Cryptogr Commun
DOI 10.1007/s12095-008-0003-x

New criteria for linear maps in AES-like ciphers

Joan Daemen · Vincent Rijmen

Received: 5 November 2007 / Accepted: 26 March 2008

Abstract In this paper, we study a class of linear transformations that are used as mixing maps in block ciphers. We address the question which properties of the linear transformation affect the probability of differentials and characteristics over Super boxes. Besides the expected differential probability (EDP), we also study the fixed-key probability of characteristics, denoted by DP[k]. We define *plateau characteristics*, where the dependency on the value of the key is very structured. Our results show that the distribution of the key-dependent probability is not narrow for characteristics in the AES Super box and hence the widely made assumption that it can be approximated by the EDP, is not justified. Finally, we introduce a property of linear maps which hasn't been studied before. We call this property *related differentials*. Related differentials don't influence the EDP of characteristics, but instead they affect the distribution of their DP[k] values.

Keywords AES-like ciphers · Linear maps · EDP

1 Introduction

Differential cryptanalysis was introduced in [7]. Additional concepts were defined in [23]. The resistance of symmetric encryption primitives like block ciphers against differential cryptanalysis depends both on the non-linear building blocks and the linear mixing maps interconnecting them. *Super boxes* are the smallest unit in

J. Daemen
STMicroelectronics, Zaventem Belgium
e-mail: joan.daemen@st.com

V. Rijmen (✉)
IAIK, Graz University of Technology, Graz, Austria
e-mail: Vincent.Rijmen@iaik.tugraz.at

V. Rijmen
ESAT/COSIC, K.U.Leuven, Belgium

which both linear and non-linear components are present. Bounds on the expected differential probability (EDP) of characteristics over Super boxes are the 2-round characteristic bounds given in the design documentation of Rijndael [15]. Bounds on the EDP of differentials have been investigated in [18, 28, 29]. In this paper, we concentrate on the role that the linear mixing maps play in the computation of these bounds.

Besides the EDP, we also study the fixed-key probability of characteristics, denoted by DP[k]. It has been reported before that the DP[k] of characteristics depends on the value of the key [4, 8, 9]. We define *plateau characteristics*, where the dependency on the value of the key is very structured. The fixed-key probability of these characteristics is either zero, or 2^h, with h a value that depends only on the characteristic and not on the key. We show that for a large class of ciphers, all two-round characteristics are plateau characteristics. Our results show that the distribution of the key-dependent probability is not narrow and hence the widely made assumption that it can be approximated by the EDP, is not justified.

Finally, we introduce a property of linear maps which hasn't been studied before. We call this property *related differentials.* Related differentials don't influence the EDP of characteristics, but instead they affect the distribution of their DP[k] values.

This paper is organized as follows. In Section 2 we repeat some basic definitions relevant for differential cryptanalysis. In Section 3 we introduce the concept of bundles and derive a bound on the EDP of differentials over Super boxes. In Section 4 we explain the difference between a bound on the EDP of characteristics or differentials and a proof of security. This completes the first part of this paper. In the second part, we introduce planar differentials and plateau characteristics in Section 5. We determine all characteristics through the AES Super box in Section 6. In the third part of the paper we find out how properties of the linear map cause non-uniformity in the distribution of the DP[k]. We introduce the concept of related differentials in Section 7 and present an algorithm to determine them in Section 8. We discuss the impact on AES in Section 9. We present our conclusions in Section 10.

2 Differential cryptanalysis

In differential cryptanalysis, we study the behavior of maps when inputs are applied in pairs that have a fixed difference. In particular, we are interested in the distribution of the output differences of the pairs. In order to resist differential attacks, there shouldn't be differentials that have a DP significantly higher than the average.

2.1 Differentials, characteristics and probability

We denote a *differential* [23] over a map by (a, b) and assume that it is clear from the context which map we mean. We call a the input difference and b the output difference.

Definition 1 The *differential probability* $\mathrm{DP}(a, b)$ of a differential (a, b) over a map f is the fraction of pairs with input difference a that have output difference b:

$$\mathrm{DP}(a, b) = 2^{-n}\#\{x|\, f(x+a) = f(x)+b\},$$

where n is the input length in bits of f.

For a keyed map, we can define a differential probability $\mathrm{DP}[k](a, b)$ for each value k of the key. We define the *expected differential probability* (EDP) of a differential as the average of the differential probability $\mathrm{DP}[k](a, b)$ over all keys.

Let $B[k]$ denote a keyed *composed function* consisting of a sequence of s steps $f^i[k]$:

$$B[k](x) = \left(f^s[k] \circ \cdots \circ f^2[k] \circ f^1[k]\right)(x). \tag{1}$$

A characteristic through $B[k]$ is a sequence of differences $Q = (a, b_1, b_2, \ldots, b_s)$. The sequence consists of an input difference a, followed by the output differences of all the steps of the composed function. A right pair of the characteristic for a given key k is a pair $\{a, x + a\}$ such that

$$\begin{aligned} f^1[k](x) + f^1[k](x+a) &= b_1 \\ \left(f^2[k] \circ f^1[k]\right)(x) + \left(f^2[k] \circ f^1[k]\right)(x+a) &= b_2 \\ &\ldots \\ B[k](x) + B[k](x+a) &= b_s. \end{aligned} \tag{2}$$

The differential probability of a characteristic $Q = (a, b_1, \ldots, b_s)$ is the fraction of pairs with input difference a that satisfy (2). A characteristic over a keyed composed map has a differential probability $\mathrm{DP}[k](Q)$ for each value k of the key. The EDP of a characteristic is the average of its $\mathrm{DP}[k]$ over all keys.

For Markov ciphers, the EDP of a characteristic Q is the product of the DP of its S-boxes [23]. A characteristic $Q = (a, b_1, \ldots, b_s)$ is *in* a differential (f, g) if $a = f$ and $(b_s = g)$. The $\mathrm{DP}[k]$ of a differential is the sum of the $\mathrm{DP}[k]$ values of all the characteristics in that differential

$$\mathrm{DP}[k](a, b_s) = \sum_{Q \text{ in } (a, b_s)} \mathrm{DP}[k](Q)\ , \tag{3}$$

and hence the same holds for the EDP.

By definition, differentials over a linear map l have $\mathrm{DP}(a, b) > 0$ if and only if $b = l(a)$. It follows that only nonlinear maps can resist differential cryptanalysis. However, when we study maps that are composed of several steps, both linear and nonlinear steps are important to characterize the differential properties of the composed map. In this paper, we study a relatively simple type of composed maps, which we call Super-boxes.

2.2 Super boxes

Let $S[x]$ denote a nonlinear transformation, or S-box, with domain and range $\mathrm{GF}(2^{\mathrm{n_s}})$. Extension to other types of fields is easy, but fields of characteristic two are by far the most used in symmetric cryptography.

Several ciphers that use S-boxes and linear transformations can be described using the structure of a *Super box*.

Definition 2 A *super box* maps a vector a of m elements a_i to an array e of m elements e_i. Each of the elements has size $\mathrm{n_s}$. A super box takes a key k of size $m \times \mathrm{n_s} = n$. It consists of the sequence of four transformations (or *steps*):

Substitution: m parallel applications of an invertible $\mathrm{n_s}$-bit S-box,

$$b = \mathrm{S}(a) \Leftrightarrow b_j = S[a_j],\ \ j = 0, 1, \ldots, m-1,$$

Mixing: a linear map,

$$c = \mathrm{M}(b),$$

Round key addition:

$$d_j = c_j + k_j,\ \ j = 0, 1, \ldots, m-1$$

Substitution: m parallel applications of a $\mathrm{n_s}$-bit S-box,

$$e = \mathrm{S}(d) \Leftrightarrow e_j = S[d_j],\ \ j = 0, 1, \ldots, m-1,$$

The S-boxes in the two S-box steps may also be all different.

2.3 Characteristics through super boxes

A differential characteristic through a super box consists of a sequence of 5 differences. The input difference is denoted by a, the difference after the first substitution step by b, the difference after the mixing step by c, the difference after the round key addition by d and the output difference of the super box by e. In a characteristic with EDP > 0, we always have $d = c$, so we often omit d from the notation and $c = \mathrm{M}(b)$. We denote these characteristics by (a, b, d, e) or by $(a, b, \mathrm{M}(b), e)$.

A characteristic over a super box can specify that one or more of the S-boxes have input difference 0. Such S-boxes have always output difference 0, with probability 1.

