A Nonlinear Mapping Based on Squaring

Denise Verbakel[1], Daniel Kuijsters[1], Silvia Mella[1], Stjepan Picek[1], Luca Mariot[2] and Joan Daemen[1]

[1] Radboud University, Digital Security Department, Nijmegen, The Netherlands
[2] Semantics, Cybersecurity and Services Group, University of Twente, The Netherlands

Abstract. Many modern symmetric cryptographic primitives operate in an iterated way: they consist of the repeated application of a relatively simple round function over a state, alternated with the addition of secret round keys or round constants. A crucial component of the round function is the nonlinear layer, usually defined via an invertible map. However, many modes of operations do not require invertibility of the underlying primitive and recently Grassi proposed the usage of non-invertible nonlinear mappings in MPC-/FHE-/ZK-friendly symmetric cryptographic primitives. In this work, we consider one of these maps. It is a simple yet efficient nonlinear map, that we call γ, based on squaring over $\mathbb{F}_q$, with q an odd prime power. We discuss for the first time the differential and linear propagation properties of such a nonlinear map and observe that they follow the same rules. This is an intriguing property that, as far as we know, only occurs with γ and the binary mapping χ_3 used in Xoodoo.

Keywords: Nonlinear layer, Squaring, Finite fields

1 Introduction

The round functions in most of the modern symmetric cryptographic primitives usually consist of a non-linear mapping and a number of linear mappings. These mappings are chosen and combined so that there is no exploitable differential propagation from input to output or exploitable correlations between input and output. The relevant properties of these mappings over binary fields have been studied extensively by an expert community of mathematicians, leading to solid designs. But, this community does not stop at the binary case and also studies similar functions over $\mathbb{F}_p$ and its extensions, with p an odd prime. For instance, Kölbl et al. designed a ternary cryptographic hash function called Troika [KTDB19]. Other examples are the MPC-/FHE-/ZK-friendly symmetric primitives defined over $\mathbb{F}_p^n$ like MiMC [AGR+], Poseidon [GKR+21], and many others.

There are interesting differences between the binary case and the odd-prime case, and to a certain extent, the fields of odd characteristics are richer in functionality than binary fields. For example, addition and subtraction are the same in $\mathbb{F}_2$. In $\mathbb{F}_p$, this is no longer the case. In $\mathbb{F}_{2^d}$, squaring is a linear operation. In $\mathbb{F}_{p^d}$ squaring is, in a certain sense, an optimally nonlinear operation. In $\mathbb{F}_2$, correlations between input and output bits have values that are rational and range from -1 to $+1$. In $\mathbb{F}_p$, correlations are complex numbers in the unit disk.

This work investigates a mapping over $\mathbb{F}_q^n$ recently proposed by Grassi [Gra22], that we call γ. We investigate the differential and linear propagation properties of such mapping, both in the forward and backward direction. Our results are useful in determining the maximum probabilities of differentials and trails and correlations of linear approximations and trails over transformations making use of this mapping in their round function, as in computer-assisted trail search [DA].

2 Preliminaries

Let $\mathbb{F}_q$ be a finite field with $q = p^d$ an odd prime power. Let $\mathbb{F}_q^n$ be a vector space of dimension n over the finite field $\mathbb{F}_q$. We denote the coordinates of a vector $x \in \mathbb{F}_q^n$ by x_i with $i \in \{0, 1, \dots n-1\}$ and call them *digits*. We denote by e_i the vector with all coordinates equal to 0 except coordinate i equal to 1. The Hamming weight $\mathrm{HW}(x)$ of a vector $x \in \mathbb{F}_q^n$ is the number of non-zero digits in the vector.

Given two vectors $x, y \in \mathbb{F}_q^n$, we denote their vector subtraction by $x - y$, hence $x - y = x + (-1)y$. We denote by $x^{\mathrm{T}}y$ the value $\sum_i x_i y_i \in \mathbb{F}_q$.

Given a vector $x \in \mathbb{F}_q^n$ its activity pattern $\widetilde{x}$ is a vector in $\mathbb{F}_q^n$ with $\widetilde{x}_i = 1$ if $x_i \neq 0$ and 0 otherwise.

3 Our non-linear mapping γ

In this work, we consider a mapping defined in [Gra22] that we will denote by $\gamma : \mathbb{F}_q^n \to \mathbb{F}_q^n$ as

$$\gamma(x) = y \text{ with } y_i = x_i + x_{i+1 \bmod n}^2 \forall i.$$

From now on, we will omit the modular reduction in the index and always assume it is reduced modulo n.

