XOR-counts and lightweight multiplication with fixed elements in binary finite fields

Lukas Kölsch

University of Rostock, Germany
lukas.koelsch@uni-rostock.de

Abstract. XOR-metrics measure the efficiency of certain arithmetic operations in binary finite fields. We prove some new results about two different XOR-metrics that have been used in the past. In particular, we disprove a conjecture from [10]. We consider implementations of multiplication with one fixed element in a binary finite field. Here we achieve a complete characterization of all elements whose multiplication matrix can be implemented using exactly 2 XOR-operations, confirming a conjecture from [2]. Further, we provide new results and examples in more general cases, showing that significant improvements in implementations are possible.

Keywords: Lightweight cryptography · Linear layer · XOR-count · Multiplication · Finite fields .

1 Introduction

In the past years, with the advent of the so called *Internet of Things*, new challenges for cryptography have emerged. Many new devices usually do not have a lot of computational power and memory, but are still required to offer some security by encrypting sensitive data. Consequentially, *lightweight cryptography* has become a major field of research in the past years, mostly focusing on symmetric-key encryption (e.g. [1, 5, 8]). In particular, linear layers (e.g. [15, 16]) and Sboxes (e.g. [3, 18]) have been thoroughly investigated as they constitute key components in classical symmetric-key ciphers like AES. The objective here is to try to minimize the cost of storage and the number of operations needed to apply a cryptographic function. Usually, the security properties of cryptographic schemes using finite fields do not depend on a specific field representation (as bit strings) in the actual implementation [4], so the choice of field implementation makes an impact on the performance of the scheme without influencing its security. It is therefore an interesting question which representation minimizes the number of operations needed.

In practice, linear layers are usually $\mathbb{F}_{2^m}$-linear mappings on $\mathbb{F}_{2^m}^n$. Recall that linear mappings are implemented as matrix multiplications. Note that we can write every $n \times n$ matrix over $\mathbb{F}_{2^m}$ as an $(mn) \times (mn)$ matrix over $\mathbb{F}_2$. As elements in $\mathbb{F}_{2^m}$ are usually represented as bit strings in computers, it is natural to consider only matrices over $\mathbb{F}_2$. Measurements of implementation costs will then only involve the number of bit-operations (XORs) needed. It is an interesting question to evaluate the efficiency of a given matrix. For that purpose two different metrics have been introduced, the *direct XOR-count* (e.g. in [12, 15, 20, 21]) and the *sequential XOR-count* (e.g. [2, 10, 23]). Roughly speaking, the direct XOR-count counts the number of non-zeros in the matrix, whereas the sequential XOR-count counts the number of elementary row operations needed to transform the matrix into the identity matrix (see Section 2 for more precise definitions). Although the sequential XOR-count of a matrix is harder to compute, it often yields a better estimation of the actual optimal number of XOR-operations needed [10], for a simple example see Example 1 in this work. When implementing a linear layer, a field representation can be chosen such that the respective matrix is optimal according to these metrics. In this way, the performance of a given linear layer can be improved (for example by choosing a field representation that results in a sparse diffusion matrix).

Our Contributions. Our goal in this work is to explore some connections and properties of the direct and sequential XOR-count metrics and then to apply these to get some theoretical results regarding optimal implementations of matrices that represent multiplication with a fixed field element $\alpha \in \mathbb{F}_{2^k}$. Optimal choices of these matrices (called *multiplication matrices*) can then be used for local optimizations of matrices over $\mathbb{F}_{2^k}$ (this approach was taken for example in [2, 10, 15, 16, 20]). Recently, the focus has shifted to global optimization, as it has become clear that local optimizations are not necessarily also globally optimal [6, 13]. However, global optimization techniques currently rely either on tools that improve the XOR-counts of matrices already known to be efficient [13] or exhaustive searches [6, 19]. In particular, theoretical results on globally optimal matrices seem to be very hard to obtain. Numerical data suggest that there is a correlation between good local optimizations and good global optimizations (see [13, Figures 2-6]). Because of this correlation, theoretical insights into local optimization are valuable for the search of globally optimal matrices.

In the second section, we compare the direct XOR-count and sequential XOR-count evaluation metrics. We prove some theoretical properties of the sequential XOR-count that can be used to improve algorithms (e.g. an algorithm presented in [2]). We also find an infinite family of matrices that have a lower direct XOR-count than sequential XOR-count, disproving a conjecture in [10]. We want to emphasize that the results presented in this section apply to all invertible matrices, not just multiplication matrices.

In the third section we provide a complete characterisation of finite field elements α where the mapping $x \mapsto \alpha x$ can be implemented with exactly 2 XOR-operations (Theorem 5), which proves a conjecture in [2]. This case is of

special interest, since for many finite fields (including the fields $\mathbb{F}_{2^n}$ with $8|n$ that are particularly interesting for many applications) there are no elements for which the mapping $x \mapsto \alpha x$ can be implemented with only 1 XOR-operation [2]. For these fields, our classification gives a complete list of elements α such that multiplication with α can be implemented in the cheapest way possible.

In the fourth section we present some more general results for multiplication matrices with higher XOR-counts. We prove that the number of XOR-operations needed to implement the mapping $x \mapsto \alpha x$ depends on the number of non-zero coefficients of the minimal polynomial of α. In particular, Theorem 6 shows that the gap between the number of XORs used in an optimal implementation and the number of XORs used in the "naive" implementation of a multiplication matrix using the rational canonical form of the mapping $x \mapsto \alpha x$ grows exponentially with the weight of the minimal polynomial of the element. This result shows that there is a large potential for improvement in the implementation of multiplication matrices. Propositions 2 and 3 imply that the bound found in Theorem 6 is optimal.

We conclude our paper with several open problems.

2 XOR-Counts

An XOR-count metric for diffusion matrices was introduced in [12] and then generalized for arbitrary matrices in [21]. It has then subsequently been studied in several works, e.g. [20, 15].

Definition 1. *The direct XOR-count (d-XOR-count) of an invertible $n \times n$ matrix M over $\mathbb{F}_2$, denoted by $\mathrm{wt}_d(M)$ is*

$$\mathrm{wt}_d(M) = \omega(M) - n,$$

where $\omega(M)$ denotes the number of ones in the matrix M.

Note that the d-XOR-count of an invertible matrix is never negative as every row of an invertible matrix needs to have at least one non-zero entry. Moreover, $\mathrm{wt}_d(M) = 0$ if and only if M has exactly one '1' in every row and column, i.e. M is a permutation matrix. The d-XOR-metric only gives an upper bound to the actual minimal implementation cost as the following example shows.

Example 1.

$$\begin{pmatrix} 1&0&0&0\\ 1&1&0&0\\ 1&1&1&0\\ 1&1&1&1 \end{pmatrix} \cdot \begin{pmatrix} a_1\\ a_2\\ a_3\\ a_4 \end{pmatrix} = \begin{pmatrix} a_1\\ a_1+a_2\\ (a_1+a_2)+a_3\\ ((a_1+a_2)+a_3)+a_4 \end{pmatrix}$$

The d-XOR-count of the matrix is 6 but it is easy to see that multiplication with this matrix can actually be implemented with only 3 XOR operations since

the results of previous steps can be reused. A metric that allows this was subsequently introduced in [10] and used in further work (e.g. [2, 6, 23]). Let us introduce some notation at first: We denote by I the identity matrix and by $E_{i,j}$ the matrix that has exactly one '1' in the i-th row and j-th column. Then $A_{i,j} := I + E_{i,j}$ for $i \neq j$ is called an *addition matrix*. Left-multiplication with $A_{i,j}$ adds the j-th row to the i-th row of a matrix, right-multiplication adds the i-th column to the j-th column. Observe that the matrices $A_{i,j}$ are self-inverse over $\mathbb{F}_2$. Let further $\mathcal{P}(n)$ be the set of $n \times n$ permutation matrices and $\mathcal{A}(n)$ the set of all $n \times n$ addition matrices $A_{i,j}$. We will omit the dimension n unless necessary.

Definition 2. *An invertible matrix M over $\mathbb{F}_2$ has a sequential XOR-count (s-XOR-count) of t if t is the minimal number such that M can be written as*

$$M = P \prod_{k=1}^{t} A_{i_k,j_k}$$

where $P \in \mathcal{P}$ and $A_{i_k,j_k} \in \mathcal{A}$. We write $\mathrm{wt}_s(M) = t$.

Note that every invertible matrix can be decomposed as a product of a permutation matrix and addition matrices in the way Definition 2 describes. Indeed, Gauss-Jordan-elimination gives a simple algorithm to do so.

In [23] a similar definition for the s-XOR-count was given that uses a representation of the form $M = \prod_{k=1}^{t} P_k A_{i_k,j_k}$ with permutation matrices P_k. Since products of permutation matrices remain permutation matrices and

$$PA_{i,j} = A_{\sigma^{-1}(i),\sigma^{-1}(j)}P \tag{1}$$

where $\sigma \in S_n$ is the permutation belonging to the permutation matrix P, this definition is equivalent to our definition.

A representation of a matrix M as a product $M = P\prod_{k=1}^{t} A_{i_k,j_k}$ is called an *s-XOR-representation* of M and an s-XOR-representation with $\mathrm{wt}_s(M)$ addition matrices is called an *optimal s-XOR-representation*. Note that optimal s-XOR-representations are generally not unique. Observe that $M = PA_{i_1,j_1} \dots A_{i_t,j_t}$ is equivalent to $MA_{i_t,j_t} \dots A_{i_1,j_1} = P$, so the s-XOR-count measures the number of column addition steps that are needed to transform a matrix into a permutation matrix. Because of equation (1) the number of column additions needed is equal to the number of row additions needed, so we may also speak about row additions.

Going back to Example 1, it is easy to find an s-XOR-representation with 3 XORs.

$$M = \begin{pmatrix} 1&0&0&0\\ 1&1&0&0\\ 1&1&1&0\\ 1&1&1&1 \end{pmatrix} = IA_{4,3}A_{3,2}A_{2,1}.$$

It is clear that we need at least 3 addition matrices since all rows but the first one need at least one update. Hence, the s-XOR-representation above is optimal and $\mathrm{wt}_s(M) = 3$.

Determining the s-XOR-count of a given matrix is generally not easy. Graph-based algorithms to find an optimal s-XOR-count have been proposed in [23] and (in a slightly different form) in [10]. The algorithms are based on the following observation. Let $G = (V, E)$ be a graph where $G = GL(n, \mathbb{F}_2)$ and $(M_1, M_2) \in E$ if $AM_1 = M_2$ for an $A \in \mathcal{A}$. Then $\mathrm{wt}_s(M) = \min_{P \in \mathcal{P}} d(M, P)$, where $d(M_1, M_2)$ denotes the distance between M_1 and M_2 in the graph G. Thus, the evaluation of the s-XOR-count can be reduced to a shortest-path-problem. Note that because the elementary matrices in $\mathcal{A}$ are all involutory, G is undirected. As the authors of [23] observe, it is possible to reduce the number of vertices by a factor $1/n!$ because matrices with permuted rows can be considered equivalent. Still $(1/n!)|GL(n, \mathbb{F}_2)| = (1/n!)(2^n - 1)(2^n - 2) \dots (2^n - 2^{n-1})$ and every vertex has $|\mathcal{A}(n)| = n^2 - n$ neighbors, so both the number of vertices and the number of edges grow exponentially. Hence, this approach is impractical unless n is small.