When computing the probability of a characteristic, only the S-boxes with nonzero input difference need to be taken into account. They are called *active S-boxes*. A well-known important property of the linear mixing map in a Super box is its (differential) branch number.

Definition 3 [12, 15] The differential *branch number* of a linear map M is defined as

$$\mathcal{B}(\mathrm{M}) = \min_{a \neq 0} \left(\mathrm{wt}(a) + \mathrm{wt}(\mathrm{M}(a))\right).$$

Here wt denotes the weight of the input and output vectors, which is the number of $\mathrm{n_s}$-bit components different from 0. The importance of the branch number follows from the following bound on the EDP of a characteristic over a Super box. Let δ be an upper bound on the DP(a, b) over one active S-box, then

$$\max_{Q, a \neq 0} \mathrm{EDP}(Q) \leq \delta^{\mathcal{B}(\mathrm{M})}. \tag{4}$$

We can partition the set of input vectors to a Super box by considering *truncated* differences [21]. All vectors in a given equivalence class have zeroes in the same positions and non-zero values in the other positions. An equivalence class is characterized by an *activity pattern*. The activity pattern has a single bit for each position indicating whether its value must be 0 (passive) or not (active). The activity pattern of a differential (a, e) is the couple of the activity patterns of a and e.

2.4 Linear maps and codes

A map M, that is linear over GF($2^{\mathrm{n_s}}$), can be written as a multiplication by an $m \times m$ matrix M_c:

$$c = \mathrm{M}(b) \Leftrightarrow c = b\,\mathrm{M}_c.$$

Then the vectors $(b\,, c)$ are codewords of a linear code with generator matrix $G = [I\ \mathrm{M}_c]$ and check matrix $H = [\mathrm{M}_c^t\ I]$. This code has length $2m$, dimension m and minimum distance equal to the branch number of M [15, 30]. We call this linear code the *associated* code of M.

3 Bundles of characteristics and the EDP of differentials over the AES super box

In this section we examine more closely some properties of the AES Super box.

3.1 The AES super box

The AES S-box operates on GF(2^8) and can be described as

$$S[x] = L(x^{-1}) + q, \tag{5}$$

Here x^{-1} denotes the multiplicative inverse of x in GF(2^8), extended with 0 being mapped to 0. L is a linear transformation over $GF(2)$ and q a constant. Note that L is not linear over GF(2^8) and can be expressed as a so-called *linearized polynomial* [15, 24]. The properties of the AES S-box have been studied in [16, 26] and in this paper we will not consider them in further detail.

The *AES super box* is a Super box where the elements are bytes and $m = 4$ and M is the multiplication by the MixColumns matrix, which we denote by $\mathrm{M_a}$:

$$c = \mathrm{M}(b) \Leftrightarrow c = b \times \begin{bmatrix} 2 & 1 & 1 & 3 \\ 3 & 2 & 1 & 1 \\ 1 & 3 & 2 & 1 \\ 1 & 1 & 3 & 2 \end{bmatrix} = b\,\mathrm{M_a}$$

The mixing map of AES has branch number 5, which is optimal for a map operating on vectors of length 4. The associated linear code is an MDS code. A mixing map with optimal branch number is called a multipermutation [32]. Observe that because we work here with row vectors, the matrix $\mathrm{M_a}$ is given in transposed form compared to the more common notation [1, 15].

If we consider two AES rounds, swap the steps ShiftRows and SubBytes in the first round, and remove all linear operations before the first application of SubBytes and after the second application of SubBytes, then we obtain a map that can also be described as 4 parallel instances of the AES super box. This map has the same distributions of DP[k] and EDP as two AES rounds.

3.2 Bundles

Since $d = c = b\mathrm{M_a}$, a characteristic through the AES Super box is fully determined by the differential (a, e) it is in and the intermediate difference b. For the EDP of a differential over the AES super box, we have:

$$\mathrm{EDP}(a, e) = \sum \mathrm{EDP}(a, b, b\mathrm{M_a}, e) = \sum_b \mathrm{EDP_S}(a, b)\mathrm{EDP_S}(b\mathrm{M_a}, e)\ . \tag{6}$$

with $\mathrm{EDP_S}(x, y)$ the EDP of a differential (x, y) over SubBytes. Since the AES S-box is invertible, $\mathrm{EDP_S}(a, b)$ and $\mathrm{EDP_S}(b\mathrm{M_a}, e)$ can be non-zero only if a and b, respectively $b\mathrm{M_a}$ and e have the same activity pattern.

If we are able to compute the number of characteristics for which both $\mathrm{EDP_S}(a, b)$ and $\mathrm{EDP_S}(b\mathrm{M_a}, e)$ are non-zero, we can use (4) to derive an upper bound on the EDP of a differential (a, e). The number of characteristics can be determined by means of *bundles*, which we define below. We start with an example.

Example 1 Consider the characteristics in a differential (a, e) with $a = [\mathrm{a}_0, 0, 0, 0]$. Then clearly we must have $b = [\mathrm{b}_0, 0, 0, 0]$ and thanks to MixColumns we have $c_0 = 2b_0$, $c_1 = b_0$, $c_2 = b_0$ and $c_3 = 3b_0$, or equivalently $c = b_0[2, 1, 1, 3]$, where $b_0[2, 1, 1, 3]$ denotes the scalar multiplication of the vector $[2, 1, 1, 3]$ with the (non-zero) scalar b_0. There are 255 characteristics in the differential, one for each nonzero value of b_0.

This can be generalized to any AES super box differential with 5 active S-boxes. If $Q = (a, b, c, e)$ and $Q' = (a, b', c', e)$ are two characteristics of the same differential with 5 active S-boxes, then there exists a γ such that $b\ = \gamma b'$, and $c = \gamma c'$. We define a *bundle* as follows.

Definition 4 The *bundle* $B(u^{\mathrm{b}})$ associated with the vector u^{b}, is the set of 255 vectors defined as follows:

$$B(u^{\mathrm{b}}) = \{\gamma u^{\mathrm{b}} | \gamma \in \mathrm{GF}(2^8) \setminus \{0\}\}\ .$$

Scalar multiplication doesn't change the activity pattern of a vector. Furthermore, MixColumns is linear over $\mathrm{GF}(2^8)$: $(\gamma b)\mathrm{M_a} = \gamma(b\mathrm{M_a})$. Hence also the activity pattern of $u^{\mathrm{c}} = u^{\mathrm{b}}\mathrm{M_a}$ is the same for all vectors u^{b} of a bundle. If $(a, u^{\mathrm{b}}, u^{\mathrm{c}}, e)$ is a characteristic through the AES super box, then $(a, b, b\mathrm{M_a}, e)$ is a characteristic through the AES super box $\forall b \in B(u^{\mathrm{b}})$. Hence, the set of characteristics in (a, e) can be partitioned into a number of classes. Each class contains the 255 characteristics $(a, b, b\mathrm{M_a}, e)$ defined by keeping a, e constant and varying b over all the values of a bundle $B(u^{\mathrm{b}})$. In the following, we use 'bundle' also to refer to such a class of characteristics. A characteristic in the bundle $B(u^{\mathrm{b}})$ of the differential (a, e) is uniquely identified by the value of γ.

3.3 Bundles and activity patterns

Since MixColumns has branch number 5, activity patterns of differentials over the AES Super box must have a minimum of five active positions or else their EDP = 0. A simple counting reveals that there are 56 activity patterns with five active positions, 28 with six active positions, eight with seven active positions and one with eight active positions.

As explained in Example 1, a differential with five active S-boxes only has a single bundle of characteristics. Table 1 lists the activity patterns with five active S-boxes and the corresponding values of $(u^{\mathrm{b}}, u^{\mathrm{c}})$ for the AES super box. The 56 activity patterns with 5 active positions can be derived by rotation of the 14 activity patterns listed.

For the bundles of a differential with 6 active positions, the u^{b} values can be found by taking (almost) all possible combinations of two u^{b} values of bundles with 5 active positions. For example, for activity pattern (1110; 1110) we combine the bundles for (1010; 1110) and (0110; 1110) as given by Table 1. This gives:

$$u^{\mathrm{b}} = [1, 0, 3, 0] + \epsilon[0, 1, 1, 0] = [1, \epsilon, 3 + \epsilon, 0]$$

$$u^{\mathrm{c}} = [1, 4, 7, 0] + \epsilon[2, 1, 3, 0] = [1 + 2\epsilon, 4 + \epsilon, 7 + 3\epsilon, 0].$$

This results in 255 different bundles, one for each nonzero value of ϵ. However, not all nonzero values of ϵ are admissible. For example, if $\epsilon = 3$ then we obtain

$$u^{\mathrm{b}} = [1, 3, 0, 0],$$

$$u^{\mathrm{c}} = [7, 7, 2, 0]$$

which doesn't have activity pattern (1110; 1110). For $u^{\mathrm{b}}, u^{\mathrm{c}}$ to have activity pattern (1110; 1110) the value of ϵ must be different from 3, $1 \cdot 2^{-1}$, 4 and $7 \cdot 3^{-1}$. Hence, a differential with 6 active S-boxes has 251 bundles.