4 Differential properties of γ

We analyzed the differential properties of the map γ. We will first define differential probability and weight for the non-binary case and then summarize our findings for γ.

4.1 Differentials, differential probability and weight

Let $x \in \mathbb{F}_q^n$ and $x^* \in \mathbb{F}_q^n$ be inputs of a transformation $\alpha : \mathbb{F}_q^n \to \mathbb{F}_q^n$ and let their difference be $a = x^* - x$. Likewise, let $y \in \mathbb{F}_q^n$ and $y^* \in \mathbb{F}_q^n$ be outputs of α and let their difference be $b = y^* - y$. The (ordered) pair $(a, b) \in \mathbb{F}_q^n \times \mathbb{F}_q^n$ containing the input and output difference is called a *differential over* α.

The *differential probability (DP)* of a differential (a, b) over the transformation α is defined as

$$\mathrm{DP}_\alpha(a, b) = \frac{\left|x \in \mathbb{F}_q^n \colon \alpha(x + a) - \alpha(x) = b\right|}{q^n}.$$

If $\mathrm{DP}_\alpha(a, b) > 0$, we say that a and b are *compatible* differences over α. We define the weight of a differential (a, b) over α with a and b compatible as:

$$\mathrm{w}_\alpha(a, b) = -\log_q(\mathrm{DP}_\alpha(a, b)).$$

4.2 Forward propagation from a given input difference

Consider the function $\beta \colon \mathbb{F}_q \to \mathbb{F}_q : x \mapsto x^2$. Given an input pair $(x + a, x)$, the corresponding output difference b is given by

$$b = (x + a)^2 - x^2 = x^2 + 2ax + a^2 - x^2 = 2ax + a^2. \tag{1}$$

This is a linear equation and for any output difference $b \in \mathbb{F}_q$ there is exactly one input pair $(x + a, x)$. Solving (2) gives $x = (2a)^{-1}(b - a^2)$ yielding the pair

$$\left(\frac{b}{2a} + \frac{a}{2}, \frac{b}{2a} - \frac{a}{2}\right).$$

It follows that the set of output differences b compatible over β with a non-zero input difference a coincides with $\mathbb{F}_q$ and they all have $\mathrm{DP}_\beta(a,b) = q^{-1}$.

For the map γ, we have

$$b_i = x_i + a_i + (x_{i+1} + a_{i+1})^2 - x_i - x_{i+1}^2 = a_i + a_{i+1}^2 + 2a_{i+1}x_{i+1}\,, \tag{2}$$

From (2) we can characterize the full difference distribution table (DDT) of γ.

Lemma 1. *An output difference b is compatible to an input difference a over γ if for every i, $b_i = a_i$ or $a_{i+1} \neq 0$, and, if so, $\mathrm{DP}(a,b) = q^{-\mathrm{HW}(a)}$.*

Therefore, for an input difference $a \in \mathbb{F}_q^n$, the compatible output differences over γ form an affine space with dimension $\mathrm{HW}(a)$. The offset and a basis with minimal Hamming weight for such affine space is given by:

- the i-th digit of the offset is equal to a_i if $a_{i+1} \neq 0$ and 0 otherwise;
- for each non-zero digit in the input difference a, the basis contains the vector e_{i-1}.

It follows that for all b compatible with an input difference a we have $\mathrm{DP}_\gamma(a,b) = q^{-\mathrm{HW}(a)}$ and likewise $\mathrm{w}_\gamma(a,b) = \mathrm{HW}(a)$ and therefore only depends on the input difference.

4.3 Backward propagation from a given output difference

For a given output difference b, the compatible input differences do not form an affine space. However, we will show in this section how to efficiently generate all compatible input differences a with $\mathrm{DP}_\gamma(a,b) \leq W$ with W some limit weight.

To this end, we introduce the concept of compatible activity pattern. We say that an activity pattern $\widetilde{a}$ is compatible with b if there exists an input difference a compatible with b that has activity pattern $\widetilde{a}$.

The generation of all compatible input differences is done in two phases: in the first phase, we generate the set of activity patterns compatible with b, and in the second phase, we determine for each compatible activity pattern the set of compatible input differences with that pattern.

We generate the compatible activity patterns in a recursive way making use of the following facts:

- if $a_i = 0$ and $b_{i-1} = 0$ then $a_{i-1} = 0$;
- if $a_i = 0$ and $b_{i-1} \neq 0$ then $a_{i-1} \neq 0$.

