KU LEUVEN

FACULTEIT INGENIEURSWETENSCHAPPEN

Linear Cryptanalysis in the Weak Key Model

Tim Beyne

Thesis voorgedragen tot het behalen
van de graad van Master of Science
in de ingenieurswetenschappen:
wiskundige ingenieurstechnieken

Promotor:
Prof. dr. ir. V. Rijmen

Assessoren:
Prof. dr. ir. L. De Lathauwer
Prof. dr. ir. F. Vercauteren

Academiejaar 2019 – 2020

Preface

This thesis is the compilation of research conducted during the academic year 2018-2019, but its motivation and goals were inspired by earlier work. Accordingly, I am profoundly indebted to Prof. Vincent Rijmen for enabling my regular visits to COSIC during 2014-2017. One of the results from those internships was the paper *Block Cipher Invariants as Eigenvectors of Correlation Matrices* (Asiacrypt; December 2018 [14]), which provided the starting point for this thesis. I am moreover grateful to Prof. Rijmen for supervising this thesis and for his diligence in proofreading it.

In addition, I would like to thank my assessors Prof. L. De Lathauwer and Prof. F. Vercauteren – for future comments and past discussions. Specifically, Theorem 3.7 is related to a conversation I had with Prof. De Lathauwer in July 2017. I discussed some of the problems that appear in Chapter 4 with Prof. Vercauteren earlier this year[1].

This is also a fitting place to acknowledge some COSIC people (in order of office distance): Liliya, Siemen, Adrián, Tomer (by reciprocity), Yu Long and Ward. I additionally thank Joan Daemen and Wouter Castryck for their interest in Chapter 3 of this thesis.

Finally, I would like to thank my parents for their abiding support. I also thank my sister, for critical remarks pertaining solely to this preface.

Tim Beyne
Leuven, June 2019

[1] In addition, he must be duly credited for posing the question "Have you considered the eigenvectors?" during a COSIC seminar I gave in September 2015, although the subject of said seminar was not related to correlation matrices.

Contents

Abstract

The theory of linear cryptanalysis is revisited and generalized in light of recent developments in symmetric-key cryptography. Trade-offs in the design of *lightweight cryptographic primitives* have enabled new attacks such as block cipher invariants, and have renewed the interest in long-standing problems such as the effective use of nonlinear approximations in cryptanalysis. These developments are intrinsically related to linear cryptanalysis in the weak key model. In addition, permutation-based cryptography – which is based on keyless primitives – is gaining traction.

In response to these urgent tendencies, the present thesis develops a pervasive generalization of linear cryptanalysis. The proposed "geometric approach" enables a uniform treatment of many variants of the classical linear attack and is suitable for use in the keyless and weak key models of analysis. The new framework additionally facilitates novel extensions to linear cryptanalysis. Furthermore, it is applied to resolve problems related to the use of nonlinear approximations. As a further contribution, the problem of proving security against linear cryptanalysis is revisited in the weak key model.

Samenvatting

De grondslagen van lineaire cryptanalyse worden herbekeken en veralgemeend in het licht van recente ontwikkelingen in de symmetrische-sleutel cryptografie. Afwegingen bij het ontwerp van *lichtgewicht cryptografische primitieven* hebben geleid tot nieuwe aanvallen zoals blokcijfer invarianten, en hebben de interesse voor klassieke problemen zoals het gebruik van niet-lineaire benaderingen in de cryptanalyse heropgewekt. Deze ontwikkelingen zijn intrinsiek verbonden met de beschrijving van lineaire cryptanalyse in het zwakke sleutel model. Bovendien wint permutatie-gebaseerde cryptografie – die gebruikmaakt van primitieven zonder sleutel – aan belang.

Als reactie op deze dwingende tendensen ontwikkelt de voorliggende thesis een doortastende veralgemening van lineaire cryptanalyse. De voorgestelde "meetkundige aanpak" laat een uniforme behandeling van een groot aantal varianten van de klassieke lineaire aanval toe en is geschikt voor gebruik in het sleutel-loze en zwakke sleutel model. Bovendien maakt het nieuwe raamwerk bijkomende uitbreidingen van lineaire cryptanalyse mogelijk en leidt het tot de oplossing van problemen gerelateerd aan het gebruik van niet-lineaire benaderingen. Ten slotte worden pogingen tot het bewijzen van veiligheid ten aanzien van lineaire cryptanalyse heroverwogen in de context van het zwakke sleutel model.

List of Figures and Tables

List of Figures

List of Tables

List of Symbols

$\|$	Represents the concatenation of bitstrings.
$\propto$	Used to indicate that two quantities are proportional.
$\cong$	Used to denote that two objects are isomorphic.
$\perp$	Denotes orthogonality of vectors in or subspaces of an inner product space.
$\lvert \cdot \rvert$	Used to denote the cardinality of a set.
$\lVert \cdot \rVert_q$	Denotes the q-norm of a vector or matrix.
$\langle \cdot, \cdot \rangle$	Inner product between vectors in an inner product space.
δ_x	Function which is zero everywhere except at x.
$\delta_{x,y}$	Equal to one when $x = y$, zero otherwise.
$\otimes$	Tensor product of vector spaces, vectors, linear maps or matrices.
$x^{\otimes n}$	n-fold tensor product of x with itself.
$\mathbb{F}_q$	The (up to isomorphism) unique finite field of order q.
$K^{m \times n}$	The set of $m \times n$ matrices over a field K.
$\operatorname{Span} S$	The subspace spannned by a set of vectors S.
$\mathbb{S}^n$	The Euclidean unit sphere in $\mathbb{R}^n$.
$A^\top$	Transpose of the matrix A (also used with vectors).
$\operatorname{vec} A$	Vectorization of a matrix A.
$P_{\mathcal{V}}$	Orthogonal projector on a vector space $\mathcal{V}$.
$\mathscr{T}^F_{\mathcal{U},\mathcal{V}}$	Approximation operator (see Definition 3.5): $\mathscr{T}^F_{\mathcal{U},\mathcal{V}} = P_{\mathcal{U}}\, T^F\, P_{\mathcal{V}}$.
$G \oplus H$	Direct sum of groups G and H.
KG	Algebra of functions $G \to K$ where K is a field and G a group (often $\mathbb{C}G$).
$\widehat{G}$	The Pontryagin dual of a finite abelian group G.
$\mathcal{F}f = \widehat{f}$	Fourier transformation (Definition 3.3) of $f \in \mathbb{C}G$.
$\Pr A$	Probability of an event A.
$\mathbb{E}\boldsymbol{X}$	Expected value of a random variable $\boldsymbol{X}$.
$\operatorname{Var}\boldsymbol{X}$	Variance of a random variable $\boldsymbol{X}$.
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2.
Φ	Cumulative normal distribution function.
$\xrightarrow{\mathcal{D}}$	Convergence in distribution.
$[n]$	$\{1, 2, \ldots, n\}$

Unlisted symbols are either standard or only used once.

Chapter 1

Introduction

Symmetric-key cryptography enables secure storage of information, confidential and authenticated communication, and many other indispensable technologies. At the core of all practical encryption schemes are carefully designed components called *primitives*. This thesis is concerned with the security assessment or *cryptanalysis* of these primitives. Specifically, in this thesis, the theory of linear cryptanalysis is reexamined and generalized in light of recent developments.

This chapter starts by delineating the context of the thesis: Section 1.1 discusses the role of primitives in symmetric-key encryption and the design requirements originating therefrom. The motivation for a general reexamination of linear cryptanalysis stems from a number of recent trends and related open problems in symmetric-key cryptography, which are reviewed in Section 1.2. Finally, Section 1.3 clarifies the overall structure of this thesis and anticipates its main contributions.

1.1 Secret-key Cryptography

As mentioned above, secret-key cryptography underpins a number of basic technologies. The most familiar functionality is provided by encryption schemes, which allow the user to transfer the confidentiality of a potentially long message (a string of bits) to the secrecy of a short key (another string of bits).

However, for most applications, confidentiality is not sufficient. It is also desirable to ensure the integrity and authenticity of messages. Authenticated encryption is generally achieved by appending a short *tag* to the ciphertext, which can be verified upon decryption. The recently concluded CAESAR competition [11] and the ongoing lightweight cryptography standardization project of the U.S. *National Institute of Standards and Technology* (NIST) [8] have spurred the development of authenticated encryption with associated data (AEAD) algorithms.

Most[1] (authenticated) encryption schemes operate by subdividing the message into blocks, which are then processed by a construction or "mode of operation" involving one or more primitives operating on fixed-size bitstrings. The most common primitives are permutations and block ciphers. A cryptographic permutation is an

[1] With the exception of some stream ciphers.

invertible function satisfying certain security requirements, which will be discussed in Section 1.1.2. Block ciphers are permutations parameterized by a key, *i.e.* each value of the key specifies an invertible function. The design of block ciphers and permutations is discussed in the next section.

Figure 1.1 shows an example of a permutation-based AEAD mode. In fact, it is the author's submission *Elephant*[2] [16, 17] (joint work with Yu Long Chen, Christoph Dobraunig and Bart Mennink) to the NIST lightweight cryptography project. The top part is responsible for encryption: n-bit message blocks are combined using exclusive-or with a keystream generated from an m-bit nonce N (a unique but otherwise arbitrary value) and a 128-bit key K. The lower part of the figure shows the generation of the t-bit tag T. The blocks A_i represent data that require authentication but not encryption, *i.e.* associated data.

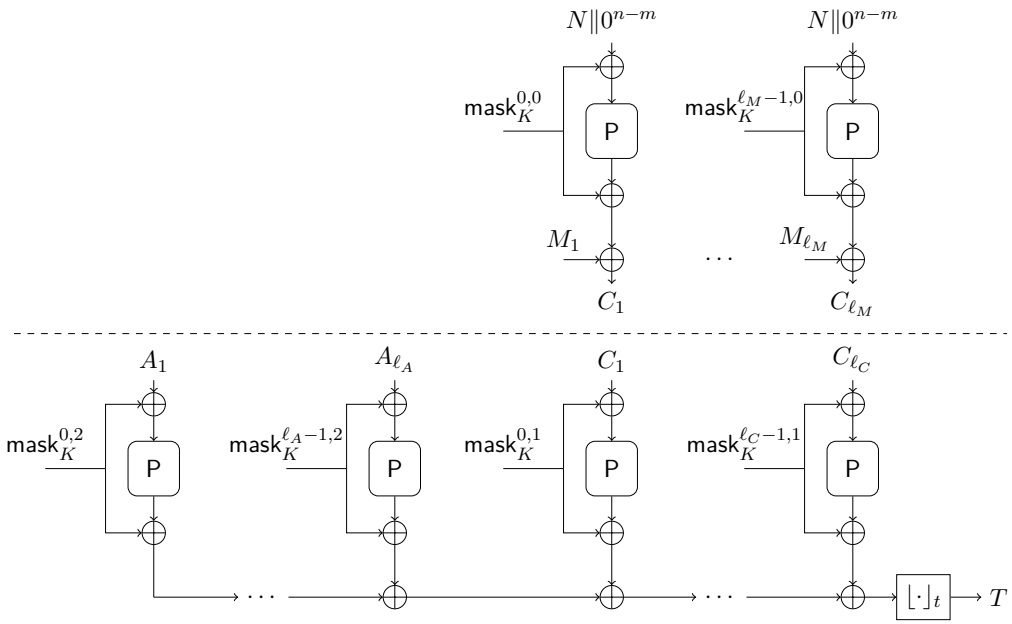


Figure 1.1: An example of an AEAD mode. The depicted mode is the author's submission to the NIST lightweight cryptography project, Elephant. Note that the first block of associated data A_1 includes the nonce N. This figure previously appeared in the Elephant specification [17, p. 5]

Remark. Hash functions are another important class of cryptographic functions that can be constructed from block ciphers and permutations. However, most of the results in this thesis relate to block ciphers and permutations *as used in encryption schemes.* ▷

[2] http://cosic.be/elephant

1.1.1 Construction of Primitives

Permutations and block ciphers are designed to operate on bitstrings of fixed length, say n bits. Hence, formally, a permutation F is a bijection $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ and a block cipher E_K is, for each k-bit key $K \in \mathbb{F}_2^k$, a permutation $E_K : \mathbb{F}_2^n \to \mathbb{F}_2^n$. Remark that fixed-length bitstrings are represented here as elements of the vector space $\mathbb{F}_2^n$. This is deliberate, since the construction of primitives often involves operations which are proper to $\mathbb{F}_2^n$, such as addition (exclusive or).

The construction of permutations and block ciphers is similar; this is not surprising, since one can think of permutations as block ciphers with a fixed key. Nevertheless, the design of block ciphers involves some additional aspects such as the choice of the key-schedule. Hence, the following discussion focuses on block ciphers.

Nearly all block cipher designs are round-based (also called iterative or iterated). This means that E_K is the composition of a sequence of permutations $F_1, F_2, \ldots, F_r$, where r is called the number of rounds. Each round function F_i may involve a different key K_i. The round keys $K_1, K_2, \ldots, K_r$ are derived from the master key K using a key scheduling algorithm. The key schedule need not be complicated; for instance, in several lightweight block ciphers, $K_i = K + c_i, i = 1, \ldots, r$ with c_i public constants.

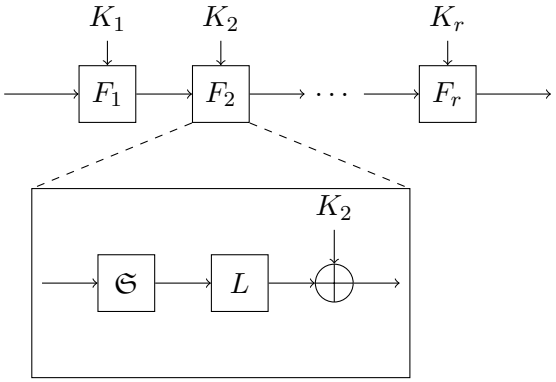


Figure 1.2: Schematic of a typical iterated block cipher (of the SPN type).

The two most popular families of iterated designs are Feistel ciphers and substitution-permutation networks (SPNs). For this thesis, the Feistel structure is of limited importance so it will not be discussed here. In SPN ciphers, the round function F_i consists of a nonlinear layer $\mathfrak{S}$ (see below), an $\mathbb{F}_2$-linear map L and a key-addition step. That is, $F_i(x) = L(\mathfrak{S}(x)) + K_i$ as in Figure 1.2. The Rijndael family of block ciphers, which includes the *advanced encryption standard* (AES), is the most important example of such a design [46]. Note that ciphers in which round keys are added to the state after each round are often referred to as *key-alternating*.

Remark. Technically, the AES is not an SPN design because its linear layer is not a *bit* permutation. That is, Figure 1.2 represents a superset of the class of SPN ciphers. Other designs such as PRESENT, Spongent-π and GIFT [7, 27, 28] are true SPNs.

Throughout this thesis, "SPN" will be used in the broad sense since this usage is common in the modern literature [32, 57, 64, 87]. ▷

The nonlinear layer $\mathfrak{S}$ consists of the parallel application of multiple nonlinear functions, each defined on a small number of bits of the state. These "small" nonlinear functions are called *substitution boxes* (S-boxes). That is, $\mathfrak{S} = (S_1, \ldots, S_m)$ such that

$$\mathfrak{S}(x_1 \| x_2 \| \cdots \| x_m) = S_1(x_1) \| S_2(x_2) \| \cdots \| S_m(x_m),$$

where $\|$ represents concatenation of bitvectors and S_i a permutation on $\mathbb{F}_2^{n/m}$.

In some SPN ciphers, the linear layer L is "strongly aligned" with the S-box layer [13]. In this thesis, such ciphers will be occasionally referred to as cell-oriented. For example, in Rijndael, the map L is linear over $\mathbb{F}_{2^{n/m}}$. Note that L need not be $\mathbb{F}_{2^{n/m}}$-linear in a cell-oriented design, one counterexample is QARMA [3].

Example. Midori-64 [6] is a cell-oriented design that recurs in several examples throughout this thesis. It has 64-bit state, which is represented as a 4×4 array of 4-bit cells. The 128-bit master key K is split into two equal parts to obtain the round keys K_0 and K_1. Figure 1.3 provides an overview. The values of the round constants $c_1, \ldots, c_{15}$ may be found in the specification [6].

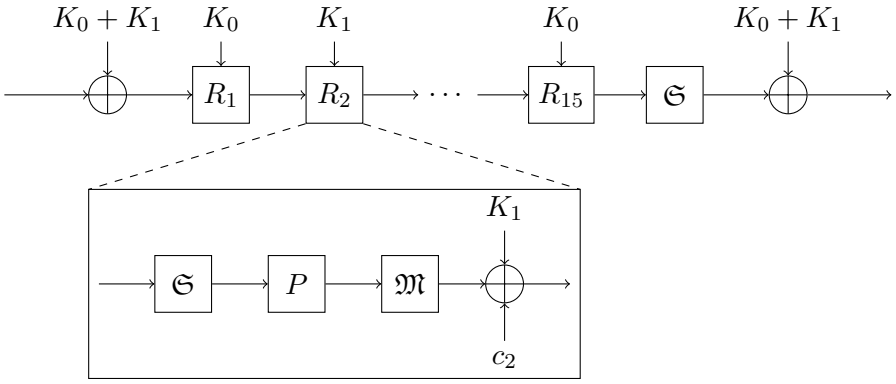


Figure 1.3: Schematic representation of Midori-64. This figure previously appeared in the author's work [14].

The nonlinear layer is as described above, with $S_1 = \ldots = S_{16} = S$. A hexadecimal representation (as a lookup table) is given in Table 1.1.

Table 1.1: Lookup table representation of the Midori-64 S-box.

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$S(x)$	c	a	d	3	e	b	f	7	8	9	1	5	0	2	4	6

The linear layer of Midori-64 consists of two transformations. The first is the `ShuffleCell` step (P in Figure 1.3), which corresponds to the following permutation of cells:

x_1	x_5	x_9	x_{13}
x_2	x_6	x_{10}	x_{14}
x_3	x_7	x_{11}	x_{15}
x_4	x_8	x_{12}	x_{16}

$\xrightarrow{P}$

x_1	x_{15}	x_{10}	x_8
x_{11}	x_5	x_4	x_{14}
x_6	x_{12}	x_{13}	x_3
x_{16}	x_2	x_7	x_9

The second transformation is the `MixColumn` operation ($\mathfrak{M}$ in Figure 1.1), which amounts to the multiplication of each state column with the block matrix

$$M = \begin{pmatrix} 0 & I & I & I \\ I & 0 & I & I \\ I & I & 0 & I \\ I & I & I & 0 \end{pmatrix},$$

where $I \in \mathbb{F}_2^{4\times 4}$ is the identity matrix. ▷

The constructions discussed above are the result of careful security evaluation using standard cryptanalytic techniques. The most popular families of such techniques are differential [21, 22] and linear [77, 78] cryptanalysis. Differential cryptanalysis will not be discussed in this thesis; Chapter 2 contains a detailed introduction to linear cryptanalysis. However, before cryptanalysis can be discussed, it is necessary to define what is meant by a "secure" primitive. This is the goal of the next section.

1.1.2 Security Model for Primitives

The security of a primitive can only be understood within the context of a mode. Hence, only attacks on the primitive that are useful to break the mode need be considered. Nevertheless, this does not rule out generic analysis because primitives are often reused and in many cases cryptanalytic results are applicable across several use-cases.

The goal of the cryptanalyst is either to recover the key (if there is one) or to exhibit a distinguisher (subject to the restrictions above). Informally, a distinguisher is an algorithm that tells apart a construction or primitive from its idealization. Cryptanalysis is always conducted within a certain security model, which specifies the capabilities of the attacker. Table 1.2 lists the most common scenarios.

Remark. The description of the "single key" model in Table 1.2 specifies that the cost of the attack is expected to be, at least approximately, uniform over the key space. This additional requirement is nonstandard, but nevertheless important to clarify the difference with the weak key model. Traditional techniques such as linear and differential cryptanalysis are usually assumed to lead to attacks that work for any key. There are many reasons to expect that, instead, these attacks work for most but not quite all of the keys [31, 44, 89]. Nevertheless, one does not say that they are in the weak key setting. ▷

Model	Description
Ciphertext only	The attacker is given the ciphertext corresponding to several unknown plaintexts. Additional information, such as the type of data that was encrypted, may be available.
Known plaintext	The attacker is given several plaintext/ciphertext pairs.
Chosen plaintext	The attacker chooses several plaintexts and receives the corresponding ciphertext.
Chosen ciphertext	In addition to choosing plaintexts, the attacker can adaptively choose several ciphertexts and receives the corresponding plaintext.
Single key	The key is unknown to the adversary. The cost and advantage of the adversary are (often implicitly) expected to be approximately uniform over the key space.
Weak key	Part of the key may be chosen by the adversary. Equivalently, the key is unknown to the adversary but the attack is only expected to work for part of the key space.
Related key [18]	The adversary has access to the block cipher under two different keys, which are somehow related. Not all relations are allowed.
Known key [68]	The key is known to the adversary. This model is useful when the cipher is used as a public permutation or in the ideal cipher model (see Section 1.1.3).

Table 1.2: Overview of cryptanalytic evaluation models. The second part of the table is only applicable to block ciphers.

From a practical point of view, some of the capabilities listed in Table 1.2 may seem excessive. On the one hand, this is consistent with a conservative approach to security. On the other hand, one can argue that practical adversaries can in fact possess an even wider range of capabilities. For example, cryptanalysts during the First and Second World Wars routinely exploited erroneously encrypted messages [60].

The cost of an attack is determined by various aspects:

- The computational cost, which is typically measured in terms of the number of block cipher calls.
- The amount of memory required to execute the attack.
- The number of known/chosen plaintexts and/or ciphertexts (“data”) that are required.

– The success probability and false positive rate. The absolute value of the difference between the success probability and the false positive rate is often called the *advantage*.

The metrics listed above are not independent; many trade-offs are possible. For primitives that are constructed by iterating a round function, it is also common to report the number of rounds broken by the attack. Indeed, attacks are often incrementally improved and initially only threaten reduced-round versions.

1.1.3 Formalizations of the Security Model

This section reviews several attempts to formalize the security of cryptographic building blocks. The branch of cryptography that deals with these formalizations is often designated by the broad term "provable security" [72].

When block ciphers are used, modes can often be proven secure in a reductionist sense. That is, it is shown that breaking the mode is hard by reducing this problem to a property of the block cipher. This model is often called the *standard model* [72, 75]. Informally, the desired property of a block cipher in the standard model is that no adversary with limited computational power can distinguish the cipher, when used with a uniform random key, from a uniform random permutation with significant advantage. Such a block cipher is said to be a *pseudorandom permutation* or PRP[3].

Remark. The phrase "limited computational power" hides a major difficulty: finding an appropriate definition of this terminology is challenging, especially in the non-asymptotic setting. For example, Bernstein and Lange [12] give examples of efficient attacks on the AES – or any other cipher – with the caveat that they require a large amount of precomputation. They suggest, among other things, to require constructive adversaries. The author of this thesis is inclined to go further, and would argue that security definitions of a purely computational nature are irreconcilable with the concrete. This is supported by the fact that there exist significant PRP distinguishers for all common block ciphers that do not require precomputation – but these are not realistic because, a priori, their result cannot be interpreted unless excessive computations are performed. ▷

A different approach to provable security is the *ideal model* [72, 75]. In this model, one shows that the mode is information-theoretically secure if it is instantiated with ideal primitives. For example, an ideal block cipher is a random permutation for every fixed value of the key. The ideal model can be used when the primitives are incompatible with standard security notions such as pseudorandomness, or if those notions are not sufficiently strong. For example, in permutation-based cryptography, the mode is built from public permutations. Finally, note that indistinguishability is not applicable to modes which are completely public such as hash functions. Other frameworks such as indifferentiability have been developed for this setting [79].

[3] In theory, the advantage is considered to be insignificant ("negligible") if it decays asymptotically faster than the reciprocal of any polynomial in the security parameter [72]. In practice, the security parameter is fixed and the notion "maximum PRP advantage" is more useful.

1.2 Open Problems and Recent Trends

The motivation for this thesis comes from a number of open problems in the cryptanalysis and design of primitives. These problems in turn originate from several recent trends in symmetric-key cryptography.

The widespread use of cryptography in resource-constrained settings such as real-time and embedded computing provides a strong motivation to improve the efficiency of primitives. The evaluation criteria, *e.g.* area, latency, throughput or energy usage, strongly depend on the application [24]. The ongoing NIST lightweight cryptography project aims to deliver a U.S. government standard for lightweight authenticated encryption and hashing within two to four years [8].

Despite significant differences in the relevant metrics, most lightweight primitives share a number of common characteristics. One tendency is the use of simple key-schedules. The energy-efficient block cipher Midori-64, which was discussed in Section 1.1.1, serves as an example. Along with other design decisions, such as the choice of the S-boxes and the linear layer, simple key-schedules have enabled new cryptanalytic techniques. In particular, block cipher invariants have led to powerful weak key attacks on lightweight ciphers, including Midori-64 [14, 51, 71, 94]. This includes the paper [14] by the author of this thesis, which received the best paper award at ASIACRYPT 2018. The approach to invariants that was introduced in the latter paper (explained in Section 2.4.2) underpins the frame of reference that will be developed in this thesis.

Invariant attacks have led to renewed interest in the weak key foundations of linear cryptanalysis, and the related open problems surrounding the use of nonlinear approximations in cryptanalysis. This is evidenced by recent work such as that of Beierle *et al.* [9], presented at FSE in March 2019.

Another trend is the increased prevalence of permutation-based cryptography. For example, many submissions to the NIST lightweight cryptography project are based on permutations as opposed to block ciphers – this includes the author's submission, Elephant [17]. Since permutations do not involve a key, a "fixed key" approach to linear cryptanalysis is desirable. This requires a reexamination of the classical approach to linear cryptanalysis, since it is justified based on key-averaging arguments. It will become clear throughout this thesis that the issues surrounding nonlinear approximations, weak keys and the fixed key model are intertwined.

Finally, a long-standing open problem in the design of symmetric-key primitives is to prove the security of a concrete construction against linear cryptanalysis. Current arguments at best succeed in demonstrating average-case (w.r.t. the key) security of block ciphers. The question then arises: can resistance against linear cryptanalysis be established in the weak key setting, or for permutations? One must then inevitably deal with the issue of *trail clustering* [43], which is closely related to nonlinear cryptanalysis.

1.3 Outline and Goals of this Thesis

This thesis aims to address the open problems that were listed in Section 1.2. By and large, its contributions fall into two categories. The first class of contributions is related to the foundations of linear cryptanalysis in the weak or fixed key setting. Specifically, a new theoretical framework for linear cryptanalysis – including generalizations such as nonlinear cryptanalysis – is proposed. The second category of results is related to proving the resistance of block ciphers against linear cryptanalysis in the weak key model.

The literature on linear cryptanalysis is reviewed in Chapter 2. The main approaches to linear cryptanalysis are discussed, including a brief discussion on key-recovery strategies. The chapter also summarizes strategies that are used to design primitives which are, at least heuristically, secure against linear attacks. Finally, several variants and generalizations of linear cryptanalysis are introduced.

Chapter 3 develops a novel theoretical framework for linear cryptanalysis, which can be termed as a "geometric approach". The main goals are the unification and generalization of the wide range of variants of linear cryptanalysis, within the weak or fixed key model. Overall, a novel way of thinking about linear cryptanalysis is sought. Several applications are discussed, including a recently proposed (previously) open problem.

The results related to provable resistance against linear cryptanalysis are contained in Chapter 4. The focus is primarily on deriving upper bounds – in the weak key model – on the correlation of linear approximations over two rounds of an SPN. A starting point is provided for extending these results to more than two rounds.

The main conclusions and realizations are summarized in Chapter 5.