Denote the number of bundles for an activity pattern with x active S-boxes by $\mathrm{BN}(x)$. For all Super boxes with the same dimensions as the AES Super box, and where the associated linear code of the mixing map is an MDS code, we have:

$$\begin{array}{lcr}
\mathrm{BN}(5) & = & 1 \\
\mathrm{BN}(6) = 255 - 4\mathrm{BN}(5) & = & 251 \\
\mathrm{BN}(7) = 255^2 - 4\mathrm{BN}(6) - 6\mathrm{BN}(5) & = & 64015 \\
\mathrm{BN}(8) = 255^3 - 4\mathrm{BN}(7) - 6\mathrm{BN}(6) - 4\mathrm{BN}(5) & = & 16323805
\end{array}$$

Table 1 Activity patterns with five active S-boxes for the AES Super box and the corresponding values of $(u^{\mathrm{b}}, u^{\mathrm{c}})$ (in hexadecimal notation)

Activity pattern	u^{b}	u^{c}
(1000;1111)	[1, 0, 0, 0]	[2, 1, 1, 3]
(1100;1011)	[2, 1, 0, 0]	[7, 0, 3, 7]
(0110;1110)	[0, 1, 1, 0]	[2, 1, 3, 0]
(0011;1011)	[0, 0, 1, 3]	[2, 0, 7, 7]
(1001;1110)	[2, 0, 0, 3]	[7, 1, 7, 0]
(1010;0111)	[2, 0, 1, 0]	[5, 1, 0, 7]
(0101;0111)	[0, 1, 0, 3]	[0, 1, 4, 7]
(0111;0101)	[0, 1, 4, 7]	[0, 9, 0, B]
(0111;1010)	[5, 1, 0, 7]	[E, 0, D, 0]
(1110;1001)	[7, 1, 3, 0]	[E, 0, 0, B]
(1011;0011)	[2, 0, 3, 7]	[0, 0, D, B]
(1110;0110)	[2, 1, 7, 0]	[0, 9, D, 0]
(1011;1100)	[7, 0, 7, 7]	[E, 9, 0, 0]
(1111;1000)	[E, 9, D, B]	[1, 0, 0, 0]

The total number of nonzero vectors of 4 bytes is $2^{32}-1$. Each bundle groups 255 such vectors, so the total number of bundles is

$$\frac{2^{32}-1}{2^8-1} = 2^{24} + 2^{16} + 2^8 + 1\ .$$

3.4 Closing

In this section, we gave a short overview of the properties of MixColumns that are relevant for the computation of the EDP of differentials. More results on two-round AES differentials can be found in [16, 18, 19, 28, 29]. In the remainder of this paper, we move our attention from the EDP to the distribution of the DP[k] of characteristics.

4 EDP versus DP

An important example of composed maps is formed by block ciphers. Modern block ciphers are designed to resist differential cryptanalysis [23]. The constructions in [25, 27] providing *provable security* against differential cryptanalysis consider only the EDP values of characteristics and differentials, hence *average-case* behavior.

Often, we are interested in the *worst-case* behavior of a cipher. In order to translate bounds on the EDP to bounds on the worst-case behavior, designers sometimes rely on the

Hypothesis 1 (Hypothesis of Stochastic Equivalence [23]) *For all differentials* (a, b)*, it holds that for most values of the key k,* $\mathrm{DP}[k](a, b) \approx \mathrm{EDP}(a, b)$.

In statistical terms, the assumption is made that all values present in the distribution DP[k] are close to the mean value of the distribution. This hypothesis has been observed to hold in toy ciphers [22], but not in ciphers that are used in practice, for instance DES [9, 20] and IDEA [10].

It follows from Definition 1 that the hypothesis can't hold for characteristics. The DP[k] of a characteristic over a map is always an integer multiple of 2^{-n}. In fact, when we are working in a field with characteristic two, the symmetry between addition and subtraction implies that DP[k] is always an integer multiple of 2^{1-n}. The EDP however, can easily take lower values. For instance, characteristics over the AES super box where all eight S-boxes have a non-zero input difference, have $EDP \leq 2^{-48}$. It follows that a bound on the EDP doesn't imply that the DP values are bounded.

This problem has of course been observed before and therefore 'folk lore' sometimes uses the hypothesis in a somewhat weaker form, namely if $\mathrm{EDP}(a, b) \ll 2^{1-n}$, then it is often assumed that for the overwhelming majority of the keys, DP[k] will be equal to either zero or 2^{1-n}, the minimum non-zero value.

In the next section we will see examples where the DP[k] values are distributed completely differently.

5 Two-round plateau characteristics

In this section we show that for a large class of ciphers, two-round characteristics have a DP$[k]$ that can take only two values.

5.1 Planar differentials and maps

As customary [7], we consider ordered pairs of inputs, but we denote them by curled braces '{}' in order to avoid confusion with differentials.

Let $F_{(a,b)}$ denote the set containing the inputs x for which the pair $\{x, x+a\}$ is a right pair for the differential (a, b). Let $G_{(a,b)}$ denote the set containing the corresponding outputs. We define planar differentials as follows:

Definition 5 ([17]) A differential (a, b) is *planar* if $F_{(a,b)}$ and $G_{(a,b)}$ form affine subspaces [2]:

$$F_{(a,b)} = u + U_{(a,b)}$$

$$G_{(a,b)} = v + V_{(a,b)}$$

with $U_{(a,b)}$ and $V_{(a,b)}$ vector spaces, u any element in $F_{(a,b)}$ and v any element in $G_{(a,b)}$.

If $F_{(a,b)}$ contains an element x, then it also contains $x+a$. Hence if $F_{(a,b)}$ is not empty, then $a \in U_{(a,b)}$. The number of elements in $F_{(a,b)}$ is $2^{\dim(U_{(a,b)})}$, so $\dim(U_{(a,b)}) = n + \log_2(\mathrm{DP}(a, b))$. Similarly, we have $b \in V_{(a,b)}$ and $\dim(V_{(a,b)}) = n + \log_2(\mathrm{DP}(a, b))$. We can now prove the following.

Lemma 1 ([17]) *The following differentials are always planar differentials:*

1. *A differential which has exactly two right pairs,*
2. *A differential which has exactly four right pairs,*
3. *A differential with* DP $= 1$.

Examples of differentials with DP $= 1$ are the trivial differential $(0, 0)$ and differentials over linear maps. If $\mathrm{DP}(a, b) = 2^{t-n}$, with $t \notin \{1, 2, n\}$, the differential may or may not be planar.

Definition 6 ([17]) A map is planar if all differentials over it are planar.

Any map for which all non-trivial differentials have $\mathrm{DP}(a, b) \leq 2^{2-n}$ is planar. Such maps are called differentially four-uniform [26]. Now we give two lemmas on planar differentials over composed maps. The first lemma applies for instance to a substitution step in a block cipher, consisting of the parallel application of some S-boxes.

Lemma 2 ([17]) *Let $y = \alpha(x)$ be a map consisting of a set of parallel maps $y_i = \alpha_i(x_i)$ with $x = (x_0, x_1, \ldots, x_t)$ and $y = (y_0, y_1, \ldots, y_t)$. A differential (a, b) for which the differentials (a_i, b_i) are planar, is planar.*

We have

$$\begin{aligned} U_{(a,b)} &= U_{(a_0,b_0)} \times U_{(a_1,b_1)} \times \cdots \times U_{(a_t,b_t)} \\ u_{(a,b)} &= \left(u_{(a_0,b_0)}, u_{(a_1,b_1)}, \ldots, u_{(a_t,b_t)}\right) \\ V_{(a,b)} &= V_{(a_0,b_0)} \times V_{(a_1,b_1)} \times \cdots \times V_{(a_t,b_t)} \\ v_{(a,b)} &= \left(v_{(a_0,b_0)}, v_{(a_1,b_1)}, \ldots, v_{(a_t,b_t)}\right) \end{aligned}$$

with × denoting the direct product [2]. The following lemma applies to a sequence of maps.