We specify our algorithm in Algorithm 1. We start with a fully unspecified activity pattern k. Then we specify whether a_{n-1} is active or not (and thus whether $k_{n-1} = 1$ or 0) and based on this we incrementally determine the activity of all other digits from a_{n-2} to a_0 using the rules given above.

Given an output difference b and a compatible input activity pattern k, all compatible input differences a with activity pattern k can be determined as follows:

- if $k_i = 0$, then $a_i = 0$;
- if $k_i = 1$ and $k_{i+1} = 0$, then $a_i = b_i$;
- if $k_i = 1$ and $k_{i+1} = 1$, then a_i can have all values.

The differentials (a,b) with given output difference b and input differences a compatible with b do not all have the same weight. We define the *minimum reverse weight* of an output difference b as:

$$\mathrm{w}_\gamma^{\mathrm{rev}}(b) = \min_{a\colon \mathrm{DP}_\gamma(a,b)>0} \mathrm{w}_\gamma(a,b)\,.$$

4.4 Computing the minimum reverse weight of an output difference

The minimum reverse weight of an output difference b is fully determined by its activity vector $\widetilde{b}$ and is given by the compatible activity patterns with minimum Hamming weight.

Algorithm 1 Generation of input activity patterns compatible with output difference b

Input: difference $b \in \mathbb{F}_q^n$ at output of γ and limit weight W
Output: list L of activity patterns k compatible with b at input of γ
Coordinates in k: $*$ denotes unspecified, 0 denotes passive, 1 denotes active

$L \leftarrow$ empty
$k \leftarrow *^n$
$k_{n-1} \leftarrow 0$; $\mathsf{buildA}(n-1, k, b, W)$
$k_{n-1} \leftarrow 1$; $\mathsf{buildA}(n-1, k, b, W)$

procedure $\mathsf{buildA}(i, k, b, W)$
  if $\mathrm{HW}(k) > W$ **then return**
  if $(i = 0)$ **then**
    if $(k_{n-1} = 1)$ OR $(\widetilde{b}_0 = k_0)$ **then** add k to L
    return
  $k' \leftarrow k$
  if $(k_i = 1)$ OR $(\widetilde{b}_{i-1} = 1)$ **then** $k'_{i-1} \leftarrow 1$; $\mathsf{buildA}(i-1, k', b, W)$
  if $(k_i = 1)$ OR $(\widetilde{b}_{i-1} = 0)$ **then** $k'_{i-1} \leftarrow 0$; $\mathsf{buildA}(i-1, k', b, W)$
  return

Let a 1-run of length ℓ in $\widetilde{b}$ be a sequence of ℓ coordinates $b_i, b_{i+1}, \cdots, b_{i+\ell-1}$ with activity 1 and such that $b_{i-1} = 0 = b_{i+\ell}$ (where indexes are considered modulo n). Namely, the sequence is preceded by at least one coordinate 0 and followed by at least one coordinate 0.

We see that for each 1-run of length ℓ in $\widetilde{b}$, the digit $\widetilde{a}_{i+\ell-1}$ must be 1 and in the sequence $\widetilde{a}_i, \widetilde{a}_{i+1}, \cdots, \widetilde{a}_{i+\ell-1}$ there can be at most a single zero digit in between two active digits. It follows that for each 1-run in $\widetilde{b}$ of length ℓ, a has at least $\ell/2$ active digits if ℓ is even and $(\ell+1)/2$ if ℓ is odd. So to determine the minimum reverse weight, we decompose its output activity pattern in a sequence of 1-runs of lengths ℓ_j yielding minimum reverse weight $\sum_j \lceil \ell_j/2 \rceil$.

5 Input-output correlation properties of γ

We analyzed the correlation properties of the map γ. We will first define linear approximations and their correlations and then summarize our findings for γ.

5.1 Linear approximations, correlation and weight

Given a complex number x, we write its complex conjugate as $\overline{x}$. In the following section we will write ω as shorthand for $e^{\frac{2\pi i}{p}}$. We will also make use of the *trace* function $\mathrm{Tr}\colon \mathbb{F}_{p^d} \to \mathbb{F}_p$ as $\mathrm{Tr}(x) = \sum_{i=0}^{d-1} x^{p^i}$.