The problem of determining the s-XOR-count is linked with the problem of optimal pivoting in Gauss-Jordan elimination since the number of additions in an optimal elimination process is clearly an upper bound of the s-XOR-count. Pivoting strategies for Gaussian elimination are a classical problem in numerical linear algebra (among lots of examples, see [14]) and the number of steps needed in a Gauss-Jordan elimination process can be used as a heuristic for the s-XOR-count.

Example 1 gives an example of a matrix with lower s-XOR-count than d-XOR-count. Considering this and the fact that the s-XOR-count of a given matrix is generally much harder to determine than the d-XOR-count, it should be clarified whether the s-XOR-count always gives a better estimation of the actual number of XOR operations needed to implement the matrix. In [10] this has been conjectured, i.e. $\mathrm{wt}_s(M) \leq \mathrm{wt}_d(M)$ for all $M \in \mathrm{GL}(n, \mathbb{F}_2)$. However, the following theorem gives a counterexample.

Theorem 1. *Let M be as follows:*

$$M = \begin{pmatrix} 1&1&0&0&0&0&0 \\ 0&1&1&0&0&0&0 \\ 0&0&1&1&0&0&0 \\ 0&0&0&1&1&0&0 \\ 0&0&0&0&1&1&0 \\ 0&0&0&0&0&1&1 \\ 1&0&0&1&0&0&1 \end{pmatrix} \in \mathrm{GL}(7, \mathbb{F}_2).$$

Then $\mathrm{wt}_s(M) > \mathrm{wt}_d(M)$.

Proof. M is invertible with $\mathrm{wt}_d(M) = 8$. Let $\mathrm{wt}_s(M) = t$, i.e. there are matrices $A_{i_k, j_k} \in \mathcal{A}$ and $P \in \mathcal{P}$ such that $\prod_{k=1}^{t} A_{i_k, j_k} \cdot M = P$. By construction, no two rows and no three rows of M add up to a row with only one non-zero entry. Every row has to be updated at least once to transform M into a permutation matrix. Since no two row vectors add up to a vector with only one non-zero entry, the first row that gets updated (row i_t) needs to get updated at least once more. But as there is also no combination of three vectors adding up to a vector with

only one non-zero entry, the second row that is updated (row i_{t-1}) also needs to be updated a second time. So two rows need to get updated at least twice, and all other 5 rows need to get updated at least once, resulting in $\mathrm{wt}_s(M) \geq 9$. □

Remark 1. Note that the structure of the counterexample can be extended to all dimensions $n \geq 7$, the middle '1' in the last row can be in any j-th column with $4 \leq j \leq n-3$. We conclude that there exists a matrix $M \in \mathrm{GL}(n, \mathbb{F}_2)$ with $\mathrm{wt}_s(M) > \mathrm{wt}_d(M)$ for all $n \geq 7$.

Studying the s-XOR-count is an interesting mathematical problem because it has some properties that can be used to get upper bounds of the actual implementation cost of potentially a lot of matrices. The actual number of XOR-operations needed is clearly invariant under permutation of rows and columns. It is therefore desirable that this property is reflected in our XOR-metrics. Obviously, this is the case for the d-XOR-count, i.e. $\mathrm{wt}_d(M) = \mathrm{wt}_d(PMQ)$ for all matrices M and permutation matrices $P, Q \in \mathcal{P}$. The following lemma shows that this also holds for the s-XOR-count. The lemma is a slight modification of a result in [2]. However the proof in [2] has a small gap, so we provide a complete proof here. We denote permutation-similarity with $\sim$, i.e. $M_1 \sim M_2$ if there exists a $P \in \mathcal{P}$ so that $M_1 = PM_2P^{-1}$.

Lemma 1. *Let $M \in GL(n, \mathbb{F}_2)$. Then $\mathrm{wt}_s(M) = \mathrm{wt}_s(PMQ)$ for $P, Q \in \mathcal{P}$. In particular, if $M_1 \sim M_2$ then $\mathrm{wt}_s(M_1) = \mathrm{wt}_s(M_2)$.*

Proof. Let $\mathrm{wt}_s(M) = t$ and $\sigma \in S_n$ be the permutation belonging to Q. Then, by shifting Q to the left

$$PMQ = PP_2 \prod_{k=1}^{t} A_{i_k,j_k} Q = PP_2 Q \prod_{k=1}^{t} A_{\sigma(i_k),\sigma(j_k)} = P' \prod_{k=1}^{t} A_{\sigma(i_k),\sigma(j_k)}$$

where $P_2, P' \in \mathcal{P}$, so $\mathrm{wt}_s(PMQ) \leq \mathrm{wt}_s(M)$. Since $M = P^{-1}(PMQ)Q^{-1}$ the same argument yields $\mathrm{wt}_s(M) \leq \mathrm{wt}_s(PMQ)$. □

Based on this result, the following normal form for permutation matrices is proposed in [2]. We introduce a notation for block diagonal matrices. Let $M_1, \ldots, M_d$ be square matrices, then we denote the block matrix consisting of these matrices by

$$\bigoplus_{k=1}^{d} M_k := \begin{pmatrix} M_1 & & 0 \\ & M_2 & \\ & & \ddots \\ 0 & & M_d \end{pmatrix}.$$

We denote by C_p the companion matrix of a polynomial $p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{F}_2[x]$, i.e.

$$C_p = \begin{pmatrix} 0 & \ldots & 0 & a_0 \\ 1 & 0 & 0 & a_1 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \ldots & 1 & a_{n-1} \end{pmatrix}.$$

Lemma 2 ([2, Lemma 2]). *Let $P \in \mathcal{P}(n)$. Then*

$$P \sim \bigoplus_{k=1}^{d} C_{x^{m_k}+1}$$

for some m_k with $\sum_{k=1}^{d} m_k = n$ and $m_1 \geq \cdots \geq m_d \geq 1$.

A permutation matrix of this structure is said to be the *cycle normal form* of P. We can then (up to permutation-similarity) always assume that the permutation matrix of the s-XOR-decomposition is in cycle normal form.

Corollary 1 ([2, Corollary 2]).

$$P \prod_{k=1}^{t} A_{i_k,j_k} \sim P' \prod_{k=1}^{t} A_{\sigma(i_k),\sigma(j_k)}$$

for some permutation $\sigma \in S_n$, where P' is the cycle normal form of P.

We say an s-XOR-representation is in cycle normal form if its permutation polynomial is in cycle normal form. Corollary 1 states that every s-XOR-representation is pemutation-similar to exactly one s-XOR-representation in cycle normal form.

The following theorem gives a connection between the s-XOR-count and optimal s-XOR-representations of a given matrix and that of its inverse.

Theorem 2. *Let M be an invertible matrix with $\mathrm{wt}_s(M) = t$ and*

$$M = P \prod_{k=1}^{t} A_{i_k,j_k} \text{ with } P = \bigoplus_{k=1}^{d} C_{x^{m_k}+1}.$$

Then $\mathrm{wt}_s(M^{-1}) = t$. Moreover,

$$M^{-1} = P A_{\sigma(i_t),\sigma(j_t)} A_{\sigma(i_{t-1}),\sigma(j_{t-1})} \cdots A_{\sigma(i_1),\sigma(j_1)}$$

for some permutation $\sigma \in S_n$ that depends only on P.

Proof. For the inverse matrix we have

$$M^{-1} = A_{i_t,j_t} \ldots A_{i_1,j_1} P^{-1} \sim P^{-1} A_{i_t,j_t} \ldots A_{i_1,j_1},$$

so $\mathrm{wt}_s(M^{-1}) \leq \mathrm{wt}_s(M)$. By symmetry, we get $\mathrm{wt}_s(M^{-1}) = \mathrm{wt}_s(M)$. Observe that $P^{-1} = P^T = \bigoplus_{k=1}^{d} C_{x^{m_k}+1}^T$ where P^T denotes the transpose of P. Let J_r be the $r \times r$ matrix with ones on the counterdiagonal, i.e. $J_{i,j} = 1$ if and only if $j = n-i+1$. Let $Q = \bigoplus_{k=1}^{d} J_{m_k} \in \mathcal{P}$. A direct calculation yields $QP^{-1}Q^{-1} = P$ and thus

$$M^{-1} \sim QP^{-1} \prod_{k=t}^{1} A_{i_k,j_k} Q^{-1} = P \prod_{k=t}^{1} A_{\sigma(i_k),\sigma(j_k)},$$

where $\sigma \in S_n$ denotes the permutation that belongs to Q. □

In particular, Theorem 2 implies that given an optimal s-XOR-representation for a matrix M, an optimal s-XOR-representation of M^{-1} can be determined with very little effort by calculation the permutation σ in the proof. Note that the statement of Theorem 2 does not exist for the d-XOR-count. Indeed, sparse matrices (i.e. matrices with low d-XOR-count) usually have dense inverse matrices (i.e. high d-XOR-count).

The next result also holds for the s-XOR-count only.

Proposition 1. *Let M, N be invertible matrices with $\mathrm{wt}_s(M) = t_1$ and $\mathrm{wt}_s(N) = t_2$. Then $\mathrm{wt}_s(MN) \leq t_1 + t_2$. In particular, $\mathrm{wt}_s(M^k) \leq |k|t_1$ for all $k \in \mathbb{Z}$.*

Proof. Let $M = P\prod_{k=1}^{t_1} A_{i_k,j_k}$ and $N = Q\prod_{k=1}^{t_2} B_{i_k,j_k}$. Then

$$MN = PQ \prod_{k=1}^{t_1} A_{\sigma(i_k),\sigma(j_k)} \prod_{k=1}^{t_2} B_{i_k,j_k},$$

where $\sigma \in S_n$ is the permutation belonging to Q. This implies $\mathrm{wt}_s(MN) \leq t_1+t_2$. The statement $\mathrm{wt}_s(M^k) \leq |k|t_1$ for $k < 0$ follows from Theorem 2. □

3 Efficient Multiplication Matrices in Finite Fields

We can consider $\mathbb{F}_{2^n}$ as the n-dimensional vector space $(\mathbb{F}_2)^n$ over $\mathbb{F}_2$. By distributivity, the function $x \mapsto \alpha x$ for $\alpha \in \mathbb{F}_{2^n}$ is linear, so it can be represented as a (left-)multiplication with a matrix in $\mathrm{GL}(n, \mathbb{F}_2)$. This matrix obviously depends on α, but also on the choice of the basis of $(\mathbb{F}_2)^n$ over $\mathbb{F}_2$. We denote the multiplication matrix that represents the function $x \mapsto \alpha x$ with respect to the basis B by $M_{\alpha,B}$. The XOR-count of $M_{\alpha,B}$ generally differs from the XOR-count of $M_{\alpha,B'}$ for different bases B, B'. Our objective here is to find the optimal basis B for a given α, in the sense that the XOR-count of $M_{\alpha,B}$ is minimized. For this, we define the XOR-count metrics from the previous section also for elements from $\mathbb{F}_{2^n}$.

Definition 3. *Let $\alpha \in \mathbb{F}_{2^n}$. We define the s-XOR-count and d-XOR-count of α as follows:*

$$\mathrm{wt}_s(\alpha) = \min_B \mathrm{wt}_s(M_{\alpha,B}), \quad \mathrm{wt}_d(\alpha) = \min_B \mathrm{wt}_d(M_{\alpha,B}),$$

where the minimum is taken over all bases of $\mathbb{F}_2^n$ over $\mathbb{F}_2$. A basis B and matrix $M_{\alpha,B}$ that satisfy the minimum are called s-XOR-optimal and d-XOR-optimal for α, respectively.