Lemma 3 ([17]) *If* (a, b) *is a planar differential of* α*, then for any pair of affine maps* L_1 *and* L_2 *with* L_1 *invertible, the differential* $(L_1(a), L_2(b))$ *is planar over* $L_2 \circ \alpha \circ L_1^{-1}$.

Examples of ciphers in which single-round differentials are planar are the AES, but also 3-Way [11], SHARK [31], Square [13], Camellia [5], Serpent [3] and Noekeon [14]. Some other popular maps that are planar, are the majority function $f(x, y, z) = xy$ xor xz xor yz and the 'if' function $g(x, y, z) = xy$ xor $(\neg x)z$.

5.2 Plateau characteristics

Similar to the concept of plateaued functions [33], for which the Walsh spectrum takes only two values (in absolute value), we define *plateau characteristics* as characteristics for which the DP[k] takes only two values (where one value is always zero). The *height* of a plateau characteristic determines how high the non-zero DP[k] value of the plateau characteristic is.

Definition 7 ([17]) A characteristic Q is a *plateau characteristic* with height $\mathrm{h}(Q)$ if and only if the following holds:

1. For a fraction $2^{n-\mathrm{h}(Q)}\mathrm{EDP}(Q)$ of the keys, $\mathrm{DP}[k](Q) = 2^{\mathrm{h}(Q)-n}$, and
2. For all other keys, $\mathrm{DP}[k](Q) = 0$.

The height of a plateau characteristic can be bounded as follows. Firstly, $\mathrm{h}(Q) \leq n$. Secondly, $\mathrm{h}(Q)$ is maximal when all but one key have DP equal to zero. Denoting the number of keys by $2^{\mathrm{n_k}}$, we obtain that in this case EDP equals $2^{-\mathrm{n_k}}$ times the non-zero DP value. Taking the logarithm, we obtain $\log_2(\mathrm{EDP}(Q)) = -\mathrm{n_k} + \mathrm{h}(Q) - n$. Hence, we have in all cases $\mathrm{h}(Q) \leq \mathrm{n_k} + n + \log_2(\mathrm{EDP}(Q))$. We can now prove the following result on an n-bit map consisting of two steps and an addition with an n-bit key in between (hence $\mathrm{n_k} = n$).

Theorem 1 ([17]) *A characteristic* $Q = (a, b, c)$ *over a map consisting of two steps with a key addition in between, in which the differentials* (a, b) *and* (b, c) *are planar, is a plateau characteristic with* $\mathrm{h}(Q) = \dim(V_{(a,b)} \cap U_{(b,c)})$.

Only if $\mathrm{h}(Q) = n \log_2(\mathrm{EDP}(Q))$, it holds that $\mathrm{DP}[k](Q) = \mathrm{EDP}(Q)$ for all keys. This can only be the case for characteristics with $\mathrm{EDP}(Q) > 2^{-n}$. This theorem is valid for all ciphers in which single-round differentials are planar and round keys are applied with XOR. This includes all ciphers mentioned in Section 5.1.

6 Characteristics over the AES super box

The AES super box satisfies the criteria of Theorem 1 and hence all characteristics Q in the AES super box are plateau characteristics. $\mathrm{DP}[k](Q)$ can be described by defining $W = V_{(a,c)} \cap U_{(c,e)}$ and $V_{(a,c)} = \mathrm{M_a}(V_{(a,b)})$, where $\mathrm{M_a}(V) = \{v\mathrm{M_a} | v \in V\}$. Table 2 gives the number of characteristics over the AES super box for each possible

Table 2 Number of characteristics (binary logarithm) per number of active S-boxes, EDP and height for the AES super box

No. active	$-\log_2(\mathrm{EDP})$	Height				
S-boxes		1	2	3	4	5
5	30	–	12.6	12.6	10.6	6.2
	31	20.9	22.1	21.2	18.1	11.0
	32	29.8	30.0	28.2	23.4	–
	33	37.1	36.9	33.7	26.4	–
	34	43.2	42.9	36.2	–	–
	35	48.0	47.5	–	–	–
6	36	20.7	15.6	8.3	3.8	–
	37	30.3	24.2	16.3	11.6	–
	38	38.6	31.5	23.1	17.5	–
	39	46.1	38.1	28.9	–	–
	40	52.6	44.0	33.4	–	–
	41	58.3	49.3	–	–	–
	42	62.7	53.4	–	–	–
7	42	27.0	15.7	5.3	–	–
	43	36.8	24.3	13.1	–	–
	44	45.3	31.7	19.5	–	–
	45	53.1	38.0	24.9	–	–
	46	60.0	43.5	–	–	–
	47	66.3	48.0	–	–	–
	48	71.7	50.9	–	–	–
	49	75.9	–	–	–	–
8	48	32.0	14.7	1.0	–	–
	49	41.9	23.7	9.0	–	–
	50	50.7	31.4	15.0	–	–
	51	58.7	38.3	–	–	–
	52	66.0	44.5	–	–	–
	53	72.7	49.9	–	–	–
	54	78.7	54.1	–	–	–
	55	83.7	–	–	–	–
	56	87.9	–	–	–	–
Total		87.9	55.0	36.6	26.6	11.0

combination of height, number of active S-boxes and EDP. We see that the EDP of the characteristics ranges from 2^{-30} to 2^{-56} and the height from 1 to 5. It follows from the data in the table that the ratio

$$\mathrm{DP}[k](Q)/\mathrm{EDP}(Q)$$

ranges from 1 to 2^{25}. We call characteristics for which the ratio is 1 *flat characteristics* because for these the equality $\mathrm{DP}[k](Q) = \mathrm{EDP}(Q)$ holds for all keys. Table 2 shows that there are in total $2^{20.9}$ flat characteristics: those with $\mathrm{EDP}(Q) = 2^{-30}$ and $\mathrm{h}(Q) = 2$, and those with $\mathrm{EDP}(Q) = 2^{-31}$ and $\mathrm{h}(Q) = 1$.

The characteristics that are the most interesting for differential attacks, are the characteristics with the highest EDP or DP. They are in the top rows of the table. We see that exactly these characteristics have the highest heights, hence the most variation between DP values for different keys. Characteristics with height 5 have a DP equal to $32/2^{32}$, which is almost three times higher than the maximal MEDP of a differential ($13.25/2^{32}$ [19, 28, 29]).

There are 72 characteristics of height 5 and EDP 2^{-30}. These characteristics have nonzero DP[k] for a fraction $2^{32-30-5} = 2^{-3}$ of all keys. For a given key this results in an expected value of 9 such characteristics with $\mathrm{DP}[k] = 2^5/2^{32}$. Similarly, there are 2^{11} characteristics of height 5 and EDP 2^{-31} resulting in an expected value of 2^7 such characteristics with $\mathrm{DP}[k] = 2^5/2^{32}$. This totals to an expected number of 137 characteristics with $\mathrm{DP}[k] = 2^5/2^{32}$ per key for the AES super box. The table shows also that it is easy to find characteristics Q_1, Q_2 with $\mathrm{EDP}(Q_1) < \mathrm{EDP}(Q_2)$ and $\mathrm{h}(Q_1) > \mathrm{h}(Q_2)$.

The results in Table 2 are due to a combination of the properties of MixColumns and the AES S-boxes. Experiments with Super boxes using different S-boxes showed that characteristics with height larger than 1 appear always. This is due to a property of MixColumns, which we investigate in the next sections.

7 Related differentials

In this section we define related differences and related differentials. We show that the existence of related differentials influences the height of the characteristics through Super boxes, and this *independent of the choice of S-boxes*.

7.1 Definitions

Definition 8 Two vectors x, $x^{\diamond}$ containing each m elements of $\mathrm{n_s}$ bits are *related* if and only if

$$x_j x_j^{\diamond}(x_j + x_j^{\diamond}) = 0, \text{ for } j = 0, 1, \ldots, m-1. \tag{7}$$

The all zero vector is trivially related to all vectors, and we exclude it from now on. If x, $x^{\diamond}$ are two related differences, then the differences $x, x + x^{\diamond}$ are also related. The following condition is equivalent to (7)

$$x_j = 0 \text{ or } x_j^{\diamond} = t_j x_j, t_j \in \{0, 1\}, \; j = 0, 1, \ldots, m-1. \tag{8}$$

Two related differences define a special type of second order differential [21]. Any second order differential defines quartets $\{p, p+x, p+x^{\diamond}, p+x+x^{\diamond}\}$. If the differences x and $x^{\diamond}$ are related, then it follows that the sets

$$\{p_j, p_j+x_j, p_j+x_j^{\diamond}, p_j+x_j+x_j^{\diamond}, \}, \quad j=0,1,\ldots,m-1.$$

contain only two different elements. This is illustrated in Fig. 1. Related differences can be combined to related differentials.

