On the Algebraic Degree of Iterated Power Functions

Clémence Bouvier[1,2], Anne Canteaut[2] and Léo Perrin[2]

[1] Sorbonne Université, France
[2] Inria, France
clemence.bouvier@inria.fr,anne.canteaut@inria.fr,leo.perrin@inria.fr

Abstract. New symmetric primitives are being designed to address a novel set of design criteria. Instead of being executed on regular processors or smartcards, they are instead intended to be run in abstract settings such as multi-party computations or zero-knowledge proof systems. This implies in particular that these new primitives are described using operations over large finite fields. As the number of such primitives grows, it is important to better understand the properties of their underlying operations.
In this paper, we investigate the algebraic degree of one of the first such block ciphers, namely MiMC. It is composed of many iterations of a simple round function, which consists of an addition and of a low-degree power permutation applied to the full state, usually $x \mapsto x^3$. We show in particular that, while the *univariate* degree increases predictably with the number of rounds, the *algebraic* degree (a.k.a multivariate degree) has a much more complex behaviour, and simply stays constant during some rounds. Such *plateaus* slightly slow down the growth of the algebraic degree.
We present a full investigation of this behaviour. First, we prove some lower and upper bounds for the algebraic degree of an arbitrary number of iterations of MiMC and of its inverse. Then, we combine theoretical arguments with simulations to prove that the upper bound is tight for up to 16265 rounds. Using these results, we slightly improve the higher-order differential attack presented at Asiacrypt 2020 to cover one or two more rounds. More importantly, our results provide some precise guarantees on the algebraic degree of this cipher, and then on the minimal complexity for a higher-order differential attack.

Keywords: symmetric cryptography, cryptanalysis, block cipher, finite field, algebraic degree, MiMC, higher order differential attack

1 Introduction

New computing environments are emerging that differ significantly from the usual computer processors, micro-controllers, or smartcards. These traditional platforms are those for which symmetric primitives have been optimized. However, the rise of environments implementing Multi-Party Computation (MPC) protocols such as smart-contracts or zero-knowledge proofs creates a new need. Indeed, symmetric primitives are still needed in these contexts, in particular to ensure computation integrity [BBHR18]. However, the basic operations provided by these platforms correspond neither to the CPU instructions (bit-wise AND, rotations, etc.) nor to the hardware components (XNOR, wires, etc.) that are used as building blocks in the usual case. Instead, the core operations that implementers can use are finite-field operations over fields $\mathbb{F}_q$ of large size q, where the size q is typically bigger than 2^{64} and is usually either a prime number or a power of 2 [AGP+19, AAB+20]. Primitives that are designed using such operations only are called *arithmetization-friendly*, see e.g. [BGL20] for a detailed survey on the arithmetization-friendly hash functions.

Designing arithmetization-friendly symmetric primitives is different from the "usual" case. Instead of using operations on $\mathbb{F}_{2^n}$ where n is a small even integer, typically $n = 4$ or 8, the underlying alphabet is now a large field whose cardinality is chosen according to some other parts in the protocol. For example, some zero-knowledge proof systems are defined over the finite field underlying a standard elliptic curve, in which case typical values of q would correspond to prime number of about 256 bits.

It would be possible to use "classical" symmetric primitives such as the AES in such contexts, but the cost of the encoding and decoding of the binary operations into finite field operations would be extremely costly (as can be seen for instance in the benchmarks presented in [BGL20]). As a consequence, dedicated primitives are designed so that they can work natively in such large fields. Furthermore, the advanced protocols running over such large fields require that the operations used have a low degree. Overall, there is a need for new symmetric cryptographic primitives operating over large (possibly prime) fields, and that rely on low degree operations.

In fact, several such proposals have been found to have significant flaws, from ad-hoc attacks relying on internal simplifications [ACG+19], to integral attacks [BCD+20]. As a consequence, it is necessary to better understand the behaviour of even the most basic cryptanalysis techniques when they are applied to arithmetization-friendly designs.

Univariate and Algebraic Degrees. In this paper, we investigate the algebraic degree of an arithmetization-friendly block cipher. The complexity of so-called *higher-order differential attacks* [Knu95] decreases with the algebraic degree, implying that it is important to understand how this quantity increases as a given round function is iterated. First, let us recall the two notions of degree which apply to a function over a finite field with characteristic 2.

Definition 1.1 (ANF and Algebraic Degree)**.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function. Its* Algebraic Normal Form (ANF) *is the representation of f as a multivariate polynomial with variables in $\mathbb{F}_2^n$, so that*

$$f(x_0, ..., x_{n-1}) = \sum_{u \in \mathbb{F}_2^n} a_u x^u \ ,$$

where $a_u \in \mathbb{F}_2$ for all u, and $x^u = \prod_{i=0}^{n-1} x_i^{u_i}$. The algebraic degree *of f is*

$$\deg^a f = \max \left\{ \mathrm{wt}(u) : u \in \mathbb{F}_2^n, a_u \neq 0 \right\} \ ,$$

where $\mathrm{wt}(u)$ is the Hamming weight of u. If $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$, then its algebraic degree, $\deg^a F$, is the maximal algebraic degree of the coordinates of F.

The algebraic degree should not be confused with the *univariate degree*, which is defined for any function $F : \mathbb{F}_q \to \mathbb{F}_q$.

Definition 1.2 (Univariate Representation and Degree)**.** *Let $q > 1$ be a prime power and let F be a function from $\mathbb{F}_q$ to $\mathbb{F}_q$. Then the* univariate polynomial representation *of F is*

$$F(x) = \sum_{i=0}^{q-1} u_i x^i \ ,$$

where $u_i \in \mathbb{F}_q$ for all integers i. Its univariate degree $\deg^u F$ *is the largest integer i for which $u_i \neq 0$.*

If $q = 2^n$, then a function $F : \mathbb{F}_q \to \mathbb{F}_q$ can be seen both as a function defined over the finite field, and as a function defined over the vector space $\mathbb{F}_2^n$ using a simple isomorphism

between $\mathbb{F}_2^n$ and $\mathbb{F}_{2^n}$. For such a function, the algebraic and univariate degrees are different quantities that are related as follows [Cha13, Page 254]:

$$\deg^a F = \max\{\mathrm{wt}(i) : i \in \mathbb{N}, u_i \neq 0\} \ ,$$

where $\{u_i, i \geq 0\}$ is the set of all coefficients in the univariate representation of F.

Our Target. In this paper, we focus on the block cipher MiMC, introduced by Albrecht *et al.* [AGR+16], which operates on $\mathbb{F}_{2^n}$. It consists of r iterations of an extremely simple round function: round i, $0 \leq i < r$, corresponds to $x \mapsto x^d + c_{i+1}$, where d is coprime with $(2^n - 1)$ in order to ensure that the round function is bijective, and where $c = (c_1, \ldots, c_r)$ is a sequence of r round constants. As a consequence, the round function of a MiMC instance is fully specified by the exponent d and by the sequence c of all round constants, and we denote such a MiMC instance $\mathsf{MIMC}_{d,c}[r]$. It is worth noting that the key is omitted in this description: indeed, as far as the algebraic degree is concerned, it can be considered to be part of the round constants.

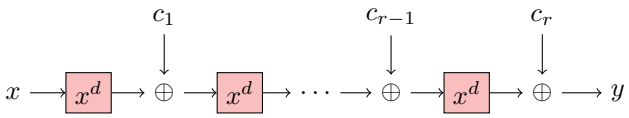

Figure 1: $\mathsf{MIMC}_{d,c}$ with r rounds.

More precisely, our aim is to investigate the security of MiMC against integral attacks, and thus its algebraic degree. We denote $(B_d^r)_{r\geq 1}$ the sequence of the maximal degrees of r rounds of MIMC_d, i.e., for any $r \geq 1$, B_d^r is the degree of $\mathsf{MIMC}_{d,c}[r]$ for at least one sequence $c = (c_1, \ldots, c_r)$ of constants:

$$B_d^r := \max_c \deg^a \mathsf{MIMC}_{d,c}[r] \ .$$

Note that, without loss of generality, we can assume that $c_r = 0$. On the other hand, it may happen that this degree is not reached for some specific sets of round constants as we will point out in Section 5.1, hence the need to take the maximum of them. Our goal is then to find the exact value of B_d^r. Indeed, a (very expensive) attack on MIMC_3 has been exhibited in [EGL+20], exploiting the fact that the number of rounds proposed by the designers is not sufficient for achieving a maximal algebraic degree. However, this weakness is based on a simple upper-bound on B_d^r and any gap between this bound and the exact value of the degree would decrease the complexity of the attack (or increase the number of rounds covered for a given complexity). Our aim is therefore to determine the exact value of B_d^r, or equivalently the minimal complexity of any attack based on higher-order differentials sych as those in [EGL+20].

A First Observation. A pattern of particular interest to us is what we call a *plateau*. To understand what it corresponds to, let us first consider a simple example. For any input x, the output of the composition of the first two rounds is

$$(x^3 + c_1)^3 + c_2 = x^9 + c_1 x^6 + c_1^2 x^3 + c_1^3 + c_2 \ . \tag{1}$$

We deduce that the composition of these two rounds is quadratic as its algebraic degree is equal to $\max\{\mathrm{wt}(i),\ i \in \{0, 3, 6, 9\}\}$, which is equal to 2. It is counter-intuitive: we would expect the algebraic degree to increase when a non-affine function is iterated. Such an event is what we call a *plateau*.

Definition 1.3 (Plateau)**.** *We say that there is a* plateau *whenever* $B_d^r = B_d^{r-1}$.

Since $(B_d^r)_{r\geq 1}$ is a non-decreasing sequence as proved later in Prop. 2.4, the existence and the frequency of plateaus are the most relevant elements when estimating the degree of MIMC_d after a large number of rounds.

Outline. Our work aims at a better understanding of these plateaus, first to identify them, and then to exploit them. In Section 2, we derive a simple method to generate the set of all exponents appearing in the univariate representation of $\mathsf{MIMC}_{3,c}[r]$ (Proposition 2.2). Then, we first bound the algebraic degree of $\mathsf{MIMC}_{3,c}[r]$ in Section 3, and identify in Section 4 a sequence of exponents that reach the upper bound.

We then perform a similar analysis of two ciphers closely related to $\mathsf{MIMC}_{3,c}[r]$, namely its inverse and $\mathsf{MIMC}_{9,c}[r]$ (Section 5). Finally, in Section 6, we use our results on the algebraic degree of $\mathsf{MIMC}_{3,c}[r]$ to slightly improve the results presented in [EGL+20] and to identify the best possible attacks exploiting the degree of the cipher. We then provide some guarantees on the lowest possible complexity for any integral attack based on the same methods as in [EGL+20].

2 Quantifying the Evolution of the Univariate Degree

In this section, we identify a process that generates the set of all the exponents appearing in the univariate form of r rounds of MIMC_d (Proposition 2.2). We then discuss several direct consequences of this observation.

2.1 Main Proposition

Recall that $\mathsf{MIMC}_{d,c}$ corresponds to the composition $F_{r-1} \circ \ldots \circ F_0$ where for any i, $0 \leq i < r$, $F_i : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$, $x \mapsto x^d \oplus c_{i+1}$, and the $c_{i+1} \in \mathbb{F}_{2^n}$ are arbitrary constants. Then, for the successive values of r, it is possible to recursively determine the list of monomials appearing in the univariate polynomial representing $\mathsf{MIMC}_{d,c}[r]$ for some c.

The following notation will be extensively used in the paper.

Definition 2.1 (Covering)**.** *For two elements x and y in $\mathbb{F}_2^n$, we say that y* is covered by *x and we write $y \preceq x$ if $y_i \leq x_i$ for all i.*

Similarly, for two integers i and j, $j \preceq i$ if the 2-adic expansion of j is covered by the 2-adic expansion of i.

Proposition 2.2. *Let n and $d < 2^n - 1$ be two integers such that $\gcd(d, 2^n - 1) = 1$. Let $\mathcal{E}_r$ be the set of exponents of the monomials appearing in the univariate polynomial $\mathsf{MIMC}_{d,c}[r]$ over $\mathbb{F}_{2^n}$ for at least one sequence c. Then, we have:*

$$\mathcal{E}_r = \{dj \bmod (2^n - 1) \text{ where } j \preceq i,\ i \in \mathcal{E}_{r-1}\} .$$

Proof. If the univariate form of $\mathsf{MIMC}_{d,c}[r-1]$ with $c = (c_2, \ldots, c_r)$ is given by

$$\mathcal{P}_{r-1}(x) = \sum_{i \in \mathcal{E}_{r-1}} \alpha_i x^i ,$$

then the univariate form of $\mathsf{MIMC}_{d,\hat{c}}[r]$ with $\hat{c} = (c_1, \ldots, c_{r-1}, c_r)$ is

$$\mathcal{P}_r(x) = \mathcal{P}_{r-1}(x^d + c_1) = \sum_{i \in \mathcal{E}_{r-1}} \alpha_i (x^d + c_1)^i .$$

But

$$(x^d + c_1)^i = \prod_{\ell \in I_i} (x^d + c_1)^{2^\ell} = \prod_{\ell \in I_i} (x^{d2^\ell} + c_1^{2^\ell})$$

where I_i corresponds to the support of the 2-adic expansion of i. Thus, the terms obtained after expansion have the following form:

$$\alpha_i \left(c_1^{\sum_{\ell \in I_i \setminus J_i} 2^\ell} \right) \left(x^{d \sum_{\ell \in J_i} 2^\ell} \right) \quad \text{where} \quad J_i \subseteq I_i \ .$$

It follows that the monomials that may appear in $\mathcal{P}_r$ are of the form $x^{dt \bmod (2^n-1)}$ with $t \preceq i$, and that the corresponding coefficient is equal to

$$p_t = \sum_{i \in E_t} \alpha_i c_1^{t \oplus i} \text{ where } E_t = \{i \in \mathcal{E}_{r-1} : t \preceq i\} \ .$$

By definition of $\mathcal{E}_{r-1}$, there exists at least one constant c such that $\alpha_i \neq 0$ for some exponent i in E_t. Then, p_t is a nonzero polynomial in c_1 and cannot vanish for all $c_1 \in \mathbb{F}_{2^n}$, implying that $x^{dt \bmod (2^n-1)}$ appears in $\mathcal{P}_r$ for at least one sequence of constants $\hat{c} = (c_1, c)$. It follows that the set of all exponents after r rounds is:

$$\mathcal{E}_r = \{(dj) \bmod (2^n - 1) \text{ where } j \preceq i, i \in \mathcal{E}_{r-1}\} \ .$$

□

The maximum algebraic degree after r rounds, B_d^r, is then the maximal weight of the elements in $\mathcal{E}_r$.

2.2 Reinterpreting Proposition 2.2

At round 0, we always have $\mathcal{E}_1 = \{0, d\}$. We can then apply Proposition 2.2 recursively to construct $\mathcal{E}_r$ from $\mathcal{E}_{r-1}$. In practice, this process revolves around two operations defined for any set of integers.

- The first multiplies each element of the input set by d modulo $(2^n - 1)$:

$$\mathsf{Mult}_d : \begin{cases} \mathbb{N}^{\mathbb{N}} & \to \mathbb{N}^{\mathbb{N}} \\ \{j_0, ..., j_{\ell-1}\} & \mapsto \{(dj_0) \bmod (2^n-1), ..., (dj_{\ell-1}) \bmod (2^n-1)\} \end{cases} ,$$

- the second returns the set of elements covered by elements in the input:

$$\mathsf{Cover} : \begin{cases} \mathbb{N}^{\mathbb{N}} & \to \mathbb{N}^{\mathbb{N}} \\ \{j_0, ..., j_{\ell-1}\} & \mapsto \{k \preceq j_i, i \in \{0, ..., \ell-1\}\} \end{cases} .$$

Using these operations, Proposition 2.2 can be re-written

$$\mathcal{E}_r = \mathsf{Mult}_d\big(\mathsf{Cover}(\mathcal{E}_{r-1})\big) \ . \tag{2}$$

Each element in $\mathcal{E}_r$ can be seen like a child of all the elements in $(\mathsf{Mult}_d \circ \mathsf{Cover})(\{j\})$ for some $j \in \mathcal{E}_{r-1}$. This view is summarized in Figure 2. While each element in $\mathcal{E}_r$ has at least one parent in $\mathcal{E}_{r-1}$, this parent might not be unique.