In order to find the matrices that optimize the s-XOR-count-metric, an exhaustive search on all matrices with low s-XOR-count is performed in [2]. In this way the s-XOR-count and an optimal s-XOR-matrix of every element $\alpha \in \mathbb{F}_{2^n}$ for $n \leq 8$ was found. Using the results presented in the previous section, the

search was restricted to matrices where the permutation matrix is in cycle normal form. The following result was used to determine whether a given matrix is a multiplication matrix for some $\alpha \in \mathbb{F}_{2^n}$ with respect to some basis B. From here on, we denote by $\chi(M) = \det(xI + M)$ the characteristic polynomial of a matrix M and by m_α the minimal polynomial of the finite field element $\alpha \in \mathbb{F}_{2^n}$. Recall that m_α is always irreducible.

Theorem 3 ([2, Theorem 1]). *Let $M \in GL(n, \mathbb{F}_2)$ and $\alpha \in \mathbb{F}_{2^n}$. Then M is a multiplication matrix for α, i.e. $M = M_{\alpha,B}$ with respect to some basis B, if and only if m_α is the minimal polynomial of M.*

Theorem 3 shows in particular that a matrix M is a multiplication for some $\alpha \in \mathbb{F}_{2^n}$ with respect to some basis B if and only if the minimal polynomial of M is irreducible. Additionally, it is clear that two field elements with the same minimal polynomial necessarily have the same XOR-counts.

Remark 2. A direct calculation of the minimal polynomial of the matrix M in Theorem 1 yields $m_M = x^7+x^6+x^5+x^4+1$ which is an irreducible polynomial. According to Theorem 3 the matrix M is a multiplication matrix for an element $\alpha \in \mathbb{F}_{2^7}$ with respect to some basis. Hence, there are elements $\alpha \in \mathbb{F}_{2^n}$ such that $\mathrm{wt}_d(\alpha) < \mathrm{wt}_s(\alpha)$. Note that this case does not have to occur for every value of n because the matrices provided in Theorem 1 might have a reducible minimal polynomial. Indeed, an exhaustive search for the cases $n = 4$ and $n = 8$ was conducted in [10], resulting in $\mathrm{wt}_s(\alpha) \leq \mathrm{wt}_d(\alpha)$ for all α in $\mathbb{F}_{2^4}$ and $\mathbb{F}_{2^8}$, respectively. We tested the examples given in Theorem 1 for $n = 16$ without finding any matrices with irreducible minimal polynomial. Hence, we conjecture that $\mathrm{wt}_s(\alpha) \leq \mathrm{wt}_d(\alpha)$ for all $\alpha \in \mathbb{F}_{2^{16}}$. It is an interesting question for which n elements with lower d-XOR-count than s-XOR-count exist.

Corollary 2. *Let $M = P\prod_{k=1}^{t} A_{i_k,j_k}$ be in cycle normal form. Then M is a multiplication matrix for $\alpha \in \mathbb{F}_{2^n}$ if and only if M^{-1} is a multiplication matrix for $\alpha^{-1} \in \mathbb{F}_{2^n}$. Moreover, M is an optimal s-XOR-matrix for α if and only if M^{-1} is an optimal s-XOR-matrix for α^{-1}.*

Proof. Let p and q be the minimal polynomial of M and M^{-1}, respectively. It is well known that q is then the reciprocal polynomial of p, that is $q(x) = x^n p(1/x)$. Moreover, p is the minimal polynomial of α if and only if q is the minimal polynomial of α^{-1}. The rest follows from Theorem 2. □

Corollary 2 allows us to determine an s-XOR-optimal matrix for α^{-1} given an s-XOR-optimal matrix M of α. Recall that the cycle normal form of M^{-1} was directly computed in Theorem 2. This allows us to cut the search space (approximately) in half for all algorithms that determine the s-XOR-count by traversing all matrices in $\mathrm{GL}(n, \mathbb{F}_2)$.

It is now an interesting question which elements $\alpha \in \mathbb{F}_{2^n}$ have multiplication matrices with low XOR-count. Obviously, the only element that can be implemented with XOR-count 0 is $\alpha = 1$. A simple upper bound on the s-XOR-count and d-XOR-count for elements can be found by considering the rational

canonical form of a matrix. Recall that a matrix $M \in \mathrm{GL}(n, \mathbb{F}_2)$ is similar to its (unique) rational canonical form. If M has an irreducible minimal polynomial m with $\deg m = k$ then there exists a $d \geq 1$ so that $kd = n$ and the rational canonical form is $\bigoplus_{i=1}^{d} C_m$. For a polynomial p we denote by $\mathrm{wt}(p)$ the weight of p, that is the number of non-zero coefficients. Note that if $2 | \mathrm{wt}(p)$ then 1 is a root of p so the only irreducible polynomial over $\mathbb{F}_2$ with even weight is $x + 1$.

Example 2. Let α be an element of $\mathbb{F}_{2^n}$ with minimal polynomial m_α and $\deg m_\alpha = k$ with $kd = n$ and $d \geq 1$. Then we can find a basis B so that $M_{\alpha,B}$ is in rational canonical form, i.e. $M_{\alpha,B} = \bigoplus_{i=1}^{d} C_{m_\alpha}$. It is easy to check that $\mathrm{wt}_s(M_{\alpha,B}) = \mathrm{wt}_d(M_{\alpha,B}) = d \cdot (\mathrm{wt}(m_\alpha) - 2)$.

This example shows in particular that all $\alpha \in \mathbb{F}_{2^n}$ with $\deg m_\alpha = n$ and $\mathrm{wt}(m_\alpha) = 3$ can be implemented with only one XOR operation. A possible basis for this case is the polynomial basis $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$.

As one row-addition on I only produces one extra '1' in the matrix, $\mathrm{wt}_d(M) = 1$ if and only if $\mathrm{wt}_s(M) = 1$, and equivalently, $\mathrm{wt}_d(\alpha) = 1$ if and only if $\mathrm{wt}_s(\alpha) = 1$. In [2] all elements that can be implemented with exactly one XOR-operation are characterized. It turns out, that these cases are exactly those covered by Example 2.

Theorem 4 ([2, Theorem 2]). *Let $\alpha \in \mathbb{F}_{2^n}$. Then $\mathrm{wt}_s(\alpha) = 1$ or $\mathrm{wt}_d(\alpha) = 1$ if and only if m_α is a trinomial of degree n.*

It is an open problem for which n irreducible trinomials of degree n exist. Among other sporadic examples, it is known that there are no irreducible trinomials of degree n if $n \equiv 0 \pmod 8$ [22], so there are no elements α with d/s-XOR-count 1 in these cases. As the case $8|n$ is especially important in practice, it is natural to consider elements that can be implemented with 2 XOR operations. In this case, s-XOR-count and d-XOR count do differ: By simply expanding the product $PA_{i_1,j_1}A_{i_2,j_2} = P(I + E_{i_1,j_1})(I + E_{i_2,j_2})$, it follows that every matrix with $\mathrm{wt}_s(M) = 2$ is of the following form:

$$M = \begin{cases} P + E_{\sigma^{-1}(i_1),j_1} + E_{\sigma^{-1}(i_2),j_2}, & i_2 \neq j_1 \\ P + E_{\sigma^{-1}(i_1),j_1} + E_{\sigma^{-1}(i_2),j_2} + E_{\sigma^{-1}(i_1),j_2}, & i_2 = j_1, \end{cases} \tag{2}$$

where σ is the permutation that belongs to P and $i_1 \neq j_1$, $i_2 \neq j_2$. In particular equation (2) shows that $\mathrm{wt}_d(M) = 2$ implies $\mathrm{wt}_s(M) = 2$, but there are some matrices with $\mathrm{wt}_s(M) = 2$ and $\mathrm{wt}_d(M) = 3$. In other words, the s-XOR-metric is a better metric for these matrices. In [2] the authors conjecture that $\mathrm{wt}_s(\alpha) = 2$ implies $\mathrm{wt}(m_\alpha) \leq 5$, i.e. the minimal polynomial is a trinomial or a pentanomial. We confirm this conjecture by giving an exact characterization of all elements with $\mathrm{wt}_s(\alpha) = 2$ and their optimal s-XOR-representation in cycle normal form in Theorem 5.

In the proof the following concept from linear algebra is used. We refer the reader to [9] for proofs and more background. Let V be a vector space over a field $\mathbb{F}$ with dimension n, $u \in V$ a vector and M an $n \times n$-matrix over $\mathbb{F}$. The monic

polynomial $g(x) \in \mathbb{F}[x]$ with the smallest degree such that $g(M)u = 0$ is called the *M-annihilator* of u. This polynomial divides any polynomial h annihilating u (i.e. $h(M)u = 0$), in particular the minimal polynomial of M. In the case that the minimal polynomial of M is irreducible the M-annihilator of every vector $u \neq 0$ is the minimal polynomial of M. So if we find a polynomial h that annihilates a vector $u \neq 0$ we know that the minimal polynomial divides h. In particular, if h is monic and the degree of h and the minimal polynomial coincide we can infer that h is the minimal polynomial of M.

Theorem 5. *Let $\alpha \in \mathbb{F}_{2^n}$. Then $\mathrm{wt}_s(\alpha) = 2$ if and only if m_α can be written in the form of a pentanomial or the trinomial appearing in Table 1.*

Table 1. Elements with minimal polynomials listed in the left column have s-XOR-count 2. The second column gives an optimal multiplication matrix and the third column points to the corresponding case in the proof.

m_α	optimal matrix representation	Case
$x^n + x^{k_1+k_2} + x^{k_1} + x^{k_2} + 1$, $k_1 + k_2 \leq n-2$	$C_{x^n+1} + E_{i_1,j_1} + E_{i_2,j_2}$	(1.3.)
$x^n + x^{n-k_1} + x^{k_2} + x^{k_2-k_1} + 1$, $k_2 > k_1$	$C_{x^n+1} + E_{i_1,j_1} + E_{i_2,j_2}$	(1.4.)
$x^n + x^{k_1+k_2} + x^{k_1} + x^{k_2} + 1$	$C_{x^n+1} + E_{i_1,j_1} + E_{j_1+1,j_2} + E_{i_1,j_2}$	(2.1.)
$x^n + x^{n_1} + x^{n_2} + x^k + 1$, $k \leq n-2$	$(C_{x^{n_1}+1} \oplus C_{x^{n_2}+1}) + E_{i_1,j_1} + E_{i_2,j_2}$	(3.2.)
$x^n + x^{n_1+k} + x^{n_2} + x^{n_1} + 1$, $0 < k < n_2$	$(C_{x^{n_1}+1} \oplus C_{x^{n_2}+1}) + E_{i_1,j_1}$ $+E_{j_1+1 \pmod{n_1},j_2} + E_{i_1,j_2}$	(4.)
$x^{n/2} + x^k + 1$	$(C_{x^{n/2}+1} \oplus C_{x^{n/2}+1}) + E_{i_1,j_1} + E_{i_2,j_2}$	(3.1.)

Proof. Let $M_{\alpha,B}$ be a multiplication matrix for some $\alpha \in \mathbb{F}_{2^n}$ and some basis $B = \{b_1, \ldots, b_n\}$. We can assume that $M_{\alpha,B}$ is in cycle normal form, $M = PA_{i_1,j_1}A_{i_2,j_2}$ with $P = \bigoplus_{k=1}^{l} C_{x^{m_k}+1}$. As a first step, we show that $l \leq 2$. Assume $l > 2$. As shown in equation (2) at most two rows of M have more than one '1' in them. So, by possibly permuting the blocks, P is a triangular block matrix, consisting of two blocks where one block is of the form C_{x^t+1}. So $\chi(C_{x^t+1}) = x^t + 1$ divides $\chi(M)$. But as minimal polynomial and characteristic polynomial share the same irreducible factors, this implies $(x+1)|m_\alpha$ which contradicts the irreducibility of m_α. So $l \leq 2$. We now deal with all possible matrices on a case by case basis, where we differentiate the cases $l \in \{1, 2\}$ and the two cases in equation (2).