Definition 9 Two differentials $(b\,, c)$, $(b^{\diamond}, c^{\diamond})$ for a linear map M are *related differentials* if and only if $c = \mathrm{M}(b)$, $c^{\diamond} = \mathrm{M}(b^{\diamond})$, the differences $b\,, b^{\diamond}$ are related differences and the differences c, $c^{\diamond}$ are related differences.

The following differentials are related differentials over the map $c = b\mathrm{M}_{\mathrm{a}}$.

$$\begin{aligned} b\ &= [0,1,4,7], & c &= b\mathrm{M}_{\mathrm{a}} &&= [0,9,0,\mathrm{B}] \\ b^{\diamond} &= [5,1,0,7], & c^{\diamond} &= b^{\diamond}\mathrm{M}_{\mathrm{a}} &&= [\mathrm{D},0,\mathrm{E},0] \\ b\ + b^{\diamond} &= [5,0,4,0], & c+c^{\diamond} &= (b\ + b^{\diamond})\mathrm{M}_{\mathrm{a}} &&= [\mathrm{D},9,\mathrm{E},\mathrm{B}]. \end{aligned} \tag{9}$$

7.2 Related differentials and plateau characteristics

Theorem 2 *Let* $Q = (a, b\,, c, e)$ *be a characteristic through a Super box with* $\mathrm{EDP}(Q) > 0$. *If* $(b\,, c)$ *is in a set of related differentials over the mixing map, then* Q *is a plateau characteristic with* $\mathrm{h}(Q) \geq 2$.

Proof Theorem 1 states that all characteristics through a Super box are plateau characteristics. Hence we only need to show that for each pair $P_x = \{x, x+a\}$, that is a right pair of Q for a certain round key value k, there exists a pair

$$P_y = \{x+a^{\diamond}, x+a^{\diamond}+a\} \neq \{x+a, x\}$$

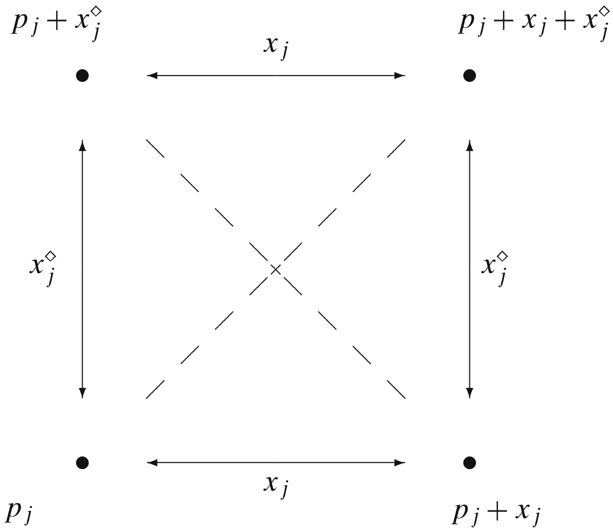

Fig. 1 A second order differential and the associated quartet can be represented by a *square*. If one of the differences x_j, $x_j^{\diamond}$ or $x_j + x_j^{\diamond}$ equals zero, then the *square collapses to a line*

that is also a right pair of Q for the same value of the round key. Let $(b^\diamond, c^\diamond)$ denote a differential that is related to (b, c). Define the difference $a^\diamond$ as the solution of:

$$\mathrm{S}(x + a^\diamond) + \mathrm{S}(x) = b^\diamond \Leftrightarrow a^\diamond = \mathrm{S}^{-1}(b^\diamond + \mathrm{S}(x)) + x. \tag{10}$$

Then the pair P_y is a right pair of Q because of the following:

- If S would be a linear map, then (10) would imply that $\mathrm{S}(x + a + a^\diamond) + \mathrm{S}(x + a) = b^\diamond$ and we would have:

  $$\mathrm{S}(x+a^\diamond)+\mathrm{S}(x+a^\diamond + a)=(\mathrm{S}(x)+b^\diamond)+(\mathrm{S}(x + a)+b^\diamond)=\mathrm{S}(x)+\mathrm{S}(x+a)=b\,. \tag{11}$$

  Since S is not linear, the equality doesn't hold for all $a^\diamond$. However, because b and $b^\diamond$ are related differences, for each j at least one of the three differences b_j, $b_j^\diamond$ and $b_j + b_j^\diamond$ equals zero. Since S uses invertible S-boxes, it follows that for each j also at least one of the three differences a_j, $a_j^\diamond$ and $a_j + a_j^\diamond$ equals zero. Hence the sets

  $$\{x_j, x_j + a_j, x_j + a_j^\diamond, x_j + a_j + a_j^\diamond\}, \quad j = 0, 1, \ldots, m - 1.$$

  contain only two different elements. Consequently, (11) holds and P_y is a right pair for the first substitution step.
- P_y is a right pair for the mixing step:

  $$\mathrm{M}(\mathrm{S}(x + a^\diamond)) + \mathrm{M}(\mathrm{S}(x + a^\diamond + a)) = \mathrm{M}(\mathrm{S}(x + a^\diamond) + \mathrm{S}(x + a^\diamond + a)) = \mathrm{M}(b) = c.$$

- Let v denote the value $\mathrm{M}(\mathrm{S}(x)) + k$, which is the output of the round key addition for the input x. Since x is in the right pair P_x, we know that $v \in F_{(c,e)}$. From the previous steps, it follows also that $\mathrm{S}(x + a^\diamond) + \mathrm{S}(x) = b^\diamond$ and

  $$\mathrm{M}(\mathrm{S}(P_y)) + k = \{v + c^\diamond, v + c^\diamond + c\}.$$

  Now we use the fact that c and $c^\diamond$ are related differences and by following a reasoning similar as for the first substitution step, we conclude that P_y is a right pair for the last substitution step and hence it is a right pair for the Super box.

□

8 Determining the related differentials

In this section, we derive an algorithm that produces for a given linear mixing map all the sets of related differentials. To simplify the description, we will assume that the linear map consists of the multiplication by an $m \times m$ matrix M_c and that the associated linear code is an MDS code. We start with an example.

8.1 First example

Assume we have a differential (b, c) with $c = b\mathrm{M_a}$ through the mixing map of the AES Super box with activity pattern (0111; 0101). We want to determine whether there exist differences $b^\diamond, c^\diamond = b^\diamond \mathrm{M_a}$ satisfying (7).

Firstly, we know that the linear code associated with MixColumns has minimal distance 5, and hence if $b_0 = c_0 = c_2 = 0$, then all other b_j, c_j are different from zero. Equation (7) doesn't put any constraints on $b^\diamond_0$, $c^\diamond_0$, and $c^\diamond_2$. From $c^\diamond = b^\diamond \mathrm{M_a}$ we can derive one equation from which these 3 elements are eliminated:

$$3c^\diamond_1 + c^\diamond_3 = 7b^\diamond_1 + 4b^\diamond_2 + b^\diamond_3.$$

Using (8) we obtain:

$$3t_5 c_1 + t_7 c_3 = 7t_1 b_1 + 4t_2 b_2 + t_3 b_3,\ t_j \in \{0, 1\} \tag{12}$$

Secondly, we know from Section 3 that there is only one bundle with activity pattern (0111; 0101). Hence:

$$(b, c) = (0, \gamma, 4\gamma, 7\gamma; 0, 9\gamma, 0, \mathrm{B}\gamma) \text{ with } \gamma \in \mathrm{GF}(2^8) \setminus \{0\}. \tag{13}$$

Combining (12) and (13) gives

$$7\gamma t_1 + 10\gamma t_2 + 7\gamma t_3 + 1\mathrm{B}\gamma t_5 + \mathrm{B}\gamma t_7 = 0.$$

For any value of γ, we obtain a system of linear equations over GF(2). The 5 unknowns are t_1, t_2, t_3, t_5, t_7. The number of independent solutions depends on the dimension of the vector space spanned by

$$\{7\gamma, 10\gamma, 7\gamma, 1\mathrm{B}\gamma, \mathrm{B}\gamma\}.$$

Note that the dimension here is determined over GF(2): linear dependencies between vectors must have binary coefficients. This dimension is always at most 4, since we know that setting all $t_j = 1$ gives the solution (b, c). In this case, the dimension equals 3, which means that there is one other solution $(b^\diamond, c^\diamond)$, hence the weight of a characteristic with (b, c) as input, respectively output difference for MixColumns, has height at least two.