Chapter 2

Linear Cryptanalysis

This chapter summarizes the main literature on linear cryptanalysis and its variants. Linear approximations are discussed in Section 2.1. The classical approach to linear key-recovery attacks is recalled in Section 2.2. Section 2.3 then discusses design strategies that can be used to develop block ciphers which are, at least heuristically, resistant against linear cryptanalysis. Finally, various extensions of and variations on the basic linear attack are reviewed in Section 2.4.

Let E_K be a block cipher on the vector space $\mathbb{F}_2^n$ and consider a nontrivial sum of plaintext and ciphertext bits. If E_K behaves as an ideal cipher, any such expression is expected to equal zero for roughly half of the plaintexts. In linear cryptanalysis, one attempts to find and exploit $\mathbb{F}_2$-linear combinations of input and output bits that are equal to each other for significantly more or less than half of the plaintexts.

2.1 Linear Approximations

This section discusses linear approximations, which are at the core of linear cryptanalysis. Specifically, several approaches to compute the probability of such approximations are discussed. The classical approach, due to Matsui [77, 78], is reviewed in Section 2.1.1. Section 2.1.2 introduces correlation matrices, which were proposed by Daemen *et al.* [40] as an alternative description of linear cryptanalysis.

2.1.1 Classical Approach

Linear cryptanalysis was introduced by Matsui and Yamagishi in the context of the FEAL cipher [78] and shortly after resulted in the cryptanalysis of the DES by Matsui [76, 77]. It was inspired by the earlier work of Tardy-Corfdir and Gilbert [93].

Consider an iterative permutation $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$. If F is a block cipher, then fix the key K – here we deviate from [76–78, 93] for the sake of generality. As briefly mentioned above, in linear cryptanalysis one attempts to find *masks* $u, v \in \mathbb{F}_2^n$ such that $v^\top F(x) = u^\top x$ holds significantly more or less often than is expected for an ideal cipher. Such a relation will be called a *linear approximation* of F. Let $\boldsymbol{x}$ be a uniformly distributed random variable on $\mathbb{F}_2^n$. The *bias* ε of a linear approximation

is the quantity

$$\varepsilon = \Pr\left[v^\top F(\boldsymbol{x}) = u^\top \boldsymbol{x}\right] - \frac{1}{2}. \tag{2.1}$$

In linear cryptanalysis, one is thus interested in finding a linear approximation such that $|\varepsilon|$ is as large as possible. Once a suitable approximation of a block cipher F has been found, this immediately yields a known-plaintext distinguisher. Section 2.2 discusses how such a distinguisher can be turned into a key-recovery attack. More generally, when F does not necessarily involve a secret key, good linear approximations often lead to attacks on the mode in which F is used.

In general, computing $|\varepsilon|$ is clearly nontrivial since the set of plaintexts $\mathbb{F}_2^n$ is large (typically, $n \geq 64$). Hence, one has to approximate ε – this is possible because F is assumed to be iterative, *i.e.* $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$. The key observation is that the biases of linear approximations of the round functions F_i can often be computed exactly. In order to glue together several one-round approximations, Matsui used a heuristic approach based on the *piling-up lemma* [77].

Lemma 2.1 (Piling-up lemma [77])**.** *Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_r$ be independent random variables on $\mathbb{F}_2$ with $c_i = 2\Pr[\boldsymbol{z}_i = 0] - 1$. Then*

$$2\Pr\Big[\sum_{i=1}^{r} \boldsymbol{z}_i = 0\Big] - 1 = \prod_{i=1}^{r} c_i.$$

Let $x_{i+1} = F(x_i)$ for $i = 1, \ldots, r$. Lemma 2.1 is used as follows: the cryptanalyst constructs a sequence of approximations $v_i^\top F_i(x_i) = u_i^\top x_i$ such that the masks are compatible, *i.e.* $u_{i+1} = v_i$ for $i = 1, \ldots, r$. For each i, let $\boldsymbol{z}_i = v_i^\top F_i(\boldsymbol{x}_i) + u_i^\top \boldsymbol{x}_i$ with $\boldsymbol{x}_1$ uniformly random. Assuming that the random variables $\boldsymbol{z}_i$ are independent, an application of Lemma 2.1 yields the desired probability. Indeed,

$$\sum_{i=1}^{r} \boldsymbol{z}_i = v_r^\top F_r(\boldsymbol{x}_r) + u_1^\top \boldsymbol{x}_1.$$

Remark. A sequence of approximations such as in the above paragraph is called a linear *trail* [40] or *characteristic* [19]. The former terminology will be used exclusively in this thesis. For reasons discussed in Section 2.1.2, the product $\prod_{i=1}^{r} c_i$ will be called the correlation of the trail. ▷

The reader will note that the variables $\boldsymbol{z}_i$ are not, in fact, independent. This raises the question why (and when) one should expect the piling-up approximation to be any good. The underlying idea in Matsui's work [77], and more explicitly in that of Biham [19], is that F involves round keys which can be assumed to be independent and uniformly distributed. For an iterative block cipher one can show that, under these assumptions, the piling-up lemma yields the exact value of

$$\mathbb{E}_{\boldsymbol{k}_1,\ldots,\boldsymbol{k}_r} \Pr_{\boldsymbol{x}} \left[v^\top E_{\boldsymbol{k}_1,\ldots,\boldsymbol{k}_r}(\boldsymbol{x}) = u^\top \boldsymbol{x} + \textstyle\sum_{i=1}^{r} v_i^\top \boldsymbol{k}_i\right]. \tag{2.2}$$

This follows, for example, immediately from the results that will be reviewed in Section 2.1.2. Nevertheless, (2.2) is not really an adequate justification of the piling-up approximation. Indeed, it only seems to raise more questions. Specifically, round

keys are not really independent and the average bias need not be equal to the bias for a specific key. In addition, in permutation-based cryptography, there is no key.

2.1.2 Correlation Matrices

This section discusses an alternative description of linear cryptanalysis due to Daemen, Govaerts and Vandewalle [40]. This approach is the main source of inspiration for the geometric approach to linear cryptanalysis and its generalizations that will be developed in Chapter 3.

Daemen *et al.*'s description of linear cryptanalysis starts from the notion of *correlation coefficients*, which is defined below.

Definition 2.1 (Correlation coefficient [40])**.** *The correlation coefficient* $\mathrm{C}(f, g)$ *between Boolean functions* $f, g : \mathbb{F}_2^n \to \mathbb{F}_2$ *is the quantity*

$$\mathrm{C}(f, g) = 2\Pr[f(\boldsymbol{x}) = g(\boldsymbol{x})] - 1,$$

for $\boldsymbol{x}$ *uniformly distributed on* $\mathbb{F}_2^n$*.*

Given a function $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$, Daemen *et al.* define a real matrix C^F with entries $\mathrm{C}(\ell_v \circ F, \ell_u)$ where $\ell_u(x) = u^\top x$ and $\ell_u(x) = u^\top x$ for $u \in \mathbb{F}_2^n$ and $v \in \mathbb{F}_2^m$:

$$C^F_{v,u} = \mathrm{C}(\ell_v \circ F, \ell_u) = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} (-1)^{v^\top F(x) + u^\top x}. \tag{2.3}$$

They call this matrix the correlation matrix of F [40]. In linear cryptanalysis, one is interested in the entries of C^F with the largest magnitude. Indeed, these correspond to the most biased linear approximations of F.

Remark. Note that, in the above paragraph, the entries of C^F are indexed by the elements of the set $\mathbb{F}_2^n$ rather than by integers $1, \ldots, 2^n$. The former notation is more natural, because C^F should be interpreted as the coordinate representation of a linear operator on the group algebra $\mathbb{C}\mathbb{F}_2^n$. ▷

There is – according to the author of this thesis – a more intuitive definition of correlation matrices. Specifically, (2.3) defines the matrix C^F by its coordinates. This obscures what is really going on. The alternative approach below first appeared in the author's work [14] at ASIACRYPT 2018.

Definition 2.2 (Transition matrix)**.** *The transition matrix* T^F *of a function* $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ *is a real* $2^m \times 2^n$ *matrix such that, if a random variable* $\boldsymbol{x}$ *has probability mass function* $p : \mathbb{F}_2^n \to [0, 1]$*, then* $F(\boldsymbol{x})$ *has probability mass function* $T^F p$*. Equivalently,* $T^F_{y,x} = \delta_{y,F(x)}$*.*

Definition 2.3 (Correlation matrix [14])**.** *The correlation matrix* C^F *of a function* F *is the coordinate representation of the linear map defined by* T^F *with respect to the character basis of the group algebra* $\mathbb{C}\mathbb{F}_2^n$*. Specifically, the character basis consists of the functions* $x \mapsto (-1)^{u^\top x}$ *with* $u \in \mathbb{F}_2^n$*.*

It is not hard to see that Definition 2.3 corresponds to (2.3), see for example [14]. An alternative formulation of Definition 2.3 is that C^F is the Fourier transformation of T^F. Due to this fact, many properties of correlation matrices are immediate consequences of the corresponding properties of T^F. A few useful results are listed in Theorem 2.1 below – these can also be found in [14, 40].

Remark. According to Definitions 2.2 and 2.3, the matrices T^F and C^F represent the same linear operator in a different basis. Alternatively, one can interpret C^F as an operator on $\mathbb{C}\widehat{\mathbb{F}_2^n}$, where $\widehat{\mathbb{F}_2^n} \cong \mathbb{F}_2^n$ denotes the Pontryagin dual of the group $\mathbb{F}_2^n$. This is more in line with the standard approach in mathematics, see Serre [92] and Diaconis [47], and will be useful in Chapter 3 where Definition 2.3 will be generalized to functions on arbitrary finite abelian groups. ▷

Theorem 2.1 (Properties of correlation matrices [14, 40])**.** *Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$. Then*

1. *For $G : \mathbb{F}_2^m \to \mathbb{F}_2^l$, $C^{G \circ F} = C^G C^F$.*
2. *If $H(x, y) = (F(x), G(y))$ with $G : \mathbb{F}_2^k \to \mathbb{F}_2^l$, then $C^H = C^F \otimes C^G$.*
3. *If F is a permutation, then C^F is an orthogonal matrix.*
4. *If $F(x) = Ax + c$ with $A \in \mathbb{F}_2^{m \times n}$ and $b \in \mathbb{F}_2^m$, then $C^F_{u,v} = (-1)^{u^\top b} \delta_{v, A^\top u}$.*

Proof. For (1), note that for the basis $\{\delta_x\}_{x \in \mathbb{F}_2^n}$ of $\mathbb{C}\mathbb{F}_2^n$, $T^G T^F \delta_x = \delta_{G(F(x))} = T^{G \circ F} \delta_x$. The result then follows by taking the Fourier transformation. Similarly, (2) follows from the trivial observation that $T^H = T^F \otimes T^G$ and the separability of the Fourier transformation. Property (3) is due to the fact that T^F is a permutation matrix and because the Fourier transformation is, up to scaling, orthogonal. Finally, (4) is an immediate consequence of $C^F \chi_u = \chi_{A^\top u + b}$ with $\chi_u(x) = (-1)^{u^\top x}$ the group character corresponding to u. □

The piling-up approximation can be seen to be a consequence of property (1) in Theorem 2.1. This observation was first made by Daemen *et al.* [40]. Indeed, if $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$, then

$$C^F_{u_r, u_0} = \sum_{u_1, \ldots, u_{r-1} \in \mathbb{F}_2^n} \prod_{i=1}^{r} C^{F_i}_{u_i, u_{i-1}}. \tag{2.4}$$

That is, the correlation of a linear approximation is the sum of the correlations of all possible trails within that approximation. Here, the correlation of a trail is computed as in Lemma 2.1. Hence, the piling-up assumption is valid whenever (2.4) contains a dominant term.

As such, the correlation matrix approach to linear cryptanalysis shows how an estimate of the correlation of a linear approximation might be refined by taking into account more than one trail. Nonetheless, (2.4) can be large even if no good trails can be found. Daemen and Rijmen call this phenomenon *trail clustering* [43]. For block ciphers, the attacker can try to cause clustering by carefully choosing

part of the key – this leads to weak key attacks. Note that this phenomenon is not theoretical, see for example the author's attack on Midori-64 [14]. Understanding the clustering of linear trails is an important motivation for the geometric approach to linear cryptanalysis that will be developed in Chapter 3.

2.2 Key-Recovery using the Last-Round-Trick

For this section, assume that a linear approximation over a block cipher E_K is given. Furthermore, suppose the corresponding absolute correlation is c – for simplicity of exposition, let us assume that c is approximately constant over the set of keys.

As mentioned in Section 2.1, such an approximation results in a distinguisher provided that c is sufficiently large. Before discussing how a distinguisher can be turned into a key-recovery attack, it is useful to make the requirement that c is "sufficiently large" more precise. Recall that a distinguisher compares E_K to an ideal block cipher, *i.e.* a collection of random permutations. Hence, the following result provides the necessary information.

Theorem 2.2 (Daemen and Rijmen [45])**.** *The correlation $\boldsymbol{c}_n$ of a nontrivial linear approximation over a uniform random n-bit permutation is asymptotically normally distributed with variance 2^{-n}. Specifically, as $n \to \infty$,*

$$2^{n/2}\boldsymbol{c}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

A more precise statement is possible: it is not hard to show that for all integers t such that $-2^{n-2} \leq t \leq 2^{n-2}$, one has [45]

$$\Pr[\boldsymbol{c}_n = 2^{2-n}t] = \frac{\binom{2^{n-1}}{2^{n-2}+t}^2}{\binom{2^n}{2^{n-1}}}.$$

For all other values of t, the probability is zero. For most practical purposes, however, the normal approximation is sufficiently accurate.

The distinguisher works by estimating the correlation of the approximation based on N known plaintext/ciphertext pairs, which we assume to be sampled at random with or without replacement. For example, when using sampling with replacement, the estimated correlation $\widehat{\boldsymbol{c}}$ will be approximately distributed as $\mathcal{N}(c, (1-c^2)/N)$ for large N. Using this fact together with Theorem 2.2, one can show that the required number of samples N to achieve a fixed success probability and false positive rate is proportional to $1/c^2$ [2, 26, 30, 90].

Remark. The normal approximation of $\boldsymbol{c}$ that is mentioned above has its limitations: if one samples (almost) the full codebook, one can determine the correlation nearly exactly. In this setting the proportionality $N \propto 1/c^2$ is no longer valid. Instead, with access to the full codebook, one can theoretically build a perfect distinguisher irrespective of the value of c. In practice, however, the cryptanalyst is forced to estimate c using the techniques discussed in Section 2.1. Hence, the model of the

cipher is imperfect and using small values of c is not feasible. An important exception is zero-correlation linear cryptanalysis ($c = 0$), which will be discussed in Section 2.4.3. Note that this issue is closely related to the remark in Section 1.1.3. ▷

Linear approximations can be used for key-recovery in essentially two ways, which are commonly referred to as Matsui's first and second algorithm [77]. The first algorithm is based on the observation that if the key is added to the state (as in a key-alternating cipher) and if a single linear trail is dominant,

$$\operatorname{sign}(c) = (-1)^b \prod_{i=1}^{r} (-1)^{v_i^\top K_i},$$

with $b \in \mathbb{F}_2$ a known constant, $K_1, \ldots, K_r$ the round keys and $v_1, \ldots, v_r$ (intermediate) masks. This is a consequence of Theorem 2.1, property (4). Hence, the attacker is able to learn some linear combination of key bits.

Matsui's second algorithm provides a more powerful approach to key-recovery. In this method, the block cipher is split into two parts and a linear approximation over the first part is used. For simplicity, assume the first part consists of the first $r - 1$ rounds. Figure 2.1 illustrates the idea: one guesses part of the last round key K_r in order to partially decrypt the last round. For each guess, the correlation of the linear approximation can be estimated. If a wrong key is used, one expects the correlation to be very small – usually, wrong guesses are modeled by a random permutation. Hence, the problem of finding the right key among all guesses of K_r is similar to distinguishing $r - 1$ rounds from a random permutation.

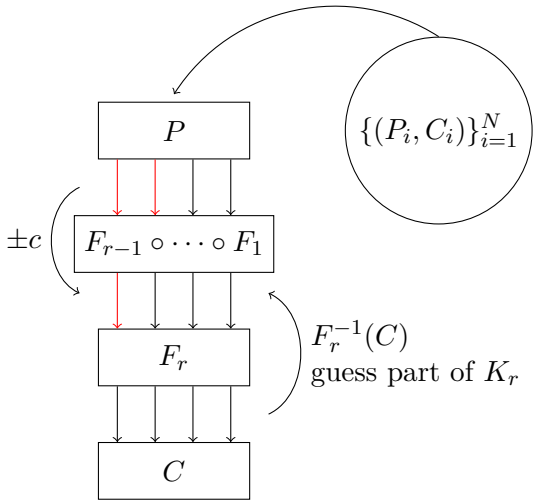


Figure 2.1: Matsui's second algorithm.

The naive approach to identify the right key is to store the estimated absolute correlation for each key guess in a large table [77, 90]. This table is sorted in descending order and the bottom part is discarded. If k bits of K_r must be guessed,

this procedure requires $\mathcal{O}(2^k N)$ time. However, since N is usually much larger than 2^k, a different approach is used in practice.

The following method was introduced by Matsui [76] and is based on the observation that the linear approximation is the sum of two disjoint parts: $u^\top P$ and $v^\top F_r^{-1}(C)$ with u the input mask and v the output mask. For simplicity, assume that the ciphertext masks contains k nonzero bits. For each observed value of the active ciphertext part, one computes the estimated correlation of $u^\top P_i$ over all plaintexts P_i having the corresponding ciphertext part. For each ciphertext part and key guess, one then computes $v^\top F_r^{-1}(C)$. From this $2^k \times 2^k$ table, the estimated correlation for each key guess can be computed. The time complexity of this approach is $\mathcal{O}(N + 2^{2k})$. Collard *et al.* [38] improve this to $\mathcal{O}(N + k\,2^k)$ using an FFT-based method for multiplication with circulant matrices.

The data requirements of Matsui's second algorithm have been analyzed under various assumptions [2, 26, 30, 90]. For c not too small, a good approximation is provided by [90]

$$N \approx \left(\frac{\Phi^{-1}(P_S) + \Phi^{-1}(1 - 2^{-a-1})}{c} \right)^2,$$

with P_S the success probability and 2^{-a} the fraction of retained keys. In prior joint work with Tomer Ashur and Vincent Rijmen, the author of this thesis has shown that this approximation remains reasonable when plaintext/ciphertext pairs are sampled without replacement [2]. The latter work also discusses the breakdown of Matsui's second algorithm when $c \leq 2^{-n/2}\,\Phi^{-1}(1 - 2^{-a-1})$. There has been some controversy regarding this subject, see Ashur's PhD thesis [1] or the note by Selçuk [91] for an overview.

2.3 Security Arguments for Linear Cryptanalysis

As discussed in Section 1.1, it is not known how to formally prove the security of the primitives that are used in practical symmetric-key cryptography. However, this does not mean that the design of primitives is arbitrary – new designs are at least expected to resist known attacks. This section discusses three approaches that are used to argue that a proposed primitive is secure against linear cryptanalysis.

Section 2.3.1 discusses the *wide trail strategy* [39, 43, 46, 88], which is the most successful approach to block cipher design. This strategy results in upper bounds on the absolute correlation of trails, but it does not guarantee that all linear approximations are weak. Section 2.3.2 reviews attempts to resolve this deficiency by bounding the variance of the correlation for a random key – or random round constants. Finally, Section 2.3.3 recalls Vaudenay's *decorrelation theory* [101] and the resulting bounds on the advantage of linear distinguishers.

2.3.1 Wide Trail Strategy

The wide trail strategy first appeared in Daemen's PhD thesis [39] and was subsequently extended by Rijmen [88]. It underpins the design of Rijndael [42, 46]. An

overview can be found in [43]. Below, the basic principle is recalled in the context of linear cryptanalysis.

Consider a substitution-permutation network as in Figure 2.2. An S-box is called *active* with respect to a linear trail if its output mask is nonzero [43]. For example, in Figure 2.2, the first S-box is active and the second is not.

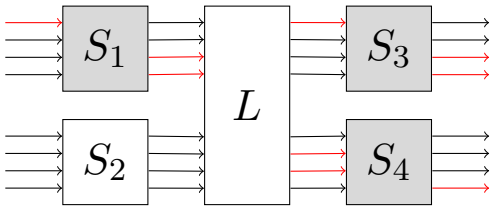


Figure 2.2: Linear trail through an eight bit function with three active S-boxes. Nonzero bits in the masks are marked in red. The linear layer is denoted by L.

Suppose that the absolute correlation of any nontrivial linear approximation of the S-box is less than c. If the masks of a linear trail are nonzero in positions corresponding to the output bits of N different S-boxes, then the correlation of this trail is clearly at most c^N. The essence of the wide trail strategy is to maximize the number of active S-boxes N [43].

If the S-boxes are permutations, the substitution step does not influence the activity pattern. The effect of the linear layer L, however, is important. Locally, the quality of the linear layer can be quantified by the linear *branch number*

$$B_L = \min_{v \in \mathbb{F}_2^n} \left[\mathrm{wt}(v) + \mathrm{wt}(L^\top v) \right],$$

where $\mathrm{wt}(v)$ denotes the number of nonzero m-bit blocks in u when the S-boxes are m bits wide. Note that B_L involves the transpose $L^\top$ rather than L due to Theorem 2.1, (4). Rijmen showed that B_L is maximized by the generator matrices of MDS codes [88]. To accurately bound the number of active S-boxes over many rounds, additional tools are required. A popular approach is to encode the problem as a mixed-integer linear programming problem. SAT and SMT solvers are also commonly used [73].

The wide trail strategy can be used to show that a primitive does not contain any good linear trails. This does not imply that all linear approximations must have low correlation. The next section discusses attempts to obtain somewhat stronger results.

2.3.2 Variance Bounds

Going beyond upper bounds on the correlation of individual trails, there is a series of papers [32, 57, 63–66, 87] concerned with upper bounding the variance of the correlation of linear approximations. Specifically, these works provide upper bounds on the average squared correlation in key-alternating block ciphers with independent round keys.

The starting point for such bounds is the observation that for a key-alternating cipher with round functions $x \mapsto F_i(x) + K_i$ for $i = 1, \dots r$, equality (2.4) becomes

$$C^{E_K}_{u_r,u_0} = \sum_{u_1,\dots,u_{r-1} \in \mathbb{F}_2^n} \prod_{i=1}^{r} (-1)^{u_i^\top K_i} C^{F_i}_{u_i,u_{i-1}}.$$

Hence, for independent and uniformly distributed round keys, one obtains

$$\mathrm{Var}[C^{E_{\boldsymbol{K}}}_{u_r,u_0}] = \sum_{u_1,\dots,u_{r-1} \in \mathbb{F}_2^n} \prod_{i=1}^{r} (C^{F_i}_{u_i,u_{i-1}})^2.$$

In order to upper bound the above quantity, one essentially needs to ensure that there are not too many trails with large correlation.

Remark. The assumption that round keys can be considered to be independent is an idealization which is invalid for all block ciphers that are used in practice. Hence, the resulting bounds are ultimately heuristic security arguments. Nevertheless, they are an improvement over bounds on individual trails.

There is a further and arguably more important limitation of variance bounds. As discussed in Section 2.2, the data requirements for linear cryptanalysis are generally proportional to $1/c^2$ with c the correlation of the underlying linear approximation. This has led some authors [57, 62, 65, a.o.] to claim that variance bounds result in provable security against linear cryptanalysis. The problem is, of course, that the squared correlation need not equal its expected value. Equivalently, weak key attacks are not covered by such arguments[1]. Chapter 4 explores approaches to bound the tail of the distribution of the correlation of linear approximations in order to resolve this issue. ▷

It can be shown that the variance does not increase when more rounds are added – provided of course that the round keys are independent. Consequently, most of the existing literature focuses on bounds for 4-round SPNs. At the core of these results are upper bounds for functions of the form

$$F_K(x) = S(L(S(x)) + K),$$

where L is a linear layer and S an S-box layer. For example, for the AES, this corresponds to the transformation applied to one column of the state after two rounds. In particular, the cell permutations and second key addition operation and linear layer can be left out as these do not influence the *maximum* absolute correlation. In this setting, one has

$$\mathrm{Var}[C^{F_{\boldsymbol{K}}}_{b,a}] = \sum_{u \in \mathbb{F}_2^n} (C^S_{b,u} C^S_{L^\top u,a})^2.$$

[1] It will be shown in Chapter 4 that variance bounds provide only limited restrictions on the number of weak keys.

Theorem 2.3 (Park *et al.* [87])**.** *Let $S = (S_1, \ldots, S_m)$ and let L be a linear map on $\mathbb{F}_2^n$ with linear branch number B_L (with respect to n/m-bit words). Then*

$$\sum_{u \in \mathbb{F}_2^n} (C^S_{b,u} C^S_{L^\top\, u,a})^2 \leq \max\left\{ \max_{1 \leq i \leq n/m} \sum_{u \in \mathbb{F}_2^m} (C^{S_i}_{u,a})^{2B_L}, \max_{1 \leq i \leq n/m} \sum_{u \in \mathbb{F}_2^m} (C^{S_i}_{b,u})^{2B_L} \right\} \leq c^{2(B_L - 1)},$$

where $c = \max_{1 \leq i \leq m} \max_{u,v \in \mathbb{F}_2^n \setminus \{0\}} |C^{S_i}_{u,v}|$ is the maximum absolute correlation of any linear approximation over any of the S-boxes $S_1, \ldots, S_m$.

Proof. For a proof of the first inequality, refer to Park *et al.* [87]; the proof of Theorem 4.1 will build on some of the ideas from [87]. The second inequality follows directly from the first, since

$$\sum_{u \in \mathbb{F}_2^m} (C^{S_i}_{u,a})^{2B_L} \leq c^{2(B_L-1)} \sum_{u \in \mathbb{F}_2^m} (C^{S_i}_{u,a})^2 = c^{2(B_L-1)}.$$

An alternate proof of the second inequality can be found in [57]. □

For the AES, $n = 32$, $m = 4$ (one state column) and $B_L = 5$. In this case, Theorem 2.3 yields the bound $48\,193\,441 \cdot 2^{-52}$. This result was improved to $31\,231\,767 \cdot 2^{-52}$ by Canteaut and Roué [32] at EUROCRYPT 2015. By means of an algorithmic search, Keliher [66] established the tight upper bound $109\,953\,193 \cdot 2^{-54}$.

In addition, one can show that for (more than) four rounds of the AES – and other SPNs with a similar cell permutation – the variance can be upper bounded by $(109\,953\,193 \cdot 2^{-54})^4 \approx 2^{-109.151}$. Of course, contrary to the two round bounds, this result assumes independent round keys.

2.3.3 Decorrelation Theory

Decorrelation theory is a design strategy that aims to achieve provable information-theoretical security. It was first proposed by Vaudenay [97, 99–101]. Several block ciphers have been based on this approach, perhaps most notably the *Decorrelated Fast Cipher* (DFC), which was an AES candidate [49]. The design of a few other block ciphers has also been based on decorrelation theory [35, 97, 98].

A block cipher $E_K : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is called perfectly dth order decorrelated if for any distinct $P_1, \ldots, P_d$, the probability distribution of $(E_{\boldsymbol{K}}(P_1), \ldots, E_{\boldsymbol{K}}(P_d))$ is uniform over the set of all d-tuples with distinct elements. More generally, the extend of dth order decorrelation may be expressed using some distance measure between d-wise distribution matrices. These are $2^{nd} \times 2^{nd}$ matrices with coordinates

$$\Pr[(E_{\boldsymbol{K}}(P_1), \ldots, E_{\boldsymbol{K}}(P_d)) = (C_1, \ldots, C_d)].$$

The construction of block ciphers with small d-wise decorrelation bias will not be discussed in detail here. One construction uses a 3-round Feistel construction instantiated with d-wise perfectly decorrelated *functions* such as polynomials of degree $d - 1$ with secret coefficients. Such constructions tend to inflate the key size,

but more recent results on Luby-Rackoff constructions such as [83] might allow for more efficient constructions.