Case 1. $M = C_{x^n+1} + E_{i_1,j_1} + E_{i_2,j_2}$, $j_1 \neq i_2 - 1$.
We investigate how the matrix operates on the basis $B = \{b_1, \ldots, b_n\}$:

$$\begin{aligned}
\alpha b_1 &= b_2 \\
&\vdots \\
\alpha b_{j_1-1} &= b_{j_1} \\
\alpha b_{j_1} &= b_{j_1+1} + b_{i_1} \qquad (3)\\
\alpha b_{j_1+1} &= b_{j_1+2} \\
&\vdots \\
\alpha b_{j_2-1} &= b_{j_2} \\
\alpha b_{j_2} &= b_{j_2+1} + b_{i_2} \qquad (4)\\
\alpha b_{j_2+1} &= b_{j_2+2} \\
&\vdots \\
\alpha b_n &= b_1.
\end{aligned}$$

Define $\gamma_1 := b_{j_1+1}$ and $\gamma_2 := b_{j_2+1}$. Then

$$b_{j_1} = \alpha^{n+j_1-j_2-1}\gamma_2, \;\; b_{j_2} = \alpha^{j_2-j_1-1}\gamma_1. \qquad (5)$$

At first, we show that the minimal polynomial has degree n. Assume $m_\alpha = x^m + \sum_{i=1}^{m-1} c_i x^i + 1$ with $c_i \in \mathbb{F}_2$ and $md = n$ with $d > 1$. In particular, $m \leq n/2$. At least one of $n+j_1-j_2$ and j_2-j_1 are greater or equal $n/2$. Assume $j_2-j_1 \geq n/2$. Then $\alpha^i\gamma_1 = b_{j_1+1+i}$ for $i < n/2$. Furthermore, $\alpha^{n/2}\gamma_1 = b_{j_1+1+n/2}$ if $j_2 - j_1 > n/2$ and $\alpha^{n/2}\gamma_1 = b_{j_1+1+n/2} + b_{i_2}$ if $j_2 - j_1 = n/2$. Consequently, $m_\alpha(\alpha)\gamma_1 = \alpha^m\gamma_1 + \sum_{i=1}^{m-1} c_i\alpha^i\gamma_1 + \gamma_1$ is a linear combination of at least one basis element and thus cannot vanish. If $n + j_1 - j_2 \geq n/2$ the same argument holds with γ_2 instead of γ_1. So $\deg m_\alpha = n$. Observe that with the equations (3), (4) and (5)

$$\alpha^{n+j_1-j_2}\gamma_2 = \gamma_1 + b_{i_1} \qquad (6)$$

$$\alpha^{j_2-j_1}\gamma_1 = \gamma_2 + b_{i_2}. \qquad (7)$$

By plugging γ_2 into the first equation and γ_1 into the second equation, we obtain

$$\alpha^n\gamma_1 + \alpha^{n+j_1-j_2}b_{i_2} + b_{i_1} + \gamma_1 = 0 \qquad (8)$$

$$\alpha^n\gamma_2 + \alpha^{j_2-j_1}b_{i_1} + b_{i_2} + \gamma_2 = 0. \qquad (9)$$

Case 1.1. $i_1 \in [j_1+1, j_2]$ and $i_2 \in [j_1+1, j_2]$.

Then $b_{i_1} = \alpha^{t_1}\gamma_1$ and $b_{i_2} = \alpha^{t_2}\gamma_1$ with $t_1 = i_1 - j_1 - 1$ and $t_2 = i_2 - j_1 - 1$ with $t_1 + t_2 < n - 1$. With equation (8), we have

$$\alpha^n\gamma_1 + \alpha^{n+j_1-j_2+t_2}\gamma_1 + \alpha^{t_1}\gamma_1 + \gamma_1 = 0$$

So the polynomial $p = x^n + x^{n+j_1-j_2+t_2} + x^{t_1} + 1$ annihilates γ_1. Hence, p is the minimal polynomial of M. But $2|\operatorname{wt}(p)$, so p is not irreducible. We conclude

that no matrix of this type can be a multiplication matrix.

Case 1.2. $i_1 \notin [j_1+1, j_2]$ and $i_2 \notin [j_1+1, j_2]$.

Then $b_{i_1} = \alpha^{t_1}\gamma_2$ and $b_{i_2} = \alpha^{t_2}\gamma_2$ with $t_1 = i_1 - j_2 - 1 \pmod n$ and $t_2 = i_2 - j_2 - 1 \pmod n$ with $t_1 + t_2 < n-1$. With equation (9), we have

$$\alpha^n\gamma_2 + \alpha^{j_2-j_1+t_1}\gamma_2 + \alpha^{t_2}\gamma_2 + \gamma_2 = 0$$

As before, the polynomial $p = x^n + x^{j_2-j_1+t_1} + x^{t_2} + 1$ annihilates γ_2, so there is no multiplication matrix of this type.

Case 1.3. $i_1 \in [j_1+1, j_2]$ and $i_2 \notin [j_1+1, j_2]$.

Then $b_{i_1} = \alpha^{t_1}\gamma_1$ and $b_{i_2} = \alpha^{t_2}\gamma_2$ with $t_1 = i_1 - j_1 - 1$ and $t_2 = i_2 - j_2 - 1 \pmod n$ with $t_1 + t_2 < n-1$. Then by equation (6)

$$\gamma_2 = \alpha^{j_2-j_1-n}\gamma_1 + \alpha^{j_2-j_1-n+t_1}\gamma_1$$

and

$$b_{i_2} = \alpha^{j_2-j_1-n+t_2}\gamma_1 + \alpha^{j_2-j_1-n+t_1+t_2}\gamma_1.$$

Using equation (8), we obtain

$$\alpha^n\gamma_1 + \alpha^{t_1+t_2}\gamma_1 + \alpha^{t_1}\gamma_1 + \alpha^{t_2}\gamma_1 + \gamma_1 = 0,$$

so $p = x^n + x^{t_1+t_2} + x^{t_1} + x^{t_2} + 1$ is the minimal polynomial of M. Note that we can choose i_1, i_2, j_1, j_2 in a way that t_1 and t_2 take any value from $\{1, \ldots, n-3\}$ as long as $t_1 + t_2 < n-1$, so every matrix with a minimal polynomial of the form $x^n + x^{a+b} + x^a + x^b + 1$ with $a + b \leq n-2$ has a multiplication matrix of this type for suitable values of i_1, j_1, i_2, j_2.

Case 1.4. $i_1 \notin [j_1+1, j_2]$ and $i_2 \in [j_1+1, j_2]$.

Then $b_{i_1} = \alpha^{t_1}\gamma_2$ and $b_{i_2} = \alpha^{t_2}\gamma_1$ with $t_1 = i_1 - j_2 - 1 \pmod n$ and $t_2 = i_2 - j_1 - 1$ with $t_1 + t_2 < n-1$. Similarly to Case 1.3, equation (6) yields

$$\gamma_1 = \alpha^{n+j_1-j_2}\gamma_2 + \alpha^{t_1}\gamma_2$$

and with equation (9)

$$\alpha^n\gamma_2 + \alpha^{j_2-j_1+t_1}\gamma_2 + \alpha^{n+j_1-j_2+t_2}\gamma_2 + \alpha^{t_1+t_2}\gamma_2 + \gamma_2 = 0,$$

so $p = x^n + x^{j_2-j_1+t_1} + x^{n+j_1-j_2+t_2} + x^{t_1+t_2} + 1 = x^n + x^{n-k_1} + x^{k_2} + x^{k_2-k_1} + 1$ with $k_1 = j_2 - j_1 - t_2 = j_2 - i_2 - 1 > 0$ and $k_2 = j_2 - j_1 + t_1$. Note that $k_2 > k_1$ for any choice of i_1, i_2, j_1, j_2. Moreover, k_1 can take on every value in $\{1, \ldots, n-3\}$ and k_2 any value greater than k_1.

Case 2. $M = C_{x^n+1} + E_{i_1,j_1} + E_{j_1+1,j_2} + E_{i_1,j_2}$.

If $j_1 = j_2$ then $\mathrm{wt}_s(M) = 1$, so we can assume $j_1 \neq j_2$. Note that the matrix operates on the basis B just as in Case 1, the only difference being that in equation (4) we have $b_{i_1} + b_{j_1+1} = b_{i_1} + \gamma_1$ instead of b_{i_2} on the right hand side. With the same argument as in Case 1 we conclude that the minimal polynomial of M has degree n.

Case 2.1. $i_1 \in [j_1+1, j_2]$.
Then $b_{i_1} = \alpha^t \gamma_1$ with $t = i_1 - j_1 - 1$. Similarly to equation (8), we obtain

$$\alpha^n \gamma_1 + \alpha^{n+j_1-j_2}\gamma_1 + \alpha^{n+j_1-j_2} b_{i_1} + b_{i_1} + \gamma_1 = 0$$

and thus

$$\alpha^n \gamma_1 + \alpha^{n+j_1-j_2}\gamma_1 + \alpha^{n+j_1-j_2+i_1-j_1-1}\gamma_1 + \alpha^{i_1-j_1-1}\gamma_1 + \gamma_1 = 0.$$

So the minimal polynomial of M is $p = x^n + x^{n+j_1-j_2} + x^{n-j_2+i_1-1} + x^{i_1-j_1-1} + 1$. Set $k_1 = i_1 - j_1 - 1$ and $k_2 = n + j_1 - j_2$ then $p = x^n + x^{k_1+k_2} + x^{k_1} + x^{k_2} + 1$ with $k_1, k_2 \in \{1, \ldots, n-1\}$ and $k_1 + k_2 < n$.
Case 2.2. $i_1 \notin [j_1+1, j_2]$.
Then $b_{i_1} = \alpha^t \gamma_2$ with $t = i_1 - j_2 - 1 \pmod n$. Similarly to equation (7), we have

$$\alpha^{j_2-j_1}\gamma_1 = \gamma_2 + \gamma_1 + \alpha^t \gamma_2.$$

Using equation (6) we obtain

$$\alpha^n \gamma_2 + \alpha^{j_2-j_1+t}\gamma_2 + \alpha^{n+j_1-j_2}\gamma_2 + \gamma_2 = 0,$$

so the minimal polynomial of M, $p = x^n + x^{j_2-j_1+t} + x^{n+j_1-j_2} + 1$, is reducible.
Case 3. $M = (C_{x^{n_1}+1} \oplus C_{x^{n_2}+1}) + E_{i_1,j_1} + E_{i_2,j_2}$, $j_1 \neq i_2 - 1$.
If both $i_1, i_2 \leq n_1$ or $i_1, i_2 > n_1$ then M is a triangular block matrix with one block being just a companion matrix. Then $(x+1)|\chi(M) = m_\alpha$, so this case cannot occur. Similarly one of j_1 and j_2 must be less or equal n_1 and the another one greater than n_1. We again investigate how M operates on the basis B:

$$\begin{aligned}
\alpha b_1 &= b_2 & \alpha b_{n_1+1} &= b_{n_1+2} \\
&\vdots & &\vdots \\
\alpha b_{j_1-1} &= b_{j_1} & \alpha b_{j_2-1} &= b_{j_2} \\
\alpha b_{j_1} &= b_{j_1+1} + b_{i_1} & \alpha b_{j_2} &= b_{j_2+1} + b_{i_2} \\
\alpha b_{j_1+1} &= b_{j_1+2} & \alpha b_{j_2+1} &= b_{j_2+2} \\
&\vdots & &\vdots \\
\alpha b_{n_1} &= b_1 & \alpha b_n &= b_{n_1+1}.
\end{aligned}$$