8.2 For any given differential

Let (b, c) be the given differential for which we want to find a related differential over a given linear map. Denote by z the number of non-zero elements in (b, c), minus m. This implies that the number of zero elements in (b, c) equals $m - z$. Denote by H the check matrix of the linear code associated to the linear map. We know that (b, c) and the related differential(s) $(b^\diamond, c^\diamond)$ must be vectors of the associated code:

$$H(b, c)^t = H(b^\diamond, c^\diamond)^t = 0.$$

This defines a first set of m constraints, one in each row of H. Secondly, for the indices j where b_j or c_j are different from zero, we get the conditions (8) on t_j. Every time we have a b_j or c_j equal to zero, we have no condition on the corresponding $b^\diamond_j$ or $c^\diamond_j$. Therefore we eliminate these unknowns from the set of conditions.

We denote by H_p the check matrix of a new linear code, where the code vectors can take any value in the positions where the activity pattern of (b, c) is zero, and where the conditions on the values in the other positions are the same as in the code associated to the linear map. We denote the elements of H_p by $h_{i,j}$ and write:

$$\sum_j h_{i,j} b_j^\diamond + \sum_j h_{i,j+m} c_j^\diamond = 0,\ i = 0, \ldots, z-1.$$

In these equations, we fill out (8) and obtain

$$\sum_{j=0, b_j \neq 0}^{m-1} h_{i,j} b_j t_j + \sum_{j=0, c_j \neq 0}^{m-1} h_{i,j+m} c_j t_{j+m} = 0,\ i = 0, \ldots, z-1.$$

The solutions t_j are the codewords of the binary code with as check matrix:

$$D_p = \begin{bmatrix} h_{0,0}b_0 & h_{0,1}b_1 & \ldots & h_{0,2m-1}c_{m-1} \\ h_{1,0}b_0 & h_{1,1}b_1 & \ldots & h_{1,2m-1}c_{m-1} \\ \ldots & \ldots & \ldots & \ldots \\ h_{z-1,0}b_0 & h_{z-1,1}b_1 & \ldots & h_{z-1,2m-1}c_{m-1} \end{bmatrix}, \tag{14}$$

except for the codeword $(1\ 1\ \ldots\ 1)$, which corresponds to the original difference (b, c). The number of independent solutions for this set of equations depends on the rank of D_p: if $\text{rank}(D_p) < z + m - 1$, then related differentials exist. Note that D_p is a matrix containing elements of $\text{GF}(2^{\text{n}_\text{s}})$, but we determine the rank over GF(2): linear dependencies must have binary coefficients.

8.3 For all differentials with the same activity pattern

The previous method can be done in parallel for all differences with a given activity pattern p. Denote by G the generator matrix of the associated linear code. All differentials can be written as a linear combination of rows of G. The differentials with activity pattern p (plus the zero vector) form a sub-space of the linear code. The sub-space contains only the code vectors that are zero in the positions where the activity pattern is zero. We denote the generator matrix for this sub-space by G_p.

We now define z parameters ϵ_k, write $(b, c) = (\epsilon_0, \ldots, \epsilon_{z-1})G_p = \epsilon G_p$ and apply the method described in Section 8.2 to determine related differentials. The elements of matrix D_p of (14) now depend on the vector ϵ. Denoting the elements of G_p by $g_{i,j}$ and the elements of $D_p(\epsilon)$ by $d_{i,j}$, we obtain:

$$d_{ij} = h_{ij} \sum_{k=0}^{z-1} \epsilon_k g_{kj}$$

The number of dependent columns in $D_p(\epsilon)$ may depend on ϵ. Any non-zero vector ϵ for which there are more than two codewords in the binary code with $D_p(\epsilon)$ as check matrix, defines the difference (b, c) for which there exists a related differential. The related differential is again determined by the codeword that is different from $(1, 1, \ldots, 1)$. Figure 2 summarizes the algorithm.

Input: $m \times m$ matrix M_c defining the linear map, activity pattern p.
Output: Related differentials (b, c) (with activity pattern p) and $(b^\diamond, c^\diamond)$.
Algorithm:

1. Compute $G = [I\ \mathrm{M}_c]$ and $H = [\mathrm{M}_c^t\ I]$. Let $z = \mathrm{wt}(p) - m$.
2. Perform elementary row operations on H to compute an equivalent matrix SH where z rows have zeroes in the $m - z$ columns corresponding to the zero bits in p. Denote the sub-matrix of SH consisting of these rows by H_p, with elements $h_{i,j}$.
3. Similarly, perform elementary row operations on G to compute an equivalent matrix TG where z rows have zeroes in the $m - z$ columns corresponding to the zero bits in p. Denote the sub-matrix of TG consisting of these rows by G_p, with elements $g_{i,j}$.
4. Define the matrix $D(\epsilon)$ as follows:
   $$d_{ij} = h_{ij} \sum_{k=0}^{z-1} \epsilon_k g_{kj},$$
   where $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_{z-1}$ is a vector of parameters.
5. Compute the values of ϵ for which the rank of $D_p(\epsilon)$ is below $z + m - 1$.
6. For each of the outputs of the previous step compute $(b, c) = \epsilon G_p$. Compute the binary codeword $t = (t_0, \ldots, t_{2m-1})$ from $D_p^t t = 0$ and $t \neq (1, \ldots, 1)$. For the positions where $b_j, c_j \neq 0$, compute $b_j^\diamond = t_j b_j$, $c_j^\diamond = t_{j+m} c_j$. Determine the remaining $b_j^\diamond, c_j^\diamond$ such that $H(b_j^\diamond, c_j^\diamond)^t = 0$.
7. Output (b, c) and $(b^\diamond, c^\diamond)$.

Fig. 2 Algorithm to compute related differentials where one of the differentials has a given activity pattern p. If the algorithm terminates without finding related differentials, then there exist none for this activity pattern

8.4 Second example

We now illustrate the algorithm described in Fig. 2. Consider the activity pattern $p = (1010; 1111)$ for the mixing map of the AES Super box. We have $z = 2$ and:

$$H = \begin{bmatrix} 2 & 3 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 3 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 2 & 3 & 0 & 0 & 1 & 0 \\ 3 & 1 & 1 & 2 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad SH = \begin{bmatrix} 7 & 0 & 7 & 1 & 2 & 3 & 0 & 0 \\ 3 & 1 & 2 & 0 & 1 & 1 & 0 & 0 \\ \mathrm{B} & 0 & 9 & 0 & 7 & 4 & 1 & 0 \\ \mathrm{E} & 0 & \mathrm{D} & 0 & 5 & 7 & 0 & 1 \end{bmatrix},$$

$$H_p = \begin{bmatrix} \mathrm{B} & 0 & 9 & 0 & 7 & 4 & 1 & 0 \\ \mathrm{E} & 0 & \mathrm{D} & 0 & 5 & 7 & 0 & 1 \end{bmatrix}$$

Further,

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 & 1 & 1 & 3 \\ 0 & 1 & 0 & 0 & 3 & 2 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 3 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 3 & 2 \end{bmatrix} = TG, \ G_p = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 & 1 & 1 & 3 \\ 0 & 0 & 1 & 0 & 1 & 3 & 2 & 1 \end{bmatrix}$$

This gives

$$D_p = \begin{bmatrix} \mathrm{B}u_0 & 0 & 9u_1 & 0 & (2u_0+u_1)7 & (u_0+3u_1)4 & u_0+2u_1 & 0 \\ \mathrm{E}u_1 & 0 & \mathrm{D}u_1 & 0 & (2u_0+u_1)5 & (u_0+3u_1)7 & 0 & 3u_0+u_1 \end{bmatrix}$$

For all values of the parameters u_0, u_1, the eight columns sum to zero. The sum of the third, the sixth and the eighth column equals

$$\begin{bmatrix} 4u_0 + (9+3.4)u_1 \\ (7+3)u_0 + (\mathrm{D}+3.7+1)u_1 \end{bmatrix} = \begin{bmatrix} 4u_0+5u_1 \\ 4u_0+5u_1 \end{bmatrix}.$$

Hence for $4u_0 = 5u_1$, these three columns are dependent. A non-trivial solution is $t_2 = t_5 = t_7 = 1, t_0 = t_4 = t_6 = 0$. This gives

$$(b^\diamond, c^\diamond) = (0, b_1^\diamond, b_2, b_3^\diamond, 0, c_1, 0, c_3)$$
$$(b, c) + (b^\diamond, c^\diamond) = (b_0, b_1^\diamond, 0, b_3^\diamond, c_0, 0, c_2, 0).$$

Filling out $4u_0 = 5u_1$, we see that we obtain again the vectors of the previous example.