Only Mult_d depends on the exponent of the round function. It trivially satisfies the following relation:

$$\mathsf{Mult}_e\big(\mathsf{Mult}_d(\mathcal{E})\big) = \mathsf{Mult}_{ed}(\mathcal{E}) \ . \tag{3}$$

It is also such that the cardinality of the output is the same as the cardinality of the input.

A simple but useful observation is that the input $\mathcal{E}$ of the cover operation is contained in its output:

$$\mathcal{E} \ \subseteq \ \mathsf{Cover}(\mathcal{E}) \ .$$

By combining this relation with Equation (2), a trivial induction using Equation (3) yields

$$\mathsf{Mult}_{d^\ell}(\mathcal{E}_{r-\ell}) \subseteq \mathcal{E}_r \ . \tag{4}$$

The simplicity of these two operations implies the following lemma.

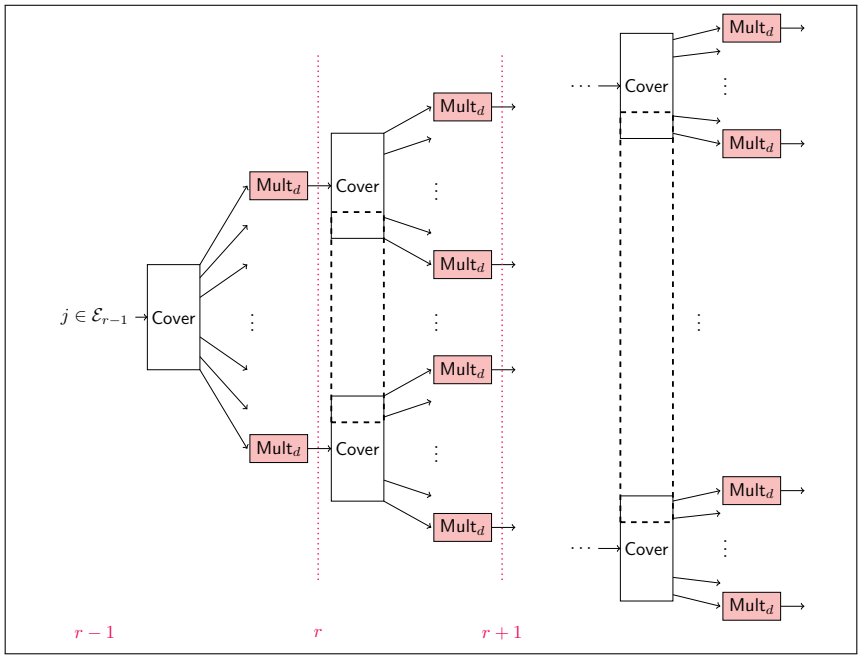

Figure 2: Getting next-round exponents.

Lemma 2.3. *The operations* Mult_2 *and* Cover *commute, i.e. for any set* $\mathcal{E}$ *of integers, we have*

$$\mathsf{Mult}_2\big(\mathsf{Cover}(\mathcal{E})\big) = \mathsf{Cover}\big(\mathsf{Mult}_2(\mathcal{E})\big) \ .$$

Proposition 2.4. *For any integer* d, $(B_d^r)_{r\geq 1}$ *is a non-decreasing sequence. Moreover, when* d *is odd, we have*

$$\mathcal{E}_{r-1} \subseteq \mathcal{E}_r, \ \forall r \geq 1 \ .$$

Proof. Let $d = \sum_{i=0}^{\mathrm{wt}(d)-1} 2^{\ell_i}$, with $0 \leq \ell_0 < \ldots < \ell_{\mathrm{wt}(d)-1}$. We will first prove by induction on r, that $\mathsf{Mult}_{2^{\ell_0}}(\mathcal{E}_r) \subseteq \mathcal{E}_{r+1}$.
It holds for $r = 1$ since $\mathcal{E}_1 = \{0, d\}$, so $\mathsf{Mult}_{2^{\ell_0}}(\mathcal{E}_1) = \{0, 2^{\ell_0} d\}$. In particular, $2^{\ell_0} \in \mathsf{Cover}(\mathcal{E}_1)$ so that:

$$\mathsf{Mult}_{2^{\ell_0}}(\mathcal{E}_1) \subseteq \mathsf{Mult}_d(\mathsf{Cover}(\mathcal{E}_1)) = \mathcal{E}_2 \ .$$

Then, let us assume that the property holds for $\mathcal{E}_r$.

$$\begin{aligned}
\mathsf{Mult}_{2^{\ell_0}}(\mathcal{E}_{r+1}) &= \mathsf{Mult}_{2^{\ell_0}}(\mathsf{Mult}_d(\mathsf{Cover}(\mathcal{E}_r))) && \text{by Equation (2),}\\
&\subseteq \mathsf{Mult}_d(\mathsf{Mult}_{2^{\ell_0}}(\mathsf{Cover}(\mathcal{E}_r))) && \text{by Equation (3),}\\
&\subseteq \mathsf{Mult}_d(\mathsf{Cover}(\mathsf{Mult}_{2^{\ell_0}}(\mathcal{E}_r))) && \text{by Lemma 2.3,}\\
&\subseteq \mathsf{Mult}_d(\mathsf{Cover}(\mathcal{E}_{r+1})) && \text{by induction hypothesis,}\\
&= \mathcal{E}_{r+2} && \text{by Equation (2).}
\end{aligned}$$

Finally, the result follows by observing that

$$\mathrm{wt}(2^{\ell_0} i \bmod (2^n - 1)) = \mathrm{wt}(i) \ .$$

In particular, if d is odd, $\ell_0 = 0$, which implies that $\mathcal{E}_{r-1} \subseteq \mathcal{E}_r$. □

Our main goal in this work is to estimate the algebraic degree of multiple iterations of MIMC_d. As a consequence, our focus is on the Hamming weight of the exponents.

Because of Lemma 2.3, we can reduce the size of $\mathcal{E}_r$ at each iteration by keeping only one representative per cyclotomic class. In other words, if $2^i x$ appears in $\mathcal{E}_r$, we can replace it with x without loosing information about the algebraic degree of the block cipher. More interestingly, if x is already in $\mathcal{E}_r$, it means we can simply remove $2^i x$ from it. In practice, this significantly simplifies the computations.

2.3 Some Simple Applications

It is possible to use Proposition 2.2 for $d = 3$ to determine the exponents in the univariate representation of two rounds of MiMC, as in (1). Using that $\mathcal{E}_1 = \{0, 3\}$, we have:

$$\begin{aligned}\mathcal{E}_2 &= \mathsf{Mult}_3\left(\mathsf{Cover}(\{0,3\})\right)\\ &= \mathsf{Mult}_3\left(\{0,1,2,3\}\right)\\ &= \{0,3,6,9\}\ .\end{aligned}$$

In fact, we can prove that there will always be such a plateau between the first and second rounds for all d of the form $d = 2^k - 1$ for some k.

Proposition 2.5. *Let $F : x \mapsto x^d$ be a permutation of $\mathbb{F}_{2^n}$ where $d = 2^k - 1$, and $\gcd(k, n) = 1$. Then, if $d^2 < 2^n - 1$, we have:*

$$\deg^a((x^d + c)^d) = \deg^a(F)\ ,$$

where c is an arbitrary constant.

Proof. Since $d = 2^k - 1$, it holds that $\mathsf{Cover}(\{d\}) = \{0, 1, ..., d\}$. As a consequence, we have that

$$\mathcal{E}_2 = \mathsf{Mult}_d(\{0, 1, ..., d\})\ .$$

In order to derive the result, it is sufficient to show that $\mathrm{wt}(dj) = \mathrm{wt}(d) = k$ for any integer $1 \leq j \leq d$. To show this, let $j \in \{0, ..., d\}$ be such that $j = \sum_{\ell=0}^{k-1} b_\ell 2^\ell$, where $b_\ell \in \{0, 1\}$ for all ℓ. We can thus write:

$$jd \ = \ (2^k - 1)\sum_{\ell=0}^{k-1} b_\ell 2^\ell \ = \ \sum_{\ell=0}^{k-1} b_\ell 2^{k+\ell} - \sum_{\ell=0}^{k-1} b_\ell 2^\ell$$

Using that $d = \sum_{\ell=0}^{k-1} 2^\ell$, we can write

$$d - \sum_{\ell=0}^{k-1} b_\ell 2^\ell \ = \ \sum_{\ell=0}^{k-1}(1 - b_\ell)2^\ell$$

from which we deduce that

$$\begin{aligned}\underbrace{(j+1)}_{j'} d &= \sum_{\ell=0}^{k-1} b_\ell 2^{k+\ell} + d - \sum_{\ell=0}^{k-1} b_\ell 2^\ell\\ &= \underbrace{\sum_{\ell=0}^{k-1} b_\ell 2^{k+\ell}}_{\mathrm{wt}=\mathrm{wt}(j)} + \underbrace{\sum_{\ell=0}^{k-1}(1 - b_\ell)2^\ell}_{\mathrm{wt}=k-\mathrm{wt}(j)}\ .\end{aligned}$$

As a consequence, the weight of dj' for any $j' \in \{1, ..., d\}$ is equal to k.

□

When $d = 3$, there exists another plateau during the first four rounds. Indeed, by using Proposition 2.2 again, we have:

$$\begin{aligned}\mathcal{E}_3 &= \mathsf{Mult}_3\left(\mathsf{Cover}(\mathcal{E}_2)\right)\\ &= \{0, 3, 6, 9, 12, 18, 24, 27\}\ ,\end{aligned}$$

which implies that the algebraic degree at the third round is $\mathrm{wt}(27) = 4$.

Using that

$$\mathcal{E}_4 = \mathsf{Mult}_3\left(\mathsf{Cover}(\mathcal{E}_3)\right)\ ,$$

we deduce that the maximum-weight exponents in $\mathcal{E}_4$ are

$$\{27, 30, 51, 54, 57, 75, 78\}\ ,$$

so that the algebraic degree is also 4 after the fourth round.

Therefore, there are two plateaus in the growth of the degree during the first four rounds, and actually, some other ones can be observed in the following rounds. Figure 3 shows the degree established using Proposition 2.2.

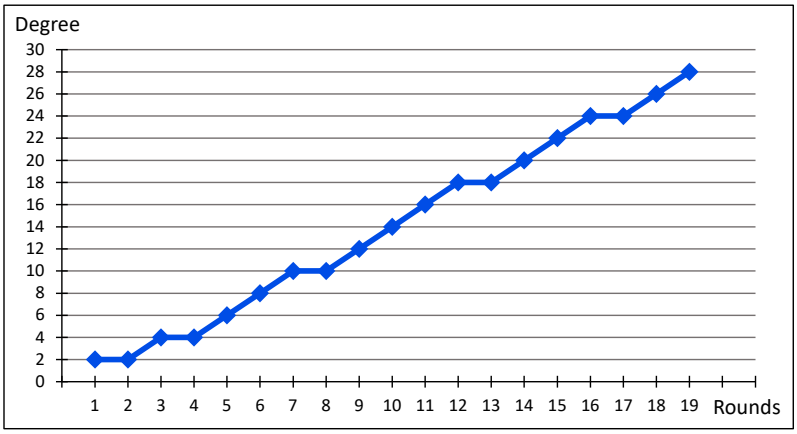

Figure 3: Algebraic degree of MIMC_3.

3 Bounding the Algebraic Degree of MIMC_3

We now mainly focus on the algebraic degree of MIMC_3 over $\mathbb{F}_{2^n}$, i.e., on the value of B_3^r. Obviously, as long as the degree of the univariate polynomial does not exceed $2^n - 1$, the algebraic degree of r rounds of MIMC_3 is upper-bounded by $\lceil \log_2(3^r) \rceil = \lceil r \log_2 3 \rceil$. But this trivial bound, used in [EGL+20] to set up integral attacks, can be easily improved by showing that the elements in $\mathcal{E}_r$ satisfy some particular properties.

3.1 Missing Exponents

Lemma 3.1. *Let $\mathcal{E}_r$ be the set of exponents in the univariate form of $\mathsf{MIMC}_3[r]$, as defined in Prop. 2.2. Then, any $i \in \mathcal{E}_r$ satisfies*

$$i \not\equiv 5, 7 \bmod 8\ .$$

Proof. We prove the result by induction on r. It holds at round 3, since $\mathcal{E}_3 = \{3k, k \in [\![0, 9]\!]\} \setminus \{15, 21\}$.

Let us now assume that the property holds for $\mathcal{E}_r$, i.e., any $i \in \mathcal{E}_r$ satisfies $i \not\equiv 5, 7 \bmod 8$. It follows that, for any $j \in \mathsf{Cover}(\mathcal{E}_r)$, we have $j \not\equiv 5, 7 \bmod 8$. Any element i in $\mathcal{E}_{r+1}$ is given by $i = 3j$ with $j \in \mathsf{Cover}(\mathcal{E}_r)$. But, if $j \bmod 8 \in \{0, 1, 2, 3, 4, 6\}$, then we necessarily have $3j \bmod 8 \in \{0, 1, 2, 3, 4, 6\}$. It follows that any $i \in \mathcal{E}_{r+1}$ is such that $i \not\equiv 5, 7 \bmod 8$. □

This lemma implies that the degree of $\mathsf{MIMC}_3[r]$ cannot exceed

$$k_r := \lfloor r \log_2 3 \rfloor$$

since $\mathcal{E}_r \subseteq \{i : i \leq 3^r\} \subseteq \{i : i < 2^{k_r+1}\}$. Indeed, the only integer $i < 2^{k_r+1}$ of weight strictly greater than k_r is $2^{k_r+1} - 1$, which does not belong to $\mathcal{E}_r$ since it satisfies $2^{k_r+1} - 1 \equiv 7 \bmod 8$.

3.2 An Upper-Bound on the Degree

We now exhibit a more accurate upper-bound on the degree, which makes use of the following result.

Lemma 3.2. *[Her36] The equation $2^x - 3^y = 5$ admits only two solutions $(x, y) = (3, 1)$ and $(5, 3)$.*

Proposition 3.3. *For all $r > 4$, we have*

$$2^{k_r+1} - 5 > 3^r,$$

where $k_r = \lfloor r \log_2 3 \rfloor$.

Proof. The proof depends on the parity of k_r.

- When $k_r = 2k + 1$, it is enough to show that

  $$3^r \notin \{2^{2k+2} - 5, 2^{2k+2} - 4, 2^{2k+2} - 3, 2^{2k+2} - 2, 2^{2k+2} - 1\},$$

  since $3^r < 2^{2k+2}$ by definition of k_r. Moreover, $2^{2k+2} - 5, 2^{2k+2} - 3, 2^{2k+2} - 2$ are not divisible by 3, and $3^r \neq 2^{2k+2} - 4$ because 3^r is not a multiple of 4. Finally, $3^r \neq 2^{2k+2} - 1$ because $2^{2k+2} - 1 \equiv 7 \bmod 8$, which is impossible for a power of 3. So $3^r < 2^{2k+2} - 5$.

- When $k_r = 2k$, we first prove that $3^r \notin \{2^{2k+1} - 4, 2^{2k+1} - 3, 2^{2k+1} - 2, 2^{2k+1} - 1\}$. Indeed, $2^{2k+1} - 4, 2^{2k+1} - 3, 2^{2k+1} - 1$ are not divisible by 3, and $3^r \neq 2^{2k+1} - 2$ because 3^r is odd. Now, according to Lemma 3.2, the equation $3^r = 2^{2k+1} - 5$ has no solution for $r > 4$.