Decorrelation theory allows one to prove security against a wide range of statistical attacks. Here, only linear cryptanalysis will be considered. In the following bound, a linear distinguisher is a hypothesis test based on the estimated correlation, which is computed from N randomly sampled (with replacement) plaintext/ciphertext pairs. The advantage of a distinguisher is the absolute value of the difference between its success probability and false positive rate.

Theorem 2.4 (Advantage of linear distinguishers [25])**.** *Let E_K be an n-bit block cipher. For any linear distinguisher $\mathcal{A}$ using at most N known plaintexts,*

$$\mathrm{Adv}_{\mathcal{A}} \leq 2\sqrt{N\,\varepsilon + \frac{N}{2^n - 1}} + 2\sqrt{\frac{N}{2^n - 1}},$$

where ε is a measure of 2-decorrelation defined by

$$\varepsilon = \max_{P_1, P_2 \in \mathbb{F}_2^n} \sum_{C_1, C_2 \in \mathbb{F}_2^n} \Bigg| \frac{\delta_{P_1 \neq P_2}\, \delta_{C_1 \neq C_2}}{2^n(2^n - 1)} + \frac{\delta_{P_1 = P_2}\, \delta_{C_1 = C_2}}{2^n} - \Pr\left[(E_{\boldsymbol{K}}(P_1), E_{\boldsymbol{K}}(P_2)) = (C_1, C_2)\right] \Bigg|.$$

Proof. Since decorrelation theory does not play an important role in this thesis, the proof will be omitted. The first term is due to the advantage of distinguishing between an approximation with expected squared correlation $\varepsilon + 1/(2^n - 1)$ and a zero-correlation approximation. The second term can be interpreted as an upper bound on the advantage of a zero-correlation distinguisher. □

The bound in Theorem 2.4 is in fact quite weak. The reason is twofold. First, one would hope to prove that a sufficiently decorrelated cipher is immune to linear cryptanalysis, regardless of the available number of known plaintexts. Such a bound can not be obtained using the proof strategy that was used for Theorem 2.4, due to the transition to zero-correlation approximations. In addition, the bound does not exclude low-advantage attacks such as those resulting from weak key classes. This has been illustrated with examples, see for instance Knudsen and Rijmen [67]. Nevertheless, it does appear to be possible to obtain stronger bounds than Theorem 2.4 by turning to higher decorrelation orders. This approach is related to the moments method that will be introduced in Section 4.2 to obtain correlation upper bounds in the weak key setting.

In practice, decorrelation theory is rarely used. This is due to its significant drawbacks: it usually requires an excessive key length, and the resulting bounds on the advantage of various distinguishers are often weak. Some of these issues might be resolved in future work, but for the moment decorrelation theory remains impractical.

2.4 Variants of Linear Cryptanalysis

Over the course of the last 25 years, many variants of linear cryptanalysis have been developed. Most of these will be mentioned in this section, and a few will be discussed in some detail.

Section 2.4.1 briefly discusses the use of multiple linear approximations. Invariants [14, 71, 94] are discussed in Section 2.4.2. Section 2.4.3 contains a concise account of zero-correlation linear cryptanalysis. Less common generalizations of linear cryptanalysis are discussed in Sections 2.4.4, 2.4.5 and 2.4.6. Section 2.4.4 discusses nonlinear approximations, which are of particular importance for Chapter 3.

Remark. As of yet, it may be unclear how some of the attacks in this section relate to linear cryptanalysis. In addition, some generalizations of linear cryptanalysis are merely hypothetical because the tools that might enable them have not yet been discovered. Chapter 3 of this thesis will provide a uniform description of all the attacks that are mentioned in this section. ▷

2.4.1 Multiple Approximations

One of the earliest generalizations of linear cryptanalysis is the use of more than one linear approximation. The main challenges – in addition to the many open questions in linear cryptanalysis – in this line of research are (1) applying techniques from multivariate statistics to build efficient distinguishers (2) understanding the joint behaviour of several linear approximations. Most of the literature is concerned with the former. This thesis is more concerned with the latter. Hence, the description below shall be brief.

Multiple linear cryptanalysis was introduced by Kaliski and Robshaw [61]. Further analysis is provided by Biryukov *et al.* [23]. In this attack, one uses several *independent* linear approximations. Usually, the sum of the squares of the estimated correlations is used as the test or ranking statistic.

Hermelin *et al.* [54, 55] introduced *multidimensional linear cryptanalysis*. In their attack, one uses a base set of linear approximations with linearly independent masks. In the distinguisher, one estimates the correlations of all linear approximations with masks in the span of the base set. Recently, Nyberg has introduced *affine linear cryptanalysis* as an improvement of this approach [85].

At FSE 2019, Biham and Perle [20] proposed *conditional linear cryptanalysis*. Contrary to multiple linear cryptanalysis, this attack explicitly takes into account dependencies between linear approximations. This leads to the best known attacks on the DES, thereby outperforming multiple and multidimensional linear cryptanalysis.

2.4.2 Invariants

Invariant subspace cryptanalysis was introduced by Leander *et al.* [71]. They showed that for `PRINTcipher`, there exists an affine space $a+V \subsetneq \mathbb{F}_2^n$ such that $E_K(a+V) = a+V$. That is, the set $a+V$ is encrypted to itself. This is usually a weak key

property, *i.e.* it only works for a subset of keys. Note that $\emptyset$ and $\mathbb{F}_2^n$ are trivially invariant under any block cipher.

At ASIACRYPT 2016, Todo, Leander and Sasaki [94] introduced another type of invariants. A Boolean function $g : \mathbb{F}_2^n \to \mathbb{F}_2$ is said to be a *nonlinear invariant* of E_K if there exists a key-dependent constant $c \in \mathbb{F}_2$ such that for all $x \in \mathbb{F}_2^n$

$$g(x) + g(E_K(x)) = c.$$

Equivalently, the set $\{x \in \mathbb{F}_2^n \mid g(x) = 0\}$ is encrypted to itself or to its complement.

At ASIACRYPT 2018, the author of this thesis showed that the invariants of a block cipher E_K can be defined in terms of the eigenvectors of its correlation matrix C^{E_K} [14]. This turns out to be useful to describe the invariants of some block ciphers.

Definition 2.4 (Block cipher invariant [14])**.** *Let E_K be a block cipher on $\mathbb{F}_2^n$. A vector $v \in \mathbb{C}\mathbb{F}_2^n$ is an invariant for E_K iff it is an eigenvector of the correlation matrix C^{E_K}.*

If one defines correlation matrices by Definition 2.3, then the well-posedness of Definition 2.4 is not surprising. Indeed, the eigenvectors of C^{E_K} are, up to a change of basis, the eigenvectors of T^F. If an eigenvector of the latter matrix is a probability mass function, then it must be a stable distribution for F by the definition of a transition matrix. In particular, an invariant set can be interpreted as the support of a uniform probability distribution on that set.

If an eigenvector of T^F does not happen to be a probability distribution, another type of property is obtained. For example, one can subtract two uniform distributions with complementary support. This gives rise to nonlinear invariants. Specifically, we have the following equivalence [14].

Theorem 2.5 (Corollary 1 from [14])**.** *Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a permutation with correlation matrix C^F. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function with correlation matrix $(\delta_0, v)^\top$. The vector v is an eigenvector of C^F with eigenvalue $(-1)^c$ iff for all $x \in \mathbb{F}_2^n$,*

$$f(x) + f(E_K(x)) = c.$$

Remark. In Theorem 2.5, f is necessarily a balanced Boolean function because the first row of its correlation matrix is equal to δ_0. The result generalizes (as does Definition 2.4) to non-invertible functions F. In addition, C^F could have eigenvalues other than ± 1. This yields new types of invariants. The general results in Chapter 3 will clarify this. ▷

2.4.3 Zero-correlation approximations

Zero-correlation linear cryptanalysis was introduced by Bogdanov and Rijmen [29]. The idea is to build a distinguisher based on one or more linear approximations with correlation zero. As discussed in Section 2.2, this only works if the correlation of the approximation is *exactly* zero.

Zero-correlation approximations can be found using a miss-in-the-middle approach based on the activity pattern of the masks. For the remainder of this thesis, it is not necessary to go into more details; a generalization of this method will be discussed in Chapter 3.

If only a single zero-correlation approximation is used, essentially half of the codebook must be available to obtain a successful distinguisher. However, the data cost can typically be reduced since in practice several zero-correlation approximations are often available.

2.4.4 Nonlinear approximations

A natural extension of linear cryptanalysis is the use of nonlinear approximations. Early attempts in this direction are the *I/O sums* of Harpes, Kramer and Massey [52] and the work of Knudsen and Robshaw [69] from EUROCRYPT 1995 and 1996 respectively. Despite this work, nonlinear approximations have mostly remained elusive. The author of this thesis believes this to be the consequence of several difficulties:

- No sufficiently powerful framework for describing nonlinear approximations has been developed. In particular, a clear approach to obtain accurate estimates of the correlation of nonlinear trails is missing.

- It is unclear which nonlinear approximations can be of practical use. A related issue is that there is no reason to expect that, *e.g.*, the correlation of a quadratic approximation might be accurately estimated by means of a limited number of "quadratic trails".

- Successfully using nonlinear approximations requires embracing the weak key or fixed key setting. Any sufficiently general approach to nonlinear cryptanalysis must make key-dependence explicit.

The recently introduced nonlinear invariant attack (discussed in Section 2.4.2) is a special case of nonlinear cryptanalysis, but has led to more significant cryptanalytic results [14, 94].

Inspired by these advances, recent work by Beierle *et al.* [9] reconsiders the use of nonlinear approximations. Their approach is to perform linear cryptanalysis on a transformed representation of the block cipher. This allows them to accurately describe the correlation of several approximations, but in some cases (Section 4.4 of their paper) the approach yields inaccurate estimates. They formulate a number of open problems related to this. One problem when working with transformations of a block cipher, is that it remains unclear how the transformation should be chosen.

Chapter 3 takes a different approach as part of a broader generalization of linear cryptanalysis. The open problems posed by Beierle *et al.* [9] will be resolved in Section 3.4.

2.4.5 Other groups

Granboulan *et al.* [50] and Baignères *et al.* [5] consider linear cryptanalysis over $\mathbb{F}_p$ with $p \neq 2$ a prime. More generally, one can consider block ciphers defined on arbitrary finite abelian groups. The theory in Chapter 3 will be developed in this general setting. Some additional aspects that occur even in the linear case but that were not addressed in earlier work [5] will be discussed.

2.4.6 Statistical cryptanalysis

Statistical cryptanalysis is often used as a collective term for all forms of cryptanalysis that rely, in one way or another, on statistics. This includes both linear and differential cryptanalysis.

The title of this section, however, specifically refers to Vaudenay's work on this subject [96]. In essence, his approach is to model the propagation of cryptanalytic properties by a Markov chain – this can be related to Chapter 3 of this thesis, despite fundamental differences. It can be argued that this particular approach to statistical cryptanalysis suffers from the same problems as nonlinear cryptanalysis (which is a special case, of course), *cf.* Section 2.4.4. Followup work [4, 58, 59] has primarily concentrated on aspects such as optimal hypothesis testing and key ranking, rather than on identifying and analyzing the properties themselves.

Chapter 3

A Geometric Approach to Linear Cryptanalysis

This chapter develops a broad generalization of linear cryptanalysis. Among other things, this includes the use of probabilistic nonlinear relations between the input and output bits of a block cipher. As the title suggests, the proposed generalization is based on an approach to linear cryptanalysis that is in some sense geometric. The starting point is that many cryptanalytic properties can be related to subspaces of an inner product space in which geometric concepts such as angles are well-defined. The reader is advised to keep this in mind.

It will be assumed that the function under analysis operates on a finite set X, which in practice is often a vector space over $\mathbb{F}_2$. Throughout this chapter, $F : X \to X$ denotes any function, such as a cryptographic permutation or a block cipher. More generally, one could consider probabilistic transitions – most of the theory below can be extended to this case, but this is left as future work.

This chapter is organized as follows. Section 3.1 discusses a brief introductory example. Some essential preliminaries are introduced in Section 3.2. The core of the geometric framework is developed in Section 3.3 and is subsequently used in Section 3.4 to resolve an open problem posed by Beierle *et al.* [9] at FSE 2019. Finally, Section 3.5 explores which types of approximations are most promising for future work and discusses how to find them using optimization techniques. Several examples are provided.

3.1 Introduction

In "classical" frequency analysis, one samples from a probability distribution at the input of F (typically, a cipher) and observes the probability distribution at the output. Let $p : X \to [0, 1]$ denote the probability mass function at the input. The output probability mass function q is then given by $q(y) = \sum_{x \in F^{-1}(y)} p(x)$. This can also be expressed as

$$q = T^F p, \tag{3.1}$$

where T^F is a $|X| \times |X|$ matrix with $T_{y,x} = \delta_{y,F(x)}$. If F is a bijection, then T^F is a permutation matrix. Note that, throughout this thesis, vector notation will be used for probability mass function such as p and q. This makes sense because the set $\mathbb{R}X$ of functions $X \to \mathbb{R}$ is an algebra[1] over $\mathbb{R}$.

The relation (3.1) can be used to compute q from p only in simple cases (for instance, when the block size n is small). In general, it is necessary to approximate (3.1) – this is the main subject of the present chapter. In addition, more general properties than probability mass functions will be considered. Specifically, all of the variants of linear cryptanalysis that were mentioned in Section 2.4 are covered by the result in this chapter.

3.2 Preliminaries

This section recalls a few results that will be used throughout this chapter. For a finite set X, let $\mathbb{C}X$ denote the vector space of functions from X to the field $\mathbb{C}$ of complex numbers. Note that $\mathbb{C}X \cong \mathbb{C}^{|X|}$. In particular, this is an inner product space with the inner product

$$\langle f, g \rangle = f^* g = \sum_{x \in X} \overline{f(x)} g(x),$$

where f^* denotes the conjugate transpose of f and $\overline{f(x)}$ is the complex conjugate of $f(x)$. It will be shown in Section 3.3 that linear cryptanalysis and its generalizations are all concerned with the propagation of low-dimensional subspaces of $\mathbb{C}X$ through a function. For this purpose, it will be useful to have a notion of similarity between two vector spaces $\mathcal{V}, \mathcal{U} \subseteq \mathbb{C}X$.

In the one-dimensional case, the natural definition is (the absolute value of) the inner product. In higher dimensions, one would like to extend this by computing inner products between several basis vectors. Definition 3.1 follows this idea, but avoids the problem of dependence on the basis.

Definition 3.1 (Principal correlations)**.** *Let $\mathcal{V}, \mathcal{U} \subseteq \mathbb{C}X$ be vector spaces and let V and U be matrices with columns that form an orthogonal basis for $\mathcal{V}$ and $\mathcal{U}$ respectively. The principal correlations between $\mathcal{V}$ and $\mathcal{U}$ are defined in the following equivalent ways:*

1. *The cosines of the principal angles between $\mathcal{V}$ and $\mathcal{U}$.*
2. *The singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d \geq 0$ of the matrix V^*U with $d = \min\{\dim \mathcal{V}, \dim \mathcal{U}\}$.*
3. *The largest d singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d \geq 0$ of $P_{\mathcal{V}} P_{\mathcal{U}}$ with $P_{\mathcal{V}}$ the orthogonal projector on $\mathcal{V}$.*

[1] Technically, only functions $X \to \mathbb{R}_{>0}$ are considered in this section. The set of such functions is a cone rather than a vector space.

Note that Definition 3.1 is well-posed, because the singular values do not depend on the choice of basis. In the one-dimensional case with $\mathcal{V} = \mathrm{Span}\{f\}$ and $\mathcal{U} = \mathrm{Span}\{g\}$, there is only one principal correlation:

$$\sigma_1 = \frac{|\langle f, g\rangle|}{\|f\|_2 \|g\|_2}.$$

In Section 3.3, further interpretations of the principal correlations will be given.

In the remainder of this section, we recall some notions related to harmonic analysis on finite abelian groups. This material can also be found in the lecture notes of Diaconis [47], or the books by Ceccherini-Silberstein *et al.* [34] and Luong [74]. For a general discussion of the representation theory of finite groups, see Serre [92]. In the following, abelian groups are written additively unless otherwise stated.

Definition 3.2 (Group characters [74, 92])**.** *Let G be a finite abelian group. By definition, a (complex) character of G is a group homomorphism $G \to \mathbb{C}^\times$. The (Pontryagin) dual of G is the group $\widehat{G}$ of all characters of G with respect to the pointwise product.*

It is easy to see that the characters indeed form a group under pointwise multiplication. Note that, more generally, characters can be defined for any group. However, extending the results below to the nonabelian case is more complicated (in this case, G has irreducible representations of degree at least two). The dual group has the following properties.

Theorem 3.1 ([47, 74, 92])**.** *Let G be a finite abelian group with dual $\widehat{G}$, then*

(1) $\widehat{G} \cong G$. In particular, $\widehat{G}$ is finite abelian.

(2) $\widehat{\widehat{G}} \cong G$, canonically through the evaluation map.

(3) If H is a finite abelian group, then $\widehat{G \oplus H} = \widehat{G} \times \widehat{H}$ with $\times$ the direct product.

(4) For the additive group of a finite field $\mathbb{F}_q \supseteq \mathbb{F}_p$ with p a prime, $\widehat{\mathbb{F}_q} = (\{x \mapsto \zeta^{\mathrm{Tr}(ux)} \mid u \in \mathbb{F}_q\}, \cdot)$ with $\zeta = e^{2\pi i/p}$ and where the trace is taken with respect to the base field $\mathbb{F}_p$.

(5) The characters of G are orthogonal in the sense that

$$\langle \chi, \psi \rangle = \sum_{x \in G} \overline{\chi(x)}\, \psi(x) = |G| \delta_{\chi,\psi}.$$

The main reason for introducing group characters is the definition of the Fourier transform, which is given below.

Definition 3.3 (Fourier transformation [47, 74])**.** *Let $f : G \to \mathbb{C}$, then the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ with*

$$\widehat{f}(\chi) = \sum_{x \in G} f(x)\, \chi(x),$$

is called the Fourier transform of f. The Fourier transformation is the invertible operator $\mathscr{F} : \mathbb{C}G \to \mathbb{C}\widehat{G}$ defined by $\mathscr{F}f = \widehat{f}$. In fact, one can check that $\mathscr{F}$ is an isomorphism of vector spaces.

Note that Definition 3.3 defines $\widehat{f}$ as a function on $\widehat{G}$, but we can think of this as a function on G after choosing an isomorphism between G and $\widehat{G}$. In general, there is no unique choice for the isomorphism.

In practice, one often wants to obtain f from $\widehat{f}$ – this is indeed possible because Definition 3.3 defines an invertible transformation. Theorem 3.2 gives an expression for the inverse. This result follows directly from the orthogonality of group characters.

Theorem 3.2 (Inverse Fourier transformation [47])**.** *Let $f : G \to \mathbb{C}$ with Fourier transformation $\widehat{f} : \widehat{G} \to \mathbb{C}$, then:*

$$f(x) = \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \widehat{f}(\chi)\, \overline{\chi(x)},$$

Theorem 3.2 implies that the Fourier transformation $\mathscr{F}$ is – up to a factor $|G|$ – a unitary transformation: $\mathscr{F}^{-1} = |G|^{-1}\mathscr{F}^*$. It follows that $\langle \mathscr{F}f, \mathscr{F}g \rangle = |G| \langle f, g \rangle$.

As will be shown in Section 3.3, the Fourier transformation is not absolutely necessary to *theoretically* describe linear cryptanalysis and its generalizations. Nevertheless, it essential from a practical point of view. This is mostly a consequence of the following property.

Theorem 3.3 (Convolution property [47, 74])**.** *Define the convolution $f * g \in \mathbb{C}G$ of two functions $f, g \in \mathbb{C}G$ by*

$$(f * g)(x) = \sum_{y \in G} f(x - y) g(y),$$

then the Fourier transformation of $f * g$ satisfies

$$\widehat{f * g}\,(\chi) = \widehat{f}(\chi)\, \widehat{g}(\chi).$$

Equivalently, the Fourier transform $\mathscr{F} : (\mathbb{C}G, *) \to (\mathbb{C}\widehat{G}, \cdot)$ is an algebra isomorphism.

3.3 General Theory

Recall from Section 3.1 that one can describe the propagation of a probability mass function $p : X \to [0, 1]$ through a function F by a matrix-vector product $T^F p$. In this section, we will consider the propagation of arbitrary subspaces $\mathcal{V} \subseteq \mathbb{C}X$. It will be shown that this is necessary to describe linear cryptanalysis and its variants in full generality.

Remark. Instead of thinking about the probability mass function p, one could consider the induced probability measure $\mu(S) = \sum_{x \in S} p(x)$ on $\{S \mid S \subseteq X\}$. More generally, for any $f \in \mathbb{C}X$, there is an associated complex-valued measure. This point of view sometimes enhances intuition. ▷

Transition matrices and correlation matrices are introduced in Section 3.3.1. When F is a "complicated" function, and n is large, it is often impractical to work directly with the transformed vector spaces $T^F\mathcal{V}$. Hence, an approximate approach is necessary. This is discussed in Section 3.3.2.

Section 3.3.3 relates functions $f : X \to Y$ to subspaces $\mathcal{V} \subseteq \mathbb{C}X$. This provides the link with the "classical" point of view, where one considers functions (typically Boolean) on the state. The relation with the principal correlations is also discussed.

Usually, F can be decomposed into a sequence of "simple" steps as $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$. In this case, one would like to combine ("pile-up") successive approximations. This is possible by means of a generalization of the piling-up lemma, which is derived in Section 3.3.4.

Section 3.3.5 briefly discusses the role of invariants and (generalized) zero-correlation approximations. Finally, Section 3.3.6 addresses the effect of sampling: in practice, one compares a model (*i.e.* theoretical approximation) of F with the behaviour of F on a relatively small sample of X.

3.3.1 Transition Matrices

Let X and Y be finite sets and $F : X \to Y$ a map between them. Given a function $f \in \mathbb{C}X$, we can ask what the corresponding $g \in \mathbb{C}Y$ is such that g satisfies the requirement

$$g(y) = \sum_{x \in F^{-1}(y)} f(x).$$

The main motivation behind this definition comes from the case where f and g are probability distributions. If f is a probability distribution and $\boldsymbol{x} \sim f$, then $F(\boldsymbol{x}) \sim g$.

Remark. The construction above is called a pushforward. Specifically, consider the complex-valued measure $\mu_f(S) = \sum_{x \in S} f(x)$ defined by f. Then the measure $\mu_g(S) = \sum_{x \in S} g(x) = \mu_f(F^{-1}(S))$ is precisely the *pushforward measure* of μ_f with respect to F. ▷

This correspondence defines a linear map $T^F : \mathbb{C}X \to \mathbb{C}Y$. Furthermore, in the δ-function basis, this map is represented by the matrix

$$T^F_{y,x} = \delta_{y,F(x)}.$$

Recall from Section 2.1, that the matrix T^F is called the transition matrix of F (*cf.* Definition 2.2). It coincides with the notion of a transition matrix in a Markov chain with deterministic transitions. In fact, the theory presented in this chapter can be generalized to arbitrary transition matrices. However, it is unclear if this would have any applications – indeed, it must be reiterated that the key is *not* to be treated as a random variable when F represents a block cipher.

For the remainder of this section, suppose that $X = G$ and $Y = H$ with G and H finite abelian groups. In this case, it is often convenient (for practical reasons) to work in the basis of group characters rather than in the basis of δ-functions. This leads to the following definition, which generalizes the notion of a correlation matrix as introduced in Definition 2.3.

Definition 3.4 (Correlation matrix)**.** *Let $F : G \to H$ be a function with G and H finite abelian groups. The correlation matrix C^F of the function F is the representation of the linear map T^F with respect to the character bases of $\mathbb{C}G$ and $\mathbb{C}H$, i.e. the Fourier transformation of the matrix T^F. Specifically,*

$$C^F_{\chi,\psi} = \frac{1}{|G|} \sum_{x \in G} \chi(F(x))\, \overline{\psi(x)}.$$

Note that we can think of the linear transformation represented by C^F either as a map $\mathbb{C}\widehat{G} \to \mathbb{C}\widehat{H}$, or (after choosing isomorphisms of G and H with their duals) as $\mathbb{C}G \to \mathbb{C}H$.

Note that Definition 3.4 is essentially the same as Definition 2.3 from Section 2.1 for the case $G = \mathbb{F}_2^n$ and $H = \mathbb{F}_2^m$, which is the definition I used in [14]. In that case, the original definition by Daemen *et al.* [40] coincides with the expression for the coordinates $C^F_{\chi,\psi}$.

The use of group characters to describe linear cryptanalysis on finite abelian groups other than $\mathbb{F}_2^n$ is not new, see the work of Granboulan *et al.* [50] and Baignères *et al.* [5]. However, they do not use correlation matrices. Moreover, the reader who is familiar with these works will note that the results in Section 3.3.3 imply that their approach is somewhat incomplete.

The following theorems state a few basic properties of transition matrices. Special cases of some of these also appear in Theorem 2.1.

Theorem 3.4 (Properties of transition matrices.)**.** *We have*

1. *For $F : X \to Y$ and $F' : Y \to Z$, $T^{F' \circ F} = T^{F'} T^F$.*
2. *If $F = (F_1, \ldots, F_n) : \prod_{i=1}^k X_i \to \prod_{i=1}^l Y_i$ with $F_i : X_i \to Y_i$, then $T^F = \bigotimes_{i=1}^n T^{F_i}$.*

3. *If $F : X \to Y$ is a bijection, T^F is a permutation matrix.*

The first two results in Theorem 3.4 also apply to correlation matrices. The only necessary change is notational. Theorem 3.5 states a few additional properties which are specific to correlation matrices. No proofs are given, but in most cases the proof from Theorem 2.1 or [40] can be generalized in a straightforward manner.

Theorem 3.5 (Properties of correlation matrices.)**.** *Let G and H be finite abelian groups and $F : G \to H$ a function. Then*

1. *If F is a bijection, then C^F is a unitary matrix.*
2. *If F is a group homomorphism, then $C^F_{\chi,\psi} = \delta_{\chi\circ F,\psi}$. Furthermore, if F is an isomorphism, then C^F is a permutation matrix.*
3. *If $F(x) = x + c$ for some constant c, then C^F is diagonal with $C^F_{\chi,\chi} = \chi(c)$.*

3.3.2 Vector Space Approximations

Given a vector space $\mathcal{V}$, the pushforward $T^F\,\mathcal{V}$ will in general not be "nice" enough to be useful. Specifically:

- Suppose that F is a (cryptographic) permutation. If $\mathcal{V}$ corresponds (per Section 3.3.3) to some function $f : X \to Y$, then $T^F\,\mathcal{V}$ will correspond to $f \circ F$. If F is a good permutation, the function $f \circ F$ is of no use in the analysis of the mode in which F is used.
- When F is a block cipher, $T^F\,\mathcal{V}$ will strongly depend on the key. This implies that $f \circ F$ can usually not even be evaluated by the attacker.

Hence, there is a need to approximate $T^F\,\mathcal{V}$ with another vector space $\mathcal{U}$. This leads to the following definition.

Definition 3.5 ((Vector space) approximation)**.** *An approximation of a function F is a pair of subspaces $\mathcal{V},\mathcal{U} \subseteq \mathbb{C}X$. The approximation map associated to $(\mathcal{U},\mathcal{V})$ is defined by*

$$\mathcal{T}^F_{\mathcal{U},\mathcal{V}} = P_{\mathcal{U}}\,T^F\,P_{\mathcal{V}}.$$

The principal correlations of $(\mathcal{U},\mathcal{V})$ are the $\min\{\dim\mathcal{U},\dim\mathcal{V}\}$ largest singular values of $\mathcal{T}^F_{\mathcal{U},\mathcal{V}}$. If F is a bijection, then these singular values correspond to the principal correlations (as introduced in Definition 3.1) between the vector spaces $\mathcal{U}$ and $T^F\,\mathcal{V}$.