We set again $\gamma_1 = b_{j_1+1}$ and $\gamma_2 = b_{j_2+1}$. Then

$$\alpha^{n_1}\gamma_1 = \gamma_1 + b_{i_1} \text{ and } \alpha^{n_2}\gamma_2 = \gamma_2 + b_{i_2}. \tag{10}$$

Case 3.1. $i_1 \in [1, n_1]$ and $i_2 \in [n_1+1, n]$.
Then $b_{i_1} = \alpha^{t_1}\gamma_1$ with $t_1 = i_1 - j_1 - 1 \pmod{n_1}$ and $b_{i_2} = \alpha^{t_2}\gamma_2$ with $t_2 = i_2 - j_2 - 1 \pmod{n_2}$. M is a block diagonal matrix: $M = (C_{x^{n_1}+1} + E_{i_1,j_1}) \oplus (C_{x^{n_2}+1} + E_{i_2,j_2}) = B_1 \oplus B_2$. Let m_M, m_{B_1}, m_{B_2} be the minimal polynomial of M, B_1 and B_2. Then $m_M = \operatorname{lcm}(m_{B_1}, m_{B_2})$ and if m_M is irreducible then $m_M = m_{B_1} = m_{B_2}$. This implies that B_1 and B_2 are multiplication matrices

with $\mathrm{wt}_s(B_1) = \mathrm{wt}_s(B_2) = 1$. From Theorem 4 we obtain that m_{B_1} and m_{B_2} are trinomials of degree n_1 and n_2, respectively. So $n_1 = n_2 = n/2$ and $m_M = x^{n/2} + x^t + 1$. Using equation (10) we can determine the choice for i_1, i_2, j_1, j_2

$$\alpha^{n/2}\gamma_1 = \gamma_1 + \alpha^{t_1}\gamma_1 \text{ and } \alpha^{n/2}\gamma_2 = \gamma_2 + \alpha^{t_2}\gamma_2.$$

Hence i_1, i_2, j_1, j_2 have to be chosen in a way that $t_1 = t_2 = t$. This is possible for every $t \in \{1, \ldots, n/2 - 1\}$.

Case 3.2. $i_1 \in [n_1 + 1, n]$ and $i_2 \in [1, n_1]$.

Then $b_{i_1} = \alpha^{t_1}\gamma_2$ with $t_1 = i_1 - j_2 - 1 \pmod{n_2}$ and $b_{i_2} = \alpha^{t_2}\gamma_1$ with $t_2 = i_2 - j_1 - 1 \pmod{n_1}$. Similarly to Case 1 we can show that the minimal polynomial of M has degree n. Applying equation (10) yields

$$\gamma_1 = \alpha^{n_2 - t_2}\gamma_2 + \alpha^{-t_2}\gamma_2$$

and

$$\alpha^{n-t_2}\gamma_2 + \alpha^{n_1-t_2}\gamma_2 + \alpha^{n_2-t_2}\gamma_2 + \alpha^{t_1}\gamma_2 + \alpha^{-t_2}\gamma_2 = 0.$$

Multiplying this equation by α^{t_2} we conclude that $p = x^n + x^{n_1} + x^{n_2} + x^{t_1+t_2} + 1$ annihilates γ_2, so $m_\alpha = p$. Note that $t_1 \in \{0, \ldots, n_2 - 1\}$ and $t_1 \in \{0, \ldots, n_1 - 1\}$ so $t_1 + t_2 \in \{0, \ldots, n - 2\}$.

Case 4. $M = (C_{x^{n_1}+1} \oplus C_{x^{n_2}+1}) + E_{i_1,j_1} + E_{j_1+1 \pmod{n_1},j_2} + E_{i_1,j_2}$.

Again, we can assume $j_1 \neq j_2$. Note that the matrix operates on the basis B just as in Case 3, the only difference being that b_{i_2} is substituted by $b_{i_1} + b_{j_1+1} = b_{i_1} + \gamma_1$. This leads to

$$\alpha^{n_2}\gamma_2 = \gamma_2 + \gamma_1 + \alpha^t\gamma_2. \tag{11}$$

With the same argument as before we conclude that the minimal polynomial of M has degree n. If $i_1 \in [1, n_1]$ then M is again a block triangular matrix with one block being a companion matrix, so this case cannot occur. So $i_1 \in [n_1+1, n]$ and $b_{i_1} = \alpha^t\gamma_2$ for $t_1 = i_1 - j_2 - 1 \pmod{n_2}$. Similarly to Case 3.2 we get

$$\gamma_2 = \alpha^{n_1-t}\gamma_1 + \alpha^{-t}\gamma_1.$$

Combining this equation with equation (11) we have

$$\alpha^{n-t}\gamma_1 + \alpha^{n_2-t}\gamma_1 + \alpha^{n_1}\gamma_1 + \alpha^{n_1-t}\gamma_1 + \alpha^{-t}\gamma_1 = 0$$

and after multiplying with α^t we conclude that $m_\alpha = x^n + x^{n_1+t} + x^{n_2} + x^{n_1} + 1$, where $t \in \{1, \ldots, n_2 - 1\}$. □

Cases 1 and 3 of Theorem 5 also provide all elements α with $\mathrm{wt}_d(\alpha) = 2$. Moreover, Theorem 4 in [2] is a slightly weaker version of Case 1.3. in Theorem 5.

Remark 3. A suitable choice for the values i_1, j_1, i_2, j_2 in the second column of Table 1 can be found in the proof of the corresponding case.

The following example shows that the cycle normal forms of optimal s-XOR-representations are generally not unique.

Example 3. Let $\alpha \in \mathbb{F}_{2^4}$ with the irreducible minimal polynomial $m_\alpha = x^4 + x^3 + x^2 + x + 1$. Then, by Theorem 5, $\mathrm{wt}_s(\alpha) = \mathrm{wt}_d(\alpha) = 2$ and $M = C_{x^4+1} + E_{2,2} + E_{3,4}$ and $M' = (C_{x^3+1} \oplus C_{x+1}) + E_{3,4} + E_{4,3}$ belong to two different optimal representations, corresponding to Case 1.4. and Case 3.2 of Theorem 5, respectively.

The following corollary is a direct result from Theorem 5 and Example 2.

Corollary 3. *Let* $\alpha \in \mathbb{F}_{2^n}$ *with* $\mathrm{wt}(m_\alpha) = 5$ *and* $\deg(m_\alpha) = n$. *Then* $\mathrm{wt}_s(\alpha) = 2$ *if* f *appears in Table 1 and* $\mathrm{wt}_s(\alpha) = 3$ *otherwise.*

Corollary 3 shows that an implementation via the rational canonical form (as in Example 2) is generally not the best way to implement multiplication in binary finite fields. However, irreducible pentanomials that do not appear in the table in Theorem 5 exist, the examples with the lowest degree are $f = x^8+x^6+x^5+x^4+1$ and its reciprocal polynomial (for a table of all s-XOR-counts of finite field elements in $\mathbb{F}_{2^n}$ for $n \leq 8$ see [2]). It is an interesting question for which field elements the "naive" representation using the rational canonical form is optimal.

4 Quantifying the Gap between the Optimal Implementation and the Naive Implementation

It is now interesting to investigate the gap between the optimal implementation and the "naive" implementation using the rational canonical form. We will give a partial answer to this question in Theorem 6. First, we need some notation and lemmas.

For a square matrix $M = (m_{r,s})$ over $\mathbb{F}_2$ and two index sequences (ordered sets) $I = (i_1, \ldots, i_{l_1}), J = (j_1, \ldots, j_{l_2})$, $l := \min(l_1, l_2)$ we denote by $M^{I,J} = (a_{r,s})$ the matrix that is constructed as follows: All rows in I and all columns in J are filled with zeroes, except the entries $a_{i_1,j_1}, \ldots, a_{i_l,j_l}$ which are set to 1. More precisely:

$$a_{r,s} = \begin{cases} 0, & r = i_k, s \neq j_k \text{ for a } k \in \{1, \ldots, l_1\} \\ 0, & r \neq i_k, s = j_k \text{ for a } k \in \{1, \ldots, l_2\} \\ 1, & r = i_k, s = j_k \text{ for a } k \in \{1, \ldots, l\} \\ m_{r,s}, & \text{otherwise.} \end{cases}$$

The following example illustrates our notation. Let $I = \{2, 4\}$ and $J = \{1, 3\}$.

$$M = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad M^{I,J} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

In the case that $l_1 \neq l_2$ the matrix $M^{I,J}$ has a zero row/column and is thus not invertible. If $l_1 = l_2$, it is easy to see that $\det(M^{I,J})$ does not depend on the

ordering of the index sets I, J and is the same as the determinant of the matrix that is created by deleting all rows of M in I and all columns of M in J. In the case that we are only concerned with the determinant, we will thus just use (unordered) index sets I, J and also talk about determinants of submatrices. If $I = \{i\}$ and $J = \{j\}$ we will also write $M^{(i,j)}$. Moreover, we denote by A_M the characteristic matrix $A_M := xI + M$ of M.

Lemma 3. *Let $M = C_{x^n+1} \in \mathrm{GL}(n, \mathbb{F}_2)$. Then we have $\mathrm{wt}(\det(A_M^{I,J})) \leq 1$ for all possible proper square submatrices $A_M^{I,J}$.*

Proof. The proof is by induction on the size of the submatrix. Clearly, $\det(A_M^{I,J}) \in \{0, 1, x\}$ if $|I| = |J| = n - 1$. Let now $|I| < n - 1$. We denote by c_{ij} the entry in the i-th row and j-th column of A_M. Then

$$c_{ij} = \begin{cases} x, & i = j, \\ 1, & i = j + 1 \pmod{n}, \\ 0, & \text{else.} \end{cases}$$

Let $i \in I$. If $i \notin J$, then $A_M^{I,J}$ has at most one non-zero entry in the i-th column. Then, by Laplace expansion along the i-th column and use of the induction hypothesis, we get $\mathrm{wt}(\det(A_M^{I,J})) \leq 1$. If $i \in J$ and $i + 1 \pmod{n} \notin I$ then the $i + 1 \pmod{n}$-th row has at most one non-zero entry and Laplace expansion along the $i + 1 \pmod{n}$-th row yields $\mathrm{wt}(\det(A_M^{I,J})) \leq 1$. We conclude that $\mathrm{wt}(\det(A_M^{I,J})) \leq 1$ for all I with $|I| < n$. □

Lemma 4. *Let $M = C_{x^n+1} + \sum_{k=1}^{t} E_{i_k,j_k}$ where i_k, j_k can be chosen arbitrarily. Then we have $\mathrm{wt}(\det(A_M^{I,J})) \leq 2^t$ for all possible proper square submatrices $A_M^{I,J}$.*