□

Proposition 3.4. *For any $r \geq 4$, the algebraic degree after r rounds of MIMC_3 satisfies*

$$B_3^r \leq 2 \times \lceil k_r/2 - 1 \rceil,$$

where $k_r = \lfloor r \log_2 3 \rfloor$.

Proof. We first show that the algebraic degree at round r is at most $k_r - 1$.
The degree cannot be k_r because all exponents of the form $2^{k_r+1} - 2^j - 1$ with $0 \leq j \leq k_r$ are either non-divisible by 3 or missing. Indeed, we know from Lemma 3.1 that, when $j \notin \{0, 2\}$, exponents $2^{k_r+1} - 2^j - 1$ are missing since $2^{k_r+1} - 2^j - 1 \bmod 8 \in \{5, 7\}$. And for $j \in \{0, 2\}$, we derive from Prop. 3.3 that

$$2^{k_r+1} - 2^j - 1 \geq 2^{k_r+1} - 5 > 3^r \ ,$$

implying that this exponent does not belong to $\mathcal{E}_r$.

Now, we prove that, when k_r is even, the degree cannot be $k_r - 1$. The only possible exponents of weight $k_r - 1$ are of the form

$$2^{k_r+1} - 2^j - 2^i - 1, \text{ with } 0 \leq i < j \leq k_r \ .$$

All such exponents are equal to 5 or 7 modulo 8 unless i or j belongs to $\{0, 2\}$. The only exponents of weight $(k_r - 1)$ which may appear in $\mathcal{E}_r$ are then of the form $(2^{k_r+1} - 2^\ell - 2)$ or $(2^{k_r+1} - 2^\ell - 5)$. But, when k_r is even, $2^{k_r+1} - 2$ and $2^{k_r+1} - 5$ are divisible by 3. It follows that neither $2^{k_r+1} - 2^\ell - 2$ nor $2^{k_r+1} - 2^\ell - 5$ can be divisible by 3.

The result then follows by observing that

$$2 \times \lceil k_r/2 - 1 \rceil = \begin{cases} k_r - 1 & \text{if } k_r \equiv 1 \bmod 2 \\ k_r - 2 & \text{if } k_r \equiv 0 \bmod 2. \end{cases}$$

□

Besides the previous upper bound, a trivial lower bound can also be exhibited. Indeed, if the univariate degree 3^r is lower than $2^n - 1$, then the monomial x^{3^r} appears in the polynomial and its coefficient is always 1, independently of the choice of the constants and therefore never vanishes. This obviously defines a trivial lower bound for B_3^r:

$$\mathrm{wt}(3^r) \leq B_3^r \leq 2 \times \lceil k_r/2 - 1 \rceil \ .$$

Figure 4 compares the observed degree with these two bounds, in the particular case where the degree of extension is $n \geq 31$. We then notice that the observed degree seems to coincide with the upper bound.

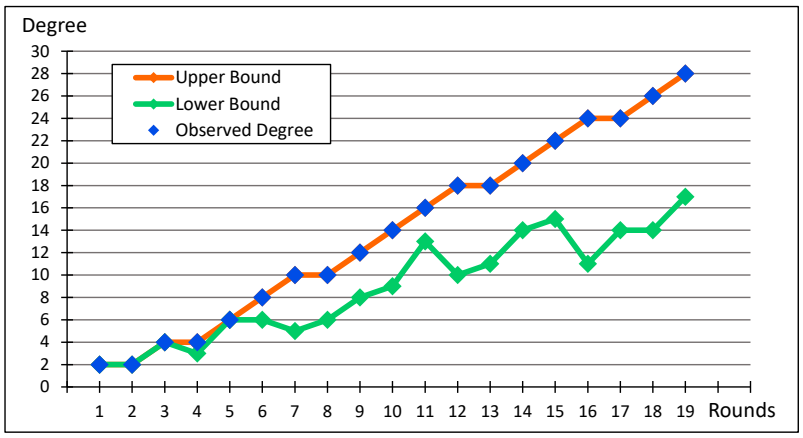

Figure 4: Comparison between the observed degree and the bounds (for $n \geq 31$).

4 Exact Degree of MIMC$_3$

While an upper bound on the algebraic degree enables an attacker to exhibit some higher-order integral attacks as in [EGL+20], it does not provide the designers with any guarantee that such attacks cannot be significantly improved. In the case of MIMC$_3$, the gap between our upper bound and the trivial lower bound raises concerns about the complexity of most efficient higher-order differential attacks that could be mounted. This issue is addressed in this section, where we show that, for all but a few round-reduced

versions of MIMC_3, the upper bound exhibited in Prop. 3.4 coincides with the exact value of B_3^r. More precisely, our approach consists in investigating Conjecture 4.1, which exhibits an exponent in the univariate polynomial representing MIMC_3 whose weight equals the upper bound.

In what follows, we let $(k_r)_{r>0}$ and $(b_r)_{r>0}$ be two sequences defined by

$$k_r = \lfloor r \log_2 3 \rfloor \quad \text{and} \quad b_r = k_r \bmod 2 \ .$$

Conjecture 4.1. *Let $(\omega_r)_{r>0}$ be the sequence of integers defined by*

$$\omega_r = 2^{k_r} - \alpha_{b_r}, \quad \textit{where} \quad \alpha_{b_r} = \begin{cases} 7 & \textit{if } b_r = 0 \\ 5 & \textit{if } b_r = 1 \ . \end{cases}$$

Then, for all $r > 0$, it holds that $\omega_r \in \mathcal{E}_r$.

While the most general case remains a conjecture at the time of writing, we show in this section that the conjecture is true for all $r < 16265$, except for a few sporadic cases for which a proof remains out of reach.

Our proof of this theorem is divided in two parts which correspond to the subsections of this section.

- We exhibit an inductive procedure establishing that, for most values of r, $\omega_r \in \mathcal{E}_r$ using the fact that $\omega_{r-\ell} \in \mathcal{E}_{r-\ell}$ for some $\ell < r$.
- We describe a MILP-based computationnally intensive procedure for proving that $\omega_r \in \mathcal{E}_r$ for some sporadic values of r, corresponding to the cases which are not covered by the inductive procedure.
- These results and algorithms are then put together in order to prove Theorem 4.10.

4.1 Properties of $(b_r)_{r>0}$ and $(k_r)_{r>0}$

It can be shown that $(k_r)_{r>0}$ is determined by the sequence $(s_r)_{r>0}$ of the *switches* from one parity to another, i.e.

$$s_1 = 0 \quad \text{and} \quad s_r = b_r \oplus b_{r-1} \ .$$

Proposition 4.2. *For any $\ell \geq 1$, and any $r > \ell$, we have*

$$k_r - k_{r-\ell} = 2\ell - \sum_{i=0}^{\ell-1} s_{r-i} \in \{k_\ell, k_\ell + 1\} \ . \tag{5}$$

Moreover,

$$b_r - b_{r-\ell} = 2 - \bigoplus_{i=0}^{\ell-1} s_{r-i}$$

Proof. By definition, $k_r = \lfloor r \log_2 3 \rfloor$. As a consequence, since $\log_2 3 \approx 1.59$, the sequence k_r increases by 1 or 2 for each increment of r. If this increase is by 1, then the parities of k_r and k_{r-1} have to be different. Otherwise, they have to be identical. Equivalently,

$$k_r - k_{r-1} \ = \ 2 - s_r \ ,$$

from which we deduce

$$k_r - k_{r-\ell} \ = \ k_r - k_{r-1} + k_{r-1} - k_{r-2} ... k_{r-\ell+1} - k_{r-\ell} \ = \ 2\ell - \sum_{i=0}^{\ell-1} s_{r-i} \ .$$

At the same time, we also have $k_r - k_{r-\ell} = \lfloor r\log_2(3)\rfloor - \lfloor (r-\ell)\log_2(3)\rfloor$. Using that

$$\lfloor x-y\rfloor \le \lfloor x\rfloor - \lfloor y\rfloor \le \lfloor x-y\rfloor + 1\,,$$

we can write

$$\lfloor \ell\log_2(3)\rfloor \;\le\; 2\ell - \sum_{i=0}^{\ell-1} s_{r-i} \;\le\; \lfloor \ell\log_2(3)\rfloor + 1\,,$$

which implies that the number of switches, i.e. the Hamming weight of the subsequences $(s_{r-i})_{0\le i<\ell}$, can take two values only. □

We then deduce the following proposition.

Proposition 4.3. *Let $r \ge 3$. Then there exists $1 \le \ell < r$ such that*

$$k_r - k_{r-\ell} = k_\ell$$

if and only if $(s_1 \dots s_r)$ is not a palindrome, i.e. if there exists i, $0 \le i < r$ such that

$$s_{r-i} \ne s_{i+1}\,.$$

Proof. From (5), we have, for any $1 \le \ell < r$,

$$\begin{aligned} k_r - k_{r-\ell} &= 2\ell - \sum_{i=0}^{\ell-1} s_{r-i} \\ \text{and } k_\ell - k_1 &= 2\ell - 2 - \sum_{j=0}^{\ell-2} s_{\ell-j} = 2\ell - 2 - \sum_{i=2}^{\ell} s_i\,. \end{aligned}$$

It follows that

$$\begin{aligned} k_r - k_{r-\ell} - k_\ell &= -k_1 + 2 - \left(\sum_{i=0}^{\ell-1} s_{r-i} - \sum_{i=2}^{\ell} s_i\right) \\ &= 1 - \left(\sum_{i=0}^{\ell-1} (s_{r-i} - s_{i+1})\right) \end{aligned}$$

where the last equality comes from the fact that $s_1 = 0$. It follows that, if $(s_1 \dots s_r)$ is a palindrome, then all terms in the sum vanish and $k_r - k_{r-\ell} = k_\ell + 1$ for all $1 \le \ell < r$. Conversely, if $(s_1 \dots s_r)$ is not a palindrome and if ℓ denotes the smallest index such that $s_{r-\ell+1} \ne s_\ell$, we obtain that

$$k_r - k_{r-\ell} - k_\ell = 1 - (-1)^{s_\ell} \in \{0, 2\}\,,$$

by observing that $s_{r-\ell+1} - s_\ell = (-1)^{s_\ell}$ since it differs from 0. Using that $k_r - k_{r-\ell} - k_\ell \in \{0, 1\}$, we deduce that $s_\ell = 0$ and $k_r - k_{r-\ell} - k_\ell = 0$. □

Remark 1. The sequence formed by the values r such that $(s_1 \dots s_r)$ is a palindrome is a subset of

$$\mathfrak{D} = \{2, 3, 5, 7, 12, 17, 29, 41, 53, 94, 147, 200, 253, 306, 359, 665, 971\}\,,$$

which corresponds to the sequence of the first denominators of the semiconvergents[1] of $\log_2 3$.

[1] The "semiconvergents" of a real number x is the sequence $(p_i/q_i)_{i\ge 0}$ such that all p_i and q_i are positive integers, and such that the sequence $(|x - p_i/q_i|)_{i\ge 0}$ is strictly decreasing.

Remark 2. For small values of r, we have computed the set $\mathcal{L}_r = \{\ell,\ 1 \leq \ell < r,\ \text{s.t. } k_{r-\ell} = k_r - k_\ell\}$ involved in the previous proposition, and we have noticed the following property, which has been checked up to $r \leq 16265$, but which remains a conjecture in the general case.

Conjecture 4.4. *Let $r \geq 3$ be such that $(s_1 \dots s_r)$ is not a palindrome. Let $\mathcal{L}_r$ and $\mathcal{P}_r$ be the two sets defined as follows:*

$$\mathcal{L}_r = \{\ell,\ 1 \leq \ell < r,\ \textit{s.t. } k_{r-\ell} = k_r - k_\ell\}$$
$$\mathcal{P}_r = \{r_i < r \textit{ s.t. } (s_1 \dots s_{r_i}) \textit{ is a palindrome}\}\ .$$

Then $\min(\mathcal{L}_r) \in \mathcal{P}_r$ *and* $\max(\mathcal{P}_r) \in \mathcal{L}_r$.

4.2 Inductive Procedure

Our objective is now to prove Prop 4.8, in which we identify a process establishing that $\omega_r \in \mathcal{E}_r$ knowing that $\omega_{r-\ell} \in \mathcal{E}_{r-\ell}$ for $\ell \in \mathcal{L}_r$. This result is valid up to a certain value of r, i.e. $r \leq 16265$, and also excludes a few sporadic cases. These constraints originate from the following two observations, which may be valid in the general case, but remain open.

Observation 4.5. *Let $r \geq 4$ be such that $s_1 ... s_r$ is a palindrome. If $r \leq 665$, then*

$$3^r > 2^{k_r} + 2^r\ .$$

Corollary 4.6. *Let $(k_r)_{r>0}$ be the sequence defined by $k_r = \lfloor r \log_2 3 \rfloor$. If $4 \leq r \leq 16265$, then*

$$3^r > 2^{k_r} + 2^r\ .$$

Proof. We prove it by induction on r.

- **For $r = 4$:** we have $k_r = 6$, and $3^4 = 81 > 80 = 2^6 + 2^4$.
- **Induction step.** We suppose that $\forall i < r$

  $$3^i > 2^{k_i} + 2^i\ .$$

  If $(s_1, \dots, s_r)$ is a palindrome, $r \in \{7, 12, 53, 359, 665\}$, we know from the previous observation that the property holds. Otherwise, there exists $\ell \in \mathcal{L}_r$, implying that

  $$3^r = 3^{r-\ell}3^\ell > (2^{k_{r-\ell}} + 2^{r-\ell})(2^{k_\ell} + 2^\ell) = 2^{k_r} + 2^{k_{r-\ell}+\ell} + 2^{k_\ell+r-\ell} + 2^r > 2^{k_r} + 2^r\ .$$

□

We also need the following observation, on the representation of all elements in $\mathbb{Z}/3^t\mathbb{Z}$ as a sum of even powers of 2.

Observation 4.7. *Let $1 \leq t \leq 21$, then*

$$\forall x \in \mathbb{Z}/3^t\mathbb{Z},\ \exists \varepsilon_2, \dots, \varepsilon_{2t+2} \in \{0,1\},\ \textit{s.t. } x = \sum_{j=2}^{2t+2} \varepsilon_j 4^j \bmod 3^t\ .$$

However, we conjecture that this result holds in general for any value of t.
We now prove that, in most cases, the fact that the exponent

$$\omega_r = 2^{k_r} - \alpha_{b_r}, \quad \text{where } \alpha_{b_r} = \begin{cases} 7 & \text{if } b_r = 0 \\ 5 & \text{if } b_r = 1\,, \end{cases}$$

belongs to $\mathcal{E}_r$ can be derived from the fact that $\omega_{r-\ell} \in \mathcal{E}_{r-\ell}$.

Proposition 4.8. *Let $(k_r)_{r>0}$ be the sequence defined by $k_r = \lfloor r \log_2 3 \rfloor$, and $(b_r)_{r>0}$ the sequence defined by $b_r = k_r \bmod 2$. Let $r \geq 4$, and $\ell \in \mathcal{L}_r$ such that one of the following situation occurs:*

(1) *$\ell = 1$,*

(2) *$\ell = 2$,*

(3) *$2 < \ell \leq 22$ such that $k_r \geq k_\ell + 3\ell + b_r + 1$, and one of the following situation occurs:*

- *ℓ is even, or*
- *ℓ is odd, with $b_{r-\ell} = \overline{b_r}$;*

(4) *$2 < \ell \leq 22$ is odd such that $k_r \geq k_\ell + 3\ell + \overline{b_r} + 5$ and $b_{r-\ell} = b_r$.*

Then $\omega_{r-\ell} \in \mathcal{E}_{r-\ell}$ implies that $\omega_r \in \mathcal{E}_r$.