If U is a basis for $\mathcal{U}$ and V is a basis for $\mathcal{V}$, then

$$\mathcal{T}^F_{\mathcal{U},\mathcal{V}} = U(U^*U)^{-1}\,[U^*T^FV]\,(V^*V)^{-1}V^*.$$

Hence, $\mathcal{T}^F_{\mathcal{U},\mathcal{V}}$ is completely determined by U^*T^FV. In practice, computing the matrix U^*T^FV is nontrivial – Section 3.3.4 generalizes the piling-up lemma for this purpose.

Definition 3.5 is quite broad: it includes linear, multiple linear, nonlinear, invariant subspaces, nonlinear invariants and many other types of approximations. An overview of special cases can be found in Section 3.6. Two basic examples are given below.

Example. Let $A, B \subseteq X$. If $f(x) = \mathbf{1}_A(x)/|A|^{1/2}$ and $g(x) = \mathbf{1}_B(x)/|B|^{1/2}$, then

$$\langle g, T^F f\rangle = \langle (T^F)^\top g, f\rangle = \frac{|F(A) \cap B|}{|F(A)|^{1/2}\,|B|^{1/2}}.$$

This value is equal to the first (and only) principal correlation of the approximation. More generally, if f and g are probability distributions, $\langle f, g\rangle$ is the expected likelihood $\mathbb{E} f(\boldsymbol{x})$ with $\boldsymbol{x}$ a random variable with probability distribution g. ▷

Example. Suppose $X = \mathbb{F}_2^n$ and let $u, v \in \mathbb{F}_2^n$ be "masks". If $f(x) = (-1)^{u^\top x}/2^{n/2}$ and $g(x) = (-1)^{v^\top x}/2^{n/2}$, then

$$\langle g, T^F\, f\rangle = \langle (T^F)^\top g, f\rangle = 2\Pr\left[u^\top \boldsymbol{x} = v^\top F(\boldsymbol{x})\right] - 1$$

with $\boldsymbol{x}$ uniformly distributed over X. Section 3.3.3 will explain the choice of f and g. ▷

To stress the difference with special cases such as linear approximations, the approximations from Definition 3.5 will occasionally be called "vector space approximations". A plausible alternative term for "approximation" would be "hull", in reference to the notion of a linear hull. This terminology will be avoided since there is no complete consensus on its definition in the case of linear cryptanalysis [82, 84].

Remark. If F is a keyed function, then the vector spaces $\mathcal{V}$ and $\mathcal{U}$ will in general depend on the key. In this thesis, the fixed-key setting is considered.[2] In practice, this means that the model of the cipher will be a function of the key. This is not a limitation (on the contrary, it is an advantage) when one is interested in evaluating the model. However, this does complicate the *selection* of the final model. Section 3.5 contains a brief summary of the issues. ▷

An important advantage of Definition 3.5 is that it is essentially basis-free. One can reasonably argue that any measure of the quality of an approximation should be independent of the choice of basis as well. The principal correlations provide one such measure, but they are not the only option. Section 3.3.4 focuses on approximating the entire map $\mathscr{T}^F_{\mathcal{U},\mathcal{V}}$ rather than only its singular values.

Two approximation measures that are related to the principal correlations are worth mentioning. The first measure is the first principal correlation σ_1. It satisfies

$$\sigma_1 = \|\mathscr{T}^F_{\mathcal{U},\mathcal{V}}\|_2 = \max_{\substack{f\in\mathcal{V}\\ g\in\mathcal{U}}} \frac{|\langle g, T^F f\rangle|}{\|f\|_2\,\|g\|_2}.$$

That is, σ_1 measures the quality of the optimal one-dimensional approximation in the vector space approximation. In multiple linear cryptanalysis, it is quite common to use the sum of the squares of the correlations of several approximations ("capacity"). This can be motivated from a statistical point of view. Baignères, Junod and

[2] The difference between the fixed-key and random-key setting may seem obvious, but lack of clearness has (continues to?) lead to substantial confusion. See for example Murphy [82].

Vaudenay [4] show that, under suitable conditions, this quantity determines the cost (data, success probability, …) of an optimal distinguisher. The capacity is closely related to

$$\sum_{i=1}^{d} \sigma_i^2 = \|\mathscr{T}^F_{\mathcal{U},\mathcal{V}}\|_F^2, \tag{3.2}$$

with $d = \min\{\dim \mathcal{V}, \dim \mathcal{U}\}$. Note that (3.2) is not quite the same as the notion of "capacity" in multiple linear cryptanalysis because (3.2) possibly contains additional terms. Section 3.3.3 should clarify this.

3.3.3 Relation with Nonlinear Functions

In Section 3.3.2, vector space approximations were defined. However, it is not yet clear how subspaces of $\mathbb{C}X$ relate to, for example, linear approximations in the classical sense. The only concrete example that has been mentioned so far is that of probability distributions.

Given a function $f : X \to Y$ (typically, $|Y| \ll |X|$), we define an associated subspace $\mathcal{V} \subseteq \mathbb{C}X$ as follows:

$$\mathcal{V} = \{g \circ f \mid g \in \mathbb{C}Y\}.$$

That is, $\mathcal{V}$ is the subspace of functions that extend a function in $\mathbb{C}Y$ to a function on $\mathbb{C}X$ by composing it with f.

Remark. Baignères, Junod and Vaudenay call functions such as f "projections" [4]. This goes back to Vaudenay's "statistical cryptanalysis" which indeed considers probability distributions on the smaller space Y. However, Baignères *et al.* do not "pull back" these distributions to X – unlike in the construction of $\mathcal{V}$ above. This leads to difficulties when they attempt to generalize the piling-up lemma. ▷

The vector space $\mathcal{V}$ is closely related to the transition matrix T^f of f. Indeed, the set of functions $\{\delta_y \circ f\}_{y \in Y}$ is an orthogonal basis for $\mathcal{V}$. Hence, we have

$$\mathcal{V} = \operatorname{Span}\{\delta_y \circ f \mid y \in Y\} = \operatorname{Row} T^f.$$

That is, the rows of T^f form an orthogonal basis for $\mathcal{V}$.

We now consider why it makes sense to say that functions $f, g : X \to Y$ correspond to approximations (in the sense of Definition 3.5) of $F : X \to X$. Let $\mathcal{V}$ and $\mathcal{U}$ be the vector spaces associated with f and g respectively. Define

$$N_{y,x} = |\{z \in X \mid f(z) = x \wedge g(F(z)) = y\}|.$$

Since the rows of T^f form an orthogonal basis for $\mathcal{V}$ and likewise for T^g and $\mathcal{U}$, we have $\mathscr{T}^F_{\mathcal{U},\mathcal{V}} = (T^g)^\top [T^g T^F (T^f)^\top] T^f$ with

$$T^g T^F (T^f)^\top = \Big[\sum_{z \in X} \delta_y(g(F(z)))\delta_x(f(z)) \Big]_{y,x} = \Big[N_{y,x} \Big]_{y,x}.$$

In particular, the principal correlations between $\mathcal{U}$ and $T^F\mathcal{V}$ are the singular values of this matrix. There are also other quantities that could be of interest, for example

$$\operatorname{Tr}\left(T^g T^F (T^f)^\top\right) = |\{x \in X \mid f(x) = g(F(x))\}|.$$

Example. Assume that F is a bijection. For $Y = \mathbb{F}_2$ and assuming f and g are balanced, we have

$$T^g T^F (T^f)^\top = \begin{pmatrix} N_{0,0} & N_{0,1} \\ N_{1,0} & N_{1,1} \end{pmatrix} = |X| \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

with $c = 2(N_{0,0} + N_{1,1})/|X| - 1$. Hence, the singular values are $\sigma_1 = 1$ and $\sigma_2 = |c|$. This example shows that if the character basis is used, then balanced nonlinear approximations can essentially be described using one-dimensional vector spaces. For $|Y| > 2$, this is no longer true. ▷

For the remainder of this section, assume $Y = H$ is a finite abelian group. In the above example, the characters of $\widehat{H}$ were composed with f to obtain an orthogonal basis for $\mathcal{V}$ (and similarly for $\mathcal{U}$). Indeed, this is possible because $\mathcal{V} = \operatorname{Span}\{\chi \circ f \mid \chi \in \widehat{H}\}$. For F bijective, the trivial character is an eigenvector of T^F so (because $|H| = 2$) this basis diagonalizes $T^g T^F (T^f)^\top$. When $|H| > 2$, the trivial character is still an eigenvector of T^F, but the other characters are not (in general).

Remark. In the above discussion, we have *not* assumed that X is an abelian group, nor have we taken Fourier transformations of functions in $\mathbb{C}X$. That is, only the choice of basis for $\mathcal{V}$ has been discussed. When $X = G$ with G a finite abelian group, we can additionally take the Fourier transformation of these basis functions. We then consider the vector space $\widehat{\mathcal{V}} = \mathscr{F}\mathcal{V} \subseteq \mathbb{C}\widehat{G}$. After choosing an isomorphism between G and $\widehat{G}$, we can think of this additional Fourier transformation as a change of basis in the ambient space $\mathbb{C}G$ of $\mathcal{V}$. ▷

It is worthwhile to consider a few special classes of functions $f : G \to H$. If f is a homomorphism, then $\chi \circ f$ with $\chi \in \widehat{H}$ is a homomorphism too. That is, $\chi \circ f \in \widehat{G}$. Hence, in this case $\mathcal{V}$ is the span of a subgroup of $\widehat{G}$.

Example. If $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a homomorphism, then the corresponding vector space $\mathcal{V}$ is given by $\operatorname{Span}\{1, \psi\}$ where 1 denotes the trivial character $x \mapsto 1$ and $\psi = (-1)^f \in \widehat{\mathbb{F}_2^n}$. Note that such a homomorphism f is necessarily of the form $f(x) = u^\top x$. ▷

If $X = \prod_{i=1}^m X_i$ (in some natural way), then it is useful to consider the correspondence $\mathbb{C}X \cong \otimes_{i=1}^m \mathbb{C}X_i$. One can then define the tensor rank of elements of $\mathcal{V} \subseteq \mathbb{C}X$. In particular, vector spaces $\mathcal{V}$ which have a basis consisting entirely of rank-one tensors form an interesting class of approximations which will be discussed in Section 3.5.

If $f(x_1, \ldots, x_m) = \sum_{i=1}^m f_i(x_i)$, then $(\chi \circ f)(x) = \prod_{i=1}^m (\chi \circ f_i)(x_i)$. That is, the vector space $\mathcal{V}$ has a basis $\{\otimes_{i=1}^m (\chi \circ f_i) \mid \chi \in \widehat{H}\}$. Such vector spaces will be

discussed in Section 3.5.2. Note that this is related to a property of correlation matrices: if $F(x_1, x_2) = F_1(x_1) + F_2(x_2)$, then the rows of C^F satisfy

$$C^F_{\cdot,\chi\otimes\psi} = C^{F_1}_{\cdot,\chi} \otimes C^{F_2}_{\cdot,\psi}.$$

The above equality is sometimes also used in the description of the correlation matrix of the primitive F, in particular when F is a Feistel cipher.

3.3.4 Piling Up Approximations

Suppose that F can be written as the composition of a sequence of simple functions: $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$. Note that we have $T^F = \prod_{i=1}^{r} T^{F_{r-i+1}}$. In general, it is too difficult to approximate F directly. Instead, one approximates each of the functions F_i separately, and then combines these approximations. Extending the terminology that is used in linear cryptanalysis, such a sequence of successive approximations will be called a trail. Formally, we have the following definition.

Definition 3.6 ((Vector space) trail)**.** *A trail of vector spaces for a function $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$ is a tuple of subspaces $\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_r \subseteq \mathbb{C}X$. The approximation map of the trail $(\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_r)$ is defined as*

$$\prod_{i=1}^{r} \mathscr{T}^{F_{r-i+1}}_{V_{r-i+1},\mathcal{V}_{r-i}} = \prod_{i=1}^{r} P_{\mathcal{V}_{r-i+1}} T^{F_{r-i+1}} P_{\mathcal{V}_{r-i}},$$

and its nonzero singular values are called principal correlations in accordance with Definition 3.5. If the vector spaces $\mathcal{V}_i$ are one-dimensional and spanned by vectors f_i for $i = 0, \ldots, r$, then we shall also say that $f_0, f_1, \ldots, f_r$ form a trail.

In some cases, the correlation of an approximation is approximately equal to the correlation of a trail. In linear cryptanalysis, this result is known as the piling-up lemma. Theorem 3.6 generalizes the piling-up lemma. In addition, it quantifies the error that is made in this approximation.

Before considering the general setting of vector space approximations, the one-dimensional case will be discussed. In this case, we obtain Corollary 3.1.

Corollary 3.1 (One-dimensional piling-up principle.)**.** *Let $f_0, f_1, \ldots, f_r \in \mathbb{C}X$ be $r+1$ unit vectors which form a trail for a function $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$. Then there exist vectors $f_i^\perp$, $i = 1, \ldots r$ such that $\langle f_i, f_i^\perp \rangle = 0$ and*

$$\langle f_r, T^F f_0 \rangle = \prod_{i=1}^{r} \langle f_i, T^{F_i} f_{i-1} \rangle + \sum_{i=1}^{r} \langle f_r, (\textstyle\prod_{j=1}^{r-i} T^{F_{r-j+1}}) f_i^\perp \rangle \langle f_i^\perp, T^{F_i} f_{i-1} \rangle \prod_{j=1}^{i-1} \langle f_j, T^{F_j} f_{j-1} \rangle.$$

Proof. The result follows from Theorem 3.6, which is stated below. Take $\mathcal{V}_i = \mathrm{Span}\{f_i\}$, then $P_{\mathcal{V}_i} = f_i f_i^*$. Then Theorem 3.6 yields

$$\begin{aligned}
&\langle f_r, T^F f_0\rangle f_r f_0^* \\
&= f_r f_0^* \prod_{i=1}^{r} \langle f_i, T^F f_{i-1}\rangle + f_r \sum_{i=1}^{r} \langle f_r, (\textstyle\prod_{j=1}^{r-i} T^{F_{r-j+1}}) f_i^{\perp}\rangle (f_i^{\perp})^* \prod_{j=1}^{i} T^{F_{i-j+1}} f_{i-j}\, f_{i-j}^* \\
&= f_r f_0^* \prod_{i=1}^{r} \langle f_i, T^F f_{i-1}\rangle + f_r f_0^* \sum_{i=1}^{r} \langle f_r, (\textstyle\prod_{j=1}^{r-i} T^{F_{r-j+1}}) f_i^{\perp}\rangle \langle f_i^{\perp}, T^{F_i} f_{i-1}\rangle \prod_{j=1}^{i-1} \langle f_j, T^{F_j} f_{j-1}\rangle,
\end{aligned}$$

which implies the result. □

Corollary 3.1 essentially states that the correlation – there is only one – of a one-dimensional approximation is equal to the product of the correlations of the intermediate approximations of a trail plus an error term. By choosing the optimal trail, one hopes to minimize the error term (but this is not guaranteed to succeed).

Now consider the general case of vector space approximations with arbitrary dimension. Figure 3.1 illustrates the geometric interpretation of the piling-up principle. The vector space $\mathcal{V}_0$ is transformed to $T^{F_1}\mathcal{V}_0$, which is then approximated by $\mathcal{V}_1$. One subsequently approximates $T^{F_2}\mathcal{V}_1$ by another vector space $\mathcal{V}_3$. As will be discussed below, each of these approximation steps corresponds to an orthogonal projection.

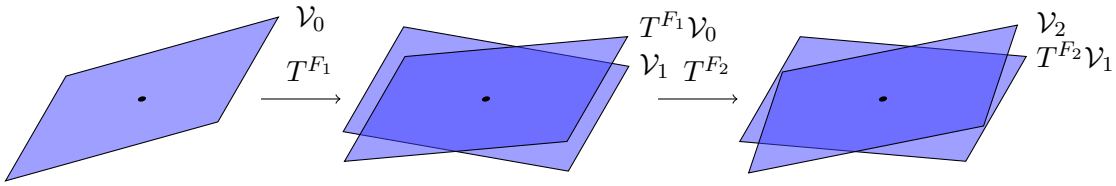


Figure 3.1: Schematic illustration of the piling-up principle.

Using Definitions 3.5 and 3.6, the general piling-up principle can be expressed as follows: the approximation matrix of a vector space approximation in the sense of Definition 3.5 can (sometimes) be approximated by the approximation matrix of a vector space trail. This is made formal, including the error term, in the next theorem. Note that Corollary 3.1 is a special case.

Theorem 3.6 (The general piling-up principle.)**.** *Let $\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_r$ be a vector space trail for a function $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$. Then*

$$\mathscr{T}^{F}_{\mathcal{V}_r,\mathcal{V}_0} = \prod_{i=1}^{r} \mathscr{T}^{F_{r-i+1}}_{\mathcal{V}_{r-i+1},\mathcal{V}_{r-i}} + \sum_{i=1}^{r} \mathscr{T}^{F_r\circ\cdots\circ F_{i+1}}_{\mathcal{V}_r,\mathcal{V}_i^{\perp}}\; T^{F_i}\; (\textstyle\prod_{j=1}^{i-1} \mathscr{T}^{T_{F_{i-j}}}_{\mathcal{V}_{i-j},\mathcal{V}_{i-j-1}}).$$

Proof. The proof considers the following sequence of projections:

$$\begin{array}{ccccccc} P_{\mathcal{V}_0} & \longrightarrow & P_{\mathcal{V}_1}T^{F_1}P_{\mathcal{V}_0} & \longrightarrow & P_{\mathcal{V}_2}T^{F_2}P_{\mathcal{V}_1}T^{F_1}P_{\mathcal{V}_0} & \longrightarrow & \cdots \\ & & + & & + & & \\ & & P_{\mathcal{V}_1^\perp}T^{F_1}P_{\mathcal{V}_0} & & P_{\mathcal{V}_2^\perp}T^{F_2}P_{\mathcal{V}_1}T^{F_1}P_{\mathcal{V}_0} & & \\ & & & & + & & \\ & & & & T^{F_2}P_{\mathcal{V}_1^\perp}T^{F_1}P_{\mathcal{V}_0} & & \end{array}$$

More formally, the result follows by showing (using induction on m) that for all $1 \leq m \leq r$,

$$T^{F_m \circ \cdots \circ F_1} P_{\mathcal{V}_0} = \prod_{i=1}^{m} P_{\mathcal{V}_{m-i+1}} T^{F_{m-i+1}} P_{\mathcal{V}_{m-i}} + \sum_{i=1}^{m} (\textstyle\prod_{j=1}^{m-i} T^{F_{m-j+1}}) P_{\mathcal{V}_i^\perp} (\prod_{j=1}^{i} T^{F_{i-j+1}} P_{\mathcal{V}_{i-j}}).$$

For $m = 1$, we have $T^{F_1}P_{\mathcal{V}_0} = P_{\mathcal{V}_1}T^{F_1}P_{\mathcal{V}_0} + P_{\mathcal{V}_1^\perp}T^{F_1}P_{\mathcal{V}_0}$. For $m+1$, the induction hypothesis then implies that

$$\begin{aligned} & T^{F_{m+1} \circ \cdots \circ F_1} P_{\mathcal{V}_0} \\ & = T^{F_{m+1}} \prod_{i=1}^{m} P_{\mathcal{V}_{m-i+1}} T^{F_{m-i+1}} P_{\mathcal{V}_{m-i}} + \sum_{i=1}^{m} (\textstyle\prod_{j=1}^{m+1-i} T^{F_{m+1-j+1}}) P_{\mathcal{V}_i^\perp} (\prod_{j=1}^{i} T^{F_{i-j+1}} P_{\mathcal{V}_{i-j}}) \\ & = (P_{\mathcal{V}_{m+1}} + P_{\mathcal{V}_{m+1}^\perp}) \prod_{i=1}^{m+1} T^{F_{m+1-i+1}} P_{\mathcal{V}_{m+1-i}} + \sum_{i=1}^{m} (\textstyle\prod_{j=1}^{m+1-i} T^{F_{m+1-j+1}}) P_{\mathcal{V}_i^\perp} (\prod_{j=1}^{i} T^{F_{i-j+1}} P_{\mathcal{V}_{i-j}}) \\ & = P_{\mathcal{V}_{m+1}} \prod_{i=1}^{m+1} T^{F_{m+1-i+1}} P_{\mathcal{V}_{m+1-i}} + \sum_{i=1}^{m+1} (\textstyle\prod_{j=1}^{m+1-i} T^{F_{m+1-j+1}}) P_{\mathcal{V}_i^\perp} (\prod_{j=1}^{i} T^{F_{i-j+1}} P_{\mathcal{V}_{i-j}}). \end{aligned}$$

This establishes the result. □

Note that the error term corresponds to the approximation matrix of a trail $\mathcal{V}_0, \ldots, \mathcal{V}_{i-1}, \mathcal{V}_i^\perp, \mathcal{V}_r$ for the function $G \circ F_i \circ F_{i-1} \circ \cdots \circ F_1$ with $G = F_r \circ F_{r-1} \circ \cdots \circ F_{i+1}$. For practical purposes, it is useful to restate Theorem 3.6 in the form of Corollary 3.2.

Corollary 3.2 (The general piling-up approximation.)**.** *Let $\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_r$ be a vector space trail for a function $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$ and let $V_0, V_1, \ldots, V_r$ be matrices whose columns are a basis for $\mathcal{V}_0, \mathcal{V}_1, \cdots, \mathcal{V}_r$ respectively. Suppose that for all $i = 1, \ldots, r$,*

$$\| \mathscr{T}_{\mathcal{V}_r, \mathcal{V}_i^\perp}^{F_r \circ \cdots \circ F_{i+1}} \; T^{F_i} \; (\textstyle\prod_{j=1}^{i-1} \mathscr{T}_{\mathcal{V}_{i-j}, \mathcal{V}_{i-j-1}}^{T_{F_{i-j}}}) \|_2 \leq \epsilon$$

Then there exists a matrix E with $\|E\|_2 \leq r\epsilon$ such that

$$V_r^* \, T^F \, V_0 = (V_r^* V_r) \prod_{i=1}^{r} (V_{r-i+1}^* V_{r-i+1})^{-1} (V_{r-i+1}^* \, T^{F_{r-i+1}} V_{r-i}) + V_r^* \, E \, V_0.$$

In particular, if the matrices V_i have orthogonal columns, then (with $\|E'\|_2 \leq \epsilon$)

$$V_r^* \, T^F \, V_0 = \prod_{i=1}^{r} V_{r-i+1}^* \, T^{F_{r-i+1}} V_{r-i} + E'.$$

Remark. Theorem 3.6 and Corollary 3.2 above are formulated in terms of the transition matrices T^F. Of course, they have an equivalent formulation in terms of the correlation matrices C^F. The only difference is notational. ▷

In practice, the choice of the vector spaces $\mathcal{V}_i$ that constitute a trail is limited in several ways. Clearly, $\dim \mathcal{V}_i$ should not be too large. In addition, $\mathcal{V}_i$ must have a basis matrix V_i such that $V_i^* T^{F_i} V_{i-1}$ can be computed efficiently.

For convenience, denote the set of vector spaces which are acceptable for the attacker by $\mathcal{T}$. The set of acceptable trails is then $\mathcal{T}^{r+1}$. For instance, in linear cryptanalysis this set contains $(r+1)$-tuples of vector spaces spanned by a small number of characters of the group $\mathbb{F}_2^n$.

It is possible that $\|E\|_2$ is large because there are *a few* other trails in $\mathcal{T}^{r+1}$ which are a better (or comparable) approximations than $\mathcal{V}_0, \mathcal{V}_1 \ldots, \mathcal{V}_r$. This is not a major problem, since one can easily combine these approximations to decrease the error term. As long as the dimension of the trails in $\mathcal{T}$ does not need to be increased too much, no practical problems occur.

An alternative possibility is that $\|E\|_2$ is large (in a relative sense) due to the lack of any good trails in $\mathcal{T}^{r+1}$. There are then two possibilities:

- The vector space approximation $\mathcal{V}_0, \mathcal{V}_r$ is bad, *i.e.* has small principal correlations. In this case, $\|E\|_2$ is still small in absolute terms. This is the goal of the designer.

- The approximation $\mathcal{V}_0, \mathcal{V}_r$ is good, but all trails in $\mathcal{T}^{r+1}$ have a large error term E. This phenomenon can be called "clustering with respect to $\mathcal{T}$".

Note that clustering is remarkable precisely because $\mathcal{V}_0, \mathcal{V}_r \in \mathcal{T}$: a "simple" (in the sense of $\mathcal{T}$) approximation of F can not be understood in terms of simple approximations of the intermediate steps F_i. Clustering can be a major problem, because it prevents obtaining an accurate estimate of $\mathscr{T}^F_{\mathcal{V}_r,\mathcal{V}_0}$ using trails in $\mathcal{T}^{r+1}$. In practice, this means that the block cipher may be insecure but this can only be detected by using a different set $\mathcal{T}'$ of acceptable vector spaces.

In general, estimating the error E (or even $\|E\|_2$) for a given trail appears to be difficult for any real-world block cipher. This is so by definition: the error E is what remains after constructing the best possible approximation that is within one's reach. If we decide to use the piling-up approach at all, showing that E is small amounts to showing that there is no trail clustering. If one can show this, then by additionally demonstrating that there are no good trails[3], the designer is able to claim (without further assumptions) that there are no good approximations with $\mathcal{V}_0, \mathcal{V}_r \in \mathcal{T}$. This problem seems to be completely open.

However, for constructions that involve a parameter such as a key, bounding E may not be entirely out of reach. One can then attempt to obtain results that are true *for most keys*. This also applies to permutations that involve constants which are chosen at random during the design phase – but this requires a leap of fate (in

[3]This should not be too hard, since by definition $\mathcal{T}$ consists of approximations that can be handled.

the designer). A first step in this direction is to show that E (or, similarly, $\mathscr{T}^F_{\mathcal{V}_r,\mathcal{V}_0}$) has small variance. As discussed in Section 2.3.2, this has been done (for the AES) in the context of linear cryptanalysis. However, as illustrated by Midori-64 [14], bounds on the variance of a distribution are not sufficient to prove security against linear cryptanalysis. The natural next step is then to obtain stronger tail bounds on E. This is the subject of Chapter 4, which investigates upper bounds on the tail probability of linear approximations.

3.3.5 Zero-Correlation Cryptanalysis and Invariants

Not all of the types of approximations covered by Definition 3.5 can be constructed using the piling-up approach. As discussed in Section 2.4.3, zero-correlation linear cryptanalysis [29] exploits linear approximations with correlation zero. This can be generalized to arbitrary approximations in the sense of Definition 3.5.

Definition 3.7 (Zero-correlation approximation)**.** *An approximation $(\mathcal{U},\mathcal{V})$ of a function F is called a zero-correlation approximation if all its principal correlations are zero. Equivalently, $\mathcal{U} \perp T^F\mathcal{V}$.*

For standard zero-correlation approximations, $\mathcal{U}$ and $\mathcal{V}$ correspond to linear approximations and such properties are usually found using a miss-in-the-middle approach based on the activity pattern of the approximation masks. The miss-in-the-middle method can be generalized as follows. Consider $(\mathcal{V}_1,\mathcal{U}_1)$ with $\mathcal{U}_1 \supseteq T^{F_1}\mathcal{V}_1$ and $(\mathcal{V}_2,\mathcal{U}_2)$ with $\mathcal{U}_2 \supseteq T^{F_2^{-1}}\mathcal{V}_2$ for invertible F_2. If $\mathcal{U}_1 \perp \mathcal{U}_2$, then $\mathcal{V}_2 \perp T^{F_2\circ F_1}\mathcal{V}_1$.