The *correlation* between two functions $f, g\colon \mathbb{F}_{p^d}^n \to \mathbb{F}_p$ is defined as:

$$\mathrm{C}(f, g) = q^{-n} \sum_{x \in \mathbb{F}_q^n} \omega^{g(x) - f(x)} .$$

For correlations of functions $f, g\colon \mathbb{F}_q^n \to \mathbb{F}_q$ we first must project the output from $\mathbb{F}_q$ to $\mathbb{F}_p$. A way to do that in a basis-agnostic way is by using the trace function:

$$\mathrm{C}(\mathrm{Tr}(uf), \mathrm{Tr}(vg)) = q^{-n} \sum_{x \in \mathbb{F}_q^n} \omega^{\mathrm{Tr}(vg(x) - uf(x))} .$$

Let α be a transformation : $\mathbb{F}_q^n \to \mathbb{F}_q^n$ with $q = p^d$. We call a pair of masks (u, v), with $u \in \mathbb{F}_q^n$ and $v \in \mathbb{F}_q^n$ a *linear approximation* over α, with u the input mask and v the output mask. The correlation of this linear approximation is the correlation between the functions $\mathrm{Tr}(u^{\mathrm{T}}x)$ and $\mathrm{Tr}(v^{\mathrm{T}}\alpha(x))$:

$$\mathrm{C}_\alpha(u, v) = q^{-n} \sum_{x \in \mathbb{F}_q^n} \omega^{\mathrm{Tr}(v^{\mathrm{T}}\alpha(x) - u^{\mathrm{T}}x)} .$$

If $\mathrm{C}_\alpha(u, v) \neq 0$, we say that masks u and v are *compatible* over α.

Correlations are, in general, complex numbers. The *linear potential (LP)* is real and related to a correlation by $\mathrm{LP}(u, v) = \mathrm{C}(u, v)\overline{\mathrm{C}}(u, v)$.

We define the weight of a linear approximation (u, v) over α with u and v compatible as

$$\mathrm{w}_\alpha(u, v) = -\log_q(\mathrm{LP}_\alpha(u, v)) .$$

5.2 Correlation properties of γ

Consider again the function β: $\mathbb{F}_{p^d} \to \mathbb{F}_{p^d} : x \mapsto x^2$. By applying Theorem 5.33 from [LN97], we obtain that the correlation between $x \mapsto vx^2$ and $x \mapsto ux$ (where $u, v \in \mathbb{F}_q$) is equal to:

$$\mathrm{C}_\beta(u, v) = \frac{1}{q} \sum_{x \in \mathbb{F}_q} \omega^{\mathrm{Tr}(vx^2 - ux)} = \begin{cases} \frac{(-1)^{d-1}}{\sqrt{q}} \omega^{\mathrm{Tr}(-u^2(4v)^{-1})} \eta(v) & \text{if } p \equiv 1 \pmod 4 \\ \frac{(-1)^{d-1}}{\sqrt{q}} i^d \omega^{\mathrm{Tr}(-u^2(4v)^{-1})} \eta(v) & \text{if } p \equiv 3 \pmod 4 \end{cases}$$

with $\eta(v) = 1$ if v is a square in $\mathbb{F}_q$ and -1 otherwise. It follows that for all $u, v \in \mathbb{F}_{p^d}$ and $v \neq 0$ we have $\mathrm{LP}_\beta(u, v) = q^{-1}$.

We can compute the correlation of linear approximations over γ from those over β:

$$\begin{aligned} \mathrm{C}_\gamma(u, v) &= q^{-n} \sum_{x \in \mathbb{F}_q^n} \omega^{\mathrm{Tr}\left(v^{\mathrm{T}}\gamma(x) - u^{\mathrm{T}}x\right)} && (3) \\ &= q^{-n} \sum_{x \in \mathbb{F}_q^n} \omega^{\mathrm{Tr}\left(\sum_i v_i(x_i + x_{i+1}^2) - u_i x_i\right)} && (4) \\ &= q^{-n} \sum_{x \in \mathbb{F}_q^n} \prod_i \omega^{\mathrm{Tr}\left((v_i - u_i)x_i + v_{i-1}x_i^2\right)} && (5) \\ &= \prod_i q^{-1} \sum_{x_i \in \mathbb{F}_q} \omega^{\mathrm{Tr}\left((v_i - u_i)x_i + v_{i-1}x_i^2\right)} && (6) \\ &= \prod_i \mathrm{C}_\beta(v_i - u_i, v_{i-1}) . && (7) \end{aligned}$$

From (3) we can characterize the full table of LPs of γ.