Proof. As a preliminary remark, we first observe that, for any even k,

$$(2^k - 1)/3 = \sum_{i=0}^{k/2-1} 2^{2i} ,$$

implying that

- for k even,
$$\frac{2^k - 7}{3} = \frac{2^k - 1}{3} - 2 = 3 + \sum_{i=2}^{k/2-1} 2^{2i}$$
- for k odd,
$$\frac{2^k - 5}{3} = \frac{2^k - 2}{3} - 1 = 1 + \sum_{i=1}^{(k-1)/2-1} 2^{2i+1} .$$

Therefore, for any $i \geq 3$,

$$\omega_i = \frac{2^{k_i} - \alpha_{b_i}}{3} = (8 - \alpha_{\overline{b_i}}) + \sum_{j=1+\overline{b_i}}^{\lfloor \frac{k_i}{2} \rfloor - 1} 2^{2j+b_i} . \tag{6}$$

The proof consists, for given r and ℓ, in exhibiting a sequence of exponents $(e_{r-\ell} \dots e_r)$ such that $e_r = \omega_r = 2^{k_r} - \alpha_{b_r}$ and each e_{r-i}, $0 < i \leq \ell$, belongs to $\mathcal{E}_{r-i}$. It is worth noting that proving that $e_{j+1} \in \mathcal{E}_{j+1}$ boils down to exhibiting some $e_j \in \mathcal{E}_j$ such that $(e_{j+1}/3) \preceq e_j$. Let us now investigate the different cases for ℓ.

(1) When $\ell = 1$, we have $k_{r-1} = k_r - 1$ and $b_{r-1} = \overline{b_r}$. By hypothesis, $\omega_{r-1} = 2^{k_r - 1} - \alpha_{\overline{b_r}}$ belongs to $\mathcal{E}_{r-1}$. Moreover, from (6), we deduce that

$$\frac{2^{k_r} - \alpha_{b_r}}{3} = (8 - \alpha_{\overline{b_r}}) + \sum_{j=1+\overline{b_r}}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} \preceq e_{r-1} ,$$

implying that $2^{k_r} - \alpha_{b_r}$ belongs to $\mathcal{E}_r$.

(2) When $\ell = 2$, we have $k_{r-2} = k_r - 3$ and $b_{r-2} = \overline{b_r}$. Therefore, by hypothesis, $\omega_{r-2} = 2^{k_r-3} - \alpha_{\overline{b_r}}$ belongs to $\mathcal{E}_{r-2}$. Let us choose

$$e_{r-1} = (8 - \alpha_{\overline{b_r}}) + \sum_{j=1+\overline{b_r}}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} + \sum_{j=2}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+\overline{b_r}} .$$

Then, we have

$$e_{r-1} = (8b_r + 8 - \alpha_{\overline{b_r}}) + \sum_{j=2}^{\lfloor \frac{k_r}{2} \rfloor - 1} (2^{2j} + 2^{2j+1}) .$$

We deduce that

$$e_{r-1}/3 = (8 - \alpha_{b_r}) + \sum_{j=2}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j} \preceq \omega_{r-2} ,$$

implying that e_{r-1} belongs to $\mathcal{E}_{r-1}$. Moreover,

$$(2^{k_r} - \alpha_{b_r})/3 = (8 - \alpha_{\overline{b_r}}) + \sum_{j=1+\overline{b_r}}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} \preceq e_{r-1} ,$$

implying that $\omega_r \in \mathcal{E}_r$.

(3) Let $\ell > 2$, such that $k_r \geq k_\ell + 3\ell + b_r + 1$. While the proposition considers two cases, we split the first one into two, so that we consider three cases:

- **(a)** ℓ is even, with $b_{r-\ell} = b_r$,
- **(b)** ℓ is even, with $b_{r-\ell} = \overline{b_r}$,
- **(c)** ℓ is odd, with $b_{r-\ell} = \overline{b_r}$.

By hypothesis, $\omega_{r-\ell} = 2^{k_r-k_\ell} - \alpha_{b_{r-\ell}}$ belongs to $\mathcal{E}_{r-\ell}$.

We now choose

$$e_{r-1} = (8 - \alpha_{\overline{b_r}}) + \sum_{j=1+\overline{b_r}}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} + S ,$$

with

$$S = \begin{cases} \displaystyle\sum_{j=1+b_r}^{2\ell-1+b_r} \varepsilon_j 2^{2j+\overline{b_r}} & \text{in Cases (a), (c)} \\ \displaystyle 2b_r + \sum_{j=1+b_r}^{2\ell-1+b_r} \varepsilon_j 2^{2j+\overline{b_r}} & \text{in Case (b)} \end{cases}$$

where the $(2\ell - 1)$ coefficients $\varepsilon_j \in \{0, 1\}$ are chosen such that $e_{r-1} \equiv 0 \bmod 3^{\ell-1}$. Indeed, it is known from Observation 4.7 that such a choice is always possible since $\ell \leq 22$.

We then use that

$$\begin{aligned} e_{r-1} &= \left((8 - \alpha_{\overline{b_r}}) + S + \sum_{j=1+\overline{b_r}}^{2\ell} 2^{2j+b_r} \right) + \sum_{j=2\ell+1}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} \\ &< 2^{4\ell+b_r+1} + \sum_{j=1}^{(k_r-\overline{b_r}-1)/2-2\ell-b_r} 2^{k_r-2j} . \end{aligned}$$

It follows that

$$\begin{aligned} 3e_{r-1} &< 3 \times \left(2^{4\ell+b_r+1} + \sum_{j=1}^{(k_r-\overline{b_r}-1)/2-2\ell-b_r} 2^{k_r-2j} \right) \\ &\leq 2^{4\ell+b_r+1} + 2^{k_r} \\ &\leq 2^{k_r-k_\ell+\ell} + 2^{k_r} \end{aligned}$$

where the last inequality comes from the hypothesis on ℓ. Moreover, since $\ell \leq 22$, we deduce from Observation 4.5 that

$$3e_{r-1} < 2^{k_r-k_\ell} \left(2^\ell + 2^{k_\ell}\right) \leq 2^{k_r-k_\ell} 3^\ell \ ,$$

which implies that $e_{r-1} < 3^{\ell-1} 2^{k_r-k_\ell}$.
We deduce that, for proving that $e_{r-1}/3^{\ell-1} \preceq \omega_{r-\ell}$, it is sufficient to show that this holds for their remainders modulo 8. This last result comes from the following facts, for each of the three cases:

- **(a)** $3^{\ell-1} \equiv 3 \bmod 8$ and $S \equiv 0 \bmod 8$, leading to $e_{r-1}/3^{\ell-1} \equiv (8 - \alpha_{b_r}) \bmod 8$
- **(b)** $3^{\ell-1} \equiv 3 \bmod 8$ and $S \equiv 2b_r \bmod 8$, leading to $e_{r-1}/3^{\ell-1} \equiv 1 \bmod 8$
- **(c)** $3^{\ell-1} \equiv 1 \bmod 8$ and $S \equiv 0 \bmod 8$, leading to $e_{r-1}/3^{\ell-1} \equiv (8 - \alpha_{\overline{b_r}}) \bmod 8$.

So, we obtain that

$$e_{r-1}/3^{\ell-1} \preceq \omega_{r-\ell} \ ,$$

implying that e_{r-1} belongs to $\mathcal{E}_{r-1}$. Moreover,

$$\frac{2^{k_r} - \alpha_{b_r}}{3} = (8 - \alpha_{\overline{b_r}}) + \sum_{j=1+\overline{b_r}}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} \preceq e_{r-1} \ ,$$

which proves that $\omega_r \in \mathcal{E}_r$.

(4) Let $\ell > 2$, such that $k_r \geq k_\ell + 3\ell + \overline{b_r} + 5$. When ℓ is odd and $b_{r-\ell} = b_r$, by hypothesis, $\omega_{r-\ell} = 2^{k_r-k_\ell} - \alpha_{b_r}$ belongs to $\mathcal{E}_{r-\ell}$. We now choose

$$e_{r-2} = (8-\alpha_{b_r}) + \sum_{j=1+b_r}^{\lfloor \frac{k_r}{2} \rfloor - 3\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor - 5 + b_r} 2^{2j+\overline{b_r}} + S + \sum_{j=1}^{\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor + 1} (2^{k_r-(6j-2)} + 2^{k_r-(6j-1)} + 2^{k_r-6j}) \ ,$$

with

$$S = 2\overline{b_r} + \sum_{j=1+\overline{b_r}}^{\lfloor \frac{k_r}{2} \rfloor - 3\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor - 5} \varepsilon_j 2^{2j+b_r}$$

where the ε_j are chosen such that $e_{r-2} \equiv 0 \bmod 3^{\ell-2}$, which is always possible from Obervation 4.7 since $\ell \leq 22$ and the number of coefficients ε_j in the sum is

$$\left\lfloor \frac{k_r}{2} \right\rfloor - 3\left\lfloor \frac{k_\ell - \ell}{6} \right\rfloor - 5 - \overline{b_r} \geq \frac{k_r - b_r - k_\ell + \ell}{2} - 5 - \overline{b_r} \geq \frac{4\ell - \overline{b_r} - b_r + 5}{2} - 5 \geq 2\ell - 3$$

Then, we have

$$\begin{aligned} e_{r-2} &< \sum_{j=1}^{\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor + 1} (2^{k_r-(6j-2)} + 2^{k_r-(6j-1)} + 2^{k_r-6j}) + 2^{k_r-6\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor - 8)} \\ &\leq \sum_{j=1}^{\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor + 1} (2^{k_r-(6j-2)} + 2^{k_r-(6j-1)} + 2^{k_r-6j}) + 2^{k_r-k_\ell+\ell-8} \ . \end{aligned}$$

It follows that

$$\begin{aligned}
9e_{r-2} &< 9 \times \left(\sum_{j=1}^{\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor+1} (2^{k_r-(6j-2)} + 2^{k_r-(6j-1)} + 2^{k_r-6j}) + 2^{k_r-k_\ell+\ell-8} \right) \\
&\leq (2^6-1) \sum_{j=1}^{\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor+1} 2^{k_r-6j} + 9 \times 2^{k_r-k_\ell+\ell-8} \\
&\leq 2^{k_r} - 2^{k_r-6\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor-5} + 9 \times 2^{k_r-k_\ell+\ell-8} \\
&< 2^{k_r} + 2^{k_r-k_\ell+\ell} \\
&\leq 2^{k_r-k_\ell} \left(2^{k_\ell} + 2^{\ell}\right) < 3^{\ell} 2^{k_r-k_\ell}
\end{aligned}$$

where the last inequality comes from Corollary 4.6 since $\ell \leq 22$. We then deduce that $e_{r-2} < 3^{\ell-2}2^{k_r-k_\ell}$.
Therefore, it is now sufficient to prove that $e_{r-2}/3^{\ell-2} \equiv 1 \bmod 8$ to order to prove that $e_{r-2}/3^{\ell-2} \preceq \omega_{r-\ell}$. This result on the remainders modulo 8 comes from the fact that $3^{\ell-2} \equiv 3 \bmod 8$ since ℓ is odd, and $S \equiv 2\overline{b_r} \bmod 8$, leading to $e_{r-2}/3^{\ell-2} \equiv 1 \bmod 8$.

So, we have

$$e_{r-2}/3^{\ell-2} \preceq \omega_{r-\ell} \ ,$$

implying that e_{r-2} belongs to $\mathcal{E}_{r-2}$.

Let now

$$e_{r-1} = (8 - \alpha_{\overline{b_r}}) + \sum_{j=1+\overline{b_r}}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} + \sum_{j=1+b_r}^{\lfloor \frac{k_r}{2} \rfloor - 3\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor - 5 + b_r} 2^{2j+\overline{b_r}} \ .$$

Then, we have

$$e_{r-1} = (8 - \alpha_{\overline{b_r}}) + b_r 2^3 + 3 \sum_{j=1+b_r}^{\lfloor \frac{k_r}{2} \rfloor - 3\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor - 5 + b_r} 2^{2j+\overline{b_r}} + \sum_{j=\lfloor \frac{k_r}{2} \rfloor - 3\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor - 3}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} \ .$$

We use that

$$\begin{aligned}
\sum_{j=\lfloor \frac{k_r}{2} \rfloor - 3\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor - 3}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} &= \sum_{i=1}^{3\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor + 3} 2^{k_r-2i} \\
&= \sum_{i=1}^{\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor + 1} 2^{k_r-6i} \left(1 + 2^2 + 2^4\right) \\
&= 3 \sum_{i=1}^{\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor + 1} 2^{k_r-6i} (1 + 2 + 2^2) \ .
\end{aligned}$$

Moreover, since

$$(8 - \alpha_{\overline{b_r}}) + b_r 2^3 = 3(8 - \alpha_{b_r}) \ ,$$

we obtain that

$$\frac{e_{r-1}}{3} = (8 - \alpha_{b_r}) + \sum_{j=1+b_r}^{\lfloor \frac{k_r}{2} \rfloor - 3\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor - 5 + b_r} 2^{2j+\overline{b_r}} + \sum_{i=1}^{\left\lfloor \frac{k_\ell-\ell}{6} \right\rfloor + 1} 2^{k_r-6i} (1 + 2 + 2^2) \preceq e_{r-2} \ ,$$

implying that e_{r-1} belongs to $\mathcal{E}_{r-1}$.

Finally,

$$\frac{2^{k_r} - \alpha_{b_r}}{3} = (8 - \alpha_{\overline{b_r}}) + \sum_{j=1+\overline{b_r}}^{\lfloor \frac{k_r}{2} \rfloor - 1} 2^{2j+b_r} \preceq e_{r-1} \ ,$$

implying that $\omega_r \in \mathcal{E}_r$.

□

4.3 MILP-based Algorithm

The induction procedure from Prop. 4.8 relies on some assumptions which are not satisfied for some values of r. These sporadic cases then need to be handled in a different (but more expensive) way.

According to Proposition 4.3, rounds r such that $(s_1 \dots s_r)$ is a palindrome are the only ones for which there is no $\ell < r$ such that $k_{r-\ell} = k_r - k_\ell$. Then, Prop. 4.8 does not apply to these values of r.

Moreover, in this proposition, we need $\ell \leq 22$, since Observation 4.7 has been proved up to $\ell \leq 22$ only. Also, there is an additional constraint on $k_r - k_\ell - 3\ell$. Let $1 \leq r \leq 16265$, such that we are not in a case of a palindromic sequence, then by computing the minimum values in $\mathcal{L}_r$, we can exhibit all the rounds for which there is no ℓ satisfying the two constraints of Prop. 4.8:

- If $r = 19, 24$, then $\min(\mathcal{L}_r)$ is respectively 7 or 12. However, the hypotheses of Prop. 4.8 are not satisfied:

$$\begin{aligned} k_{19} = 30 < 33 = k_7 + 3 \times 7 + b_{19} + 1, \\ k_{24} = 38 < 56 = k_{12} + 3 \times 12 + b_{24} + 1. \end{aligned}$$

- If r belongs to the set

$$\{665\lambda + 53\mu, 0 \leq \lambda \leq 24, 0 \leq \mu \leq 6\} \cup \{359 + 665\lambda + 53\mu, 0 \leq \lambda \leq 23, 0 \leq \mu \leq 5\} \ ,$$

  we have $\min(\mathcal{L}_r) \geq 53$.

Let us recall that, when the univariate degree 3^r is lower than $2^n - 1$, we necessarily have $3^r \in \mathcal{E}_r$. Consequently, we will search for ℓ such that $3^{r-\ell} \in \mathcal{E}_{r-\ell}$ implies that $\omega_r \in \mathcal{E}_r$. This can be done by exhibing a sequence of operations, composed of Cover and Mult_3, which generates ω_r from $3^{r-\ell}$. These functions a priori need to be iterated ℓ times but, since $x \in \mathsf{Cover}(x)$, it is possible to ignore some calls to Cover.