Proof. The proof is by induction on t. The case $t = 0$ is covered by Lemma 3. Let now $t \geq 1$. Let $M' = C_{x^n+1} + \sum_{k=1}^{t-1} E_{i_k,j_k}$, so that $M = M' + E_{i,j}$ with $i = i_k$, $j = j_k$. If $i \in I$ or $j \in J$ we have $A_M^{I,J} = A_{M'}^{I,J}$ and thus $\mathrm{wt}(\det(A_M^{I,J})) = \mathrm{wt}(\det(A_{M'}^{I,J})) \leq 2^{t-1}$. If $i \notin I$ and $j \notin J$ then $A_M^{I,J} = A_{M'}^{I,J} + E_{i,j}$ and thus Laplace expansion along the i-th row yields $\det(A_M^{I,J}) \leq \det(A_{M'}^{I,J}) + \det(A_{M'}^{I\cup\{i\},J\cup\{j\}})$ and thus

$$\mathrm{wt}(\det(A_M^{I,J})) \leq \mathrm{wt}(\det(A_{M'}^{I,J})) + \mathrm{wt}(\det(A_{M'}^{I\cup\{i\},J\cup\{j\}})) \leq 2^{t-1} + 2^{t-1} = 2^t$$

by induction hypothesis. □

Corollary 4. *Let $M = C_{x^n+1} + \sum_{k=1}^{t} E_{i_k,j_k}$ where i_k, j_k can be chosen arbitrarily. Then $\mathrm{wt}(\chi(M)) \leq 2^t + 1$.*

Proof. The proof is by induction on t. The case $t = 0$ holds because $\chi(C_{x^n+1}) = x^n + 1$ by definition of the companion matrix. Let now $t \geq 1$ and $M' = C_{x^n+1} + \sum_{k=1}^{t-1} E_{i_k,j_k}$. Laplace expansion along the i_t-th row yields $\chi(M) = \det(A_M) = \chi(M') + \det(A_{M'}^{(i_t,j_t)})$. We conclude with Lemma 4 and the induction hypothesis that $\mathrm{wt}(\chi(M)) \leq 2^{t-1} + 1 + 2^{t-1} = 2^t + 1$. □

Theorem 6. *Let $\alpha \in \mathbb{F}_{2^n}$ be not contained in a proper subfield of $\mathbb{F}_{2^n}$ and let $M_{\alpha,B}$ be a multiplication matrix of α with respect to some basis B. Then* $\mathrm{wt}_d(M_{\alpha,B}) = t$ *implies* $\mathrm{wt}(m_\alpha) \leq 2^t + 1$.

Proof. Let B be an optimal (regarding the d-XOR-count) basis and $M := M_{\alpha,B} = \bigoplus_{k=1}^{l} C_{x^{m_k}+1} + \sum_{r=1}^{t} E_{i_r,j_r}$ be an optimal multiplication matrix. The case $l = 1$ is covered in Corollary 4, so we only consider $l > 1$ for the rest of the proof. Since α is not contained in a proper subfield of $\mathbb{F}_{2^n}$, the minimal polynomial of M coincides with its characteristic polynomial. We call the sets

$$\{1, \ldots, m_1\}, \{m_1 + 1, \ldots, m_2\}, \ldots, \{\sum_{k=1}^{l-1} m_k + 1, \ldots, \sum_{k=1}^{l} m_k\}$$

the l *blocks* of M. We can decompose $M = M_1 + M'$ with $M_1 = \bigoplus_{k=1}^{l} C_{x^{m_k}+1} + \sum_{r=1}^{t_1} E_{i_r,j_r}$ and $M' = \sum_{r=1}^{t_2} E_{i_r,j_r}$ in a way that all pairs (i_r, j_r) in M_1 are in the same block and all pairs (i_r, j_r) in M' are in different blocks. M_1 is a block diagonal matrix and with Corollary 4 we get

$$\mathrm{wt}(\chi(M_1)) \leq \prod_{k=1}^{l} (2^{s_k} + 1) \text{ with } \sum_{k=1}^{l} s_k = t_1 \tag{12}$$

where s_k denotes the number of pairs (i_r, j_r) that are in the k-th block. We call $B_1, \ldots, B_l$ the l blocks of M_1 and $m_1, \ldots, m_l$ the size of these blocks. Note that $\chi(M)$ is irreducible which implies that M is not a block triangular matrix and thus $t_2 \geq l$. So we can write $M' = M_2 + M_3$ in a way that (after a suitable permutation of blocks) $M_1 + M_2$ looks like this:

$$M_1 + M_2 = \begin{pmatrix} B_1 & 0 & \ldots & E_{i_l,j_l} \\ E_{i_1,j_1} & B_2 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & E_{i_{l-1},j_{l-1}} & B_l \end{pmatrix}. \tag{13}$$

From this, we infer by Laplace expansion along the i_l-th row

$$\chi(M_1 + M_2) = \chi(M_1) + \det(A^{(i_l,v)}_{M_1+M_2}), \tag{14}$$

where $v = \sum_{k=1}^{l-1} m_k + j_l$. We now determine $\mathrm{wt}(\det(A^{(i_l,v)}_{M_1+M_2}))$. We get

$$\det(A^{(i_l,v)}_{M_1+M_2}) = \det \begin{pmatrix} B_1^{(i_l,\emptyset)} & 0 & \ldots & 0 & E_{i_l,j_l} \\ E_{i_1,j_1} & B_2 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & E_{i_{l-2},j_{l-2}} & B_{l-1} & 0 \\ 0 & \ldots & 0 & E_{i_{l-1},j_{l-1}} & B_l^{(\emptyset,j_l)} \end{pmatrix}$$

$$= \det \begin{pmatrix} B_1^{(i_l,\emptyset)} & 0 & \dots & E_{i_{l-1},j_{l-1}} & * \\ E_{i_1,j_1} & B_2 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & E_{i_{l-2},j_{l-2}} & B_{l-1} & 0 \\ 0 & \dots & 0 & 0 & B_l^{(i_{l-1},j_l)} \end{pmatrix}$$

by swapping the i_l-th row with the $\sum_{k=1}^{l-1} m_k + i_{l-1}$-th row. This operation can now be repeated for the upper-left $l-1$ blocks, the result is the following block diagonal matrix

$$\det(A_{M_1+M_2}^{(i_l,v)}) = \det \begin{pmatrix} B_1^{(i_l,j_1)} & * & \dots & 0 & 0 \\ 0 & B_2^{(i_1,j_2)} & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & 0 & B_{l-1}^{(i_{l-2},j_{l-1})} & * \\ 0 & \dots & 0 & 0 & B_l^{(i_{l-1},j_l)} \end{pmatrix}.$$

Lemma 4 then implies $\mathrm{wt}(\det(A_{M_1+M_2}^{(i_l,v)})) \leq \prod_{k=1}^{l} 2^{s_k} = 2^{t_1}$. Equations (12) and (14) yield

$$\mathrm{wt}(\chi(M_1+M_2)) \leq \prod_{k=1}^{l} (2^{s_k}+1) + 2^{t_1}. \tag{15}$$

We now investigate the determinant of the square submatrices of M_1+M_2. Let I, J be index sets and set $I = \dot{\bigcup}_r I_r$ and $J = \dot{\bigcup}_r J_r$ where I_r and J_r contain the indices that belong to the r-th block. Observe that $|I| = |J|$. Let us first look at the case $I = I_r$ and $J = J_r$ for some r. Using Lemma 4

$$\mathrm{wt}(\det(A_{M_1+M_2}^{I,J})) \leq 2^{s_r} \prod_{\substack{k \in \{1,\dots,l\} \\ k \neq r}} (2^{s_k}+1) + 2^{t_1}.$$

Similarly, if $|I_r| = |J_r|$ for all $1 \leq r \leq l$ then

$$\mathrm{wt}(\det(A_{M_1+M_2}^{I,J})) \leq \prod_{r:I_r \neq \emptyset} 2^{s_r} \prod_{r:I_r=\emptyset} (2^{s_k}+1) + 2^{t_1}. \tag{16}$$

Let us now assume that there is a block r with $|I_r| \neq |J_r|$. We can assume w.l.o.g. that $r = 1$ and $p := |I_1| < |J_1|$. If $i_1 + m_1, i_l \in I$ or $v, j_1 \in J$ then equation (13) implies $\det(A_{M_1+M_2}^{I,J}) = 0$. We order $I = (a_1, \dots, a_t)$ and $J = (b_1, \dots, b_t)$ in ascending order. Then

$$\det(A_{M_1+M_2}^{I,J}) = \det \begin{pmatrix} B_1^{I_1,J_1} & A \\ C & D \end{pmatrix},$$

with $A = (a_{r,s}) \in \mathbb{F}_2^{m_1 \times (n-m_1)}$, $C = (c_{r,s}) \in \mathbb{F}_2^{(n-m_1)\times m_1}$ with

$$a_{r,s} = \begin{cases} 1, & \text{for } (r,s) = (i_l, v), \\ 0, & \text{else,} \end{cases} \quad c_{r,s} = \begin{cases} 1, & \text{for } (r,s) = (a_k, b_k), k > p, \\ 1, & \text{for } (r,s) = (i_1, j_1), \\ 0, & \text{else.} \end{cases}$$

Swapping the i_l-th row with the a_{p+1}-th row, we obtain

$$\det(A^{I,J}_{M_1+M_2}) = \det \begin{pmatrix} B_1^{I_1 \cup \{i_l\}, J_1} & 0 \\ * & D' \end{pmatrix}$$

and thus $\det(A^{I,J}_{M_1+M_2}) = \det(B_1^{I_1\cup\{i_l\},J_1}) \det(D')$. Observe that $\det(A^{I,J}_{M_1+M_2}) = 0$ if $|I_1| \neq |J_1| + 1$. Moreover, $\det(D') = \det(C^{I',J'}_{M_1+M_2})$ where $\{1,\dots,m_1\}$ is a subset of I' and J'. In particular, the number of indices in I' and J' belonging to the first block is the same. By induction, equation (16) and Lemma 4, we get

$$\mathrm{wt}(\det(A^{I,J}_{M_1+M_2})) = \mathrm{wt}(\det(B_1^{I_1\cup\{i_l\},J_1}) \det(D')) \leq 2^{s_1} \prod_{k=2}^{l} (2^{s_k}+1) + 2^{t_1}. \quad (17)$$

Equations (16) and (17) imply that for arbitrary index sets I, J, there exists an $r \in \{1,\dots,l\}$ such that

$$\mathrm{wt}(\det(A^{I,J}_{M_1+M_2})) \leq 2^{s_r} \prod_{\substack{k\in\{1,\dots,l\} \\ k\neq r}} (2^{s_k}+1) + 2^{t_1}. \quad (18)$$

As in the proof of Lemma 4, for arbitrary index sets I, J and $i, j \in \{1,\dots,n\}$ there is an $r \in \{1,\dots,l\}$ such that

$$\begin{aligned} \mathrm{wt}(\det(A^{I,J}_{M_1+M_2+E_{i,j}})) &\leq \mathrm{wt}(\det(A^{I,J}_{M_1+M_2})) + \mathrm{wt}(\det(A^{I\cup\{i\},J\cup\{j\}}_{M_1+M_2})) \\ &\leq 2 \cdot \left(2^{s_r} \prod_{\substack{k\in\{1,\dots,l\} \\ k\neq r}} (2^{s_k}+1) + 2^{t_1} \right) \end{aligned}$$

and, inductively, for an arbitrary matrix $M_3 = \sum_{k=1}^{z} E_{i_k,j_k}$ with z non-zero entries

$$\begin{aligned} \mathrm{wt}(\det(A^{I,J}_{M_1+M_2+M_3})) &\leq 2^z \left(2^{s_r} \prod_{\substack{k\in\{1,\dots,l\} \\ k\neq r}} (2^{s_k}+1) + 2^{t_1} \right) \\ &< 2^z \left(\prod_{k=1}^{l} (2^{s_k}+1) + 2^{t_1} \right). \end{aligned} \quad (19)$$

