.
2. For $i \in \{0, 1, \dots, k-1\}$, $a_{(i+1) \bmod k} \neq a_i$. Assume otherwise. Let $i \in \{0\} \cup [k-1]$ be such that $a_{(i-1) \bmod k} \neq a_i = a_{(i+1) \bmod k}$. Clearly, such an i exists, else $a_1 = a_2 = \dots = a_k$ which contradicts space-optimality. We are now going to argue that there exist two distinct elements in the domain $\mathbb{Z}_2 \times \mathbb{Z}_{m_2} \times \mathbb{Z}_{m_3}$ such that they have the same successor in C , thus contradicting that C is generated by a space-optimal counter. For the sake of simplicity let us assume $a_i = 0$. The other case is similar. Since $a_i = 0$, we have $a_{i-1} = 1$ and $a_{i+1} = 0$. Thus, we have the tuple $\langle 0, b_{i+1}, c_{i+1} \rangle$ following the tuple $\langle 0, b_i, c_i \rangle$ in C . We claim that the successor of the tuple $\langle 1, b_{i+1}, c_{i-1} \rangle$ is also $\langle 0, b_{i+1}, c_{i+1} \rangle$. It is easily seen because $c_{i+1} = c_i$.

The second item above says that the period of each element in the first coordinate is 2. That is, in the sequence C the value in the first coordinate repeats every second step. We will now argue that the value in the second coordinate (respectively, third coordinate) repeats every $2m_2$ steps (respectively, $2m_3$ steps). We will only establish the claim for the second coordinate, as the argument for the third coordinate is similar. First of all observe that C induces a cycle on the elements of \mathbb{Z}_{m_2} . In particular, Consider the sequence $b_0, b_2, b_4, \dots, b_\ell$ such that b_ℓ is the first time when an element in the sequence has appeared twice. First of all, for C to be a valid cycle b_ℓ must be equal to b_0 . Secondly, $b_0, b_2, \dots, b_{\ell-2}$ are the only distinct elements that appear in the second coordinate throughout C . This is because the change in the second coordinate depends,

deterministically, on what value it holds prior to the change. Thus, for C to be space-optimal $\ell/2$ must be equal to m_2 . Hence, an element in the second coordinate repeats every $2m_2$ steps in C .

Therefore, for the cyclic sequence C , we get that $k = \text{lcm}(2m_2, 2m_3) = 2m_2m_3/\text{gcd}(m_2, m_3)$. Hence if m_2 and m_3 are not co-prime then $k < 2m_2m_3$ implying C to be not space-optimal.

We now proceed to the case $m_1 = 3$. As mentioned in the beginning, we will show that from any space-optimal hierarchical counter over $\mathbb{Z}_3 \times \mathbb{Z}_{m_2} \times \mathbb{Z}_{m_3}$ we can obtain a space-optimal hierarchical counter over $\mathbb{Z}_2 \times \mathbb{Z}_{m_2} \times \mathbb{Z}_{m_3}$.

Let $C = (\langle a_0, b_0, c_0 \rangle, \dots, \langle a_{k-1}, b_{k-1}, c_{k-1} \rangle)$ be the cyclic sequence associated with a space-optimal hierarchical counter T over $\mathbb{Z}_3 \times \mathbb{Z}_{m_2} \times \mathbb{Z}_{m_3}$. We query the first coordinate, say x_1 , at the root of T . Let T_0, T_1 , and T_2 be the three subtrees rooted at $x_1 = 0, 1$, and 2 , respectively. Without loss of generality, we assume that the roots of T_0 and T_2 are labeled by the second coordinate, say x_2 , and the root of T_1 is labeled by the third coordinate, say x_3 . We say that there is a *transition* from T_i to T_j if there exists a leaf in T_i labeled with an instruction to change x_1 to j . We will use the shorthand $T_i \rightarrow T_j$ to denote this fact. For C to be space-optimal there must be a transition from T_1 to either T_0 or T_2 or both. We now observe some structural properties of T . We first assume that there is a transition from T_1 to T_0 . (The other case when there is a transition from T_0 to T_2 will be dual to this.)