It is possible to encode the existence of such a sequence of operations as a MILP problem that is then solved using `PySCIPOpt` [GAB+20a, GAB+20b], an off-the-shelf solver. This encoding works as follows.

Integers are represented via their binary representation over n bits. We use two sets of intermediate variables for each round r, namely $(a_i^r)_{0 \leq i < n}$ and $(b_i^r)_{0 \leq i < n}$, corresponding to the integers a^r and b^r which are such that

$$b^r = \mathsf{Cover}(a^r) \text{ and } a^{r+1} = \mathsf{Mult}_3\,(b^r) \ .$$

The relation $b^r = \mathsf{Cover}(a^r)$ is easily encoded as a set of MILP equations since it corresponds to $b_i^r \leq a_i^r$ for all $i \in \{0, ..., n-1\}$. In order to ensure that $a^{r+1} = 3b^r$, we use a bitwise description of the multiplication by 3 that can be found for instance in [BFL+21]. By setting $a^{r-\ell} = 3^{r-\ell}$ and $a^r = \omega_r$, we have that the existence of a solution to all the previously

described equations is equivalent to the fact that a^r is in $(\mathsf{Mult}_3 \circ \mathsf{Cover})^\ell(\{3^{r-\ell}\})$, meaning that it is indeed in $\mathcal{E}_r$.

This technique is rather slow, and it cannot be applied to large values of r. Indeed, the experiments were ran on a cluster with Intel Xeon Gold 5218 processor and 192GB of RAM, for which we were limited to one week of computation. However, it plays a crucial role in our proof of Theorem 4.10.

Table 1 provides the values of all $r \leq 16265$ corresponding to the length of a palindromic sequence for which it has been checked with our MILP-based algorithm that $\omega_r \in \mathcal{E}_r$.

Table 1: Lengths r of palindromic sequences for which it has been proved with a MILP algorithm that $\omega_r \in \mathcal{E}_r$, using that $3^{r-\ell} \in \mathcal{E}_{r-\ell}$.

r	7	12	53	359	665
$2^{k_r} - \alpha_{b_r}$	$2^{11} - 5$	$2^{19} - 5$	$2^{84} - 7$	$2^{569} - 5$	$2^{1054} - 7$
ℓ	2	3	4	6	7

Similarly, Table 2 covers the first values of r for which there is no ℓ satisfying one of the situations of Prop. 4.8.

Table 2: ℓ such that $3^{r-\ell} \in \mathcal{E}_{r-\ell}$ implies $\omega_r \in \mathcal{E}_r$.

r	19	24	$53k$	$359 + 53k (k = 1, 3, 5)$
$2^{k_r} - \alpha_{b_r}$	$2^{30} - 7$	$2^{38} - 7$	$2^{k_r} - 7$	$2^{k_r} - 5$
ℓ	4	3	$5 (k = 3, 5)$ - $6 (k = 2, 4, 6)$	$9 (k = 1)$ - $7 (k = 3, 5)$

The first value of r for which the cost becomes too high to obtain a result from the solver is $r = 465$. Thus, if $R = \{665\lambda + 53\mu,\ 0 \leq \lambda \leq 23, 0 \leq \mu \leq 5\}$, then, up to 16265, the only rounds for which we cannot definitively prove the presence of maximum-weight exponents are the following ones (in red in Figure 5):

$$\mathcal{F} = \big((359 + R) \cup (665 + R) \cup (718 + R)\big) \backslash \mathcal{V}\,, \qquad \text{where } \mathcal{V} = \{359, 412, 518, 624, 665\}\ .$$

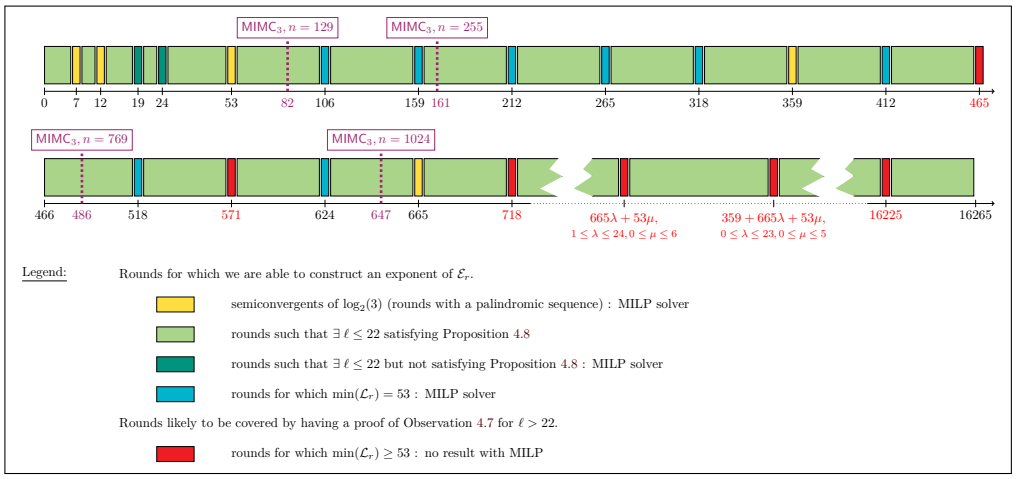

Figure 5: Rounds for which we are able to exhibit a maximum-weight exponent.

The last relevant point we need is to check that the rounds belonging to $\mathcal{F}$ do not raise any problem to build a recurrence on elements of $\boldsymbol{R} = \{4 \leq r \leq 16265 \text{ s.t. } r \notin \mathcal{F}\}$. The following observation has been checked by computer, by looping through all $\ell \in \mathcal{L}_r$.

Observation 4.9. *For any $r \in \boldsymbol{R}$, there exists $\ell \in \mathcal{L}_r$ such that $r - \ell$ belongs to $\boldsymbol{R}$.*

4.4 Combining Both Steps

As a consequence, we are now able to construct, by induction, maximum-weight exponents for all rounds until 464, and for almost all rounds until 16265.

Theorem 4.10. *Let $\boldsymbol{R}$ be the set $\{4, ..., 16265\}\backslash\mathcal{F}$, where $\mathcal{F} = \big((359 + R) \cup (665 + R) \cup (718 + R)\big)\backslash\mathcal{V}$ with*

$$R = \{665\lambda + 53\mu,\ 0 \leq \lambda \leq 23, 0 \leq \mu \leq 5\},$$
$$\mathcal{V} = \{359, 412, 518, 624, 665\}\ .$$

Then $\omega_r \in \mathcal{E}_r$ for all $r \in \boldsymbol{R}$.

Proof. (of Theorem 4.10) We prove the result by induction on r. Let $(\mathcal{H}_r)$ be the following hypothesis:

$$(\mathcal{H}_r) : \forall 4 \leq i < r, i \in \boldsymbol{R},\ \omega_i = 2^{k_i} - \alpha_{b_i} \in \mathcal{E}_i$$

- **For $r = 5$:**
$$(\mathcal{H}_5) : \forall 4 \leq i < 5, i \in \boldsymbol{R},\ \omega_i \in \mathcal{E}_i$$
is satisfied since:
$$2^{k_4} - \alpha_{b_4} = 2^6 - 7 = 57 \in \mathcal{E}_4\ .$$

- **Induction step.** We assume that $(\mathcal{H}_r)$ is satisfied, then we will show that $(\mathcal{H}_{r+1})$ is also satisfied. If $(s_1, \ldots, s_r)$ is a palindrome, or if there is no ℓ that satisfies the conditions of Prop 4.8 then $\omega_r \in \mathcal{E}_r$ as summarized in Table 1 and Table 2.
Otherwise, according to Proposition 4.3, we know that there exists $\ell \in \mathcal{L}_r$. Moreover, we know from Observation 4.9 that there is always a round $r - \ell \in \boldsymbol{R}$ so that we can use Prop. 4.8, and prove that we have $\omega_r \in \mathcal{E}_r$ since $\omega_{r-\ell} \in \mathcal{E}_{r-\ell}$.

□

Corollary 4.11. *Let $r \in \boldsymbol{R}$ be an integer, then the algebraic degree after r rounds of* MIMC_3 *satisfies:*

$$B_3^r = 2 \times \lceil k_r/2 - 1 \rceil\ ,$$

where $k_r = \lfloor r \log_2 3 \rfloor$.

Proof. Proposition 3.4 proves that $2 \times \lceil k_r/2 - 1 \rceil$ is an upper bound on the degree, and Theorem 4.10 exhibits some exponents that reach the degree at each round when $r \in \boldsymbol{R}$. □

Figure 6 compares the exact value of B_3^r given in Corollary 4.11, with the bound given in [EGL+20].

By observing that $\lceil k_r/2 - 1 \rceil \leq \lceil k_{r-1}/2 - 1 \rceil + 1$, we deduce that, between two consecutive rounds, the degree increases by 2 or remains stable, in which case we have a plateau. More precisely, there is a plateau in the growth of the algebraic degree between rounds r and $r + 1$, i.e. $B_3^r = B_3^{r+1}$, when k_r is odd and k_{r+1} is even. This implies that two consecutive plateaus correspond to 3 switches in the parity of $(k_r)_{r>0}$. From Prop. 4.2, we know that

$$\sum_{i=1}^{\ell} s_{r+i} \in \{2\ell - 1 - k_\ell, 2\ell - k_\ell\}\ . \tag{7}$$

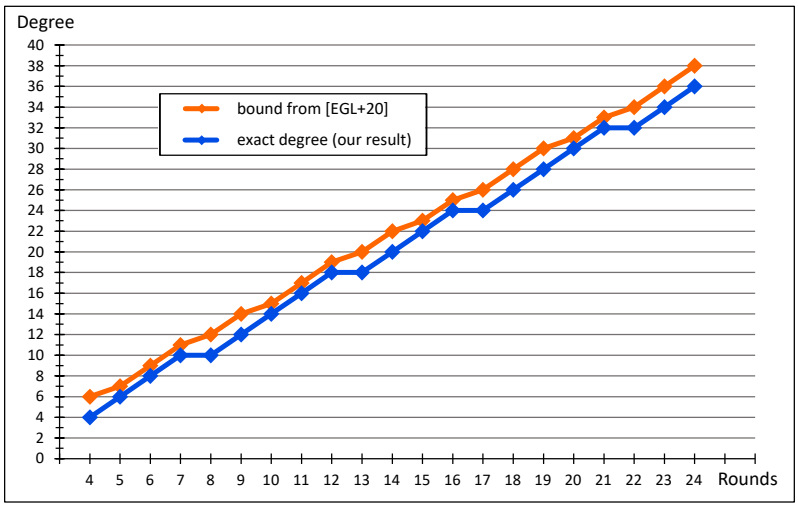

Figure 6: Comparison of the exact value of B_3^r given in Corollary 4.11 with previous work.

We deduce that $\sum_{i=1}^{4} s_{r+i} \leq 2$ and $\sum_{i=1}^{8} s_{r+i} \geq 4$, which implies that, if there is a plateau between rounds r and $(r+1)$, the next plateau starts at round $(r+4)$, $(r+5)$ or $(r+6)$. By using (7) for $2 \leq \ell \leq 7$, we deduce that there are exactly three possible patterns for a subsequence of $(s_r)_{r>0}$ starting by 1 and having Hamming weight 3, namely

$$\begin{aligned} s_{r+1} \dots s_{r+5} &= 10101 \\ s_{r+1} \dots s_{r+6} &= 101001 \\ s_{r+1} \dots s_{r+6} &= 100101 \end{aligned}$$

Although Corollary 4.11 does not allow to cover all the rounds, the number of rounds of MIMC_3 we are interested in is largely covered ($\simeq 80$), as shown in Figure 5. For the hash functions, as we will see in Subsection 6.2, we need to cover 486 and 687 rounds. In these cases, we have the exact value of B_3^r for all rounds needed, except for $r \in \{465, 571\}$. Recalling that $(B_3^r)_{r\geq 1}$ is a non decreasing sequence (Prop. 2.4) and that B_3^r is upper bounded by $2 \times \lceil k_r/2 - 1 \rceil$ (Prop. 3.4), we have:

$$\begin{aligned} 734 = B_3^{464} \leq B_3^{465} \leq 736 \ , \\ 902 = B_3^{570} \leq B_3^{571} \leq 904 \ . \end{aligned}$$

5 Generalization to Other Permutations

In this section, we discuss the algebraic degree of $\mathsf{MIMC}_{9,c}[r]$ (Sec. 5.1) and more generally, of $\mathsf{MIMC}_{d,c}[r]$, with $d = 2^j + 1$ (Sec. 5.2). Interestingly, this analysis also enlightens the influence of the choice of the round constants c on the degree of $\mathsf{MIMC}_{3,c}[r]$. Sec. 5.3 focuses on the decryption function and studies the degree of the inverse $\mathsf{MIMC}_{3,c}[r]^{-1}$.

5.1 Degree of MIMC_9 and Influence of the Round Constants on the Degree of MIMC_3

The value of B_3^r determined in the previous sections corresponds to the maximal algebraic degree of $\mathsf{MIMC}_{3,c}[r]$ over all possible choices for the round constants c. However,

as shown in Prop. 2.2, the exponents in $\mathcal{E}_r$ are of the form dj, $j \preceq i$ with $i \in \mathcal{E}_{r-1}$. Therefore, it may happen that a monomial with a given exponent dj originates from several values $i \in \mathcal{E}_{r-1}$. In this case, its coefficient is a sum of terms depending on the round constants, which may vanish.

An interesting approach when analyzing the influence of the round constants on the degree of MIMC_3 consists in comparing the algebraic degree of the transformation describing MIMC_9 and the one describing two rounds of MIMC_3. Indeed, using x^9, as round function, is equivalent to using x^3 with all constants c_i, i odd, equal to zero. On Figure 7, we can thus see that the maximal algebraic degree at round r for MIMC_9 does not always correspond to the maximal algebraic degree at round $2r$ for MIMC_3.

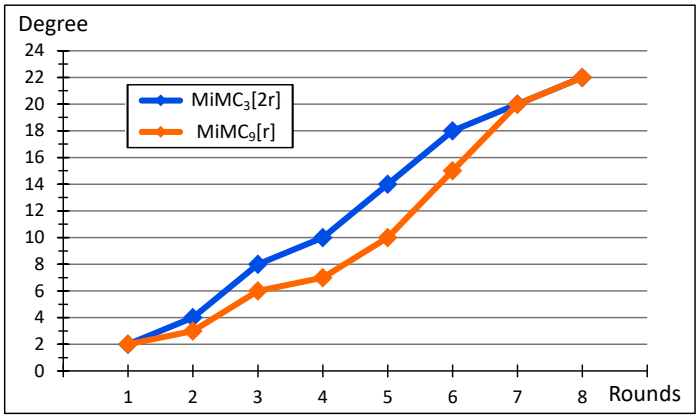

Figure 7: Comparison of algebraic degree for rounds r of MIMC_9 and for rounds $2r$ of MIMC_3 ($n = 23$).

For instance, if we consider the polynomial representing $\mathsf{MIMC}_3[4]$, its maximal algebraic degree is 4, while after 2 rounds of MIMC_9, it is 3. Consequently, the degree of $\mathsf{MIMC}_{3,c}$ at round 4 may drop from 4 to 3, because the coefficients of the maximum-weight exponents only depend on the constants with odd indices:

$$27 : c_1^{18} + c_3^2 \qquad 30 : c_1^{17} \qquad 51 : c_1^{10} \qquad 54 : c_1^9 + c_3 \qquad 57 : c_1^8 \qquad 75 : c_1^2 \qquad 78 : c_1$$

More generally, the coefficients of monomials with exponents not divisible by 9 always admit as a factor a linear combination of constants with odd indices.