Lemma 2. *An input mask u is compatible to an output mask v over γ if for every i, $u_i = v_i$ or $v_{i-1} \neq 0$, and, if so,* $\mathrm{LP}(u, v) = q^{-\mathrm{HW}(v)}$.

Clearly, Lemma 1 and Lemma 2 are very alike and therefore propagation of differences and masks over γ follow similar laws. Concretely, let $\pi : \mathbb{F}_q^n \to \mathbb{F}_q^n : x \mapsto y$ with $\forall i$: $y_{-i} = x_i$. Then we have

$$\text{for } v = \pi(a), u = \pi(b) : \mathrm{LP}_\gamma(u, v) = \mathrm{DP}_\gamma(a, b) .$$

So masks propagate as differences, taking into account following correspondence:

- output masks play the role of input differences and vice versa;
- indexes shall be reversed: index i in a mask corresponds to index $-i$ in a difference.

For a nonlinear mapping this is an intriguing property that, as far as we know, occurs only in γ and the mapping χ_3 [DHVK18].

It follows that we can extend the results obtained in Section 4 to masks. In particular, for a given output mask, we can build the affine space of compatible input masks as in Section 4.2. Moreover, for a given input mask, the compatible output masks can be found by applying Algorithm 1. For a given input masks u, there can be several compatible output masks v. Among them, there will be one realizing the minimum value of $\mathrm{w}(u, v)$. The *minimum reverse weight* of u is defined as

$$\mathrm{w}_\gamma^{\mathrm{rev}}(u) = \min_{v:\ \mathrm{LP}_\gamma(u,v)>0} \mathrm{w}_\gamma(u, v)\,.$$

and is determined by the number of 1-runs in u and their weight, as in Section 4.4.

6 Non-invertibility and imbalance

A non-zero input difference a can lead to a zero output difference if 0 is in the affine space of compatible output differences, or equivalently, if its offset is 0. This can only happen if, for all positions, both a_i and a_{i+1} are active. Therefore, the input differences a that can lead to a collision are those with all coordinates active. There are $(q-1)^n$ such differences and for all of them $\mathrm{DP}(a, 0) = q^{-n}$.

Similarly, a non-zero output mask v can only be imbalanced if 0 is in the affine space of compatible input masks, or equivalently, if its offset is 0. This can only happen if, for all positions, both v_i and v_{i+1} are active. Therefore the output masks v that can lead to a collision are those with all coordinates active. There are $(q-1)^n$ such masks and for all of them $\mathrm{LP}(a, 0) = q^{-n}$.

The collision probability of a mapping is the probability that when randomly choosing two different inputs, the outputs collide. A permutation naturally has collision probability 0. A random transformation over $\mathbb{F}_q^n$ has collision probability q^{-n}: the probability that the two chosen inputs have the same image. For γ, the collision probability is the number of colliding pairs divided by the total number of pairs:

$$\frac{(q-1)^n}{\binom{q^n}{2}} = \frac{2(q-1)^n}{q^n(q^n-1)} \approx \frac{2(q-1)^n}{q^{2n}}$$

So the collision probability of γ is that of a random transformation times a factor $2\left(1-\frac{1}{q}\right)^n$.

References

[AGR+] M. R. Albrecht, L. Grassi, C. Rechberger, A. Roy, and T. Tiessen. MiMC: Efficient encryption and cryptographic hashing with minimal multiplicative complexity. Advances in Cryptology - ASIACRYPT 2016.

[DA] J. Daemen and G. Van Assche. Differential propagation analysis of keccak. Fast Software Encryption - 19th International Workshop, FSE 2012.

[DHVK18] J. Daemen, S. Hoffert, G. Van Assche, and R. Van Keer. The design of Xoodoo and Xoofff. *IACR Transactions on Symmetric Cryptology*, 2018(4):1–38, December 2018.

[GKR+21] L. Grassi, D. Khovratovich, C. Rechberger, A. Roy, and M. Schofnegger. Poseidon: A new hash function for zero-knowledge proof systems. 30th USENIX Security Symposium, 2021.

[Gra22] L. Grassi. Bounded surjective quadratic functions over f_p^n for mpc-/zk-/fhe-friendly symmetric primitives. Cryptology ePrint Archive, Paper 2022/1313, 2022. https://eprint.iacr.org/2022/1313.

[KTDB19] S. Kölbl, E. Tischhauser, P. Derbez, and A. Bogdanov. Troika: a ternary cryptographic hash function. *Designs, Codes and Cryptography*, 88(1):91–117, August 2019.

[LN97] Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Cambridge University Press, 1997.