- (i) If $T_1 \rightarrow T_0$, then $T_2 \not\rightarrow T_0$. Suppose not, and let $\langle 2, x_2, * \rangle$ goes to $\langle 0, x'_2, * \rangle$. Since $T_1 \rightarrow T_0$, we have $\langle 1, *, x_3 \rangle$ goes to $\langle 0, *, x'_3 \rangle$. It is easily seen that both $\langle 2, x_2, x'_3 \rangle$ and $\langle 1, x'_2, x_3 \rangle$ have the same successor in C which is $\langle 0, x'_2, x'_3 \rangle$. (Notice similarly we could also argue that if $T_1 \rightarrow T_2$, then $T_0 \not\rightarrow T_2$.)
- (ii) If $T_1 \rightarrow T_0$, then $T_1 \not\rightarrow T_2$. Suppose not, then from item (i) we know that $T_2 \not\rightarrow T_0$ and $T_0 \not\rightarrow T_2$. Thus, in the sequence C two tuples with first coordinate 0 or 2 is separated by a tuple with the first coordinate 1. Since there are exactly equal number of tuples with a fixed first coordinate, C can not be space-optimal.
- (iii) If $T_1 \rightarrow T_0$, then $T_0 \not\rightarrow T_0$ and $T_1 \not\rightarrow T_1$. The argument is similar to item (i).

We now prune T to obtain a space-optimal hierarchical counter T' over $\mathbb{Z}_2 \times \mathbb{Z}_{m_2} \times \mathbb{Z}_{m_3}$ as follows. We first remove the whole subtree T_2 , i.e., only T_0 and T_1 remain incident on x_1 . We don't make any change to T_1 . Now we change instructions at the leaves in T_0 to complete the construction of T' . Let ℓ be a leaf in T_0 such that it is labeled with the assignment: $x_1 \leftarrow 2$ and $x_2 \leftarrow b$ for some $b \in \mathbb{Z}_{m_2}$. As a thought experiment, let us traverse T_2 in T until we reach a leaf in T_2 with assignments of the form $x_1 \leftarrow 1$ and $x_2 \leftarrow b'$ for some $b' \in \mathbb{Z}_{m_2}$. In this case, we change the content of the leaf ℓ in T_0 in T' to $x_1 \leftarrow 1$ and $x_2 \leftarrow b'$. The fact that T' is space-optimal and hierarchical follows easily from properties (i) - (iii) of T , and it being space-optimal and hierarchical. This completes the proof. \square

Acknowledgments

Authors would like to thank Gerth Stølting Brodal for bringing the problem to our attention, and to Petr Gregor for giving a talk on space-optimal counters in our seminar, which motivated this research. Authors thank Meena Mahajan and Venkatesh Raman for pointing out the result in [38]. We also thank anonymous reviewers for helpful suggestions that improved the presentation of the paper.

References

- [1] D. J. Amalraj, N. Sundararajan, and Goutam Dhar. Data structure based on Gray code encoding for graphics and image processing. In *Proceedings of the SPIE: International Society for Optical Engineering*, pages 65–76, 1990.
- [2] David A. Mix Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1 . *Journal of Computer and System Sciences*, 38:150–164, 1989.
- [3] Michael Ben-Or and Richard Cleve. Computing algebraic formulas using a constant number of registers. *SIAM J. Comput.*, 21(1):54–58, 1992. URL: <https://doi.org/10.1137/0221006>, doi:10.1137/0221006.
- [4] James R. Bitner, Gideon Ehrlich, and Edward M. Reingold. Efficient generation of the binary reflected Gray code and its applications. *Commun. ACM*, 19(9):517–521, September 1976.
- [5] Prosenjit Bose, Paz Carmi, Dana Jansens, Anil Maheshwari, Pat Morin, and Michiel H. M. Smid. Improved methods for generating quasi-Gray codes. In *Algorithm Theory - SWAT 2010, 12th Scandinavian Symposium and Workshops on Algorithm Theory, Bergen, Norway, June 21-23, 2010. Proceedings*, pages 224–235, 2010. URL: https://doi.org/10.1007/978-3-642-13731-0_22, doi:10.1007/978-3-642-13731-0_22.
- [6] Gerth Stølting Brodal, Mark Greve, Vineet Pandey, and Srinivasa Rao Satti. Integer representations towards efficient counting in the bit probe model. *J. Discrete Algorithms*, 26:34–44, 2014. URL: <https://doi.org/10.1016/j.jda.2013.11.001>, doi:10.1016/j.jda.2013.11.001.
- [7] Harry Buhrman, Richard Cleve, Michal Koucký, Bruno Loff, and Florian Speelman. Computing with a full memory: catalytic space. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 857–866, 2014. URL: <http://doi.acm.org/10.1145/2591796.2591874>, doi:10.1145/2591796.2591874.
- [8] C. C. Chang, H. Y. Chen, and C. Y. Chen. Symbolic Gray code as a data allocation scheme for two-disc systems. *The Computer Journal*, 35(3):299–305, 1992. doi:10.1093/comjnl/35.3.299.
- [9] M. S. Chen and K. G. Shin. Subcube allocation and task migration in hypercube multiprocessors. *IEEE Transactions on Computers*, 39(9):1146–1155, 1990. doi:10.1109/12.57056.
- [10] Richard Cleve. *Methodologies for Designing Block Ciphers and Cryptographic Protocols*. PhD thesis, University of Toronto, April 1989.
- [11] Martin Cohn. Affine m-ary Gray codes. *Information and Control*, 6(1):70–78, 1963.
- [12] Don Coppersmith and Edna Grossman. Generators for certain alternating groups with applications to cryptography. *SIAM J. Appl. Math.*, 29(4):624–627, December 1975.
- [13] C. Ding, D. Pei, and A. Salomaa. *Chinese Remainder Theorem: Applications in Computing, Coding, Cryptography*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 1996.
- [14] David S. Dummit and Richard M. Foote. *Abstract Algebra*. John Wiley & Sons, 2004.