8.5 A combinatorial bound

If we want to check the existence of related differentials for a given map, then in principle we need to repeat the algorithm of Fig. 2 for all possible activity patterns. We present here an observation that reduces the number of activity patterns that need to be considered if the linear code associated with the map is an MDS code.

The three differentials (b, c), $(b^\diamond, c^\diamond)$ and $(b + b^\diamond, c + c^\diamond)$ correspond to vectors of an MDS code with minimal distance $m + 1$. Hence they can have at most $m - 1$ components equal to zero. On the other hand, from (7) we see that we need to distribute at least $2m$ zeroes over these three differentials. A simple counting argument results in the following bound.

Lemma 4 *If (b, c), $(b^\diamond, c^\diamond)$ are related differentials over a linear map with an associated code that is an MDS code with length $2m$ and distance $m + 1$, then*

$$\min\left\{\mathrm{wt}(b, c), \mathrm{wt}(b^\diamond, c^\diamond), \mathrm{wt}(b + b^\diamond, c + c^\diamond)\right\} \leq m + \lfloor m/3 \rfloor.$$

This means that if related differentials exist, they will be revealed when we check all the differentials with weights up to $m + \lfloor m/3 \rfloor$. For instance, if $m = 4$, then two of the three differentials need to have zeroes in at least 3 positions. Consequently, if we check all activity patterns of weight 5 for the existence of related differentials, then we have determined all the related differentials. Table 3 lists for $m = 4, 5, 6, 7, 8, 9$ the possible distributions of $2m$ zeroes and the differentials that need to be checked for related differentials.

Table 3 Possible distributions of $2m$ zeroes over 3 vectors, where each vector counts at most $m-1$ zeroes

m	Possible distributions of zeroes	Weights of activity patterns to be checked
4	(3,3,2)	5
5	(4,4,2), (4,3,3)	6
6	(5,5,2), (5,4,3), (4,4,4)	7, 8
7	(6,6,2), (6,5,3), (6,4,4), (5,5,4)	8, 9
8	(7,7,2), (7,6,3), (7,5,4), (6,6,4), (6,5,5)	9, 10
9	…	10, 11, 12

The last column gives the weights of the activity patterns that need to be checked in order to determine all sets of related differentials

9 Implications for AES-like super boxes

9.1 Related differentials over circulant matrices

The fact that MixColumns has related differentials is no coincidence. This can be understood easily if we write out the equivalent of Table 1 for a general 4 × 4 circulant matrix. Denote the matrix and its inverse by

$$\mathrm{M}_c = \begin{bmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{bmatrix}, \qquad \mathrm{M}_c^{-1} = \begin{bmatrix} e & f & g & h \\ h & e & f & g \\ g & h & e & f \\ f & g & h & e \end{bmatrix}. \tag{15}$$

Table 4 gives the bundles for a mixing map using this matrix. Looking at the second and third row, respectively 12th and 13th row, we notice that they define related differentials. All rotations of these bundles also define related differentials, as do all scalar multiples.

Table 4 Bundles for a mixing map based on the circulant matrix defined in (15)

u^{b}				u^{c}			
1	0	0	0	a	b	c	d
a	0	c	0	a^2+c^2	$ab+cd$	0	$ad+bc$
0	b	0	d	0	$ab+cd$	b^2+d^2	$ad+bc$
a	0	0	d	a^2+bd	$ab+cd$	d^2+ac	0
a	b	0	0	a^2+bd	0	b^2+ac	$ad+bc$
0	b	c	0	c^2+bd	$ab+cd$	b^2+ac	0
0	0	c	d	c^2+bd	0	d^2+ac	$ad+bc$
e^2+fh	$ef+gh$	h^2+eg	0	e	0	0	h
e^2+fh	0	f^2+eg	$eh+fg$	e	f	0	0
g^2+fh	$ef+gh$	f^2+eg	0	0	f	g	0
g^2+fh	0	h^2+eg	$eh+fg$	0	0	g	h
e^2+g^2	$ef+gh$	0	$eh+fg$	e	0	g	0
0	$ef+gh$	f^2+h^2	$eh+fg$	0	f	0	h
e	f	g	h	1	0	0	0

9.2 Related differentials in MixColumns

Besides the related differentials described in the previous section, MixColumns has eight more. Table 5 lists the four pairs of bundles from which all related differentials can be derived by means of rotation and/or multiplication by a scalar. We know from Section 3 that a differential with weight 5 is determined uniquely by its activity pattern. This implies that 3/7 of the differentials with $\mathrm{wt}(b, c) = 5$ is part of a set of related differentials.

Four rounds of AES are, up to a linear transformation, equivalent to a large Super box structure where the S-boxes are exactly the Super boxes we described before. The mixing transformation of this large Super box structure is equivalent to the sequence Shiftrows followed by MixColumns followed by ShiftRows. Also this map has related differentials, and their activity patterns are the same as the activity patterns of the differentials in Table 5. Hence, also a large set of characteristics over 4 rounds of AES has height at least 2.

9.3 Avoiding related differentials

There exist 4 × 4 matrices over $\mathrm{GF}(2^8)$ without related differentials, even matrices with special structure. For instance matrices with the Hadamard structure, as in Anubis [6]. Denote a 4 × 4 Hadamard matrix and its inverse as follows:

$$\mathrm{M}_c = \begin{bmatrix} a & b & c & d \\ b & a & d & c \\ c & d & a & b \\ d & c & b & a \end{bmatrix}, \qquad \mathrm{M}_c^{-1} = \begin{bmatrix} e & f & g & h \\ f & e & h & g \\ g & h & e & f \\ h & g & f & e \end{bmatrix}. \tag{16}$$

Table 6 gives the bundles for this matrix. It can be seen that in general, there are no related differentials with five active positions. From Lemma 4 we know that this means there are in general no related differentials. Anubis uses:

$$\mathrm{M}_{\mathrm{Anubis}} = \begin{bmatrix} 1 & 2 & 4 & 6 \\ 2 & 1 & 6 & 4 \\ 4 & 6 & 1 & 2 \\ 6 & 4 & 2 & 1 \end{bmatrix}.$$

Table 5 The sets of related differentials over MixColumns

b	c	$b^\diamond$	$c^\diamond$	$b + b^\diamond$	$c + c^\diamond$
[0, 1, 4, 7]	[0, 9, 0, B]	[5, 1, 0, 7]	[E, 0, D, 0]	[5, 0, 4, 0]	[E, 9, D, B]
[0, 1, 0, 3]	[0, 1, 4, 7]	[2, 0, 1, 0]	[5, 1, 0, 7]	[2, 1, 1, 3]	[5, 0, 4, 0]
[7, 0, 7, 7]	[9, E, 0, 0]	[7, 7, 7, 0]	[0, 0, 9, E]	[0, 7, 0, 7]	[9, E, 9, E3]
[0, 3, 2, 0]	[7, 0, 7, 1]	[2, 0, 0, 3]	[7, 1, 7, 0]	[2, 3, 2, 3]	[0, 1, 0, 1]

Table 6 Bundles for a mixing map based on the Hadamard matrix defined in (16)

u^{b}				u^{c}			
1	0	0	0	a	b	c	d
a	b	0	0	a^2+b^2	0	$ac+bd$	$ad+bc$
0	0	c	d	c^2+d^2	0	$ac+bd$	$ad+bc$
a	0	c	0	a^2+c^2	$ab+cd$	0	$ad+bc$
0	b	0	d	b^2+d^2	$ab+cd$	0	$ad+bc$
a	0	0	d	a^2+d^2	$ab+cd$	$ac+bd$	0
0	b	c	0	b^2+c^2	$ab+cd$	$ac+bd$	0
f^2+g^2	$ef+gh$	$eg+fh$	0	0	f	g	0
e^2+h^2	$ef+gh$	$eg+fh$	0	e	0	0	h
f^2+h^2	$ef+gh$	0	$eh+fg$	0	f	0	h
e^2+g^2	$ef+gh$	0	$eh+fg$	e	0	g	0
g^2+h^2	0	$eg+fh$	$eh+fg$	0	0	g	h
e^2+f^2	0	$eg+fh$	$eh+fg$	e	f	0	0
e	f	g	h	1	0	0	0

This matrix has related differentials: ([0, 0, 4, 6]; [4, 0, 8, E]) and ([8, E, 4, 0]; [4, 6, 0, 0]) (and their sum: ([8, E, 0, 6]; [0, 6, 8, E])). However, if the four 6'es are replaced by 9s, then there are no related differentials.