We now show by induction that we have for $z \geq 1$

$$\mathrm{wt}(\chi(M_1 + M_2 + M_3)) < 2^z \left(\prod_{k=1}^{l} (2^{s_k} + 1) + 2^{t_1} \right). \tag{20}$$

The case $z = 1$ is dealt with using equations (15) and (18):

$$\begin{aligned} \mathrm{wt}(\chi(M_1 + M_2 + M_3)) &\leq \mathrm{wt}(\chi(M_1 + M_2)) + \mathrm{wt}(\det(A^{i_1,j_1}_{M_1+M_2})) \\ &< 2 \left(\prod_{k=1}^{l} (2^{s_k} + 1) + 2^{t_1} \right). \end{aligned}$$

Let now $z > 1$ and $M'_3 = \sum_{k=1}^{z-1} E_{i_k,j_k}$. With the induction hypothesis and equation (19) we conclude

$$\begin{aligned} \mathrm{wt}(\chi(M_1 + M_2 + M_3)) &\leq \mathrm{wt}(\chi(M_1 + M_2 + M'_3)) + \mathrm{wt}(\det(A^{i_1,j_1}_{M_1+M_2+M'_3})) \\ &< 2^z \left(\prod_{k=1}^{l} (2^{s_k} + 1) + 2^{t_1} \right), \end{aligned}$$

proving equation (20). Note that the bound in equation (20) depends only on the parameters l, t_2 and $s_k, k = 1, \ldots, l$ where $\sum_{k=1}^{l} s_k = t_1$ and $t_1 + t_2 = t = \mathrm{wt}(M)$. For $t_2 > l$ we have

$$\mathrm{wt}(\chi(M_1 + M_2 + M_3)) < 2^{t_2 - l} \left(\prod_{k=1}^{l} (2^{s_k} + 1) + 2^{t_1} \right).$$

Using equation (15), a matrix N with values $l_N = t_2$ and $s_k = 0$ for $k > l$ yields

$$\begin{aligned} \mathrm{wt}(\chi(N)) &\leq \prod_{k=1}^{l_N} (2^{s_k} + 1) + 2^{t_1} \\ &= 2^{t_2 - l} \prod_{k=1}^{l} (2^{s_k} + 1) + 2^{t_1}. \end{aligned}$$

In particular, the upper bound given in equation (20) is always worse than the one given in equation (15) and we can focus on the case $M_3 = 0$ (or, equivalently, $t_2 = l$) for the rest of this proof. In other words, we just have to find the parameters that give the maximum weight estimation in equation (15). A direct calculation yields

$$\prod_{k=1}^{l} (2^{s_k} + 1) \leq (2^{t_1} + 1) \cdot 2^{l-1},$$

i.e. the choice $s_1 = t_1$, $s_i = 0$ for $i > 1$ is optimal. Plugging these parameters into equation (15), we get

$$\mathrm{wt}(\chi(M)) \leq 2^{t_1 + l - 1} + 2^{l-1} + 2^{t_1} = 2^{t-1} + 2^{l-1} + 2^{t-l}. \tag{21}$$

Obviously, the maximum of $2^{l-1} + 2^{t-l}$ for $2 \leq l \leq t$ is attained at $l = t$. The result follows from equation (21). □

We now show that the bound given in Theorem 6 is optimal by giving two examples where the upper bound is attained. Note that the proof of Theorem 6 implies that this can only occur if the number of blocks of the optimal multiplication matrix is 1 or t. We will give examples for both cases in Propositions 2 and 3.

Theorem 7 ([11, Theorem 3.5], [7, Theorem 4.3.9]). *Let R be a (commutative) Euclidean domain and $A \in R^{n\times n}$. Then A can be transformed into an upper triangular matrix using elementary row operations (i.e. a sequence of left-multiplications with matrices $I + rE_{i,j}$ with $r \in R$ and $i \neq j$).*

Proposition 2. *Let $\alpha \in \mathbb{F}_{2^n}$ with an irreducible minimal polynomial f with $\mathrm{wt}(f) = 2^t + 1$ of the form*

$$f = x^n + \prod_{j=1}^{t} \left(x^{i_j} + 1\right)$$

for arbitrary values of $i_j \in \mathbb{N}$ with $\sum_{j=1}^{t} i_j \leq n - t$. Then there exists a basis B such that the matrix $M := M_{\alpha,B}$ satisfies $\mathrm{wt}_s(M) = \mathrm{wt}_d(M) = t$.

Proof. We show that the matrix $M = C_{x^n+1} + \sum_{k=1}^{t-1} E_{j_k+i_k+1,j_k} + E_{i_t+n-j_t,j_t}$ where the j_k are chosen arbitrarily under the conditions that $j_{k+1} \geq i_k + j_k + 1$ for all $k = 1, \ldots, t-1$ and $i_t < j_1$ has the desired property. It is clear that $\mathrm{wt}_s(M) = \mathrm{wt}_d(M) = t$. Let $B = \{b_1, \ldots, b_n\}$ be some basis of $(\mathbb{F}_2)^n$ over $\mathbb{F}_2$. We investigate how M (viewed as a transformation matrix) operates on this basis:

$$\begin{aligned}
Mb_1 &= b_2 \\
&\vdots \\
Mb_{j_1-1} &= b_{j_1} \\
Mb_{j_1} &= b_{j_1+1} + M^{i_1} b_{j_1+1} \qquad (22)\\
Mb_{j_1+1} &= b_{j_1+2} \\
&\vdots \\
Mb_{j_2-1} &= b_{j_2} \\
Mb_{j_2} &= b_{j_2+1} + M^{i_2} b_{j_2+1} \qquad (23)\\
Mb_{j_2+1} &= b_{j_2+2} \\
&\vdots \\
Mb_n &= b_1.
\end{aligned}$$

Set $n_i = j_i - j_{i-1}$ for $2 \leq i \leq t$ and $n_1 = n + j_1 - j_t$. Note that $\sum_{i=1}^{t} n_i = n$ and $Mb_{j_k} = M^{n_i} b_{j_{k-1}+1}$. With this and the equations of type (22) and (23) we

obtain the following set of equations:

$$\begin{pmatrix} M^{n_2} & M^{i_1}+1 & 0 & \dots & 0 \\ 0 & M^{n_3} & M^{i_2}+1 & \dots & 0 \\ & & \ddots & \ddots & \\ 0 & \dots & 0 & M^{n_t} & M^{i_{t-1}}+1 \\ M^{i_t}+1 & 0 & \dots & 0 & M^{n_1} \end{pmatrix} \begin{pmatrix} b_{j_1+1} \\ b_{j_2+1} \\ \vdots \\ \vdots \\ b_{j_t+1} \end{pmatrix} = 0. \tag{24}$$

We denote by A the matrix in equation (24). A is a matrix over $\mathbb{F}_2[M]$. It is clear that $\mathbb{F}_2[M]$ is isomorphic to the usual polynomial ring $\mathbb{F}_2[x]$ and thus a Euclidean domain. Using the Leibniz formula for determinants, we obtain $\det(A) = f(M)$. By Theorem 7, we can transform A into an upper triangular matrix A' using only elementary row operations. In particular $\det(A') = \prod_{i=1}^{n} a'_{i,i} = \det(A) = f(M)$ where the $a'_{i,i}$ denote the entries on the diagonal of A'. Since f is irreducible, we obtain $a_{k,k} = f(M)$ for one $1 \leq k \leq n$ and $a_{i,i} = 1$ for all $i \neq k$, i.e.

$$\begin{pmatrix} 1 & & & & * \\ & \ddots & & & \\ & & f(M) & * & * \\ & & & \ddots & \\ 0 & & & & 1 \end{pmatrix} \begin{pmatrix} b_{j_1+1} \\ \vdots \\ b_{j_k+1} \\ \vdots \\ b_{j_t+1} \end{pmatrix} = 0.$$

It is clear that all entries $a'_{k,k+1}, \dots, a'_{k,n}$ can be eliminated by further row additions. Hence, we obtain $f(M)b_{j_k+1} = 0$, i.e. f is the M-annihilator of b_{j_k+1}. As f is irreducible this implies that the minimal polynomial of M is f and thus M is a multiplication matrix of α.

□

Proposition 3. *Let $\alpha \in \mathbb{F}_{2^n}$ with an irreducible minimal polynomial f with $\mathrm{wt}(f) = 2^t + 1$ of the form*

$$f = \prod_{j=1}^{t} (x^{n_j} + 1) + x^k$$

for arbitrary values of n_j and $k \leq n - t$ with $\sum_{j=1}^{t} n_j = n$. Then there exists a basis B such that the matrix $M := M_{\alpha,B}$ satisfies $\mathrm{wt}_s(M) = \mathrm{wt}_d(M) = t$.

Proof. The proof is similar to the proof of the previous lemma. Define $\hat{n}_l = \sum_{u=1}^{l-1} n_u$ for $1 \leq l \leq t$. Let r_l be chosen arbitrarily such that $1 \leq r_l \leq n_l$ for $1 \leq l \leq t$ and $\sum_{l=1}^{t} r_l = k$. Further let $j_l := \hat{n}_l + r_l$ for all $1 \leq l \leq t$ and $s_l := i_l + r_{l+1} + 1 \pmod{n_{l+1}}$ for $l < t$ and $s_t := i_t + r_1 + 1 \pmod{n_1}$.

Define now $M = \bigoplus_{i=1}^{t} C_{x^{n_i}+1} + \sum_{k=1}^{t} E_{\hat{n_k}+s_k,j_k}$. Obviously, $\mathrm{wt}_s(M) = \mathrm{wt}_d(M) = t$. Let $B = \{b_1, \dots, b_n\}$ be some basis of $(\mathbb{F}_2)^n$ over $\mathbb{F}_2$. We investigate

how M (viewed as a transformation matrix) operates on this basis:

$$\begin{array}{lll}
Mb_1 = b_2 & Mb_{n_1+1} = b_{n_1+2} & \dots\ Mb_{n_{t-1}+1} = b_{n_t+2} \\
\vdots & \vdots & \vdots \\
Mb_{j_1-1} = b_{j_1} & Mb_{j_2-1} = b_{j_2} & Mb_{j_t-1} = b_{j_t} \\
Mb_{j_1} = b_{j_1+1} + M^{i_1} b_{j_2+1} & Mb_{j_2} = b_{j_2+1} + M^{i_2} b_{j_3+1} & Mb_{j_t} = b_{j_t+1} + M^{i_t} b_{j_1+1} \\
Mb_{j_1+1} = b_{j_1+2} & Mb_{j_2+1} = b_{j_2+2} & Mb_{j_t+1} = b_{j_t+2} \\
\vdots & \vdots & \vdots \\
Mb_{n_1} = b_1 & Mb_{n_2} = b_{n_1+1} & \dots\ Mb_n = b_{n_{t-1}+1}.
\end{array}$$

Clearly, $Mb_{j_k} = M^{n_k} b_{j_k+1}$, so we get the following set of equations:

$$\begin{pmatrix}
M^{n_1}+1 & M^{i_1} & 0 & \dots & 0 \\
0 & M^{n_2}+1 & M^{i_2} & \dots & 0 \\
 & & \ddots & \ddots & \\
0 & \dots & 0 & M^{n_{t-1}}+1 & M^{i_{t-1}} \\
M^{i_t} & 0 & \dots & 0 & M^{n_t}+1
\end{pmatrix}
\begin{pmatrix}
b_{j_1+1} \\ b_{j_2+1} \\ \vdots \\ \vdots \\ b_{j_t+1}
\end{pmatrix} = 0.$$

The determinant of the matrix is exactly $f(M)$. We can now repeat the arguments from the proof of Proposition 2 and obtain that M is a multiplication matrix for α. □

Observe that the polynomials in Propositions 2 and 3 are generalizations of Case 1.3. and Case 3.2. in Theorem 5.