Exploring various types of generalized (in the sense of Definition 3.7) zero-correlation approximations is left as future work. One example of such a property will be encountered in Section 3.4.

Invariants are another special type of approximations. Such properties were already discussed in Section 2.4.2, albeit in a less general context. The following definition provides a further extension to Definition 2.4.

Definition 3.8 (Invariant)**.** *Let $F : X \to X$ be a function. An approximation $(\mathcal{V},\mathcal{V})$ such that $T^F\mathcal{V} \subseteq \mathcal{V}$ will be called an invariant for F.*

Definition 3.8 essentially states that the invariants of F correspond precisely to the invariant subspaces of T^F. It should be noted that, since T^F is diagonalizable, such a $\mathcal{V}$ is necessarily spanned by one or more eigenvectors of T^F. This leads to a formulation which is only slightly more general than Definition 2.4.

In all existing examples of nontrivial invariants, such as those discussed in Section 2.4.2, one has $\dim\mathcal{V} = 1$. As in the following example, those can be combined to form invariant subspaces of a larger dimension. In fact, as shown below, this observation can sometimes reveal additional weak keys.

Example. Consider the round transformations of Midori-64 (see Section 1.1.1): denote the "MixColumn" operation by M and the S-box by S. In [14], it was shown that $v^{\otimes 4}$ with

$$v = 1/2 \cdot (0,0,0,0,0,0,0,0,0,0,-1,-1,0,0,1,-1)^\top$$

is an eigenvector of $[C^S]^{\otimes 4} C^M C^K [C^S]^{\otimes 4} C^M$. The coordinates of the above vector are given in the lexicographically ordered character basis. One can show that the same property also holds for $w^{\otimes 4}$ with

$$w = 1/2 \cdot (0,0,0,0,0,0,0,0,0,0,-1,1,0,0,-1,-1)^\top$$

provided that $K_{4i-2} = K_{4i} = 0$ for $i = 1, \dots 4$. Hence, one has the invariant $\mathcal{V} = \mathrm{Span}\{v^{\otimes 4}, w^{\otimes 4}\}$ for 2^8 different keys K. Furthermore, for any $K' \in \mathbb{F}_2^{16}$ with $K'_{4i-2} = K'_{4i} = 1$, we have $C^{K'} v^{\otimes 4} = w^{\otimes 4}$. Hence, $C^{K'}\mathcal{V} = \mathcal{V}$ for any K' satisfying $K'_{4i-2} = K'_{4i}$.

The same analysis leads to new weak keys for full-state Midori-64 under the same changes to the round constants as in [14]. Let R denote the round function of Midori-64:

$$C^R = [C^M]^{\otimes 4} C^P [C^S]^{\otimes 16}.$$

For each column of the state, one can rely on either of the invariants $v^{\otimes 4}$ or $w^{\otimes 4}$. This leads to a strictly larger class of weak keys than reported in [14], for which

$$C^{K'} C^R C^K C^R C^{K'} v^{\otimes 16} = \pm v^{\otimes 16}.$$

The number of weak keys can be counted as

$$\sum_{k=0}^{4} \binom{4}{k} 2^{16k+8(4-k)+32} = 2^{64} \sum_{k=0}^{4} \binom{4}{k} 2^{8k} \approx 1.016 \cdot 2^{96}.$$

▷

3.3.6 Sampling Effect

The preceding sections were concerned with the construction of a model for F. This is entirely independent from the evaluation of this model, which is typically done using a relatively small sample of chosen or known plaintexts/ciphertexts. This section briefly discusses the case of approximations defined by functions $X \to Y$. Note that the choice of the actual distinguisher (*e.g.* ranking, hypothesis testing…), is not within the intended scope of this thesis.

Consider an approximation defined by two functions $f, g : X \to Y$. Let $N = T^g T^F (T^f)^\top$, *i.e.* $N_{y,x} = |\{z \in X \mid f(z) = x \wedge g(F(z)) = y\}|$. In all attacks, the distinguishing quantity (*i.e.* a test statistic) is derived from N. Let $\widehat{\boldsymbol{N}}$ denote the estimate of N obtained by evaluating F on $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M$, which are sampled with replacement from X:

$$\widehat{\boldsymbol{N}}_{y,x} = \sum_{i=1}^{M} \delta_{x,f(\boldsymbol{x}_i)} \delta_{y,g(F(\boldsymbol{x}_i))}. \tag{3.3}$$

By the multivariate central limit theorem [48], $\widehat{\boldsymbol{N}}$ is asymptotically (for large M) normally distributed. In fact, one expects a normal approximation to be quite good when M is large and $|Y|$ is sufficiently small. The convergence rate will not be

discussed here. Should it be an issue, one can always use the fact that the exact distribution of $\widehat{\boldsymbol{N}}$ is multinomial.

The average value of $\widehat{\boldsymbol{N}}$ is clearly equal to $(M/|X|)\,N$. Let Σ be the covariance matrix of $\mathrm{vec}(\widehat{\boldsymbol{N}})$. That is,

$$\Sigma_{(y,x),(y',x')} = \mathbb{E}\widehat{\boldsymbol{N}}_{y,x}\widehat{\boldsymbol{N}}_{y',x'} - \left[\frac{M}{|X|}\right]^2 N_{y,x}N_{y',x'}.$$

From (3.3), we get $\mathbb{E}\widehat{\boldsymbol{N}}_{y,x}\widehat{\boldsymbol{N}}_{y',x'} = (M/|X|)N_{y,x}\,\delta_{y,y'}\delta_{x,x'} + (M(M-1)/|X|^2)N_{y,x}N_{y',x'}$. Hence,

$$\Sigma = \frac{M}{|X|}\left[\mathrm{diag}\,(\mathrm{vec}(N)) - \frac{1}{|X|}\,N \otimes N\right].$$

In practice, N is not known exactly. Instead, one estimates N (or its Fourier transformation) using, for example, the piling-up approach. Obtaining a better estimate for Σ, or more generally for the distribution of $\widehat{\boldsymbol{N}}$, requires an improved model of the cipher – for example by increasing the dimension of the trail. One should take into account that the model is imperfect when deciding on a statistical distinguisher (but again, this will not be discussed).

Remark. In multiple linear cryptanalysis, one estimates (part of) the Fourier transformation of N. There are many papers that discuss the distribution of the resulting estimates. The discussion above should cover all cases. In fact, the essence of the results above (for linear approximations) can also be found in the work of Murphy [81].

Of course, as discussed in Section 3.3.3, a multiple linear approximation with masks $u_1, \ldots, u_l$ can only be completely described when *all* linear approximations with masks in $\mathrm{Span}\{u_1, \ldots, u_l\}$ are considered. In many cases, many of these linear approximations can be assumed to have a negligible contribution. ▷

Remark. The discussion above is *completely unrelated* to the distribution of $\widehat{\boldsymbol{N}}$ when F involves a key which is taken to be random. In the single-key setting, knowing the "right-key distribution" is mostly useful to obtain better estimates of the cost of an attack. Of course, if the key-dependence is strong then it makes sense to take this into account in the distinguisher. One could imagine doing the same thing in the case of weak key-dependence – but most attacks just use the *average* of the sum of squared correlations (when hypothesis testing is used) or are nonparametric (ranking). ▷

3.3.7 Relation to Other Techniques

As mentioned in Section 3.3.2, many variants of linear cryptanalysis are based on approximations of the type considered in Definition 3.5. Table 3.1 provides a summary, but note that it is not exhaustive. The table shows that the theory presented in this chapter provides a uniform description of the variants of linear cryptanalysis that that were mentioned in Section 2.4. Note that Table 3.1 does not list any techniques based on approximations with X or Y not vector spaces over $\mathbb{F}_2$. For example, Baigères *et al.* [5] discuss linear cryptanalysis in the setting with

$X = \mathbb{F}_p^n$ and $Y = \mathbb{F}_p$. One could also consider the case where $X = \mathbb{F}_p^n$ but Y is not a vector space over $\mathbb{F}_p$ – I am not aware of any attempts in this direction.

Although the theory developed this chapter provides generality, this is not its sole purpose. Another important goal is to enable the use of many unexplored types of approximations. Two examples, sparse and low-rank approximations, are listed in Table 3.1 and will be discussed in Section 3.5.

Table 3.1: Classification of several techniques. Multiple and multidimensional cryptanalysis are essentially equivalent as far as the modelling of the cipher is concerned (at least when one starts from the approach outlined in Section 3.3.3).

		Technique	Refs.	Exact	$\dim \mathcal{V}$
Probability		Invariant subspaces/sets	[71]	✔	1
		Integral cryptanalysis*	[41, 70]	✔	1
		Statistical saturation†	[36, 37]	✖	≥ 1
$f : X \to Y$	Linear	Linear cryptanalysis	[77]	✖	1
		Zero-correlation ——	[29]	✔	≥ 1
		Multiple ————	[23, 61]	✖	> 1
		Multidimensional —	[55]	✖	> 1
		Conditional ———	[20]	✖	> 1
	Nonlinear	Nonlinear invariants	[94]	✔	1
		I/O sums	[52]	✖	1
		Partitioning attacks	[53]	✖	≥ 1
		Sparse approximation	§ 3.5.1	✖	≥ 1
		Low-rank ————	§ 3.5.2	✖	≥ 1

* To some extent; if one does not consider zero-sum properties.

† They use the transition matrix approach of Vaudenay, which is related to the piling-up approach of Section 3.3.4 but without several essential aspects such as the construction described in Section 3.3.3.

3.4 Application to an Open Problem of Beierle *et al.*

This section applies the theory developed in Section 3.3 to a problem posed by Beierle, Canteaut and Leander [9, Section 4.4]. They consider a nonlinear approximation over two rounds of Midori-64, restricted to a single column of the state. The correlation matrix corresponding to this map is given by

$$C^{E_{K_1,K_2}} = C^M [C^S]^{\otimes 4} C^{K_2} C^M [C^S]^{\otimes 4} C^{K_1}.$$

Recall from Section 2.4.4 that Beierle *et al.* [9] describe nonlinear approximations using linear properties of a transformed representation of the cipher. The details of their approach will not be discussed here; the framework developed in this chapter will be used instead. The nonlinear property proposed by Beierle *et al.* amounts to a one-dimensional approximation $(u \otimes v^{\otimes 3}, u \otimes v^{\otimes 3})$, with

$$\begin{aligned} u &= 1/4 \cdot (0,1,0,-1,0,1,0,-1,0,-1,0,1,0,-1,0,-3)^\top \\ v &= 1/2 \cdot (0,0,0,-1,0,0,0,-1,0, \quad 0,0,1,0, \quad 0,0,-1)^\top. \end{aligned}$$

Note that the coordinates of the above vectors are given in the character basis with lexicographic ordering.

3.4.1 Description of the Problem

Beierle *et al.* estimate the correlation of the two-round approximation described above by iterating the following one-round trail, which has correlation at least $\pm 9/32$:

$$u \otimes v \otimes v \otimes v \xrightarrow[\pm 1 \text{ or } \pm 1/2]{[C^S]^{\otimes 4} C^{K_i}} u \otimes v \otimes v \otimes v \xrightarrow[9/16]{C^M} u \otimes v \otimes v \otimes v. \tag{3.4}$$

This was expected to hold whenever $K_i \in \mathbb{F}_2^4 \times \mathcal{K}^3$ for $i = 1, 2$ and with $\mathcal{K} = \{(0,0,x,y) \mid x, y \in \mathbb{F}_2\}$. Remark that, in general, computing the correlation over C^M is nontrivial. Beierle *et al.* essentially compute the inner product directly, which is feasible in this case since the ambient space is only of dimension 2^{16}.

Based on this trail, one expects an absolute correlation of at least $(9/32)^2 \approx 2^{-3.66}$ over E_{K_1,K_2}. However, Beierle *et al.* observe that this approximation is not accurate. Specifically:

- When $K_2 \in (\mathbb{F}_2^4 \setminus \mathcal{K}) \times \mathcal{K}^3$, the correlation is found to equal zero.
- For other keys, the correlation takes on various values, but is always significantly larger than the estimated minimum of $2^{-3.66}$. Specifically, for $K_1, K_2 \in \mathcal{K}^4$, the correlation ranges from $35/64$ to $40/64 = 5/8$. For other keys, it lies between $39/256$ and $65/256$.

In their conclusion, the authors remark that understanding this phenomenon is "a major open problem". Section 3.4.2 aims to clarify these observations.

3.4.2 Solution to the Problem

In light of the results in Section 3.3.4, the observant reader will object to the use of the trail (3.4). Indeed, the piling-up approximation presupposes a dominant trail, but in this example the choice of the trail is essentially arbitrary. In such cases, one must tread carefully.

The first step of the trail, *i.e.* the approximation of $[C^S]^{\otimes 4} C^{K_i}(u \otimes v \otimes v \otimes v)$ by $u \otimes v \otimes v \otimes v$, is the most problematic. If the absolute correlation of this approximation is $1/2$ – which is the case for most keys –, then there is another approximation which also achieves correlation $\pm 1/2$. Furthermore, there is no particular reason to assume that this second approximation leads to a worse approximation over the linear layer.

Since multiplication with $[C^S]^{\otimes 4} C^{K_i}$ does not result in a substantially more complicated state – specifically, it preserves the rank – it seems more reasonable to approximate only *after* the application of the linear layer. In general, this is a nontrivial problem since $C^M(u \otimes v \otimes v \otimes v)$ need not be a low rank tensor. However, for the linear layer of Midori-64, the following result can be applied.

Theorem 3.7. *For any integer $n \geq 1$, let $M_n \in \mathbb{F}_2^{4n \times 4n}$ be the matrix*

$$M_n = \begin{pmatrix} 0 & I & I & I \\ I & 0 & I & I \\ I & I & 0 & I \\ I & I & I & 0 \end{pmatrix},$$

with $I \in \mathbb{F}_2^{n \times n}$. Then C^{M_n} has a symmetric polyadic decomposition

$$C^{M_n} = 2^{-n} \sum_{(i_1, i_2, \ldots, i_n) \in [4]^n} \left(\prod_{j=1}^n \lambda_{i_j}\right) \left[\, \otimes_{j=1}^n A_{i_j} \right]^{\otimes 4},$$

with $\lambda_1 = \lambda_2 = \lambda_3 = 1$, $\lambda_4 = -1$ and

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad A_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$
$$A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad A_4 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

That is, C^{M_n} has tensor rank at most 4^n.

Proof. In the following proof, the entries of C^{M_n} are indexed by elements of $\mathbb{F}_2^n$ (as opposed to $\widehat{\mathbb{F}_2^n}$). The decomposition can be derived by induction on n. For $n = 1$, one can check that

$$C^{M_1} = \frac{1}{2}\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes 4} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^{\otimes 4} + \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^{\otimes 4} - \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\otimes 4} \right].$$

For $n > 1$, Theorem 2.1, (4) implies that

$$C^{M_n}_{u_1,u_2,u_3,u_4,v_1,v_2,v_3,v_4} = \delta\left(\sum_{j=1}^4 \begin{pmatrix} u_j \\ u_j \\ u_j \\ u_j \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} \right) = \prod_{i=1}^4 \delta\big(u_i + v_i + \textstyle\sum_{j=1}^4 u_j\big).$$

Let $u_i^{(1)}$ denote the leftmost bit of u_i and let $u_i^{(2,\ldots,n)}$ denote the other $n-1$ bits. Then

$$\delta\big(u_i+v_i+\textstyle\sum_{j=1}^{4} u_j\big) = \delta\big(u_i^{(1)}+v_i^{(1)}+\sum_{j=1}^{4} u_j^{(1)}\big)\,\delta\big(u_i^{(2,\ldots,n)}+v_i^{(2,\ldots,n)}+\sum_{j=1}^{4} u_j^{(2,\ldots,n)}\big).$$

It suffices to find the decomposition of the first term. Indeed, if

$$C^{M_{n-1}} = 2^{-n+1} \sum_{(i_1,i_2,\ldots,i_{n-1})\in[4]^{n-1}} \big(\textstyle\prod_{j=1}^{n-1}\lambda_{i_j}\big)\,\big[\,\otimes_{j=1}^{n-1} A_{i_j}\big]^{\otimes 4},$$

then we have

$$\begin{aligned}
&C^{M_n}_{u_1,u_2,u_3,u_4,v_1,v_2,v_3,v_4}\\
&= 2^{-n+1}\prod_{i=1}^{4}\delta\big(u_i^{(1)}+v_i^{(1)}+\textstyle\sum_{j=1}^{4}u_j^{(1)}\big)\\
&\qquad\sum_{(i_1,i_2,\ldots,i_{n-1})\in[4]^{n-1}} \textstyle\prod_{j=1}^{n-1}\lambda_{i_j}(A_{i_j}^{\otimes 4})_{u_1^{(2,\ldots,n)},\ldots,u_4^{(2,\ldots,n)},v_1^{(2,\ldots,n)},\ldots,v_4^{(2,\ldots,n)}}.
\end{aligned}$$

The decomposition of the first factor above is equivalent to the base case, *i.e.* $n=1$. Finally, the result follows by the definition of the Kronecker product. □

Let $K_1 = k_1\|k_2\|\cdots\|k_{16}$ and $K_2 = k_1'\|k_2'\|\cdots\|k_{16}'$. Remark that v is invariant under the key-addition operation for keys in $\mathcal{K}$. Hence the key bits $k_5,\ldots,k_{16}$ only influence the sign of the correlation. A convenient strategy to compute the correlation or to identify the best trails is to propagate $u\otimes v^{\otimes 3}$ in both the forward direction (under the map $C^M[C^S]^{\otimes 4}C^{K_1}$) and backward direction (under the map $C^{K_2}[C^S]^{\otimes 4}C^M$). In the forward direction, we have

$$\begin{aligned}
&C^M[C^S]^{\otimes 4}C^{K_1}(u\otimes v\otimes v\otimes v)\\
&= \nu\, C^M[(C^S C^{k_1\|\cdots\|k_4}u)\otimes v\otimes v\otimes v]\\
&= \nu/2\,(0,0,0,1,0,0,0,1,0,0,0,-1,0,0,0,1)^\top\otimes\Big(\sum_{i=1}^{16} c_i\, v_i^{\otimes 3}\Big),
\end{aligned}$$

where the last equality follows from Theorem 3.7. The vectors v_i and corresponding coefficients c_i are listed in Table 3.2. The sign ν is given by $\nu = \prod_{i=2}^{4}(-1)^{k_{4i-1}+k_{4i}}$. Observe that the first factor (up to a factor $-1/2$) is equal to v.

For the backward direction, it holds that

$$\begin{aligned}
&C^{K_2}[C^S]^{\otimes 4}C^M(u\otimes v\otimes v\otimes v)\\
&= \nu'/2\,(0,0,0,1,0,0,0,(-1)^{k_2'},0,0,0,(-1)^{1+k_1'},0,0,0,(-1)^{k_1'+k_2'})^\top\\
&\otimes\Big(\sum_{i=1}^{8} c_i'\bigotimes_{j=1}^{3}(C^{k_{4j}'\|\cdots\|k_{4j+4}'}v_i')\Big).
\end{aligned}$$

In the above, $\nu' = (-1)^{k_3'+k_4'}$. Table 3.3 lists the coefficients c_i' and vectors v_i'.

Table 3.2: Vectors v_i and corresponding coefficients in the forward decomposition. The notation $\kappa_i = (-1)^{k_i}$ is used.

i	$2\,v_i^\top$	$\kappa_4\,c_i$
1	$(0,0,0,\ 1,0,0,0,-1,0,0,0,\ 1,0,0,0,\ 1)$	$1/32\,(3\kappa_1\kappa_2\kappa_3+\kappa_1\kappa_2-\kappa_1\kappa_3-\kappa_1+2\kappa_2)$
2	$(0,0,0,\ 1,0,0,0,-1,0,0,0,-1,0,0,0,-1)$	$-1/32\,(\kappa_1\kappa_3-2\kappa_2\kappa_3+\kappa_1+2\kappa_2+\kappa_3+1)$
3	$(0,0,0,-1,0,0,0,-1,0,0,0,\ 1,0,0,0,-1)$	$-1/16\,\kappa_3\,(3\kappa_1\kappa_2+\kappa_1+\kappa_2+1)$
4	$(0,0,0,-1,0,0,0,-1,0,0,0,-1,0,0,0,\ 1)$	$1/32\,(3\kappa_1\kappa_2\kappa_3-\kappa_1\kappa_2+2\kappa_1-\kappa_3+1)$
5	$(0,0,\ 1,0,0,0,-1,0,0,0,\ 1,0,0,0,\ 1,0)$	$1/32\,(2\kappa_1\kappa_2+\kappa_1\kappa_3-\kappa_1-\kappa_3-1)$
6	$(0,0,\ 1,0,0,0,-1,0,0,0,-1,0,0,0,-1,0)$	$-1/32\,(3\kappa_1\kappa_2\kappa_3-\kappa_1\kappa_2+\kappa_1\kappa_3-2\kappa_2\kappa_3-\kappa_1)$
7	$(0,0,-1,0,0,0,-1,0,0,0,\ 1,0,0,0,-1,0)$	$-1/32\,(3\kappa_1\kappa_2\kappa_3+\kappa_1\kappa_2-2\kappa_2\kappa_3-2\kappa_2+\kappa_3-1)$
8	$(0,0,-1,0,0,0,-1,0,0,0,-1,0,0,0,\ 1,0)$	$-1/16\,(\kappa_2-1)$
9	$(0,\ 1,0,0,0,-1,0,0,0,\ 1,0,0,0,\ 1,0,0)$	$1/32\,\kappa_1\,(3\kappa_2\kappa_3+\kappa_2-\kappa_3+1)$
10	$(0,\ 1,0,0,0,-1,0,0,0,-1,0,0,0,-1,0,0)$	$-1/32\,(2\kappa_1\kappa_2+\kappa_1\kappa_3+\kappa_1-\kappa_3+1)$
11	$(0,-1,0,0,0,-1,0,0,0,\ 1,0,0,0,-1,0,0)$	$1/16\,\kappa_3\,(3\kappa_1\kappa_2-1)$
12	$(0,-1,0,0,0,-1,0,0,0,-1,0,0,0,\ 1,0,0)$	$1/32\,(3\kappa_1\kappa_2\kappa_3+\kappa_1\kappa_2-2\kappa_1\kappa_3+\kappa_3+1)$
13	$(\ 1,0,0,0,-1,0,0,0,\ 1,0,0,0,\ 1,0,0,0)$	$-1/32\,(\kappa_1\kappa_3-\kappa_1-\kappa_3+1)$
14	$(\ 1,0,0,0,-1,0,0,0,-1,0,0,0,-1,0,0,0)$	$-1/32\,\kappa_1\,(3\kappa_2\kappa_3-\kappa_2-\kappa_3-1)$
15	$(-1,0,0,0,-1,0,0,0,\ 1,0,0,0,-1,0,0,0)$	$-1/32\,(3\kappa_1\kappa_2\kappa_3-\kappa_1\kappa_2-\kappa_3-1)$
16	$(-1,0,0,0,-1,0,0,0,-1,0,0,0,\ 1,0,0,0)$	$1/16\,(\kappa_1-1)$

Remark that the occurrence of correlation zero when $K_2 \in (\mathbb{F}_2^4 \setminus \mathcal{K}) \times \mathcal{K}^3$ is an immediate consequence of the above. Indeed, it holds that

$$\begin{aligned}&(0,0,0,1,0,0,0,(-1)^{k_2'},0,0,0,(-1)^{1+k_1'},0,0,0,(-1)^{k_1'+k_2'})\\&\cdot(0,0,0,1,0,0,0,1,0,0,0,-1,0,0,0,1)^\top\\&=1+(-1)^{k_1'}+(-1)^{k_2'}+(-1)^{k_1'+k_2'},\end{aligned}$$

which equals zero unless $k_1' = k_2' = 0$.

For $K_2 \in \mathcal{K}^4$, the total correlation is given by

$$\nu\nu'\left\langle \sum_{i=1}^{16} c_i\, v_i^{\otimes 3},\ \sum_{j=1}^{8} c_j'\ \bigotimes_{l=1}^{3}(C^{k_{4l}'\|\cdots\|k_{4l+4}'}v_j')\right\rangle = \nu\nu' \sum_{i=1}^{16}\sum_{j=1}^{8} c_i c_j' \prod_{l=1}^{3}\langle v_i, C^{k_{4l}'\|\cdots\|k_{4l+4}'}v_j'\rangle.$$

In principle, the above expression can be used to compute the correlation for arbitrary keys. This is feasible, but offers little insight. Instead, it is preferable to estimate the correlation based on a trail of low-rank approximations.

It turns out that a single rank-one trail suffices to obtain a reasonably accurate estimate when $K_2 \in \mathcal{K}^4$. Remark that $|\langle v_i, C^{k_{4l}'\|\cdots\|k_{4l+4}'}v_j'\rangle| \le 1/2$ unless $i = 3$ and $j = 1$, in which case it equals one. That is, the term $c_3\,c_1'$ has weight one whereas the other terms are multiplied by a factor of at most 2^{-3}.

Table 3.3: Vectors v'_i and corresponding coefficients in the backward decomposition. The notation $\kappa_i = (-1)^{k'_{4j+i}}$ is used.

i	$4\,(C^{k'_{4j}\|\cdots\|k'_{4j+4}} v'_i)^\top$	c'_i
1	$2(0,\ 0,\ 0,\ -1,\ 0,\ 0,\ 0,\ -1,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ -1)$	$-3/8\kappa_3\kappa_4$
2	$2(0,\ 0,-\kappa_3,-\kappa_2\kappa_3,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,-\kappa_3,\kappa_3\kappa_4)$	$1/8$
3	$(0,\ 0,\ 2\kappa_3,\ 0,-2,\ 0,\ 0,\ 0,-1,\ \kappa_4,\ \kappa_3,\ \kappa_3\kappa_4,\ 1,\ \kappa_4,-\kappa_3,\kappa_3\kappa_4)$	$1/8$
4	$(0,\ 0,\ 0,\ 0,\ 0,-2\kappa_4,\ 2\kappa_3,\ 0,\ 1,-\kappa_4,-\kappa_3,\ \kappa_3\kappa_4,\ 1,\ \kappa_4,\ \kappa_3,\kappa_3\kappa_4)$	$1/8$
5	$(0,\ \kappa_4,\ 0,-\kappa_3\kappa_4,-1,\ 2\kappa_4,\ \kappa_3,2\kappa_3\kappa_4,\ 0,-\kappa_4,\ 0,\ \kappa_3\kappa_4,-1,\ 0,\ \kappa_4,\ 0)$	$-1/8$
6	$(0,-\kappa_4,\ 0,-\kappa_3\kappa_4,\ 1,\ 0,\ \kappa_3,\ 0,-1,\ 2\kappa_4,\ \kappa_3,\ 0,\ 0,\ -\kappa_4,\ 2\kappa_3,\kappa_3\kappa_4)$	$1/8$
7	$(0,-\kappa_4,\ 0,\ \kappa_3\kappa_4,\ 1,\ 0,-\kappa_3,\ 0,\ 1,\ 0,\ \kappa_3,2\kappa_3\kappa_4,-2,\ \kappa_4,\ 0,\kappa_3\kappa_4)$	$1/8$
8	$(0,\ \kappa_4,\ 0,\ \kappa_3\kappa_4,\ 1,\ 0,\ \kappa_3,\ 0,\ 0,-\kappa_4,\ 2\kappa_3,\ \kappa_3\kappa_4,\ 1,-2\kappa_4,-\kappa_3,\ 0)$	$-1/8$

Remark that $v_3 = v'_1 = v$, so the paragraph above suggests the use of the intermediate approximation $v^{\otimes 4}$. It is not a coincidence that $v^{\otimes 4}$ itself is an invariant for the weak-key class $\mathcal{K}^4$. The absolute value of the corresponding correlation is equal to

$$\begin{aligned}|c_3\, c'_1| &= 3/32\,|3\,(-1)^{k_1+k_2} + (-1)^{k_1} + (-1)^{k_2} + 1| \\ &= 3/8\,|\delta_{k_1}\delta_{k_2} + 1/2\,(-1)^{k_1+k_2}| \\ &= \begin{cases} 9/16 \text{ if } k_1 = k_2 = 0 \\ 3/16 \text{ otherwise.} \end{cases}\end{aligned}$$

These values are close to the experimental values reported by Beierle *et al.* [9]. Specifically, the relative error is less than 10% in the first case and less than 30% in the second case; in Section 3.4.3, this will be improved by taking into account additional trails. In addition, one can show that the sign of the correlation is given by

$$\prod_{i=1}^{4} (-1)^{k'_{4i-1}+k'_{4i}+k_{4i-1}+k_{4i}}.$$

In conclusion, for $K_2 \in \mathcal{K}^4$ (that is, when the correlation is not zero), the nonlinear approximation considered by Beierle *et al.* is dominated by the following trail:

$$u \otimes v^{\otimes 3} \xrightarrow[\pm 3/4 \text{ or } \pm 1/4]{[C^S]^{\otimes 4} C^{K_1}} v^{\otimes 4} \xrightarrow[1]{C^M} v^{\otimes 4} \xrightarrow[\pm 1]{[C^S]^{\otimes 4} C^{K_2}} v^{\otimes 4} \xrightarrow[3/4]{C^M} u \otimes v^{\otimes 3}.$$

One might have anticipated the above trail without relying on Theorem 3.7: the choice of $v^{\otimes 4}$ as an intermediate state is a natural choice, since it is invariant under the round function and because $\langle u, v\rangle = 3/4$ and $\langle v, C^{0\|0\|k_3\|k_4} u\rangle = \pm 1/4$. Nevertheless, an intuitive selection of trails is error prone. Theorem 3.7 essentially provides an automatic and general way to analyze rank one approximations over two rounds of

Midori-64. Since the tensor rank of C^M is at most 256, one can always understand such approximations using a relatively low-dimensional vector space trail spanned by rank one states. This is quite a remarkable result; *e.g.* it includes all two-round linear approximations.