We have already shown in Subsection 3.1 that for MIMC_3, the exponents equal to 5 and 7 modulo 8 are missing. For MIMC_9, we can similarly prove Lemma 5.1.

Lemma 5.1. *Let $\mathcal{E}_r$ be the set of exponents in the univariate form of $\mathsf{MIMC}_9[r]$, as defined in Prop. 2.2. Then, any $i \in \mathcal{E}_r$ satisfies*

$$i \bmod 8 \in \{0, 1\} .$$

Proof. We prove the result by induction on r. First, it holds at round 2, since $\mathcal{E}_2 = \{0, 9, 72, 81\}$.

Then let us assume that the property holds for $\mathcal{E}_r$, i.e., any $i \in \mathcal{E}_r$ satisfies $i \bmod 8 \in \{0, 1\}$. It follows that, for any $j \in \mathsf{Cover}(\mathcal{E}_r)$, $j \bmod 8 \in \{0, 1\}$. Since $9j \bmod 8 \in \{0, 1\}$ for any $j \bmod 8 \in \{0, 1\}$, we deduce that any $i \in \mathcal{E}_{r+1}$ is such that $i \bmod 8 \in \{0, 1\}$. □

5.2 Other Quadratic Functions

The mappings x^3 and x^9 are specific cases of Gold functions [Gol68], i.e. of x^d, with d of the form $2^j + 1$. Let us investigate the general case for such permutations.

Proposition 5.2. *[McE87] The mapping x^d with $d = 2^j + 1$ is a permutation in $\mathbb{F}_{2^n}$ if and only if $n/\gcd(j, n)$ is odd.*

We can generalize Lemma 5.1 to any permutation x^d, where $d = 2^j + 1$.

Proposition 5.3. *Let $\mathcal{E}_r$ be the set of exponents in the univariate form of $\mathsf{MIMC}_d[r]$, where $d = 2^j + 1$. Then, any $i \in \mathcal{E}_r$ satisfies*

$$i \bmod 2^j \in \{0, 1\} .$$

Proof. We prove it by induction on r. First, it holds at round 2, since

$$\mathcal{E}_2 = \{0, 2^j + 1, 2^{2j} + 2^j, 2^{2j} + 2^{j+1} + 1\} .$$

Then, let us assume that the property holds for $\mathcal{E}_r$, i.e., any $i \in \mathcal{E}_r$ can be written $i = a2^j + \varepsilon$ with $\varepsilon \in \{0, 1\}$. Let $i' \preceq i$. Then, $i' = ai2^j + \varepsilon'$ with $a' \preceq a$ and $\varepsilon' \leq \varepsilon$. Moreover,

$$\begin{aligned} di' &= (2^j + 1)(a'2^j + \varepsilon') = 2^j(a'2^j + \varepsilon' + 1) + \varepsilon' \\ &\equiv \varepsilon' \bmod 2^j . \end{aligned}$$

Then, it follows that any $\ell = di' \in \mathcal{E}_{r+1}$ is such that $\ell \bmod 2^j \in \{0, 1\}$.

□

Proposition 5.3 hows that the proportion of exponents which may appear in the univiate form of MIMC_{2^j+1} decreases when j increases.

Corollary 5.4. *The maximal algebraic degree after r rounds of $\mathsf{MIMC}_d[r]$, with $d = 2^j + 1$, satisfies:*

$$B_d^r \leq \lfloor r \log_2 d \rfloor - j + 1 .$$

Proof. Any exponent $i \in \mathcal{E}_r$ is such that $\mathrm{wt}(i) \leq \lfloor r \log_2 d \rfloor$. Moreover, from Prop. 5.3 any $i \in \mathcal{E}_r$ satisfies $i \bmod 2^j \in \{0, 1\}$, i.e., $i = [*****\underbrace{00\ldots00}_{j-1}*]$. The weight of the exponents, and consequently the degree, is at most $\lfloor r \log_2 d \rfloor - j + 1$. □

Knowing an explicit formula for the Hamming weight of multiples of $(2^j + 1)$, we could improve this bound on the degree using the following one:

$$B_{2^j+1}^r \leq \max\{\, \mathrm{wt}((2^j + 1)2^j\ell), \mathrm{wt}((2^j + 1)(2^j\ell + 1)) : \ell = 0, \ldots, ((2^j + 1)^{r-1} - 1)/2^j \,\} .$$

5.3 On the Algebraic Degree of MIMC_3^{-1}

We are now interested in the algebraic degree of the inverse transformation. MIMC_3^{-1} is obtained by reversing the order of the round constants and by replacing the round function by $F^{-1}(x) = x^s$ where

$$s = \frac{2^{n+1} - 1}{3} = \sum_{i=0}^{(n-1)/2} 2^{2i}$$

(see e.g. [Nyb94, Prop. 5]).

On Figure 8 we observe two significant facts, on which we will focus:

1. Whatever the extension degree is, there is a plateau between the first two rounds (see Section 5.3.1).
2. The degree grows rapidly up to $n - 2$ and then there is a large plateau that increases with the size of the field on the last rounds (see Section 5.3.2).

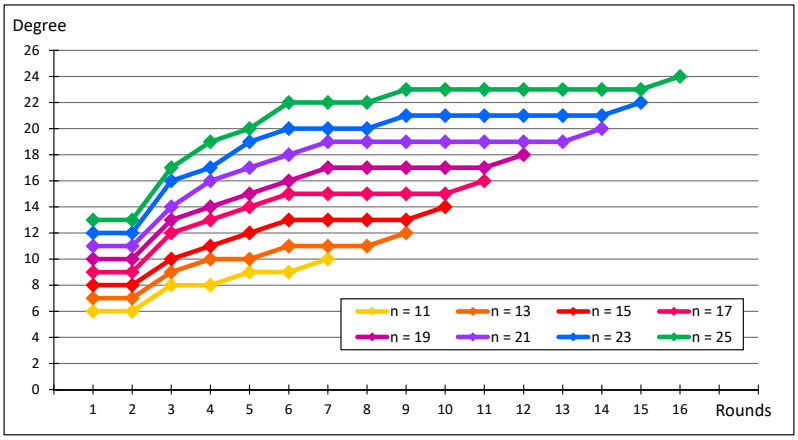

Figure 8: Algebraic degree of MIMC_3^{-1}

5.3.1 A Plateau Between Rounds 1 and 2

The plateau between the first two rounds is counter-intuitive: since the algebraic degree is already high in the first round, we would expect an explosion of the degree in the second round. In this subsection we will see that such an event is due to the particular shape of the exponent $s = (2^{n+1} - 1)/3$.

Proposition 5.5. *Let $j \preceq s$. Then for all j such that $\mathrm{wt}(j) \geq 2$, we have:*

$$\mathrm{wt}(js \bmod 2^n - 1) \in \begin{cases} [\mathrm{wt}(j) - 1, (n-1)/2] \text{ if } \mathrm{wt}(j) \equiv 2 \bmod 3 \ , \\ [\mathrm{wt}(j), (n-1)/2] \text{ if } \mathrm{wt}(j) \equiv 0 \bmod 3 \ , \\ [\mathrm{wt}(j), (n+1)/2] \text{ if } \mathrm{wt}(j) \equiv 1 \bmod 3 \ . \end{cases}$$

Proof. In this proof, we use in particular that for any triple of even integers $i_1 < i_2 < i_3$:

$$s(2^{i_1} + 2^{i_2} + 2^{i_3}) = 2^{i_1} + \sum_{\ell=i_1/2}^{(i_2-2)/2} 2^{2\ell+1} + \sum_{\ell=i_2/2}^{(i_3-2)/2} 2^{2\ell} \bmod 2^n - 1 \ . \tag{8}$$

Indeed, we can check that

$$\begin{aligned} 3 \times \left(2^{i_1} + \sum_{\ell=i_1/2}^{(i_2-2)/2} 2^{2\ell+1} + \sum_{\ell=i_2/2}^{(i_3-2)/2} 2^{2\ell} \right) &= 2^{i_1} + \sum_{\ell=i_1/2}^{(i_2-2)/2} 2^{2\ell+1} + \sum_{\ell=i_2/2}^{(i_3-2)/2} 2^{2\ell} \\ &\quad + 2^{i_1+1} + \sum_{\ell=(i_1+2)/2}^{i_2/2} 2^{2\ell} + \sum_{\ell=i_2/2}^{(i_3-2)/2} 2^{2\ell+1} \\ &= 2^{i_1} + 2^{i_1+1} + 2^{i_1+1}\left(2^{i_2-i_1} - 1\right) + \left(2^{i_3} - 2^{i_2}\right) \\ &= 2^{i_1} + 2^{i_2} + 2^{i_3} \ . \end{aligned}$$

We will investigate three different cases, depending on the value of $wt(j)$ mod 3.

(a) First, let us take an integer j such that $\mathrm{wt}(j) \equiv 2 \bmod 3$. We let $\mathrm{wt}(j) = 2 + 3k$ and $j = \sum_{m=0}^{3k+1} 2^{i_m}$, where the i_m are even since $j \preceq s$, and $2 \leq i_1 < \ldots < i_{3k+1} \leq n-1$.

Then let us show that:

$$sj = \sum_{\ell=0}^{(i_0-2)/2} 2^{2\ell} + 2^{i_1} + \sum_{m=1}^{k} \left(\sum_{\ell=i_{3m-2}/2}^{(i_{3m-1}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3m-1}/2}^{(i_{3m}-2)/2} 2^{2\ell} + 2^{i_{3m+1}} \right) + \sum_{\ell=i_{3k+1}/2}^{(n-3)/2} 2^{2\ell+1} \bmod 2^n - 1 \,, \quad (9)$$

We prove it by induction on k.

- **For $k = 0$:**
  let $j = 2^{i_0} + 2^{i_1}$ where $2 \leq i_0 < i_1 < n$, we have $\mathrm{wt}(j) = 2$. Then,

$$\begin{aligned} s(2^{i_0} + 2^{i_1}) &= \sum_{\ell=0}^{(n-1)/2} 2^{2\ell+i_0} + \sum_{\ell=0}^{(n-1)/2} 2^{2\ell+i_1} \\ &= \sum_{\ell=i_0/2}^{i_1/2-1} 2^{2\ell} + \sum_{\ell=i_1/2}^{(n-1+i_0)/2} 2^{2\ell+1} + \sum_{\ell=(n+1+i_0)/2}^{(n-1+i_1)/2} 2^{2\ell} \\ &= \sum_{\ell=i_0/2}^{i_1/2-1} 2^{2\ell} + \sum_{\ell=i_1/2}^{(n-3)/2} 2^{2\ell+1} + \sum_{\ell=0}^{i_0/2} 2^{2\ell} + \sum_{\ell=i_0/2}^{i_1/2-1} 2^{2\ell+1} \\ &= \sum_{\ell=0}^{i_0/2-1} 2^{2\ell} + 2^{i_1} + \sum_{\ell=i_1/2}^{(n-3)/2} 2^{2\ell+1} \bmod (2^n - 1) \,, \end{aligned}$$

  implying that $\mathrm{wt}(sj \bmod 2^n - 1) = (n + i_0 - i_1 + 1)/2$.

- **Induction step.** Let us assume that the property holds for k, i.e., for any $j_0 = \sum_{m=0}^{3k+1} 2^{i_m}$ such that $\mathrm{wt}(j_0) = 2 + 3k$, sj_0 satisfies (9). Then, let $j = j_0 + 2^{i_{3k+2}} + 2^{i_{3k+3}} + 2^{i_{3k+4}}$, $\mathrm{wt}(j) = 2 + 3(k+1)$ and:

$$\begin{aligned} sj &= sj_0 + s(2^{i_{3k+2}} + 2^{i_{3k+3}} + 2^{i_{3k+4}}) \\ &= sj_0 + \left(2^{i_{3k+2}} + \sum_{\ell=i_{3k+2}/2}^{(i_{3k+3}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3k+3}/2}^{(i_{3k+4}-2)/2} 2^{2\ell} \right) \\ &= \sum_{\ell=0}^{(i_0-2)/2} 2^{2\ell} + 2^{i_1} + \sum_{m=1}^{k} \left(\sum_{\ell=i_{3m-2}/2}^{(i_{3m-1}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3m-1}/2}^{(i_{3m}-2)/2} 2^{2\ell} + 2^{i_{3m+1}} \right) \\ &+ \sum_{\ell=i_{3k+1}/2}^{(i_{3k+2}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3k+2}/2}^{i_{3k+3}/2} 2^{2\ell} + \sum_{\ell=i_{3k+3}/2}^{(i_{3k+4}-2)/2} 2^{2\ell} + \sum_{\ell=i_{3k+3}/2}^{(n-3)/2} 2^{2\ell+1} \\ &= \sum_{\ell=0}^{(i_0-2)/2} 2^{2\ell} + 2^{i_1} + \sum_{m=1}^{k} \left(\sum_{\ell=i_{3m-2}/2}^{(i_{3m-1}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3m-1}/2}^{(i_{3m}-2)/2} 2^{2\ell} + 2^{i_{3m+1}} \right) \\ &+ \sum_{\ell=i_{3k+1}/2}^{(i_{3k+2}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3k+2}/2}^{(i_{3k+3}-2)/2} 2^{2\ell} + 2^{i_{3k+4}} + \sum_{\ell=i_{3k+4}/2}^{(n-3)/2} 2^{2\ell+1} \,. \end{aligned}$$

Then we get:

$$sj = \sum_{\ell=0}^{(i_0-2)/2} 2^{2\ell} + 2^{i_1} + \sum_{m=1}^{k+1} \left(\sum_{\ell=i_{3m-2}/2}^{(i_{3m-1}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3m-1}/2}^{(i_{3m}-2)/2} 2^{2\ell} + 2^{i_{3m+1}} \right) + \sum_{\ell=i_{3k+4}/2}^{(n-3)/2} 2^{2\ell+1} \bmod 2^n - 1 \ .$$

It follows from (9) that, when $\mathrm{wt}(j) = 2 + 3k$, the Hamming weight of sj is:

$$\begin{aligned} \mathrm{wt}(sj \bmod 2^n - 1) &= \frac{i_0}{2} + 1 + \sum_{m=1}^{k} \left(\frac{i_{3m-1} - i_{3m-2}}{2} + \frac{i_{3m} - i_{3m-1}}{2} + 1 \right) \\ &\quad + \frac{n - 1 - i_{3k+1}}{2} \\ &= \frac{1}{2} \left(n + 2k + 1 - \sum_{m=0}^{k} (i_{3m+2} - i_{3m+1}) \right) \ . \end{aligned}$$

Obviously, $i_{3m+1} - i_{3m} \geq 2$, implying that

$$\mathrm{wt}(sj \bmod 2^n - 1) \leq (n-1)/2 \ .$$

Moreover, the largest value for $\sum_{m=0}^{k}(i_{3m+2} - i_{3m+1})$ is obtained when all other distances between two consecutive elements among $i_0, \ldots, i_{3k+1}$ are minimized, i.e., equal to 2, leading to $\sum_{m=0}^{k}(i_{3m+2} - i_{3m+1}) \leq (n-1) - 4k$. We then deduce that

$$\mathrm{wt}(sj \bmod 2^n - 1) \geq 3k + 1 = \mathrm{wt}(j) - 1 \ .$$

(b) Let us now consider j such that $\mathrm{wt}(j) \bmod 3 = 0$. We let $\mathrm{wt}(j) = 3k$, and $j = \sum_{m=0}^{3k-1} 2^{i_m}$, where the i_m are even, and $2 \leq i_1 < \ldots < i_{3k-1} \leq n-1$. Then, we will show by induction on k that

$$sj = \sum_{m=0}^{k-1} \left(2^{i_{3m}} + \sum_{\ell=i_{3m}/2}^{(i_{3m+1}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3m+1}/2}^{(i_{3m+2}-2)/2} 2^{2\ell} \right) \bmod 2^n - 1 \ . \qquad (10)$$