Note that irreducible polynomials of the types mentioned in Propositions 2 and 3 do exist, examples up to $t = 8$, corresponding to polynomials of weight $2^t + 1$, are compiled in Table 2. The table lists in the second column values for i_l and n that belong to an irreducible polynomial of the type of Proposition 2 and in the third column the values for n_l and k that belong to an irreducible polynomial of the type of Proposition 3. The values listed were found with a simple randomized algorithm. They generally do not correspond to the irreducible polynomial of that type with the least degree. Propositions 2 and 3 together with Theorem 6 show that the gap between the number of XORs used in the optimal implementation and the number of XORs used in the naive implementation of a multiplication matrix using the rational canonical form grows exponentially with the weight of the minimal polynomial of the element.

Propositions 2 and 3 show that there are elements $\alpha \in \mathbb{F}_{2^n}$ with $\mathrm{wt}(m_\alpha) = 2^t + 1$ and $\mathrm{wt}_s(\alpha) = t$. We believe that this upper bound is strict, i.e. the bound is the same for s-XOR-count and d-XOR-count.

Conjecture 1. Let $\alpha \in \mathbb{F}_{2^n}$ be not contained in a proper subfield of $\mathbb{F}_{2^n}$ and $M_{\alpha,B}$ a multiplication matrix of α with respect to some basis B. Then $\mathrm{wt}_s(M_{\alpha,B}) = t$ implies $\mathrm{wt}(\chi(M)) \leq 2^t + 1$.

Table 2. Irreducible polynomials of the form described in Propositions 2 and 3.

t	values for $i_1, \ldots, i_t; n$	values for $n_1, \ldots, n_t; k$
2	1,2;5	2,4;1
3	1,2,4;10	4,5,6;1
4	3,5,6,12;30	2,3,6,10;1
5	1,2,4,9,17;39	12,13,15,19,23;9
6	1,12,16,24,31; 123	13,22,26,27,28,30;23
7	2,30,47,56,60,64,91; 357	25,114,174,231,279,281,331;196
8	23,28,41,59,62,106,141,153; 628	44,148,195,357,363,368,386,480;240

5 Open Problems

Our investigations open up many possibilities for future research. While Theorem 1 shows that there is an infinite family of matrices with higher s-XOR-count than d-XOR-count, a more precise classification of these cases as well as finding upper/lower bounds is desirable. Because of the nature of the s-XOR-count, answers to these problems would also give insight into optimal Gauss elimination strategies over $\mathbb{F}_2$.

Problem 1. Classify the matrices $M \in \mathrm{GL}(n, \mathbb{F}_2)$ with $\mathrm{wt}_d(M) < \mathrm{wt}_s(M)$.

Problem 2. Find bounds c, C so that $c\,\mathrm{wt}_d(M) \leq \mathrm{wt}_s(M) \leq C\,\mathrm{wt}_d(M)$ for all matrices $M \in \mathrm{GL}(n, \mathbb{F}_2)$.

Finding out if/how the bounds c, C depend on n and $\mathrm{wt}_s(M)$ would greatly improve the understanding of the two XOR-metrics.

As observed in Section 3, there are elements $\alpha \in \mathbb{F}_2$ where the optimal implementation of the mapping $x \mapsto \alpha x$ is the rational canonical form in both of the investigated metrics. These elements are (compared to elements with minimal polynomials of the same weight) the most expensive to implement. A more thorough understanding of these elements would be helpful.

Problem 3. Classify the minimal polynomials $m_\alpha \in \mathbb{F}_2[x]$ for which the optimal multiplication matrix is in rational canonical form.

We also want to repeat a problem about elements in subfields mentioned in [2].

Problem 4. Let $\alpha \in \mathbb{F}_{2^n}$ be contained in a subfield $\mathbb{F}_{2^l}$ with $ld = n$. Let M_l be an optimal multiplication matrix of α regarding d- or s-XOR-count. Is $M = \bigoplus_{k=1}^{d} M_l$ then an optimal multiplication matrix of $\alpha \in \mathbb{F}_{2^n}$ regarding d- or s-XOR-count?

In Sections 3 and 4 we limited ourselves to optimal XOR-implementations of matrices that are multiplication matrices for a fixed field element (which are exactly those with irreducible minimal polynomial). Investigating a more general case is also an interesting problem.

Problem 5. Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a bijective linear mapping and $M_{f,B} \in \mathrm{GL}(n, \mathbb{F}_2)$ the matrix that belongs to f with respect to the basis B. Find a basis B such that the matrix $M_{f,B}$ is the optimal d/s-XOR-count matrix.

In particular, finding optimal matrices $M_{f,B}$ where f denotes the mapping induced by a linear layer of a cryptographic scheme is a very interesting problem.

Note. In December 2018, after the submission of this paper, Mesnager, Kim, Jo, Choe, Han and Lee [17] have independently proven the conjecture of Beierle, Kranz and Leander in [2], i.e. they prove that $\mathrm{wt}_s(M_{\alpha,B}) = 2$ implies $\mathrm{wt}(m_\alpha) \leq 5$. This result is implied by Theorem 5 in this work. The proof techniques used in [17] are similar to the ones used in this paper.

Acknowledgments. The author wishes to thank the anonymous referees for their comments that improved especially the introduction considerably and helped to set this work into context with existing literature.

I also thank Gohar Kyureghyan for many discussions and help with structuring this paper.

References

1. Babbage, S., Dodd, M.: The MICKEY Stream Ciphers, pp. 191–209. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-68351-3_15
2. Beierle, C., Kranz, T., Leander, G.: Lightweight multiplication in $GF(2^n)$ with applications to MDS matrices. In: Proceedings, Part I, of the 36th Annual International Cryptology Conference on Advances in Cryptology — CRYPTO 2016 - Volume 9814. pp. 625–653. Springer-Verlag New York, Inc., New York, NY, USA (2016). https://doi.org/10.1007/978-3-662-53018-4_23
3. Canright, D.: A very compact S-Box for AES. In: Rao, J.R., Sunar, B. (eds.) Cryptographic Hardware and Embedded Systems – CHES 2005. pp. 441–455. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)
4. Daemen, J., Rijmen, V.: Correlation analysis in $GF(2^n)$. In: Junod, P., Canteaut, A. (eds.) Advanced Linear Cryptanalysis of Block and Stream Ciphers. Cryptology and information security. pp. 115–131. IOS Press (2011)
5. De Cannière, C., Preneel, B.: Trivium, pp. 244–266. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-68351-3_18
6. Duval, S., Leurent, G.: MDS matrices with lightweight circuits. IACR Transactions on Symmetric Cryptology **2018**(2), 48–78 (Jun 2018). https://doi.org/10.13154/tosc.v2018.i2.48-78
7. Hahn, A., O'Meara, T.: The classical Groups and K-Theory. Springer Berlin Heidelberg, Berlin, Heidelberg (1989)
8. Hell, M., Johansson, T., Meier, W.: Grain; a stream cipher for constrained environments. Int. J. Wire. Mob. Comput. **2**(1), 86–93 (May 2007). https://doi.org/10.1504/IJWMC.2007.013798
9. Hoffman, K., Kunze, R.: Linear algebra. Prentice-Hall, Englewood Cliffs, New Jersey (1961)

10. Jean, J., Peyrin, T., Sim, S.M., Tourteaux, J.: Optimizing implementations of lightweight building blocks. IACR Transactions on Symmetric Cryptology **2017**(4), 130–168 (2017)
11. Kaplansky, I.: Elementary divisors and modules. Trans. Amer. Math. Soc. **66**, 464–491 (1949). https://doi.org/10.1090/S0002-9947-1949-0031470-3
12. Khoo, K., Peyrin, T., Poschmann, A.Y., Yap, H.: FOAM: Searching for hardware-optimal SPN structures and components with a fair comparison. In: Batina, L., Robshaw, M. (eds.) Cryptographic Hardware and Embedded Systems – CHES 2014. pp. 433–450. Springer Berlin Heidelberg, Berlin, Heidelberg (2014)
13. Kranz, T., Leander, G., Stoffelen, K., Wiemer, F.: Shorter linear straight-line programs for mds matrices. IACR Transactions on Symmetric Cryptology **2017**(4), 188–211 (Dec 2017). https://doi.org/10.13154/tosc.v2017.i4.188-211, https://tosc.iacr.org/index.php/ToSC/article/view/813
14. LaMacchia, B.A., Odlyzko, A.M.: Solving large sparse linear systems over finite fields. In: Menezes, A.J., Vanstone, S.A. (eds.) Advances in Cryptology-CRYPT0' 90. pp. 109–133. Springer Berlin Heidelberg, Berlin, Heidelberg (1991)
15. Li, Y., Wang, M.: On the construction of lightweight circulant involutory MDS matrices. In: Revised Selected Papers of the 23rd International Conference on Fast Software Encryption - Volume 9783. pp. 121–139. FSE 2016, Springer-Verlag New York, Inc., New York, NY, USA (2016). https://doi.org/10.1007/978-3-662-52993-5_7
16. Liu, M., Sim, S.M.: Lightweight MDS generalized circulant matrices. In: Peyrin, T. (ed.) Fast Software Encryption. pp. 101–120. Springer Berlin Heidelberg, Berlin, Heidelberg (2016)
17. Mesnager, S., Kim, K.H., Jo, D., Choe, J., Han, M., Lee, D.N.: A proof of the Beierle-Kranz-Leander conjecture related to lightweight multiplication in $\mathbb{F}_{2^n}$. http://arxiv.org/abs/1812.09666 (2018)
18. Saarinen, M.J.O.: Cryptographic analysis of all 4 × 4-bit S-Boxes. In: Miri, A., Vaudenay, S. (eds.) Selected Areas in Cryptography. pp. 118–133. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
19. Sajadieh, M., Mousavi, M.: Construction of lightweight mds matrices from generalized feistel structures. IACR Cryptology ePrint Archive **2018**, 1072 (2018)
20. Sarkar, S., Sim, S.M.: A deeper understanding of the XOR count distribution in the context of lightweight cryptography. In: Proceedings of the 8th International Conference on Progress in Cryptology — AFRICACRYPT 2016 - Volume 9646. pp. 167–182. Springer-Verlag, Berlin, Heidelberg (2016). https://doi.org/10.1007/978-3-319-31517-1_9
21. Sim, S.M., Khoo, K., Oggier, F., Peyrin, T.: Lightweight MDS involution matrices. In: Leander, G. (ed.) Fast Software Encryption. pp. 471–493. Springer Berlin Heidelberg, Berlin, Heidelberg (2015)
22. Swan, R.G.: Factorization of polynomials over finite fields. Pacific J. Math. **12**(3), 1099–1106 (1962)
23. Zhao, R., Wu, B., Zhang, R., Zhang, Q.: Designing optimal implementations of linear layers (full version). Cryptology ePrint Archive, Report 2016/1118 (2016)