3.4.3 Further Analysis

The analysis in Section 3.4.2 can be refined by taking into account additional trails. In fact, when considering all trails, one obtains the exact value of the correlation. It follows from the discussion in the previous section that (for $K_2 \in \mathcal{K}^4$) the absolute correlation is given by

$$\Big| \sum_{i=1}^{16} c_i \sum_{j=1}^{8} c'_j \prod_{l=1}^{3} \langle v_i, C^{k'_{4l} \| \cdots \| k'_{4l+4}} v'_j \rangle \Big|,$$

with c_i, v_i listed in Table 3.2 and c'_j, v'_j in Table 3.3. It turns out that ten terms in the outer sum are nonzero, namely those with $i \in \{1, 2, 3, 4, 5, 6, 9, 10, 11, 12\}$. In addition, for each i, typically only a few terms of the inner sum can be nonzero for some key. A symbolic computation using SAGE (source code listed online[4]) then yields the following expression for the absolute correlation

$$\begin{aligned}
&\big| 303/1024\,(-1)^{k_1+k_2+k_3+s_3+s_4} + 189/2048\,[1+(-1)^{k_1}](-1)^{k_3+s_3+s_4} \\
&+ 47/512\,(-1)^{k_2+k_3+s_3+s_4} + 69/2048\,(-1)^{k_1+k_2+k_3+s_4} + 17/2048\,(-1)^{k_1+k_2+s_4} \\
&- 1/128\,(-1)^{k_3+s_4} + 7/1024\,(-1)^{k_2+s_3+s_4} + 5/1024\,(-1)^{s_4} \\
&+ 9/2048\,[1-(-1)^{k_3}]\,(-1)^{k_1+k_2+k_3+s_3} \\
&+ 7/2048\,[1+(-1)^{s_3}+(-1)^{k_1}-(-1)^{s_3+s_4}]\,(-1)^{k_1+s_3+s_4} \\
&- 3/1024\,(-1)^{k_2+k_3+s_3} + 3/2048\,[1+(-1)^{k_3}]\,(-1)^{k_3+s_3} \big|,
\end{aligned}$$

with $s_3 = \sum_{i=2}^{4} k'_{4i-1}$ and $s_4 = \sum_{i=2}^{4} k'_{4i}$. One can check that the expression above indeed corresponds exactly to the values observed by Beierle *et al.*.

[4] https://homes.esat.kuleuven.be/~tbeyne/masterthesis/midori64.html

3.5 Practical Aspects

This section discusses important practical questions: which approximation types can be used in practice, and how does one find the best among these approximations? An important aspect in this context is the choice of the set of acceptable trails, *i.e.* $\mathcal{T}$ in the notation from Section 3.3.4. For simplicity of notation, it will be assumed that the vector spaces in a trail are all chosen from the same set $\mathcal{T}$. The choice of $\mathcal{T}$ is essentially restricted by the requirement that it must be feasible to approximate the solution to the optimization problem

$$\max_{\mathcal{V}_0,\ldots,\mathcal{V}_r \in \mathcal{T}^{r+1}} \mathcal{J}(\prod_{i=1}^{r} \mathcal{T}_{\mathcal{V}_{r-i+1},\mathcal{V}_{r-i}}^{F_{r-i+1}}), \tag{3.5}$$

where $\mathcal{J}$ is some cost function, *e.g.* the 2-norm or the Frobenius norm. A minimum requirement is that it must be possible to evaluate the cost function. This leads to the following constraints:

- The dimension of the vector spaces in $\mathcal{T}$ should be low, to avoid large memory usage and computational cost.
- There should a basis matrix V_i of $\mathcal{V}_i$ such that is possible to compute $V_{i+1}^* T^{F_i} V_i$. Of course, this also depends on the choice of F_i.
- If the function F involves an unknown key, it is typically desired that the solution to (3.5) is valid for a significant fraction of keys.

In permutation-based cryptography, the last restriction listed above does not apply. This paves the way for a potentially wide variety of different types of approximations. A few choices of $\mathcal{T}$ are proposed below.

3.5.1 Linear and Sparse Approximations

In linear cryptanalysis, one uses a basis consisting of character functions of $\mathbb{C}G$. Equivalently, when working in $\mathbb{C}\widehat{G}$, such a basis is a subset of $\{\delta_\chi \mid \chi \in \widehat{G}\}$. A natural variation on this is to use *sparse* basis functions. In a block cipher, it is probably more interesting to use a low-dimensional vector space spanned by characters – that is, linear cryptanalysis using several one-dimensional trails. However, for keyless permutations more accurate results can be obtained using a sparse basis, because this allows taking into account clustering over a small number of rounds.

In many constructions, it should be feasible to propagate sparse vectors exactly over a few rounds. Indeed, the correlation matrix of the nonlinear layer is typically a tensor product of several small matrices. Hence, products with sparse vectors can be computed using a divide-and-conquer approach (similar to the FFT). Applying the linear layer amounts to permuting a sparse vector, which is a cheap operation. Of course, one still needs to solve a (difficult) optimization problem of the type (3.5); a heuristic approach may however suffice. Further development of this method is left as future work.

3.5.2 Rank One Approximations

A different class of approximations consists in choosing basis vectors of tensor rank one (in some well-defined sense). Such approximations already made a brief appearance in Section 3.4 and are a particularly attractive choice when the function of interest F is cell-oriented. That is, we assume that F operates on a group $X = G = \oplus_{i=1}^{m} G_i$; however, note that the choice of the group G is not necessarily unique. As mentioned in Section 3.3.3, a function $f : \prod_{i=1}^{m} X \to Y$ such that $f(x_1, \ldots, x_m) = \sum_{i=1}^{m} f_i(x_i)$ can be associated with a vector space spanned by rank-one tensors in $\otimes_{i=1}^{m} \mathbb{C}G_i$. One can give several intuitive reasons why such approximations might be useful:

- Computing inner products between rank one tensors reduces to computing m inner products of short vectors.
- Due to the cell-oriented structure of the cipher, it is possible to compute the exact result of the S-box layer on such states – but the effect of the linear layer is nontrivial.
- For several block ciphers, there exist rank one tensors which are invariants as in [14].
- A rank one tensor represents a state with independent cells. Hence, one can think of a low rank approximation as a mixture of states with independent cells.

Consider the one-dimensional case, *i.e.* $\dim \mathcal{V}_1 = \ldots = \dim \mathcal{V}_r = 1$. In this case, (3.5) can be interpreted as an optimization problem over the product of spheres $\prod_{i=1}^{m} \mathbb{S}^{|G_i|}$, which is a manifold. It is again preferable to work in the Fourier basis, since this makes it easy to look only for balanced approximations and, simultaneously, to eliminate the trivial approximation δ_{χ_0}. One then obtains the optimization problem

$$\max_{\substack{f_i^{(j)} \in \{0\} \times \mathbb{S}^{|G_i|-1} \\ \text{for } j=1,\ldots,m \\ i=0,\ldots,r}} \sum_{i=1}^{r} \log |\langle \otimes_{j=1}^{m} f_i^{(j)},\, C^{F_i} \otimes_{j=1}^{m} f_{i-1}^{(j)} \rangle|, \tag{3.6}$$

where $f_i^{(j)}$ should be interpreted as an element of $\mathbb{C}\widehat{G}_i$, given by its coordinates in the basis $\{\delta_\chi \mid \chi \in \mathbb{C}\widehat{G}\}$ with δ_{χ_0} considered as the first basis element. Note that $f_i^{(j)} \in \{0\} \times \mathbb{S}^{|G_i|-1}$ then amounts to requiring that cell j in round i is active and balanced. One can require that the cell is inactive by setting $f_i^{(j)} = \delta_{\chi_0}$. The use of the logarithm in the cost function helps to avoid excessively small objective and gradient values.

Remark. The problem (3.6) does not impose that the vectors $f_i^{(j)}$ correspond to nonlinear Boolean functions as outlined in Section 3.3.3. In this sense, (3.6) is a relaxation of a problem which is in principle discrete. As illustrated below, this does not appear to be a problem in practice – rounding the solution is sufficient. ▷

The optimization problem (3.6) may have many local optima and one runs into trouble when m or the number of rounds r is large. Fortunately, it is not necessary to solve the full-round problem in order to obtain useful results. This will be illustrated using the block cipher Midori-64. The problem (3.6) will be solved for $r = 1$ and $C^{F_1} = [C^S]^{\otimes 4} C^M [C^S]^{\otimes 4}$ or equivalently $C^{F_1} = C^M$. As illustrated below, even this limited case has interesting applications. Remark that the key addition is omitted: this implies that the obtained property need not hold for many keys, but it also simplifies the problem significantly. Equivalently, the block cipher is treated as a permutation – an improved method that explicitly takes into account key-dependence is left as future work. The cost function can be evaluated efficiently by using the polyadic decomposition of C^M from Theorem 3.7. For other linear layers, one either has to compute such a decomposition or use a more direct approach. For Midori-64, the latter is feasible but significantly more expensive.

The following example serves as a proof-of-concept of the method sketched above. Another example with more practical relevance is provided in Section 3.5.3. The optimization problem (3.6) was solved using (Riemannian) conjugate gradient based on the PYMANOPT library [95] (full implementation available online[5]).

Example. Consider the optimization problem

$$\max_{\substack{f_0^{(j)},\, f_1^{(j)} \in \{0\} \times \mathbb{S}^{15} \\ \text{for } j=1,\ldots,4}} \log |\langle \otimes_{j=1}^4 C^S f_1^{(j)},\, C^M \otimes_{j=1}^4 C^S f_0^{(j)} \rangle|.$$

Remark that the choice of C^S does not matter. Several solutions to this problem are known from the analysis of invariants in Midori-64. These correspond to a cost of zero, *i.e.* correlation one. The conjugate gradient method converges quickly and yields other optimal approximations, depending on the initial guess. To search specifically for invariants, one can set $f_0^{(j)} = f_1^{(j)}$ for $j = 1, \ldots 4$. This indeed results in additional invariants which are, contrary to previous examples, not symmetric. One example is

$$f_0^{(1)} = f_0^{(3)} = (0, 1/4, 0, 1/4, 1/4, 0, -1/4, 1/2, 0, -1/4, 0, -1/4, 1/4, 1/2, -1/4, 0)^\top$$
$$f_0^{(2)} = f_0^{(4)} = (0, -1/4, 1/2, 1/4, 1/4, 0, 1/4, 0, 1/4, 0, 1/4, 0, -1/2, 1/4, 0, -1/4)^\top.$$

One can check that this defines an invariant for 2^8 weak keys. However, due its asymmetry, it is of limited practical interest. Symmetry can be required by setting $f_0^{(1)} = f_0^{(2)} = f_0^{(3)} = f_0^{(4)}$, but all such invariants were already classified in the author's earlier work [14]. ▷

3.5.3 Another Approximation over Midori-64

This section contains another example of a rank one approximation over Midori-64. The example is based on known invariants of the Midori-64 round function, but is not an invariant itself. In the following, $X = \mathbb{F}_2^{64}$ and we work in $\mathbb{C}\widehat{X}$.

[5] https://homes.esat.kuleuven.be/~tbeyne/masterthesis/manifold.html

From previous work, it is known that the vectors $v^{\otimes 16}$, $\delta_{\chi_5}^{\otimes 16}$ with $v = C^S \delta_{\chi_5}$ determine an invariant of Midori-64 for slightly modified round constants. In particular, we have $C^M \delta_{\chi_5}^{\otimes 4} = \delta_{\chi_5}^{\otimes 4}$ and $C^M v^{\otimes 4} = v^{\otimes 4}$. One can additionally verify that the best rank one approximation of $C^M[v^{\otimes 2} \otimes \delta_{\chi_0}^{\otimes 2}]$ is given by $v^{\otimes 2} \otimes \delta_{\chi_0}^{\otimes 2}$ and has correlation

$$\langle v^{\otimes 2} \otimes \delta_{\chi_0}^{\otimes 2}, C^M[v^{\otimes 2} \otimes \delta_{\chi_0}^{\otimes 2}] \rangle = 1/4.$$

Due to the symmetry of C^M, the same holds when the order of the factors v and δ_{χ_0} is permuted. To obtain more weak keys, at the cost of a lower correlation, one can split this nonlinear approximation into 2^8 approximations which are linear at the input side. Indeed, v is a linear combination of $\delta_{\chi_{\mathtt{A}}}$, $\delta_{\chi_{\mathtt{B}}}$, $\delta_{\chi_{\mathtt{E}}}$ and $\delta_{\chi_{\mathtt{F}}}$.

Figure 3.2 shows the resulting trails over 6.5 rounds of Midori-64. Some transitions require assumptions on the key. Hence, the trails are only valid for 2^{96} keys – but one can easily modify the first state to obtain similar trails for 2^{16} different classes of weak keys. Equivalently, without whitening keys, the trail is valid for 2^{112} weak keys.

Each of the 2^8 trails shown in Figure 3.2 has absolute correlation 2^{-24}. As discussed above, for some keys clustering results in a correlation of 2^{-16}. More generally, the degree of clustering depends on the number of solutions $i, j, k, l \in \{\mathtt{A}, \mathtt{B}, \mathtt{E}, \mathtt{F}\}$ of

$$\begin{aligned} &i^\top(K_{1,1} + K_{1,2} + C_{2,1} + C_{2,2}) + j^\top(K_{1,7} + K_{1,8} + C_{2,7} + C_{2,8}) + \\ &k^\top(K_{1,9} + K_{1,10} + C_{2,9} + C_{2,10}) + l^\top(K_{1,15} + K_{1,16} + C_{2,15} + C_{2,16}) = 0, \end{aligned} \tag{3.7}$$

where K_1 is the second round key and C_2 the round constant added in the second round. The second subscripts refer to the indices of bits within these values.

For all-zero round keys, experimentally (using a sample of 2^{44} plaintexts/ciphertexts) a correlation of $2^{-15.59}$ was observed. Note that the Midori-64 specification [6] bounds the correlation of linear six round trails by 2^{-30}; but such trails have only 30 active S-boxes, as opposed to 48 in Figure 3.2.

Since the correlation of the property in Figure 3.2 is quite large, and since the assumptions on the key do not yet conflict, it is worthwhile to attempt an extension to more rounds. Figure 3.3 shows an extension to 10.5 rounds, but this extension does require some changes[6] to the (default) round constants. Another disadvantage of the property in Figure 3.3 is that the weak key density (for correlation 2^{-24}) is reduced by a factor 2^{-12}. This is due to a second clustering requirement, but now on K_0 rather than K_1. Note that the weak-key density decreases by 2^{-12} rather than 2^{-16}, because there is some overlap between restrictions imposed by the clustering and the invariant requirements. Hence, the number of weak keys for Figure 3.3 is at least 2^{84}. A more careful analysis of the condition (3.7) could reveal additional weak keys, but this will be left open for now.

[6]The requirement on the round constants is that the round constants in round 5 and 9 are equal in the first and third columns, and that the two relevant bits for the invariant are the same for the first (top) two cells in these columns.

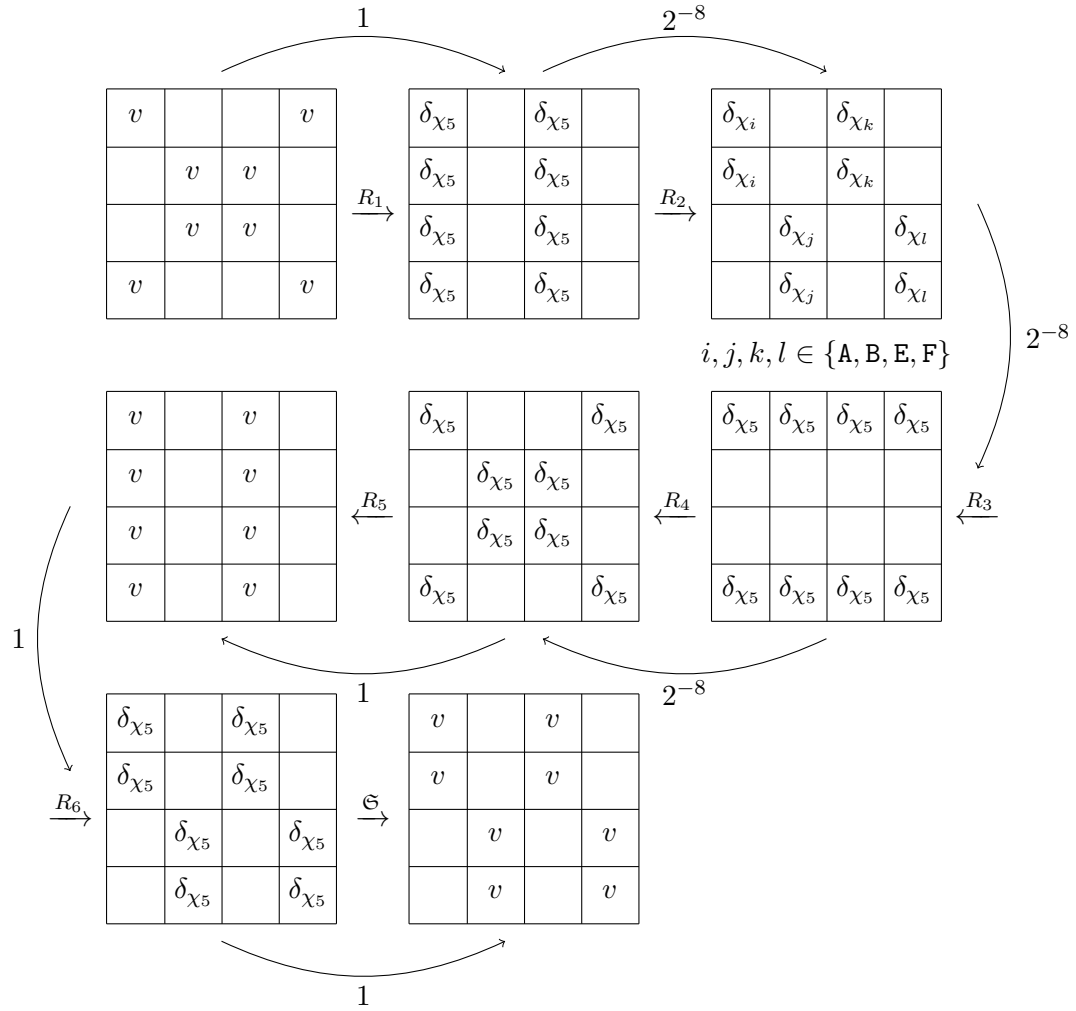


Figure 3.2: Nonlinear approximations over 6.5 rounds of Midori-64 with absolute correlation 2^{-24} for 2^{96} (extends to 2^{112}) weak keys. For some keys, these trails can be combined leading to an absolute correlation of 2^{-16}.

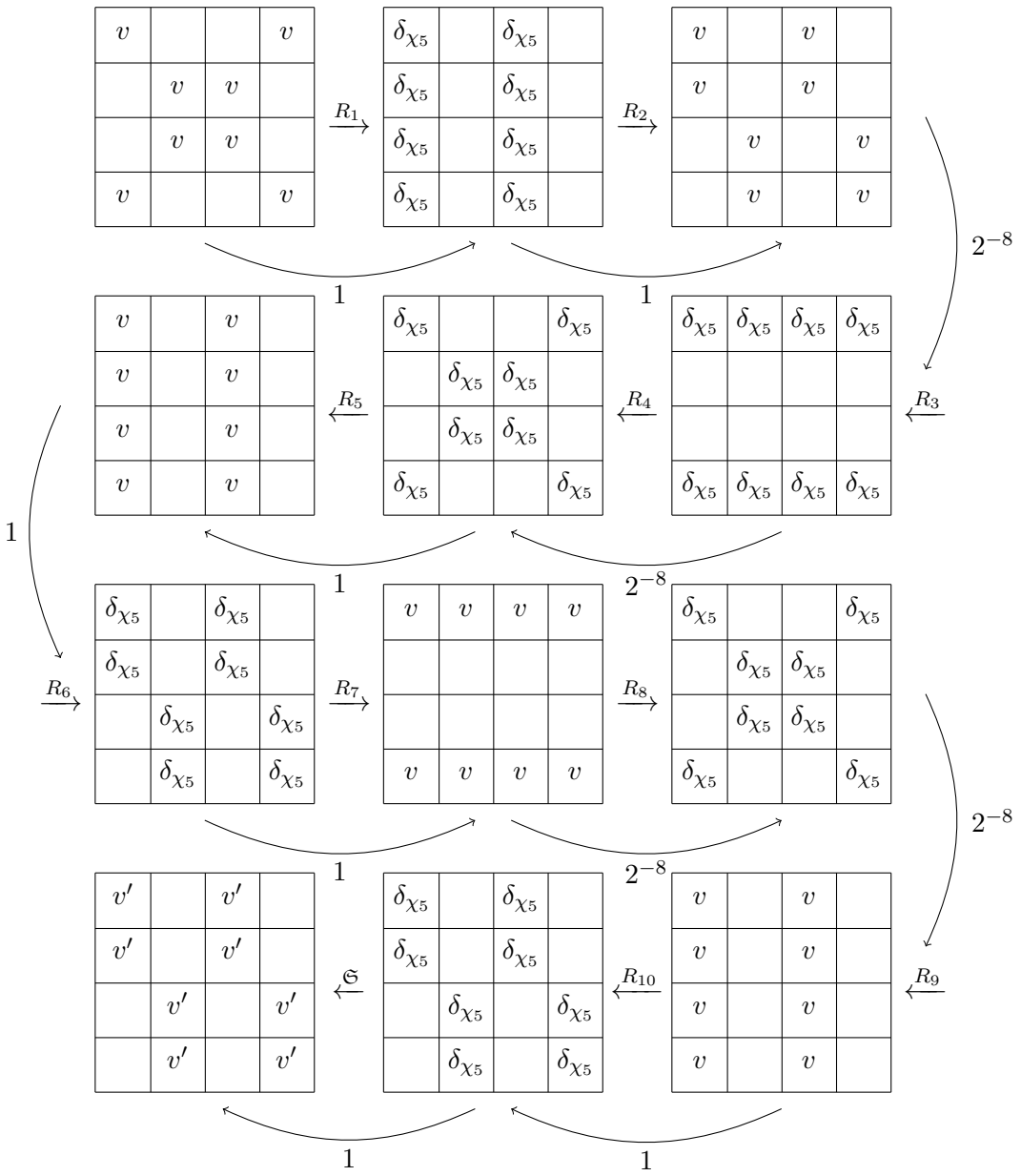


Figure 3.3: Nonlinear property over 10.5 rounds of Midori-64 (with modified constants) with correlation 2^{-24} for 2^{84} weak keys.

3.6 Conclusion

A general framework for the analysis of linear and nonlinear approximations in the fixed and weak key settings was introduced. As illustrated by Table 3.1, this "geometric approach" provides a uniform description of many variants of linear cryptanalysis.

Linear approximations were generalized by considering pairs of vector spaces in the inner product space $\mathbb{C}G$ or $\mathbb{C}\widehat{G}$. These vector spaces can be related to sets, probability distributions or nonlinear (vectorial) Boolean functions. The classical notion of correlation generalizes to the cosines of the principal angles between such vector spaces. The standard piling-up approach was extended to this general setting. It was shown how the piling-up approximation can be interpreted as a sequence of successive orthogonal projections. This gave rise to the concept of vector space trails. In addition, the possibility of more general invariants and zero-correlation approximations has been raised.

In Section 3.4, an open problem of Beierle *et al.* was resolved by consistently applying the theoretical framework that was introduced in this chapter. Finally, the choice of the type of approximations was discussed from a practical point of view. As a proof of concept, it was shown how rank one approximations can be found by solving a manifold optmization problem. In addition, good rank-one approximations for 6.5 and 10.5 rounds of Midori-64 were given as an example.

The primary goal of this chapter was to lay the foundations of a new theoretical framework for linear cryptanalysis and its variants. Future work should focus on the application of that theory to the cryptanalysis of primitives. On the one hand, this will allow a more detailed analysis of cryptanalytic properties that are already in use. On the other hand, the newly introduced types of approximations should enable more powerful attacks on block ciphers (particularly in the weak key setting) and cryptographic permutations.

Chapter 4

Clustering of Linear Trails

This chapter takes up the problem of linear trail clustering. As discussed at the end of Section 3.3.4, this is equivalent to providing upper bounds on the correlation of linear approximations in the weak key setting. The first part of this chapter considers linear approximations over two rounds of a block cipher. Specifically, the introductory Section 4.1 illustrates some of the difficulties related to weak key upper bounds. The main result, Theorem 4.1, is derived and illustrated in Section 4.2. Finally, Section 4.3 provides provisional results related to linear approximations over more than two rounds.

4.1 Introduction

The focus of this chapter is exclusively on block ciphers. Specifically, as for the variance bounds that were discussed in Section 2.3.2, the focus is on linear approximations over functions of the form

$$F_K(x) = \mathfrak{S}(L(\mathfrak{S}(x)) + K),$$

with $\mathfrak{S} = (S_1, \ldots, S_m)$. These functions arise when considering linear approximations over two rounds of an SPN. The possibility of obtaining results for more than two rounds will be discussed in Section 4.3.

The correlation of a linear approximation over F_K with masks $a = a_1 \| \cdots \| a_m$ and $b = b_1 \| \cdots \| b_m$ can be expressed as

$$c = \langle C^{\mathfrak{S}^{-1}} \delta_{\chi_b}, C^K C^{L^\top} C^{\mathfrak{S}} \delta_{\chi_a} \rangle = \langle \otimes_{i=1}^m (C^{S_i}_{\cdot,\chi_{b_i}})^\top, C^K C^{L^\top} \otimes_{i=1}^m C^{S_i}_{\cdot,\chi_{a_i}} \rangle. \tag{4.1}$$

Recall from Section 2.2 that the required amount of known plaintexts to achieve a fixed success probability and false positive rate is proportional to $1/c^2$ with c the correlation of the underlying linear approximation. Since c depends on the key, the variance bounds from Section 2.3.2 appear to provide only an average-case (for a random key) lower bound on the data complexity. Indeed, by Jensen's inequality, $\mathbb{E}[1/\boldsymbol{c}^2] \geq 1/\mathbb{E}\boldsymbol{c}^2$.