- [15] Tomáš Dvořák, Petr Gregor, and Václav Koubek. Generalized Gray codes with prescribed ends. *Theor. Comput. Sci.*, 668:70–94, 2017. URL: <https://doi.org/10.1016/j.tcs.2017.01.010>, doi:10.1016/j.tcs.2017.01.010.
- [16] Gideon Ehrlich. Loopless algorithms for generating permutations, combinations, and other combinatorial configurations. *J. ACM*, 20(3):500–513, 1973. URL: <http://doi.acm.org/10.1145/321765.321781>, doi:10.1145/321765.321781.
- [17] Ivan Flores. Reflected number systems. *IRE Transactions on Electronic Computers*, EC-5(2):79–82, 1956.
- [18] Gudmund Skovbjerg Frandsen, Peter Bro Miltersen, and Sven Skyum. Dynamic word problems. *J. ACM*, 44(2):257–271, 1997. URL: <http://doi.acm.org/10.1145/256303.256309>, doi:10.1145/256303.256309.
- [19] Michael L. Fredman. Observations on the complexity of generating quasi-Gray codes. *SIAM J. Comput.*, 7(2):134–146, 1978. URL: <https://doi.org/10.1137/0207012>, doi:10.1137/0207012.
- [20] Zachary Frenette. Towards the efficient generation of Gray codes in the bitprobe model. Master’s thesis, University of Waterloo, Waterloo, Ontario, Canada, 2016.
- [21] E. Gad, M. Langberg, M. Schwartz, and J. Bruck. Constant-weight Gray codes for local rank modulation. *IEEE Transactions on Information Theory*, 57(11):7431–7442, 2011. doi:10.1109/TIT.2011.2162570.
- [22] F. Gray. Pulse code communication, 1953. US Patent 2,632,058. URL: <http://www.google.com/patents/US2632058>.
- [23] Petr Gregor and Torsten Mütze. Trimming and gluing Gray codes. In *34th Symposium on Theoretical Aspects of Computer Science, STACS 2017, March 8-11, 2017, Hannover, Germany*, pages 40:1–40:14, 2017. URL: <https://doi.org/10.4230/LIPIcs.STACS.2017.40>, doi:10.4230/LIPIcs.STACS.2017.40.
- [24] Felix Herter and Günter Rote. Loopless Gray code enumeration and the tower of bucharest. In *8th International Conference on Fun with Algorithms, FUN 2016, June 8-10, 2016, La Maddalena, Italy*, pages 19:1–19:19, 2016.
- [25] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck. Rank modulation for flash memories. *IEEE Transactions on Information Theory*, 55(6):2659–2673, 2009. doi:10.1109/TIT.2009.2018336.
- [26] James T. Joichi, Dennis E. White, and S. G. Williamson. Combinatorial Gray codes. *SIAM J. Comput.*, 9(1):130–141, 1980. URL: <https://doi.org/10.1137/0209013>, doi:10.1137/0209013.
- [27] Donald E. Knuth. *The Art of Computer Programming. Volume 4A: Combinatorial Algorithms, Part 1*. Addison-Wesley Professional, 2011.
- [28] Serge Lang. *Linear Algebra*. Undergraduate Texts in Mathematics. Springer New York, 1987.
- [29] Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 1996. doi:10.1017/CB09780511525926.