- **For $k = 1$:** let $j = 2^{i_0} + 2^{i_1} + 2^{i_2}$ where i_1 and i_2 are two even integers such that $2 \leq i_0 < i_1 < i_2 < n$. Then, we know from (8) that

$$sj = 2^{i_0} + \sum_{\ell=i_0}^{(i_1-2)/2} 2^{2\ell+1} + \sum_{\ell=i_1/2}^{(i_2-2)/2} 2^{2\ell} \bmod 2^n - 1 \ .$$

- **Induction step.** Let us assume that the property holds for k, that is: for any $j_0 = \sum_{m=0}^{3k-1} 2^{i_m}$, sj_0 satisfies (10). Then, for $j = j_0 + 2^{i_{3k}} + 2^{i_{3k+1}} + 2^{i_{3k+2}}$, we deduce from (8) that

$$\begin{aligned} sj &= sj_0 + s\left(2^{i_{3k}} + 2^{i_{3k+1}} + 2^{i_{3k+2}}\right) \\ &= sj_0 + 2^{i_{3k}} + \sum_{\ell=i_{3k}/2}^{(i_{3k+1}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3k+1}/2}^{(i_{3k+2}-2)/2} 2^{2\ell} \\ &= \sum_{m=0}^{k} \left(2^{i_{3m}} + \sum_{\ell=i_{3m}/2}^{(i_{3m+1}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3m+1}/2}^{(i_{3m+2}-2)/2} 2^{2\ell} \right) \bmod 2^n - 1 \ . \end{aligned}$$

Therefore, if $\mathrm{wt}(j) = 3k$, we have

$$\mathrm{wt}(sj \bmod 2^n - 1) = \frac{1}{2}\left(2k + \sum_{m=0}^{k-1}(i_{3m+2} - i_{3m})\right) \in [\mathrm{wt}(j), (n-1)/2] \ .$$

Obviously, $i_{3m+2} - i_{3m} \geq 4$, implying that

$$\mathrm{wt}(sj \bmod 2^n - 1) \geq 3k = \mathrm{wt}(j) \ .$$

Moreover, the largest value for $\sum_{m=0}^{k-1}(i_{3m+2} - i_{3m})$ is obtained when all other distances between two consecutive elements among $i_0, \ldots, i_{3k-1}$ are minimized, i.e., equal to 2, leading to $\sum_{m=0}^{k-1}(i_{3m+2} - i_{3m}) \leq (n-1) - 2k$. We then deduce that

$$\mathrm{wt}(sj \bmod 2^n - 1) \leq (n-1)/2 \ .$$

(c) Finally, let us consider j such that $\mathrm{wt}(j) \bmod 3 = 1$. We let $\mathrm{wt}(j) = 1 + 3k$, and $j = \sum_{m=0}^{3k} 2^{i_m}$, where the i_m are even, and $2 \leq i_0 < \ldots < i_{3k} \leq n-1$. Now, we will prove by induction on k that

$$sj = \sum_{\ell=0}^{(i_0-2)/2} 2^{2\ell+1} + \sum_{m=0}^{k-1}\left(\sum_{\ell=i_{3m}/2}^{(i_{3m+1}-2)/2} 2^{2\ell} + 2^{i_{3m+2}} + \sum_{\ell=i_{3m+2}/2}^{(i_{3m+3}-2)/2} 2^{2\ell+1}\right) + \sum_{\ell=i_{3k}/2}^{(n-1)/2} 2^{2\ell} \bmod 2^n - 1 \ . \quad (11)$$

- **For $k = 1$:** let $j = 2^{i_0} + 2^{i_1} + 2^{i_2} + 2^{i_3}$ with i_0, i_1, i_2, i_3 even such that $2 \leq i_0 < i_1 < i_2 < i_3 < n$. Then, we deduce from (8) that

$$\begin{aligned}
sj &= 2^{i_0} + \sum_{\ell=i_0/2}^{(i_1-2)/2} 2^{2\ell+1} + \sum_{\ell=i_1/2}^{(i_2-2)/2} 2^{2\ell} + \sum_{\ell=0}^{(n-1)/2} 2^{2\ell+i_3} \\
&= 2^{i_0} + \sum_{\ell=i_0/2}^{(i_1-2)/2} 2^{2\ell+1} + \sum_{\ell=i_1/2}^{(i_2-2)/2} 2^{2\ell} + \sum_{\ell=i_3/2}^{(n-1)/2} 2^{2\ell} + \sum_{\ell=0}^{(i_3-2)/2} 2^{2\ell+1} \\
&= \sum_{\ell=0}^{(i_0-2)/2} 2^{2\ell+1} + \sum_{\ell=i_0/2}^{i_1/2} 2^{2\ell} + \sum_{\ell=i_1/2}^{(i_2-2)/2} 2^{2\ell} + \sum_{\ell=i_1/2}^{(i_3-2)/2} 2^{2\ell+1} + \sum_{\ell=i_3/2}^{(n-1)/2} 2^{2\ell} \\
&= \sum_{\ell=0}^{(i_0-2)/2} 2^{2\ell+1} + \sum_{\ell=i_0/2}^{(i_1-2)/2} 2^{2\ell} + 2^{i_2} + \sum_{\ell=i_2/2}^{(i_3-2)/2} 2^{2\ell+1} + \sum_{\ell=i_3/2}^{(n-1)/2} 2^{2\ell} \bmod 2^n - 1 \ .
\end{aligned}$$

- **Induction step.** Let us assume that the property holds for k, i.e., for any $j_0 = \sum_{m=0}^{3k} 2^{i_m}$, sj_0 satisfies (11). Then, for $j = j_0 + 2^{i_{3k+1}} + 2^{i_{3k+2}} + 2^{i_{3k+3}}$,

we have:

$$\begin{aligned} sj &= sj_0 + 2^{i_{3k+1}} + \sum_{\ell=i_{3k+1}/2}^{(i_{3k+2}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3k+2}/2}^{(i_{3k+3}-2)/2} 2^{2\ell} \\ &= \sum_{\ell=0}^{(i_0-2)/2} 2^{2\ell+1} + \sum_{m=0}^{k-1} \left(\sum_{\ell=i_{3m}/2}^{(i_{3m+1}-2)/2} 2^{2\ell} + 2^{i_{3m+2}} + \sum_{\ell=i_{3m+2}/2}^{(i_{3m+3}-2)/2} 2^{2\ell+1} \right) \\ &\quad + \sum_{\ell=i_{3k}/2}^{(n-1)/2} 2^{2\ell} + 2^{i_{3k+1}} + \sum_{\ell=i_{3k+1}/2}^{(i_{3k+2}-2)/2} 2^{2\ell+1} + \sum_{\ell=i_{3k+2}/2}^{(i_{3k+3}-2)/2} 2^{2\ell} \\ &= \sum_{\ell=0}^{(i_0-2)/2} 2^{2\ell+1} + \sum_{m=0}^{k} \left(\sum_{\ell=i_{3m}/2}^{(i_{3m+1}-2)/2} 2^{2\ell} + 2^{i_{3m+2}} + \sum_{\ell=i_{3m+2}/2}^{(i_{3m+3}-2)/2} 2^{2\ell+1} \right) \\ &\quad + \sum_{\ell=i_{3k+3}/2}^{(n-1)/2} 2^{2\ell} \bmod 2^n - 1 \, . \end{aligned}$$

Then, if $\mathrm{wt}(j) = 3k+1$, we have

$$\begin{aligned} \mathrm{wt}(sj \bmod 2^n - 1) &= \frac{1}{2} \left(n + 1 + 2k + i_1 + \sum_{m=1}^{k-1} (i_{3m+1} - i_{3m-1}) - i_{3k-1} \right) \\ &= \frac{1}{2} \left(n + 1 + 2k - \sum_{m=1}^{k} (i_{3m-1} - i_{3m-2}) \right) \, . \end{aligned}$$

Obviously, $i_{3m-1} - i_{3m-2} \geq 2$, implying that

$$\mathrm{wt}(sj \bmod 2^n - 1) \leq (n+1)/2 \, .$$

Moreover, the largest value for $\sum_{m=1}^{k} (i_{3m-1} - i_{3m-2})$ is obtained when all other distances between two consecutive elements among $i_0, \ldots, i_{3k}$ are minimized, i.e., equal to 2, leading to $\sum_{m=1}^{k} (i_{3m-1} - i_{3m-2}) \leq (n-1) - 4k$. We then deduce that

$$\mathrm{wt}(sj \bmod 2^n - 1) \geq 3k + 1 = wt(j) \, .$$

□

As an immediate consequence, we obtain the following corollary.

Corollary 5.6. *There is a plateau on the first two rounds of* MIMC_3^{-1}*, i.e.:*

$$B_s^1 = B_s^2 = \frac{n+1}{2} \, .$$

Since there is a plateau between the first and second round for both MIMC_3 and MIMC_3^{-1}, we may wonder whether this corresponds to a more general phenomenon, since in Section 2.3 we also proved that $B_d^1 = B_d^2$, when $d = 2^k - 1$. However, there is not necessarily a plateau for MIMC_d^{-1}. Indeed, in $\mathbb{F}_{2^{11}}$, we have $15 = 2^4 - 1$, so according to Prop. 2.5 we have $B_{15}^1 = B_{15}^2$. But for MIMC_{15}^{-1}, the inverse of 15 is 273, so the algebraic degree of the first round is $\mathrm{wt}(273) = 3$, while it is 5 after two rounds (for example $\mathrm{wt}(273 \times 273 \bmod 2^{11} - 1) = 5$).

5.3.2 Influence of the Encryption Degree

Studying the algebraic degree of MIMC_3^{-1} over iterations is much more difficult than for MIMC_3 since the underlying round function x^s has a much higher degree. However, the following result from [BC13] shows how the encryption degree influences the decryption degree.

Proposition 5.7. *[BC13] For any $i \in [1, n-1]$ if the degree of the encryption function is strictly less than $(n-1)/i$, the degree of the decryption function is strictly less than $n-i$.*

Based on this result, we can exhibit a lower bound on the number of rounds needed by the decryption function to reach degree $(n-i)$ (for some round constants).

Corollary 5.8. *Let r_{n-i} denote the smallest value of r such that $B_s^r \geq n-i$ for $1 \leq i \leq (n-1)/4$. Then*

$$r_{n-i} \geq \left\lceil \frac{1}{\log_2 3} \left(2 \left\lceil \frac{1}{2} \left\lceil \frac{n-1}{i} \right\rceil \right\rceil + 1 \right) \right\rceil .$$

Most notably,

$$r_{n-2} \geq \left\lceil \frac{1}{\log_2 3} \left(2 \left\lceil \frac{n-1}{4} \right\rceil + 1 \right) \right\rceil .$$

Proof. From the previous proposition, we know that, if $B_s^r \geq n-i$, then $B_3^r \geq (n-1)/i$. Since $i \leq (n-1)/4$, $B_3^r \geq 4$, and then $r \geq 4$, implying that Proposition 3.4 applies. We then deduce that

$$2 \times \left\lceil \frac{\lfloor r \log_2 3 \rfloor}{2} - 1 \right\rceil \geq \frac{n-1}{i} .$$

It follows that

$$\frac{\lfloor r \log_2(3) \rfloor - 1}{2} \geq \left\lceil \frac{\lfloor r \log_2 3 \rfloor}{2} - 1 \right\rceil \geq \left\lceil \frac{1}{2} \left\lceil \frac{n-1}{i} \right\rceil \right\rceil .$$

Therefore,

$$r \geq \left\lceil \frac{1}{\log_2 3} \left(2 \left\lceil \frac{1}{2} \left\lceil \frac{n-1}{i} \right\rceil \right\rceil + 1 \right) \right\rceil .$$

□

As an illustration, for $n = 25$, Corollary 5.8 applied with $1 \leq i \leq 6$, leads to the upper bound depicted on Fig. 9.

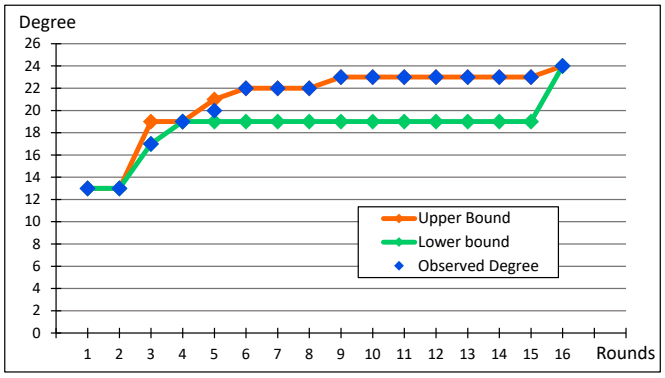

Figure 9: Bounds on the algebraic degree of MIMC_3^{-1} for $n = 25$.

6 Higher-Order Differential Attacks

In this section, we focus on attacks based on some algebraic properties of the cipher, most notably on higher-order differential attacks exploiting the algebraic degree of the primitive. Indeed, a distinguisher[2] can be exhibited using that $\bigoplus_{x\in\mathcal{V}} F(x) = 0$ for any affine subspace $\mathcal{V} \subset \mathbb{F}_2^n$ with $\dim \mathcal{V} \geq \deg^a(F) + 1$. Since almost all permutations of $\mathbb{F}_2^n$ have algebraic degree $(n-1)$ (see e.g. [Wel69, Das02, KP02]), an iterated cipher needs to have enough rounds to reach the maximal algebraic degree in order to be indistinguishable from a random permutation.

6.1 Secret-Key Zero-Sum Distinguisher

It has been shown in [EGL+20, Prop. 2] that the maximal algebraic degree for MIMC_3 and for its inverse MIMC_3^{-1} can be reach only when $r \geq \lceil \log_3(2^{n-1}+1) \rceil$. Prop. 3.4 enables us to slightly improve this bound.

Proposition 6.1. *For any $r < \lceil \log_3 2^n \rceil$, the algebraic degree of MIMC_3 is at most $(n-3)$ and the algebraic degree of MIMC_3^{-1} is at most $(n-2)$.*

Proof. From Prop 3.4, if the degree of r rounds of MIMC_3 is $(n-1)$ for some round constants, then

$$\left\lceil \frac{k_r}{2} - 1 \right\rceil \geq \frac{n-1}{2} ,$$

which implies that $k_r \geq n$, i.e., $\log_2 3^r \geq n$. It follows that, for $r < \lceil \log_3 2^n \rceil$, $\deg^a \mathsf{MIMC}_3[r] \leq (n-2)$. Using that the upper bound in Prop. 3.4 is always even, we derive that $\deg^a \mathsf{MIMC}_3[r] \leq (n-3)$. Moreover, as already observed in [BC13, EGL+20], a permutation of $\mathbb{F}_2^n$ has degree $(n-1)$ if and only if its inverse has degree $(n-1)$. Thus, $\deg^a \mathsf{MIMC}_3^{-1}[r] \leq (n-2)$. □

Therefore, the number of rounds covered by a zero-sum distinguisher against MIMC_3 or MIMC_3^{-1} is slightly higher than predicted in [EGL+20]; and for all values of r covered by Corollary 4.11, which includes the parameters studied in [EGL+20], i.e. $n \in \{127, 129, 255\}$, we derive that *this is the highest number of rounds which can be covered by such a distinguisher*. Moreover, this distinguisher against MIMC_3 has complexity at most 2^{n-2}, instead of 2^{n-1}. As noted in [EGL+20], such a zero-sum distinguisher for $(r-1)$ rounds of MIMC_3^{-1} can be extented to a key-recovery attack over r rounds.

Another observation is that, in many cases, the complexity of the distinguisher can be reduced to 2^{n-4} by removing the last round, as stated in the following proposition.