Average-case bounds do not rule out weak key attacks. A strong upper bound on c that holds for all keys is desired, but this is too hopeful. Instead, one can

consider probabilistic bounds: for how many keys can the correlation be larger than a given value? This corresponds to a tail bound on the probability distribution of the correlation $\boldsymbol{c}$ for a random key. If an upper bound on the variance is available, Chebyshev's inequality provides one such bound:

$$\Pr\left[|\boldsymbol{c}| \geq t\sqrt{\mathbb{E}\boldsymbol{c}^2}\right] \leq 1/t^2.$$

However, the Chebyshev bound is rather weak – indeed, Theorem 2.2 implies the superexponentially decreasing bound $2e^{-t^2/2}$ for random permutations. Nevertheless, as illustrated in the following section, the Chebyshev bound is sometimes tight.

4.1.1 Tightness of the Chebyshev bound: Midori-64

It will be shown that for a linear approximation over two rounds of Midori-64 with $a_1 = \ldots a_m = \mathsf{5} = b_1 = \ldots = b_m$, the Chebyshev bound can be tight. As discussed in Section 3.5.3, for 2^8 out of 2^{16} keys, $\delta_{\chi_5}^{\otimes 4}$ is an invariant for $[C^S]^{\otimes 4}C^K C^M [C^S]^{\otimes 4}$. The branch number of the linear layer M equals four, so Theorem 2.3 yields $\mathbb{E}\boldsymbol{c}^2 \leq 2^{-6}$. Hence, $\Pr\left[|\boldsymbol{c}| = 1\right] \leq 2^{-6}$ – not yet tight. However, in this case, the variance bound from Theorem 2.3 can be improved when all S-boxes are active. Let $f = C^{\mathfrak{S}}_{\cdot,\chi_a}$ and $g^\top = C^{\mathfrak{S}}_{\chi_b,\cdot}$, then

$$\mathbb{E}\boldsymbol{c}^2 = \sum_{u\in\mathbb{F}_2^n} f_{\chi_u}^2\, g_{\chi_{Mu}}^2 \leq \sum_{i=1}^{l} c_i^2 N_i,$$

with $0 = c_1 < c_2 < \cdots < c_l$ and N_i an upper bound on the number of linear trails with correlation between c_i and c_{i+1}. This approach is feasible because many choices of the masks a and b result in the same values of $c_1, \ldots, c_l$ and $N_1, \ldots, N_l$.

For Midori-64, one has five possible distributions of $f_{\chi_u}^2$ and $g_{\chi_u}^2$, see Table 4.1. Hence, there are only ten possible distributions for the total correlation. The maximum variance is 2^{-8}, corresponding to the first row of Table 4.1. This yields the tight bound $\Pr\left[|\boldsymbol{c}| = 1\right] \leq 2^{-8}$.

Table 4.1: Possible distributions of linear trail correlations over four parallel Midori-64 S-boxes. The first row results in the optimal bound: $256 \cdot (2^{-8})^2 = 2^{-8}$.

c^2	2^{-8}	2^{-10}	2^{-12}	2^{-14}	2^{-16}
	256				
	128	512			
	64	512	1024		
	32	384	1536	2048	
	16	256	1536	4096	4096

4.1.2 Alternative S-boxes for Midori-64

To obtain a better bound using the enumeration technique from Section 4.1.1, each column and row of the correlation matrix C^S of the S-box should be supported on a set of size exceeding four. Indeed, this ensures that there can not be as many as 256 trails with correlation 2^{-8}; instead, there will be more trails but with significantly smaller correlations. The resulting S-boxes will serve as a useful reference in Section 4.2.

It is not hard to show that only eight out of 302 classes of affinely-equivalent [33] 4-bit S-boxes satisfy the above condition, while also achieving optimal nonlinearity. Table 4.2 provides an example. The variance upper bound is as determined by the

Table 4.2: S-box satisfying the conditions outlined in Section 4.1.2.

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$S(x)$	0	1	2	3	4	6	8	c	5	9	b	d	e	a	7	f

the last row of Table 4.1:

$$16 \cdot 2^{-16} + 256 \cdot 2^{-20} + 1536 \cdot 2^{-24} + 4096 \cdot 2^{-28} + 4096 \cdot 2^{-32} = 625 \cdot 2^{-20}.$$

In fact, for the S-boxes in each of these eight classes, one can additionally show that no vector $v^{\otimes 4}$ with v a column of C^S or $C^{S^{-1}}$ is an eigenvector of C^M. This follows from the author's previous work [14, Theorem 9].

4.2 Moments Method

One can sometimes improve upon Chebyshev's bound by considering higher moments. Indeed, by applying Markov's inequality to $|\boldsymbol{c}|^q$, one obtains (suppose $q \geq 2$)

$$\Pr\left[|\boldsymbol{c}| \geq t\sqrt[q]{\mathbb{E}|\boldsymbol{c}|^q}\right] \leq 1/t^q.$$

Since $|\boldsymbol{c}| \leq 1$, such bounds are only useful when the qth moment $\mathbb{E}|\boldsymbol{c}|^q$ is not too large. Specifically, an improvement over the Chebyshev bound is only achieved when comparing $|\boldsymbol{c}|$ with correlations exceeding $(\mathbb{E}|\boldsymbol{c}|^q/\mathbb{E}\boldsymbol{c}^2)^{1/(q-2)}$.

The moment method consists in upper bounding the absolute qth moments $\mathbb{E}|\boldsymbol{c}|^q$ with $\boldsymbol{c}$ the correlation as defined by (4.1) for a random key. For technical reasons, it will be assumed that q is even. Section 4.2.1 proceeds by deriving a general upper bound, which will be applied to the specific case of (4.1) in Section 4.2.2.

4.2.1 General Upper Bound

The main result of this section (Theorem 4.1) considers arbitrary rank-one approximations. That is, upper bounds are derived for

$$\mathbb{E}\,\langle \boldsymbol{f}, C^L \boldsymbol{g}\rangle^q = \mathbb{E}\,\langle \otimes_{i=1}^m \boldsymbol{f}^i, C^L \otimes_{i=1}^m \boldsymbol{g}^i\rangle^q,$$

where $\boldsymbol{f}^i$ and $\boldsymbol{g}^i$, $i = 1, \ldots, m$ are independent random variables. The proof of Theorem 4.1 below makes use of the following consequence of Hölder's inequality. It also appears in the proof of Theorem 2.3 by Park *et al.* [87].

Lemma 4.1 (Park *et al.* [87])**.** *For $i = 1, \ldots, l$, let $x_1^{(i)}, \ldots, x_k^{(i)}$ be non-negative real numbers. It holds that*

$$\sum_{i=1}^{l} \prod_{j=1}^{k} x_i^{(j)} \leq \max_{1 \leq j \leq k} \sum_{i=1}^{l} \left[x_i^{(j)}\right]^k.$$

Proof. The following proof is given by Park *et al.* [87, Lemmas 1 and 2]. Hölder's inequality implies that

$$\sum_{i=1}^{l} \prod_{j=1}^{k} x_i^{(j)} \leq \prod_{j=1}^{k} \Big(\sum_{i=1}^{l} \left[x_i^{(j)}\right]^k \Big)^{1/k}.$$

The result follows by upper bounding each factor in the resulting product by the largest factor. □

The main result will now be stated. To simplify notation, the convention will be used that C^L represents an operator on $\mathbb{C}\mathbb{F}_2^n$ rather than $\mathbb{C}\widehat{\mathbb{F}_2^n}$. As discussed in Chapter 3, this can be done by identifying $\mathbb{F}_2^n$ with $\widehat{\mathbb{F}_2^n}$ through the isomorphism $u \mapsto \chi_u$.

Theorem 4.1 (Moment upper bound)**.** *Let $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a linear map such that $L^\top$ has branch number d. Suppose further that $\boldsymbol{f} = \otimes_{i=1}^m \boldsymbol{f}^i$ and $\boldsymbol{g} = \otimes_{i=1}^m \boldsymbol{g}^i$ with $\{\boldsymbol{f}^i\}_{i=1}^m$ and $\{\boldsymbol{g}^i\}_{i=1}^m$ independent random variables on $\mathbb{R}\mathbb{F}_2^n$. Then for any even integer $q > 1$, it holds that*

$$\mathbb{E}\,\langle \boldsymbol{f}, C^L \boldsymbol{g} \rangle^q \leq \sqrt{\alpha_{q,1}\beta_{q,1}}^{N-d} \max\{\alpha_{q,d}, \beta_{q,d}\},$$

with $N = |\{i \in [m] \mid \boldsymbol{f}^i \not\equiv \delta_0\}| + |\{i \in [m] \mid \boldsymbol{g}^i \not\equiv \delta_0\}|$ the number of active cells, and (with $V = \mathbb{F}_2^{n/m}$)

$$\alpha_{q,e} = \max_{j \in [m]} \sum_{w_1, \ldots, w_q \in V \setminus \{0\}} \Big| \mathbb{E} \prod_{i=1}^{q} \boldsymbol{f}_{w_i}^j \Big|^e,$$

$$\beta_{q,e} = \max_{j \in [m]} \sum_{w_1, \ldots, w_q \in V \setminus \{0\}} \Big| \mathbb{E} \prod_{i=1}^{q} \boldsymbol{g}_{w_i}^j \Big|^e.$$

Proof. The proof is similar to that of Theorem 2.3 due to Park *et al.* [87], but considerably more technical. Let $M = L^\top$. The qth moment can be expressed as

$$\mathbb{E}\,\langle \boldsymbol{f}, C^L \boldsymbol{g} \rangle^q = \mathbb{E} \prod_{i=1}^{q} \sum_{u \in \mathbb{F}_2^n} \boldsymbol{f}_u \boldsymbol{g}_{Mu}, \tag{4.2}$$

Expanding the right hand side of (4.2) yields

$$\mathbb{E}\,\langle \boldsymbol{f}, C^L \boldsymbol{g} \rangle^q = \sum_{u_1, \ldots, u_q \in \mathbb{F}_2^n} \mathbb{E} \prod_{i=1}^{q} \boldsymbol{f}_{u_i} \boldsymbol{g}_{Mu_i} = \sum_{u_1, \ldots, u_q \in V^m} \mathbb{E} \prod_{i=1}^{q} \prod_{j=1}^{m} \boldsymbol{f}_{u_{i,j}}^j \boldsymbol{g}_{v_{i,j}}^j,$$

where for $u_i \in V^m \cong \mathbb{F}_2^n$, we write $u_i = (u_{i,1}, \ldots, u_{i,m})$ with $u_{i,j} \in V$ and similarly for $v_i = Mu_i$. Now let $A = \{j \in [m] \mid \boldsymbol{f}^j \not\equiv \delta_0\}$ and $B = \{j \in [m] \mid \boldsymbol{g}^j \not\equiv \delta_0\}$. Remark that, by definition of the branch number, $|A| + |B| \geq d$. It then follows that

$$\mathbb{E}\,\langle \boldsymbol{f}, C^L\boldsymbol{g}\rangle^q = \sum_{u_1,\ldots,u_q \in V^m} \mathbb{E} \prod_{i=1}^{q} \Big(\prod_{j\in B} \boldsymbol{f}^j_{u_{i,j}}\Big)\Big(\prod_{j\in A} \boldsymbol{g}^j_{v_{i,j}}\Big)\Big(\prod_{j\in[m]\setminus B} \delta_{u_{i,j}}\Big)\Big(\prod_{j\in[m]\setminus A} \delta_{v_{i,j}}\Big).$$

In the right-hand side above, one can exclude all terms of the sum which do not satisfy the conditions

$$\begin{aligned} v_{i,j} &\neq 0 \text{ for } j \in A \quad \text{and} \quad v_{i,j} = 0 \text{ for } j \in [m] \setminus A, \\ u_{i,j} &\neq 0 \text{ for } j \in B \quad \text{and} \quad u_{i,j} = 0 \text{ for } j \in [m] \setminus B. \end{aligned}$$

Let Z denote the set of pairs (u_i, v_i) such that the above conditions are satisfied. Then one can write

$$\mathbb{E}\,\langle \boldsymbol{f}, C^L\boldsymbol{g}\rangle^q = \sum_{(u_1,v_1),\ldots,(u_q,v_q)\in Z} \mathbb{E} \prod_{i=1}^{q} \Big(\prod_{j\in B} \boldsymbol{f}^j_{u_{i,j}}\Big)\Big(\prod_{j\in A} \boldsymbol{g}^j_{v_{i,j}}\Big).$$

By the triangle inequality (recall that q is even),

$$\begin{aligned} \mathbb{E}\,\langle \boldsymbol{f}, C^L\boldsymbol{g}\rangle^q &\leq \sum_{(u_1,v_1),\ldots,(u_q,v_q)\in Z} \Big|\mathbb{E} \prod_{i=1}^{q} \Big(\prod_{j\in B} \boldsymbol{f}^j_{u_{i,j}}\Big)\Big(\prod_{j\in A} \boldsymbol{g}^j_{v_{i,j}}\Big)\Big| \\ &= \sum_{(u_1,v_1),\ldots,(u_q,v_q)\in Z} \Big(\prod_{j\in B} \Big|\mathbb{E} \prod_{i=1}^{q} \boldsymbol{f}^j_{u_{i,j}}\Big|\Big)\Big(\prod_{j\in A} \Big|\mathbb{E} \prod_{i=1}^{q} \boldsymbol{g}^j_{v_{i,j}}\Big|\Big), \end{aligned}$$

where we have used the independence of the random variables $\{\boldsymbol{f}^j\}_{j=1}^m$ and $\{\boldsymbol{g}^j\}_{j=1}^m$.

The remaining work is similar to the reasoning of Park *et al.* [87], but notationally somewhat heavier. Recall that $|A| + |B| \geq d$. Hence, we can apply Lemma 4.1 with $k \geq d$. Taking a large value of k is in principle beneficial, but it is also necessary to restrict the set Z. As will be shown below, this requires $k = d$.

Let $r = |A| + |B| - d$ and choose sets $A^* \subseteq A$ and $B^* \subseteq B$ with $|A^*| + |B^*| = r$. For arbitrary constants $v^*_{i,j} \in V$ with $j \in A^*$ and $u^*_{i,j} \in V$ with $j \in B^*$, let

$$\bar{Z}_{u^*_i,v^*_i} = \{(u_i, v_i) \in Z \mid \forall j \in A^*, j' \in B^* : v_{i,j} = v^*_{i,j} \wedge u_{i,j'} = u^*_{i,j'}\}.$$

That is, $\bar{Z}_{u^*_i,v^*_i}$ is obtained by fixing the value of some of the coordinates of the elements of Z. The idea behind this is as follows: for all distinct $(u_i, v_i), (u'_i, v'_i) \in \bar{Z}_{u^*_i,v^*_i}$, we have $v_{i,j} \neq v'_{i,j}$ when $j \notin A^*$ and $u_{i,j} \neq u'_{i,j}$ when $j \notin B^*$. Indeed, this is true by the definition of the branch number. Using these definitions, one obtains

$$\mathbb{E}\,\langle \boldsymbol{f}, C^L\boldsymbol{g}\rangle^q \leq \sum_{\substack{u^*_1,\ldots,u^*_q \\ v^*_1,\ldots,v^*_q}} \sum_{(u_l,v_l)\in\bar{Z}_{u^*_l,v^*_l} : l\in[q]} \Big(\prod_{j\in B} \Big|\mathbb{E} \prod_{i=1}^{q} \boldsymbol{f}^j_{u_{i,j}}\Big|\Big)\Big(\prod_{j\in A} \Big|\mathbb{E} \prod_{i=1}^{q} \boldsymbol{g}^j_{v_{i,j}}\Big|\Big),$$

where the first sum ranges over all possible $|V|^{rq}$ assignments to the variables $u^*_{i,j}$ and $v^*_{i,j}$. Lemma 4.1 can now be applied to obtain the upper bound

$$\sum_{(u_l,v_l)\in\bar{Z}_{u_l^*,v_l^*}:l\in[q]} \Big(\prod_{j\in B\setminus B^*}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{f}^j_{u_{i,j}}\Big|\Big)\Big(\prod_{j\in A\setminus A^*}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{g}^j_{v_{i,j}}\Big|\Big)$$

$$\le \max\Big\{\max_{j\in A\setminus A^*}\sum_{(u_l,v_l)\in\bar{Z}_{u_l^*,v_l^*}:l\in[q]}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{f}^j_{u_{i,j}}\Big|^d,\ \max_{j\in B\setminus B^*}\sum_{(u_l,v_l)\in\bar{Z}_{u_l^*,v_l^*}:l\in[q]}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{g}^j_{v_{i,j}}\Big|^d\Big\}$$

$$\le \max\Big\{\max_{j\in A\setminus A^*}\sum_{u_{1,j},\ldots,u_{q,j}\in V_0}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{f}^j_{u_{i,j}}\Big|^d,\ \max_{j\in B\setminus B^*}\sum_{v_{1,j},\ldots,v_{q,j}\in V_0}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{g}^j_{v_{i,j}}\Big|^d\Big\}.$$

Denote the right hand side of the above inequality by $\mathcal{U}$, then

$$\begin{aligned}\mathbb{E}\,\langle \boldsymbol{f}, C^L\boldsymbol{g}\rangle^q &\le \mathcal{U}\sum_{\substack{u_1^*,\ldots,u_q^*\\ v_1^*,\ldots,v_q^*}}\Big(\prod_{j\in B^*}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{f}^j_{u^*_{i,j}}\Big|\Big)\Big(\prod_{j\in A^*}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{g}^j_{v^*_{i,j}}\Big|\Big),\\ &\le \mathcal{U}\Big(\prod_{j\in B^*}\sum_{u^*_{1,j},\ldots,u^*_{q,j}\in V_0}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{f}^j_{u^*_{i,j}}\Big|\Big)\Big(\prod_{j\in A^*}\sum_{v^*_{1,j},\ldots,v^*_{q,j}\in V_0}\Big|\mathbb{E}\prod_{i=1}^{q}\boldsymbol{g}^j_{v^*_{i,j}}\Big|\Big)\\ &\le \alpha_{q,1}^{|A^*|}\beta_{q,1}^{|B^*|}\max\{\alpha_{q,d},\beta_{q,d}\}.\end{aligned}$$

The choice of A^* and B^* is still free, up to the constraint $|A^*|+|B^*|=r$. The value $\alpha_{q,e}$ is typically a property of the S-box, and $\beta_{q,e}$ is a property of its inverse. Hence, $|A^*|=r/2$ and $|B^*|=r/2$ is usually a reasonable choice. This results in the bound

$$\mathbb{E}\,\langle \boldsymbol{f}, C^L\boldsymbol{g}\rangle^q \le \sqrt{\alpha_{q,1}\beta_{q,1}}^{|A|+|B|-d}\ \max\{\alpha_{q,d},\beta_{q,d}\}.$$

□

Remark. Theorem 4.1 makes no assumptions on L other than that its linear branch number equals d. If one additionally assumes that L is $\mathbb{F}_{2^{n/m}}$-linear, (somewhat) stronger bounds may be obtained by using ideas from the variance bounds of Canteaut and Roué [32]. This will be left as future work. In any case, only relatively small improvements are expected. For example, for $q=2$, both bounds agree in several important cases (such as for the AES). ▷

Broadly speaking, Theorem 4.1 shows that the qth moment can be larger (*i.e.* more clustering can occur) if more cells are active. This is due to the first factor in the upper bound: it is exponential in the difference between the number of active cells and the branch number.

Also note that, although Theorem 4.1 does not apply to mixed moments[1], it depends (for $q>2$) on mixed-moments of the vectors $\boldsymbol{f}^i$ and $\boldsymbol{g}^i$, $i=1,\ldots,q$. This provides a weak link with nonlinear approximations of the S-box.

[1]The result can be extended based on the same proof ingredients, but this makes the notation heavier without yielding significant results. This may change if the analysis is extended to more than two rounds(see Section 4.3).

4.2.2 Application to Linear Approximations

This section applies Theorem 4.1 to the case of linear approximations over two rounds of an SPN. For simplicity, assume that all S-boxes are equal: $S_1 = \ldots = S_m = S$. That is, consider $\boldsymbol{g}^i = C^{\boldsymbol{K}} C^S \delta_{a_i}$ and $\boldsymbol{f}^i = C^{\boldsymbol{K'}} [C^S]^\top \delta_{b_i}$ with $\boldsymbol{K}$ and $\boldsymbol{K'}$ independent and uniformly distributed. Hence,

$$\alpha_{q,e} = \max_{a \in V_0} \sum_{w_1,\ldots,w_q \in V_0} \left| \mathbb{E} \prod_{i=1}^{q} (-1)^{w_i^\top \boldsymbol{K}} C^S_{w_i,a} \right|^e = \max_{a \in V_0} \sum_{\substack{w_1,\ldots,w_q \in V_0 \\ w_1+\cdots+w_q=0}} \prod_{i=1}^{q} |C^S_{w_i,a}|^e,$$

$$\beta_{q,e} = \max_{b \in V_0} \sum_{w_1,\ldots,w_q \in V_0} \left| \mathbb{E} \prod_{i=1}^{q} (-1)^{w_i^\top \boldsymbol{K'}} C^S_{b,w_i} \right|^e = \max_{b \in V_0} \sum_{\substack{w_1,\ldots,w_q \in V_0 \\ w_1+\cdots+w_q=0}} \prod_{i=1}^{q} |C^S_{b,w_i}|^e.$$

The cost of naively computing the values $\alpha_{q,e}$ is bounded by $\mathcal{O}(|V_0|^q)$, hence feasible only for q not too large. An easy optimization is to iterate only over all q-combinations (with repetition) from elements of V_0. Further improvements are possible if the correlation only takes a small number of different values (as is often the case); this will not discussed here since the naive approach suffices for small values of q.

The remainder of this section computes concrete values of the bound provided by Theorem 4.1 for several block ciphers. First, consider ciphers with 4-bit S-boxes. Table 4.3 gives concrete values for $\alpha_{q,e}$ and $\beta_{q,e}$ for the Midori-64 S-box as well as the alternative S-boxes discussed in Section 4.1.2. Linear layers with linear branch numbers $d = 4$ or $d = 5$ (MDS) are considered.

Table 4.3: Values of $\alpha_{q,e} = \beta_{q,e}$ with $q \in \{2, 4, 6, 8\}$ and $e \in \{1, 5\}$ for the Midori-64 S-box and any of the alternative S-boxes from Section 4.1.2.

	Midori-64 S-box			S-box from Section 4.1.2		
q	$\alpha_{q,1} = \beta_{q,1}$	$\alpha_{q,4} = \beta_{q,4}$	$\alpha_{q,5} = \beta_{q,5}$	$\alpha_{q,1} = \beta_{q,1}$	$\alpha_{q,4} = \beta_{q,4}$	$\alpha_{q,5} = \beta_{q,5}$
2	1	2^{-6}	2^{-8}	1	$2^{-6.978}$	$2^{-8.994}$
4	5.5	2^{-10}	2^{-14}	5.5	$2^{-12.869}$	$2^{-16.966}$
6	46	2^{-14}	2^{-20}	46	$2^{-18.679}$	$2^{-24.917}$
8	410.5	2^{-18}	2^{-26}	410.5	$2^{-24.408}$	$2^{-32.845}$

As anticipated in Section 4.1.1, when all S-boxes are active, the result for the Midori-64 S-box and $d = 4$ does not not improve upon the Chebyshev bound. The choice of the S-boxes was discussed in Section 4.1.2. Another relevant question is whether or not a linear layer with branch number $d = 5$ can result in stronger bounds.

Figure 4.1 shows the resulting tail bounds on the correlation for both choices of the S-box and $d = 5$. With many active S-boxes, little or no improvement over the Chebyshev bound is obtained. In fact, the situation is aggravated when the improved variance bounds from Sections 4.1.1 and 4.1.2 are taken into account. For a small number of active S-boxes, the bounds in Figure 4.1 are more interesting. Indeed, one can conclude that such approximations can only exhibit strong clustering for

a small fraction of keys. For example, Figure 4.1 rules out the existence of perfect linear approximations with less than seven active S-boxes (when the S-boxes from Section 4.1.2 are used). Faster decay of the tail can likely be obtained by taking into account higher moments (Figure 4.1 relies only on $q \leq 8$).

Tail bounds for the block ciphers AES [46] and SKINNY-128 [10] are shown in Figure 4.2. Table 4.4 lists the corresponding values of $\alpha_{q,e}$ and $\beta_{q,e}$. Note that presented bounds rely only on the second and forth moment; considering higher moments can yield further improvements. The comparatively poor bounds for SKINNY are expected, given its lightweight design choices. In particular, SKINNY uses a linear layer with branch number $d = 4$ and an S-box with significantly lower nonlinearity than AES. With respect to the number of active S-boxes, the same conclusion as above applies.

Table 4.4: Values of $\alpha_{q,e} = \beta_{q,e}$ with $q \in \{2, 4\}$ for SKINNY ($e \in \{1, 4\}$) and AES ($e \in \{1, 5\}$). For SKINNY, the reported values are such that $\sqrt{\alpha_{q,1}\beta_{q,1}}^{N-d} \max\{\alpha_{q,4}, \beta_{q,4}\}$ is maximal – rather than the worst case values of $\alpha_{q,1}$ and $\alpha_{q,4}$ separately.

	SKINNY		AES	
q	$\alpha_{q,1} = \beta_{q,1}$	$\alpha_{q,4} = \beta_{q,4}$	$\alpha_{q,1} = \beta_{q,1}$	$\alpha_{q,5} = \beta_{q,5}$
2	1	2^{-6}	1	$2^{-26.478}$
4	2	2^{-10}	$2^{7.021}$	$2^{-49.171}$

4.2.3 Future Work related to the Moments Method

The examples in Section 4.2.2 show that Theorem 4.1 provides effective tail bounds on the correlation of linear approximations with few active S-boxes, but not when many S-boxes are active. It was shown in Section 4.1.1 that when all S-boxes are active, a relatively straightforward combinatorial enumeration of linear trails can yield better variance bounds than results based on the branch number (*i.e.* Theorem 4.1 with $q = 2$). Hence, an interesting direction for future work is to investigate the extension of this approach to higher moments. This may help to cover the remaining case, *i.e.* linear approximations involving many active S-boxes.

One important limitation of Theorem 4.1, and even more so for the enumerative approach from Section 4.1.1, is that it does not consider the relation between the linear and nonlinear layer. This is problematic, because many weak key approximations ultimately rely on structural properties that only become worrisome when the linear and nonlinear layer are analyzed jointly. An alternative approach – not discussed in this thesis – based on averaging-out key-dependencies might resolve this: for a weak key class $\mathcal{K}$ with correlation at least c, one has

$$c \leq \frac{1}{|\mathcal{K}|} \sum_{K \in \mathcal{K}} |\langle f, C^K g \rangle| = \langle f, \widehat{\varphi}_{\mathcal{K}} g \rangle \text{ with } \varphi_{\mathcal{K}}(K) = \begin{cases} \operatorname{sign}\langle f, C^K g \rangle / |\mathcal{K}| & \text{if } K \in \mathcal{K} \\ 0 & \text{otherwise.} \end{cases}$$

However, for this to work, assumptions must be made about the structure of $\mathcal{K}$.