- [30] H. M. Lucal. Arithmetic operations for digital computers using a modified reflected binary code. *IRE Transactions on Electronic Computers*, EC-8(4):449–458, 1959.
- [31] J. Ludman. Gray code generation for mpsk signals. *IEEE Transactions on Communications*, 29(10):1519–1522, 1981. doi:10.1109/TCOM.1981.1094886.
- [32] Torsten Mütze. Proof of the middle levels conjecture. *Proceedings of the London Mathematical Society*, 112(4):677–713, 2016.
- [33] Torsten Mütze and Jerri Nummenpalo. Efficient computation of middle levels Gray codes. In *Algorithms - ESA 2015 - 23rd Annual European Symposium, Patras, Greece, September 14-16, 2015, Proceedings*, pages 915–927, 2015.
- [34] Torsten Mütze and Jerri Nummenpalo. A constant-time algorithm for middle levels Gray codes. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 2238–2253, 2017.
- [35] Patrick K. Nicholson, Venkatesh Raman, and S. Srinivasa Rao. A survey of data structures in the bitprobe model. In *Space-Efficient Data Structures, Streams, and Algorithms - Papers in Honor of J. Ian Munro on the Occasion of His 66th Birthday*, pages 303–318, 2013. URL: https://doi.org/10.1007/978-3-642-40273-9_19, doi:10.1007/978-3-642-40273-9_19.
- [36] Albert Nijenhuis and Herbert S. Wilf. *Combinatorial Algorithms*. Academic Press, 1978.
- [37] M. Ziaur Rahman and J. Ian Munro. Integer representation and counting in the bit probe model. *Algorithmica*, 56(1):105–127, 2010. URL: <https://doi.org/10.1007/s00453-008-9247-2>, doi:10.1007/s00453-008-9247-2.
- [38] Mikhail Raskin. A linear lower bound for incrementing a space-optimal integer representation in the bit-probe model. In *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland*, pages 88:1–88:12, 2017.
- [39] Dana S. Richards. Data compression and Gray-code sorting. *Inf. Process. Lett.*, 22(4):201–205, 1986. URL: [https://doi.org/10.1016/0020-0190\(86\)90029-3](https://doi.org/10.1016/0020-0190(86)90029-3), doi:10.1016/0020-0190(86)90029-3.
- [40] J. P. Robinson and M. Cohn. Counting sequences. *IEEE Transactions on Computers*, C-30(1):17–23, 1981. doi:10.1109/TC.1981.6312153.
- [41] Carla Savage. A survey of combinatorial Gray codes. *SIAM review*, 39(4):605–629, 1997.
- [42] Victor Shoup. Searching for primitive roots in finite fields. *Mathematics of Computation*, 58:369 – 380, 1992.
- [43] Igor Shparlinski. On finding primitive roots in finite fields. *Theoretical Computer Science*, 157(2):273 – 275, 1996.
- [44] Igor Shparlinski. *Finite Fields: Theory and Computation*, volume 477 of *Mathematics and its Applications*. Springer Netherlands, 1999.
- [45] Igor E. Shparlinski. Finding irreducible and primitive polynomials. *Applicable Algebra in Engineering, Communication and Computing*, 4(4):263–268, 1993.

- [46] Andrew Chi-Chih Yao. Should tables be sorted? *J. ACM*, 28(3):615–628, 1981. URL: <http://doi.acm.org/10.1145/322261.322274>, doi:10.1145/322261.322274.
- [47] Y. Yehezkeally and M. Schwartz. Snake-in-the-box codes for rank modulation. *IEEE Transactions on Information Theory*, 58(8):5471–5483, 2012. doi:10.1109/TIT.2012.2196755.