Proposition 6.2. *Let $\mathcal{R} = \lceil \log_3 2^n \rceil$. For any $r < \mathcal{R} - 1$, the algebraic degree of MIMC_3 is at most $(n-5)$, unless $k_{\mathcal{R}} = k_{\mathcal{R}-1}$ is even and $k_{\mathcal{R}-2}$ is odd (which equivalently means that there is a plateau between rounds $(\mathcal{R}-2)$ and $(\mathcal{R}-1)$).*

Proof. Recall that a plateau between rounds i and $(i+1)$ corresponds to the situation where k_{i-1} is odd and k_i even, i.e. $b_{i-1}b_i = 10$ (see Section 4.4). As stated in the previous proposition, there is no plateau between rounds $(\mathcal{R}-1)$ and $\mathcal{R}$, implying that $b_{\mathcal{R}-1}b_{\mathcal{R}} \neq 10$. Therefore, two situations may occur.

(i) $b_{\mathcal{R}-1}b_{\mathcal{R}} = 00$. In this case, there is a plateau between rounds $(\mathcal{R}-2)$ and $(\mathcal{R}-1)$ if and only if $b_{\mathcal{R}-2} = 1$.

[2] A *distinguisher* is any property that should not be expected from an ideal object, here a permutation picked uniformly at random from the set of all permutations of $\mathbb{F}_2^n$. The existence of a distinguisher is an undesirable property for a cryptographic primitive.

(ii) $b_{\mathcal{R}-1}b_{\mathcal{R}} \in \{01, 11\}$. The only possibility corresponding to a plateau between rounds $(\mathcal{R}-2)$ and $(\mathcal{R}-1)$ is then $b_{\mathcal{R}-2}b_{\mathcal{R}-1}b_{\mathcal{R}} = 101$, which is impossible because it would imply the existence of two consecutive switches in $(b_r)_{r>0}$, i.e. $s_{\mathcal{R}-1}s_{\mathcal{R}} = 11$, while we known from Prop. 4.2 that

$$s_{\mathcal{R}-1}s_{\mathcal{R}} \in \{3-k_2, 4-k_2\} = \{0, 1\} .$$

Therefore, Case (i) is the only situation where we may have $B_3^{\mathcal{R}-2} = n-3$. In all other cases, $B_3^{\mathcal{R}-2} \leq n-5$. □

As an example, for $n = 127$, $\mathcal{R} = 81$, and we can check from Table 3 that we are in a case where $B_3^{\mathcal{R}-2} = B_3^{\mathcal{R}-1} = n-3$. While [EGL+20] exhibits a distinguisher with complexity 2^{125} for 78 rounds, we show that it actually covers 80 rounds. For $n = 129$, $\mathcal{R} = 82$, so we have a distinguisher of complexity 2^{127} for 81 rounds (instead of 80 in [EGL+20]), and of complexity 2^{125} for 80 rounds.

Table 3: Comparison of degree for MIMC_3.

r	77	78	79	80	81	82
$\lfloor \log_2(3^r+1) \rfloor$	122	123	125	126	128	129
B_3^r	120	122	124	124	126	128

For $n = 129$, we compare our results with those of Eichlseder et al. [EGL+20] in Table 4, where we use the same notation: "KR" for Key-Recovery, "KK" for Known-Key distinguisher, and SK for Secret-Key distinguisher. Overall, our careful study of the algebraic degree allows us to improve their attacks.

Table 4: Complexity of attacks on MIMC_3.

Type	n	Rounds	Time	Data	Source
SK	129	80	2^{128} XOR	2^{128}	[EGL+20]
	n	$\lceil \log_3(2^{n-1}-1) \rceil - 1$	2^{n-1} XOR	2^{n-1}	
	129	81	2^{128} XOR	2^{128}	New
	n	$\lceil \log_3 2^n \rceil - 1$	2^{n-1} XOR	2^{n-1}	
	129	81 (MIMC_3)	2^{127} XOR	2^{127}	New
	n	$\lceil \log_3 2^n \rceil - 1$ (MIMC_3)	2^{n-2} XOR	2^{n-2}	
	129	80 (MIMC_3)	2^{125} XOR	2^{125}	New
	n	$\lceil \log_3 2^n \rceil - 2$ (MIMC_3)	2^{n-2} or 2^{n-4} XOR	2^{n-2} or 2^{n-4}	
KK	129	160	-	2^{128}	[EGL+20]
	n	$2 \cdot \lceil \log_3(2^{n-1}-1) \rceil - 2$	-	2^{n-1}	
	129	162	-	2^{128}	New
	n	$2 \cdot \lceil \log_3 2^n \rceil - 2$	-	2^{n-1}	
KR	129	82	$2^{122.64}$	2^{128}	[EGL+20]
	n	$\lceil n \cdot \log_3 2 \rceil$	$2^{n-1-(\log_2 \lceil n \log_3 2 \rceil)}$ or $2^{n-(\log_2 \lceil n \log_3 2 \rceil)}$	2^{n-1}	
	129	82	$2^{121.64}$	2^{128}	New
	n	$\lceil n \cdot \log_3 2 \rceil$	$2^{n-1-(\log_2 \lceil n \log_3 2 \rceil)}$	2^{n-1}	

6.2 Known-Key Zero-Sum Distinguisher

Using a subspace of dimension $n-1$, the number of rounds we can distinguish is $\mathcal{R}-1$ for both MIMC_3, and MIMC_3^{-1}. As a consequence, there is a known-key zero-sum

distinguisher as defined in [AM09] on almost twice the number of rounds, starting from the middle of the primitive.

Such a known-key distinguisher can be applied to the hash function proposed in [AGR+16] hash functions, based on the use of MIMC_3 within the sponge framework, as depicted on Figure 10, where r is the rate and c the capacity.

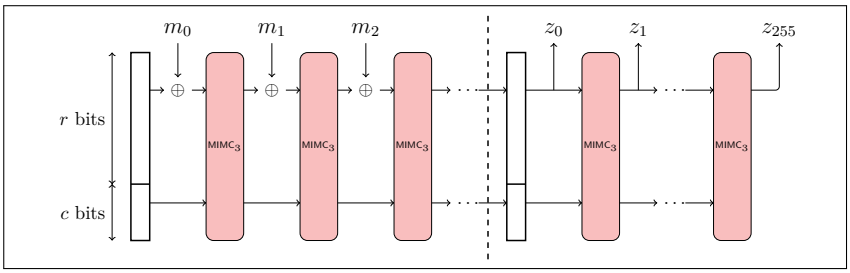

Figure 10: Hash function in sponge framework.

While there is a 0-sum distinguisher on $2\mathcal{R} - 2$ rounds when the dimension of the subspace $\mathcal{V}$ is $n - 1$, we are also interested in reducing the size of the subspace, in order to decrease the data complexity.

First, let us consider the hash function using MIMC_3 with an extension degree $n = 1025$, which corresponds to 647 rounds. In this case, the last plateau for MIMC_3 is between rounds $\mathcal{R} - 4$ and $\mathcal{R} - 3$, where the degree is equal to $n - 7$. Furthermore, for MIMC_3^{-1}, we know that $r_{n-2} \geq 324$, where r_{n-2} is the first round where the degree reaches $n - 2$. It follows that, if we operate on a subspace $\mathcal{V}$ of dimension $n - 2$, we would reduce by a quarter the number of rounds for which we can set up a distinguisher, as seen in Figure 11.

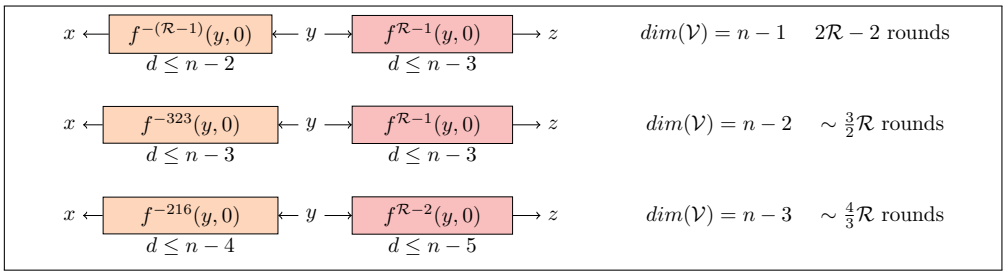

Figure 11: 0-sum with hash function (with $n = 1025$).

7 Conclusion

Symmetric primitives designed over a large field are inherently different from the "usual" ones that are defined over $(\mathbb{F}_2)^n$. Due to its simplicity, MiMC is an interesting target to investigate the security leven offered by such algorithms. Yet, despite this simplicity, tightly quantifying its security against higher-order differentials required us to develop new mathematical tools to track the evolution of the exponents that appear in the univariate representation of the encryption function as the round function is iterated. In the end, we have managed to evaluate the exact algebraic degree of up to more than 16000 rounds of this block cipher. Overall, our results contribute to a better understanding of the behaviour of symmetric primitives defined over large finite fields.

References

[AAB+20] Abdelrahaman Aly, Tomer Ashur, Eli Ben-Sasson, Siemen Dhooghe, and Alan Szepieniec. Design of symmetric-key primitives for advanced cryptographic protocols. *IACR Trans. Symm. Cryptol.*, 2020(3):1–45, 2020.

[ACG+19] Martin R. Albrecht, Carlos Cid, Lorenzo Grassi, Dmitry Khovratovich, Reinhard Lüftenegger, Christian Rechberger, and Markus Schofnegger. Algebraic cryptanalysis of STARK-friendly designs: Application to MARVELlous and MiMC. In Steven D. Galbraith and Shiho Moriai, editors, *ASIACRYPT 2019, Part III*, volume 11923 of *LNCS*, pages 371–397. Springer, Heidelberg, December 2019.

[AGP+19] Martin R. Albrecht, Lorenzo Grassi, Léo Perrin, Sebastian Ramacher, Christian Rechberger, Dragos Rotaru, Arnab Roy, and Markus Schofnegger. Feistel structures for MPC, and more. In Kazue Sako, Steve Schneider, and Peter Y. A. Ryan, editors, *ESORICS 2019, Part II*, volume 11736 of *LNCS*, pages 151–171. Springer, Heidelberg, September 2019.

[AGR+16] Martin R. Albrecht, Lorenzo Grassi, Christian Rechberger, Arnab Roy, and Tyge Tiessen. MiMC: Efficient encryption and cryptographic hashing with minimal multiplicative complexity. In Jung Hee Cheon and Tsuyoshi Takagi, editors, *ASIACRYPT 2016, Part I*, volume 10031 of *LNCS*, pages 191–219. Springer, Heidelberg, December 2016.

[AM09] Jean-Philippe Aumasson and Willi Meier. Zero-sum distinguishers for reduced Keccak-f and for the core functions of Luffa and Hamsi. Rump session of Cryptographic Hardware and Embedded Systems-CHES, 2009.

[BBHR18] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Scalable, transparent, and post-quantum secure computational integrity. Cryptology ePrint Archive, Report 2018/046, 2018. https://eprint.iacr.org/2018/046.

[BC13] Christina Boura and Anne Canteaut. On the Influence of the Algebraic Degree of F^{-1} on the Algebraic Degree of $G \circ F$. *IEEE Trans. Inf. Theory*, 59(1):691–702, 2013.

[BCD+20] Tim Beyne, Anne Canteaut, Itai Dinur, Maria Eichlseder, Gregor Leander, Gaëtan Leurent, María Naya-Plasencia, Léo Perrin, Yu Sasaki, Yosuke Todo, and Friedrich Wiemer. Out of oddity - new cryptanalytic techniques against symmetric primitives optimized for integrity proof systems. In Daniele Micciancio and Thomas Ristenpart, editors, *CRYPTO 2020, Part III*, volume 12172 of *LNCS*, pages 299–328. Springer, Heidelberg, August 2020.

[BFL+21] Olivier Bronchain, Sebastian Faust, Virginie Lallemand, Gregor Leander, Léo Perrin, and François-Xavier Standaert. Moe: Multiplication operated encryption with trojan resilience. *IACR Transactions on Symmetric Cryptology*, 2021(1):78–129, Mar. 2021.

[BGL20] Eli Ben-Sasson, Lior Goldberg, and David Levit. STARK friendly hash – survey and recommendation. Cryptology ePrint Archive, Report 2020/948, 2020. https://eprint.iacr.org/2020/948.

[Cha13] Pascale Charpin. *Handbook of Finite Fields*, chapter PN and APN functions. CRC Press, 2013.

[Das02] Pinaki Das. The number of permutation polynomials of a given degree over a finite field. *Finite Fields and Their Applications*, 8(4):478–490, 2002.

[EGL+20] Maria Eichlseder, Lorenzo Grassi, Reinhard Lüftenegger, Morten Øygarden, Christian Rechberger, Markus Schofnegger, and Qingju Wang. An algebraic attack on ciphers with low-degree round functions: Application to full MiMC. In Shiho Moriai and Huaxiong Wang, editors, *ASIACRYPT 2020, Part I*, volume 12491 of *LNCS*, pages 477–506. Springer, Heidelberg, December 2020.

[GAB+20a] Gerald Gamrath, Daniel Anderson, Ksenia Bestuzheva, Wei-Kun Chen, Leon Eifler, Maxime Gasse, Patrick Gemander, Ambros Gleixner, Leona Gottwald, Katrin Halbig, Gregor Hendel, Christopher Hojny, Thorsten Koch, Pierre Le Bodic, Stephen J. Maher, Frederic Matter, Matthias Miltenberger, Erik Mühmer, Benjamin Müller, Marc E. Pfetsch, Franziska Schlösser, Felipe Serrano, Yuji Shinano, Christine Tawfik, Stefan Vigerske, Fabian Wegscheider, Dieter Weninger, and Jakob Witzig. The SCIP Optimization Suite 7.0. Technical report, Optimization Online, March 2020.

[GAB+20b] Gerald Gamrath, Daniel Anderson, Ksenia Bestuzheva, Wei-Kun Chen, Leon Eifler, Maxime Gasse, Patrick Gemander, Ambros Gleixner, Leona Gottwald, Katrin Halbig, Gregor Hendel, Christopher Hojny, Thorsten Koch, Pierre Le Bodic, Stephen J. Maher, Frederic Matter, Matthias Miltenberger, Erik Mühmer, Benjamin Müller, Marc E. Pfetsch, Franziska Schlösser, Felipe Serrano, Yuji Shinano, Christine Tawfik, Stefan Vigerske, Fabian Wegscheider, Dieter Weninger, and Jakob Witzig. The SCIP Optimization Suite 7.0. ZIB-Report 20-10, Zuse Institute Berlin, March 2020.

[Gol68] Robert Gold. Maximal recursive sequences with 3-valued recursive crosscorrelation functions. *IEEE Transactions on Information Theory*, 14:154–156, 1968.

[Her36] Aaron Herschfeld. The equation $2^x - 3^y = d$. *Bull. Amer. Math. Soc.*, 42(4):231–234, 04 1936.

[Knu95] Lars R. Knudsen. Truncated and higher order differentials. In Bart Preneel, editor, *FSE'94*, volume 1008 of *LNCS*, pages 196–211. Springer, Heidelberg, December 1995.

[KP02] Sergei Konyagin and Francesco Pappalardi. Enumerating permutation polynomials over finite fields by degree. *Finite Fields and Their Applications*, 8(4):548–553, 2002.

[McE87] Robert J. McEliece. *Finite Fields for Computer Scientists and Engineers*. Springer Verlag, 1987.

[Nyb94] Kaisa Nyberg. Differentially uniform mappings for cryptography. In Tor Helleseth, editor, *EUROCRYPT'93*, volume 765 of *LNCS*, pages 55–64. Springer, Heidelberg, May 1994.

[Wel69] Charles Wells. The degrees of permutation polynomials over finite fields. *Journal of Combinatorial Theory*, 7(1):49–55, 1969.