10 Conclusions

In this paper, we studied Super boxes, which are a building block of several modern block ciphers. We concentrated on the properties of the linear mixing map in relation to the resistance of a Super box against differential cryptanalysis. Besides the previously described branch number and its impact on the EDP of differentials over Super boxes, we also described a new property which has an impact on the distribution of the DP[k] values, namely related differentials.

We showed that all characteristics over Super boxes are plateau characteristics: the distribution of their DP[k] values is bipolar: either zero or 2^{h-32}. The value h is called the height of the characteristic. The height is always an integer value, constant over all values of the round key. It turns out that for the AES Super box, the height can be as large as 5, implying that DP[k] can be 16 times higher than what would be expected for an ideal map of the same dimensions. This is a surprisingly large difference.

The presence of related differentials in MixColumns is one reason why such high values for the height occur for so many characteristics. We studied how related differentials can be discovered for any given linear map.

There remain further questions to be answered about plateau characteristics. We believe that an analysis of resistance against differential cryptanalysis needs to take into account more than the average behavior of a key-dependent map. There are currently no attacks on block ciphers known which exploit non-uniformities in the distribution of the DP[k] values. However, in case of iterated mappings without a key, like for instance hash functions, they may well be very relevant.

References

1. Specification for the Advanced Encryption Standard (AES). Federal Information Processing Standards Publication 197 (2001)
2. American Mathematical Society. Algebra, ISBN 0821816462 (1999)
3. Anderson, R.A., Biham, E., Knudsen, L.R.: Serpent. Proc. of the 1st AES candidate conference, CD-1: Documentation, August 20–22, Ventura (1998)
4. Aoki, K.: Maximum non-averaged differential probability. Selected Areas in Cryptography SAC '98, LNCS 1556, pp. 118–130. Springer-Verlag (1998)
5. Aoki, K., Ichikawa, T., Kanda, M., Matsui, M., Moriai, S., Nakajima, J., Tokita, T.: Camellia: a 128-bit block cipher suitable for multiple platforms—Design and analysis. In: Stinson, D., Tavares, S. (eds.) Selected Areas in Cryptography 2000, LNCS 2012, pp. 39–56. Springer-Verlag (2000)
6. Barreto, P., Rijmen, V.: The Anubis block cipher. First open NESSIE Workshop, Leuven, November 13–14, http://paginas.terra.com.br/informatica/paulobarreto/AnubisPage.html (2000)
7. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like Cryptosystems. J. Cryptol. **4**(1), 3–72 (1991)
8. Ben-Aroya, I., Biham, E.: Differential cryptanalysis of Lucifer. In: Stinson, D. (ed.) Advances in Cryptology, Proc. Crypto'93, LNCS 773, pp. 187–199. Springer-Verlag (1994)
9. Canteaut, A.: Differential cryptanalysis of Feistel ciphers and differentially δ-uniform mappings. Workshop record of Selected Areas in Cryptography SAC '97, pp. 172–184 (1997)
10. Daemen, J., Govaerts, R., Vandewalle, J.: Weak keys of IDEA. In: Stinson, D. (ed.) Advances in Cryptology, Proc. Crypto'93, LNCS 773, pp. 224–231. Springer-Verlag (1994)
11. Daemen, J., Govaerts, R., Vandewalle, J.: A new approach to block cipher design. In: Anderson, R. (ed.) Proc. of Fast Software Encryption 1993, LNCS 809, pp. 18–32. Springer-Verlag (1994)
12. Daemen, J.: Cipher and hash function design. Strategies based on linear and differential cryptanalysis. Ph.D. thesis, Katholieke Universiteit Leuven (1995)
13. Daemen, J., Knudsen, L.R. Rijmen, V.: The block cipher square. In: Biham, E. (ed.) Fast Software Encryption '97, LNCS 1267, pp. 149–165. Springer-Verlag (1997)
14. Daemen, J., Peeters, M., Van Assche G., Rijmen, V.: Nessie proposal: the block cipher Noekeon. (Submitted to Nessie)
15. Daemen, J., Rijmen, V.: The Design of Rijndael—AES, The Advanced Encryption Standard. Springer-Verlag (2002)
16. Daemen, J., Rijmen, V.: Understanding two-round differentials in AES. Security and Cryptography for Networks 2006 (SCN 2006), LNCS 4116, pp. 78–94. Springer-Verlag (2006)
17. Daemen, J., Rijmen, V.: Plateau characteristics. IET Inf. Secur. **1**(1), 11–18 (2007)
18. Keliher, L.: Refined analysis of bounds related to linear and differential cryptanalysis for the AES. Advanced Encryption Standard—AES, 4th international conference (AES 2004), LNCS 3373, pp. 42–57. Springer-Verlag (2005)
19. Keliher, L., Sui, J.: Exact maximum expected differential and linear probability for 2-round advanced encryption standard (AES). IET Inf. Secur. **1**(2), 53–57 (2007)
20. Knudsen, L.R.: Iterative characteristics of DES and s^2-DES. In: Brickell, E.F. (ed.) Advances in Cryptology, Proc. CRYPTO'92, LNCS 746, pp. 497–511. Springer-Verlag (1993)
21. Knudsen, L.R.: Truncated and higher order differentials. In: Preneel, B. (ed.) Fast Software Encryption '94, LNCS 1008, pp. 196–211. Springer-Verlag (1995)
22. Knudsen, L.R., Mathiassen, J.E.: On the role of key schedules in attacks on iterated ciphers. ESORICS 2004, LNCS 3193, pp. 322–334. Springer-Verlag (2004)
23. Lai, X., Massey, J.L., Murphy, S.: Markov ciphers and differential cryptanalysis. In: Davies, D.W. (ed.) Advances in Cryptology, Proc. Eurocrypt'91, LNCS 547, pp. 17–38. Springer-Verlag (1991)
24. Lidl, R., Niederreiter, H.: Introduction to Finite Fields and Their Applications. Cambridge University Press, 1986 (Reprinted 1988)
25. Matsui, M.: New block encryption algorithm misty. In: Biham, E. (ed.) Fast Software Encryption '97, LNCS 1267, pp. 64–74. Springer-Verlag (1997)
26. Nyberg, K.: Differentially uniform mappings for cryptography. In: Helleseth, T. (ed.) Advances in Cryptology, Proc. Eurocrypt'93, LNCS 765, pp. 55-64. Springer-Verlag (1993)
27. Nyberg, K., Knudsen, L.R.: Provable security against a differential attack. J. Cryptol. **8**(1), 27–38 (1995)

28. Park, S., Sung, S.H., Chee, S., E-J. Yoon, Lim, J.: On the security of Rijndael-like structures against differential and linear cryptanalysis. In: Zheng, Y. (ed.) Advances in Cryptology, Proceedings of Asiacrypt '02, LNCS 2501, pp. 176–191. Springer-Verlag (2002)
29. Park, S., Sung, S.H., Lee, S., Lim, J.: Improving the upper bound on the maximum differential and the maximum linear hull probability for SPN structures and AES. In: Johansson, T. (ed.) Fast Software Encryption '03, LNCS 2887, pp. 247–260. Springer-Verlag (2003)
30. Rijmen, V.: Cryptanalysis and design of iterated block ciphers. Doctoral Dissertation, October 1997, K.U. Leuven
31. Rijmen, V., Daemen, J., Preneel, B., Bosselaers, A., De Win E.: The cipher SHARK. In: Gollmann, D. (ed.) Fast Software Encryption '96, LNCS 1039, pp. 99–111. Springer-Verlag (1996)
32. Vaudenay, S.: On the need for multipermutations: cryptanalysis of MD4 and SAFER. In: Preneel, B. (ed.) Fast Software Encryption '94, LNCS 1008, pp. 286–297. Springer-Verlag (1995)
33. Zheng, Y., Zhang, X.M.: Plateaued functions. Advances in Cryptology, ICICS '99, LNCS 1726, pp. 284–300. Springer-Verlag (1999)