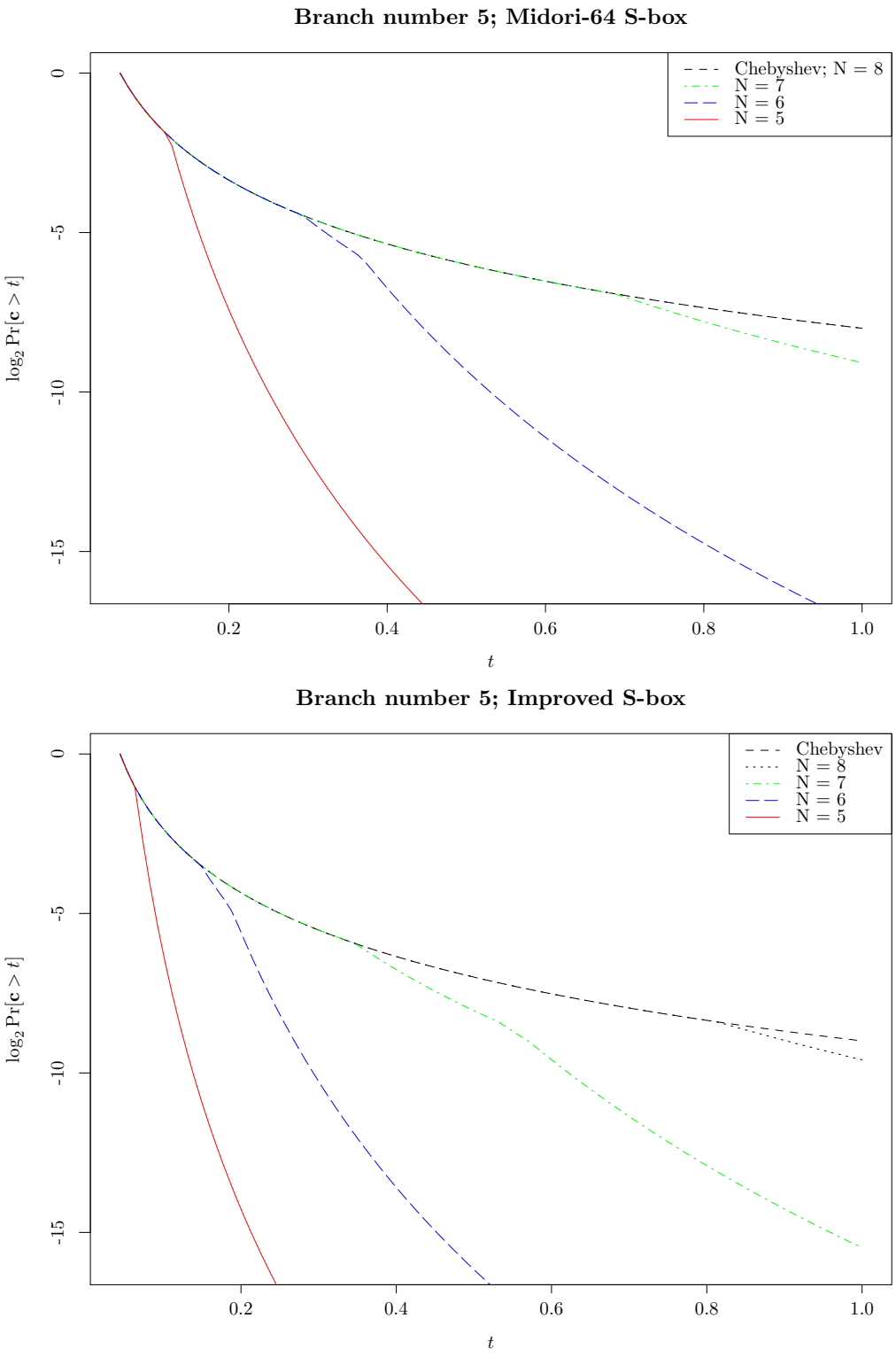


Figure 4.1: Upper bounds on the tail of the correlation distribution for two-round SPNs for branch number five. The S-box is either the Midori-64 S-box (top) or any of the S-boxes discussed in Section 4.1.2 (bottom).

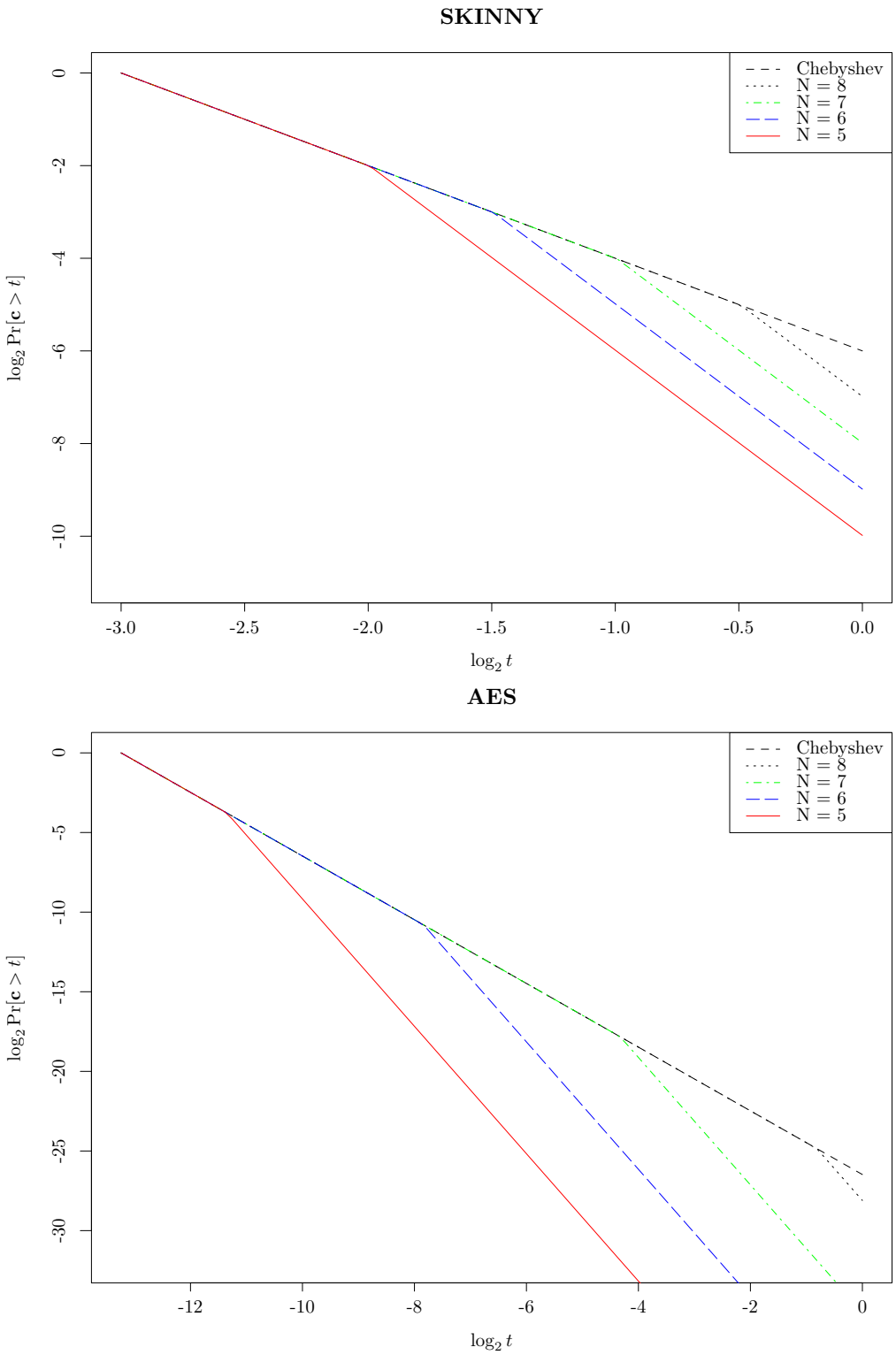


Figure 4.2: Upper bounds on the tail of the correlation distribution for two-round `SKINNY` (top) and AES (bottom). `SKINNY` uses a linear layer with branch number $d = 4$; for AES, $d = 5$.

4.3 Hypercontractivity

Ideally, one would like to extend the bounds discussed in the previous sections beyond two rounds. This is nontrivial even when round keys may be assumed to be independent. Indeed, whereas the variance can be shown not to increase with additional rounds, higher moments *can* increase. This section discusses how the notion of *hypercontractivity* [80, 86] can potentially be used to resolve this problem. As of yet, the results in this section are conditional.

4.3.1 Definitions and Basic Results

In the following, denote the q-norm of a random variable $\boldsymbol{X}$ by $\|\boldsymbol{X}\|_q = \sqrt[q]{\mathbb{E}|\boldsymbol{X}|^q}$. One can show that this indeed defines a norm, subject to the restriction that $\|\boldsymbol{X}\|_q < \infty$. The latter condition is, of course, satisfied for all random variables in this chapter.

Definition 4.1 (Hypercontractivity [86])**.** *Let $1 \leq p \leq q$ and $\rho \in (0, 1)$ be real numbers. A random variable $\boldsymbol{X}$ is (p, q, ρ)-hypercontractive iff for all $a \in \mathbb{R}$,*

$$\|1 + \rho a\boldsymbol{X}\|_q \leq \|1 + a\boldsymbol{X}\|_p.$$

In the context of linear approximations, the goal will be to show that the correlation $\boldsymbol{c}$ is hypercontractive with $p = 2$. The intuition is that, when ρ is not too small, the distribution of the correlation is "reasonable" in the sense that its variance will be a meaningful measure of dispersion. A few basic properties of hypercontractive random variables are listed in Theorem 4.2.

Theorem 4.2 (Hypercontractivity properties [86])**.** *Let $\boldsymbol{X}$ be a (p, q, ρ)-hypercontractive random variable. Then*

1. *For all $\rho' \in (0, \rho]$, $\boldsymbol{X}$ is (p, q, ρ')-hypercontractive.*
2. *The mean of $\boldsymbol{X}$ is zero.*
3. *For all $a \in \mathbb{R}$, $a\boldsymbol{X}$ is (p, q, ρ)-hypercontractive.*
4. *$\rho \leq \min\{\sqrt{(p-1)/(q-1)}, \|\boldsymbol{X}\|_p / \|\boldsymbol{X}\|_q\}$.*

Hypercontractivity is closely related to the moments method. Indeed, by Theorem 4.2, a $(2, q, \rho)$-hypercontractive random variable $\boldsymbol{X}$ satisfies $\|\boldsymbol{X}\|_q \leq \|\boldsymbol{X}\|_2 / \rho$. Conversely, one can show that any *symmetric* random variable is $(2, q, \rho)$-hypercontractive with $1/\rho = \|\boldsymbol{X}\|_q \sqrt{q-1}$ [86, §10.2, Theorem 12].

It can be shown that optimal $(2, q, \rho)$-hypercontractivity, *i.e.* $\rho = 1/\sqrt{q-1}$, is achieved for Rademacher and Gaussian random variables [86]. Theorem 4.3 states that the correlation of any nonlinear approximation over a uniform random permutation is also optimally hypercontractive.

Theorem 4.3. *Let $\boldsymbol{F}$ be a uniform random permutation on $\mathbb{F}_2^n$. The correlation of any nonlinear expression of the form $f(x) + g(\boldsymbol{F}(x))$, with f and g balanced Boolean functions, is a $(2, q, 1/\sqrt{q-1})$-hypercontractive random variable for all $q \geq 2$.*

Proof. A proof can be found in Appendix A.1. □

4.3.2 Towards Upper Bounds for Multiple Rounds

Suppose that one has established $(2, q, \rho)$-hypercontractivity for the correlation of linear approximations over round functions $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ depending on independent random keys. What can then be concluded about linear approximations over $\boldsymbol{F}_2 \circ \boldsymbol{F}_1$? Theorem 4.4 provides a preliminary result in this direction. Note that $\operatorname{supp} \boldsymbol{X}$ refers to the set of values that a random variable $\boldsymbol{X}$ can take with nonzero probability.

Theorem 4.4. *Let $\boldsymbol{F}_1, \boldsymbol{F}_2$ be independent random functions with corresponding random correlation matrices $C^{\boldsymbol{F}_1}$ and $C^{\boldsymbol{F}_2}$. Let $a, b \in \mathbb{F}_2^n$. If*

1. *for all $f^\top \in \operatorname{supp} C^{\boldsymbol{F}_2}_{\chi_b, \cdot}$, the random variable $\langle f, C^{\boldsymbol{F}_1}_{\cdot, \chi_a} \rangle$ is $(2, q, \rho_1)$-hypercontractive,*
2. *for all $g^\top \in \operatorname{supp} C^{\boldsymbol{F}_1}_{\cdot, \chi_a}$, the random variable $\langle C^{\boldsymbol{F}_2}_{\cdot, \chi_b}, g \rangle$ is $(2, q, \rho_2)$-hypercontractive,*

then $C^{\boldsymbol{F}_2 \circ \boldsymbol{F}_1}_{\chi_b, \chi_a}$ is $(2, q, \rho_1\rho_2)$-hypercontractive.

Proof. The proof relies on a significant amount of background material, so it has been moved to Appendix A.2. □

Theorem 4.4 makes minimal assumptions about $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$, so it is not surprising that the hypercontractivity parameter ρ decreases. Nevertheless, one may still obtain stronger tail bounds if it can be shown *e.g.* that the variance decreases. One should consequently think of Theorem 4.4 as a result that shows to what extend hypercontractivity can deteriorate when composing two non-ideal functions.

The main difficulty with Theorem 4.4 is that it requires knowledge about (hypercontractivity of) the correlation of some nonlinear approximations over $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$. The moment bound provided by Theorem 4.1 is not sufficiently general to provide this information. Obtaining the relevant extensions is left as future work.

4.4 Conclusion

In this chapter, advances were made in the search for upper bounds on the correlation of linear approximations. Providing such bounds is equivalent to ruling out clustering of linear trails. In Section 4.1, it was shown that the variance bounds from Section 2.3.2 can be used to obtain weak key bounds through Chebyshev's inequality. It was illustrated, using Midori-64, that the Chebyshev bound is sometimes optimal.

In Section 4.2, the variance approach was generalized by considering higher moments. This resulted in Theorem 4.1, which provides an upper bound on the even moments of the correlation of rank-one approximations over a linear layer with a given linear branch number. The applications of Theorem 4.1 in Section 4.2.2 illustrated that, for some ciphers such as the AES, strong clustering is not possible when few S-boxes are active. However, it was noted that Theorem 4.1 generally does not improve upon Chebyshev's inequality when all S-boxes are active. Some suggestions were made to address this in future work.

Finally, Section 4.3 discussed preliminary results on using hypercontractivity to extend the moments method from Section 4.2 beyond two rounds. Theorem 4.3 states

that correlations of (non)linear approximations over uniform random permutations are indeed optimally hypercontractive. The main result, Theorem 4.4, shows to what extend hypercontractivity may deteriorate when composing non-ideal functions. Future work is necessary to investigate the assumptions of this result.

Chapter 5

Conclusion

The object of this thesis was the reexamination and generalization of the theory of linear cryptanalysis. This goal was motivated by the growing importance of lightweight and permutation-based cryptography, as is demonstrated by the ongoing NIST lightweight cryptography standardization project [8]. The design choices resulting therefrom have encouraged the development of block cipher invariants [14, 51, 71, 94] and have led to renewed interest in old ideas such as the pursuit of nonlinear approximations [9, 52, 69]. These attacks and others are entangled with linear cryptanalysis in the weak model and, as shown in this thesis, can be described using a single theoretical framework. This framework is rooted in the author's work on block cipher invariants [14] from ASIACRYPT 2018. In addition, this thesis considers the problem of proving security against linear cryptanalysis in the weak key setting – thereby extending variance bounds.

After given a substantive overview of the state of the art in Chapter 2, a "geometric approach" to linear cryptanalysis and its generalizations was developed in Chapter 3. The essence of this approach is to describe (multiple) linear and nonlinear approximations by pairs of low-dimensional subspaces of $\mathbb{C}G$, where G is a finite abelian group (usually $\mathbb{F}_2^n$). Such vector spaces can be propagated through an iterative function by means of successive orthogonal projections on intermediate vector spaces. This process generalizes the classical piling-up principle. The theory developed in Sections 3.1 to 3.3 was then used to resolve an open problem posed by Beierle *et al.* [9] at FSE 2019. Further applications were discussed in Section 3.5.

The issue of correlation upper bounds in the weak key setting was taken up in Chapter 4. Specifically, the study of probabilistic upper bounds on the correlation of linear approximations over two rounds of SPNs was initiated. The limitations of variance bounds in obtaining such results were discussed in Section 4.1. Improved bounds, based on higher moments of the correlation, were derived in Section 4.2 and applied to several examples including the AES. Finally, a prospective approach to the extension of these results beyond two rounds was developed in Section 4.3. Future work should focus on obtaining stronger results for approximations with many active S-boxes. In addition, as suggested by the conditional results in Section 4.3, correlation bounds for some nonlinear approximations should be derived.

Appendices

Appendix A

Hypercontractivity

This appendix contains the proofs of Theorem 4.3 and Theorem 4.4. Some additional background information on hypercontractivity is also included.

A.1 Proof of Theorem 4.3

Lemma A.1. *Let $\boldsymbol{F}$ be a uniform random permutation and let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{2^{n-1}}$ be a random sample without replacement from the multiset $\{0^{(2^{n-1})}, 1^{(2^{n-1})}\}$. The correlation $C(f, g \circ \boldsymbol{F})$ of an approximation $f(x) + g(\boldsymbol{F}(x))$ with f and g balanced Boolean functions satisfies*

$$C(f, g \circ \boldsymbol{F}) = \frac{1}{2^{n-2}} \left(\sum_{k=1}^{2^{n-1}} \boldsymbol{X}_k \right) - 1.$$

Proof. By the definition of the correlation coefficient $C(f, g \circ \boldsymbol{F})$:

$$\begin{aligned} C(f, g \circ \boldsymbol{F}) &= \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) + g(\boldsymbol{F}(x))} \\ &= \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} [4g(\boldsymbol{F}(x))f(x) - 2g(\boldsymbol{F}(x)) - 2f(x) + 1] \\ &= \frac{1}{2^{n-2}} \left(\sum_{\substack{x \in \mathbb{F}_2^n \\ f(x)=1}} g(\boldsymbol{F}(x)) \right) - 1. \end{aligned}$$

The result follows by the balancedness of f and g. □

The proof of Theorem 4.3 further requires the following lemma, which is due to Wassily Hoeffding [56].

Lemma A.2 (Hoeffding [56])**.** *Let $P \subset \mathbb{R}$ be a finite population (multiset). Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ be a random sample without replacement from P. Likewise, let*

$\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_N$ be a random sample with replacement from P. Then for any continuous convex function $f : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E} f\left(\sum_{i=1}^{N} \boldsymbol{X}_i\right) \leq \mathbb{E} f\left(\sum_{i=1}^{N} \boldsymbol{Y}_i\right).$$

Proof. A proof may be found in Hoeffding's work [56]. □

Proof of Theorem 4.3. Let $\rho = 1/\sqrt{q-1}$ and $a \in \mathbb{R}$. By Lemma A.1, we have

$$a + \rho C(f, g \circ \boldsymbol{F}) = a + \rho \left[\frac{1}{2^{n-2}}\left(\sum_{k=1}^{2^{n-1}} \boldsymbol{X}_k\right) - 1\right],$$

where $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{2^{n-1}}$ is a random sample without replacement from the multiset $\{0^{(2^{n-1})}, 1^{(2^{n-1})}\}$. Since the real function $x \mapsto |x|^q$ is continuous and convex for $q \geq 1$, we have

$$\|a + \rho C(f, g \circ \boldsymbol{F})\|_q \leq \left\| a + \rho \left[\frac{1}{2^{n-2}}\left(\sum_{k=1}^{2^{n-1}} \boldsymbol{Y}_k\right) - 1\right] \right\|_q,$$

where $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{2^{n-1}}$ is a random sample with replacement from the multiset $\{0^{(2^{n-1})}, 1^{(2^{n-1})}\}$. The right hand side above can be rewritten as

$$\begin{aligned} a + \rho \left[\frac{1}{2^{n-2}}\left(\sum_{k=1}^{2^{n-1}} \boldsymbol{Y}_k\right) - 1\right] &= a + \rho \frac{1}{2^{n-2}}\left(\sum_{k=1}^{2^{n-1}} \boldsymbol{Y}_k - 1/2\right) \\ &= a + \rho \frac{1}{2^{n-1}}\left(\sum_{k=1}^{2^{n-1}} (-1)^{\boldsymbol{Y}_k}\right). \end{aligned}$$

Since the Rademacher random variables $(-1)^{\boldsymbol{Y}_k}$ are mutually independent and $(2, q, \rho)$-hypercontractive [86, §10.1], so is their sum (see Section A.2). It follows that

$$\|a + \rho C(f, g \circ \boldsymbol{F})\|_q \leq \|a + C(f, g \circ \boldsymbol{F})\|_2.$$

This establishes $(2, q, \rho)$-hypercontractivity. □

A.2 Proof of Theorem 4.4

The proof relies on multilinear polynomials in orthogonal ensembles. Note that many of the results below appear in the work of Mossel, O'Donnell and Oleszkiewicz [80]. The proof of Theorem 4.4 does not require the full generality of the background material presented here. Nevertheless, the general theory is instructive and enables other results that are not discussed in this thesis.

Definition A.1 (Sequence of orthonormal ensembles [80])**.** *An orthonormal ensemble is a set $\boldsymbol{\mathcal{X}}_i = \{\boldsymbol{X}_{i,j}\}_{j \in [m]}$ of orthonormal random variables with $\boldsymbol{X}_{i,1} = 1$. A tuple $\boldsymbol{\mathcal{X}} = (\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_n)$ with $\boldsymbol{\mathcal{X}}_i$ an orthonormal ensemble, will be called a sequence of ensembles. $\boldsymbol{\mathcal{X}}$ is called independent if the ensembles $\boldsymbol{\mathcal{X}}_i$ are mutually independent.*

Note that an orthogonal ensemble can trivially be converted into an orthonormal ensemble, and consequently the results that are listed below carry over to the non-normalized case. In order to define hypercontractivity of sequences of ensembles, it is useful to introduce the following notation for multilinear polynomials.

Definition A.2 (Multilinear polynomial [80])**.** *A multilinear polynomial Q over a doubly-indexed set $\mathcal{X} = \{X_{i,j}\}_{i\in[n],j\in[m]}$ of variables is a polynomial of the form*

$$Q(\mathcal{X}) = \sum_{\sigma\in[m]^n} c_\sigma \, \mathcal{X}^\sigma,$$

with $\mathcal{X}^\sigma = \prod_{i=1}^n X_{i,\sigma_i}$ and $c_\sigma \in \mathbb{R}$. The degree of σ is defined as $|\sigma| = |\{i \in [n] \mid \sigma_i \neq 1\}|$.

To generalize the notion of hypercontractivity, the so-called noise operator T_ρ is defined. Below, a basic (but somewhat uninformative) definition is given.

Definition A.3 (Noise operator [80])**.** *For any $\rho \in [0,1]$, the operator T_ρ is defined by its action on the set of multilinear polynomials (Definition A.2):*

$$(T_\rho Q)(\mathcal{X}) = \sum_{\sigma\in[m]^n} \rho^{|\sigma|} \, c_\sigma \, \mathcal{X}^\sigma.$$

Definition A.4 (Hypercontractivity for sequences of ensembles [80])**.** *Let $1 \leq p \leq q$ and $\rho \in (0,1)$ be real numbers. A sequence $\boldsymbol{\mathcal{X}}$ of orthonormal ensembles is (p,q,ρ)-hypercontractive iff for all multilinear polynomials in $\boldsymbol{\mathcal{X}}$,*

$$\|(T_\rho Q)(\boldsymbol{\mathcal{X}})\|_q \leq \|Q(\boldsymbol{\mathcal{X}})\|_p.$$

Note that Definition 4.1 corresponds to the special case of a sequence containing one ensemble, which in turn contains only two random variables (one of which is constant). Two essential properties of hypercontractive sequences of orthonormal ensembles will now be stated.

Theorem A.1 (Union of sequences [80])**.** *Let $\boldsymbol{\mathcal{X}} = (\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_{n_1})$ and $\boldsymbol{\mathcal{Y}} = (\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_{n_2})$ be independent sequences of orthonormal ensembles. If $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are (p,q,ρ)-hypercontractive, then $\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}} = (\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_{n_1}, \boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_{n_2})$ is (p,q,ρ)-hypercontractive.*

Theorem A.2 (Moments of multilinear polynomials [80])**.** *Let $\boldsymbol{\mathcal{X}}$ be a $(2,q,\rho)$-hypercontractive sequence of orthonormal ensembles, then for any multilinear polynomial Q of degree d,*

$$\|Q(\boldsymbol{\mathcal{X}})\|_q \leq \rho^{-d} \, \|Q(\boldsymbol{\mathcal{X}})\|_2.$$

Theorem A.1 results in a useful necessary condition for a sequence $\boldsymbol{\mathcal{X}}$ to be hypercontractive. Specifically, $\boldsymbol{\mathcal{X}}$ is (p,q,ρ)-hypercontractive if

– The orthonormal ensembles $\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_n$ in $\boldsymbol{\mathcal{X}}$ are mutually independent.

– For each orthonormal ensemble $\boldsymbol{\mathcal{X}}_i$, all linear combinations of the random variables in $\boldsymbol{\mathcal{X}}_i$ are (p, q, ρ)-hypercontractive.

A variation on Theorem A.2 shows that a low-degree polynomial without constant term of a $(2, q, \rho)$-hypercontractive sequence yields a $(2, q, \rho^d)$-hypercontractive random variable. Since this variant is not mentioned in [80,86], it is stated in Lemma A.3 below together with a proof.

Lemma A.3. *Let $\boldsymbol{\mathcal{X}}$ be a $(2, q, \rho)$-hypercontractive sequence of orthogonal ensembles. Then for any multilinear polynomial Q of degree d without constant term, the random variable $Q(\boldsymbol{\mathcal{X}})$ is $(2, q, \rho^d)$-hypercontractive.*

Proof. The proof is very similar to that of Theorem A.2, see Appendix A in [80]. Let $Q(\boldsymbol{\mathcal{X}}) = \sum_{i=1}^{d} Q^{=i}(\boldsymbol{\mathcal{X}})$ where $Q^{=i}$ is a homogeneous multilinear polynomial of degree i. For any a, we have

$$\begin{aligned}
\|a + \rho^d Q(\boldsymbol{\mathcal{X}})\|_q &= \|T_\rho \left(a + \sum_{i=1}^{d} \rho^{d-i}\, Q^{=i}(\boldsymbol{\mathcal{X}})\right)\|_q \\
&\leq \|a + \sum_{i=1}^{d} \rho^{d-i}\, Q^{=i}(\boldsymbol{\mathcal{X}})\|_2 \\
&= \left(a^2 + \sum_{i=1}^{d} \rho^{2(d-i)}\, \|Q^{=i}(\boldsymbol{\mathcal{X}})\|_2^2\right)^{1/2} \\
&\leq \left(a^2 + \sum_{i=1}^{d} \|Q^{=i}(\boldsymbol{\mathcal{X}})\|_2^2\right)^{1/2} \\
&= \|a + \sum_{i=1}^{d} Q^{=i}(\boldsymbol{\mathcal{X}})\|_2.
\end{aligned}$$

The third and fifth steps follow from the orthogonality of $Q^{=i}(\boldsymbol{\mathcal{X}})$ and $Q^{=j}(\boldsymbol{\mathcal{X}})$ when $i \neq j$. □

Before stating the proof of Theorem 4.4, a useful lemma (a simple consequence of Theorem A.2) will be derived. The proof of the lemma is similar to that of Theorem A.1, which can be found in Appendix A of [80].

Lemma A.4. *Let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be independent sequences of orthonormal ensembles which are respectively $(2, q, \rho_1)$- and $(2, q, \rho_2)$-hypercontractive. Let Q be a multilinear polynomial in $\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}}$. If Q is of degree d_1 in $\boldsymbol{\mathcal{X}}$ and of degree d_2 in $\boldsymbol{\mathcal{Y}}$, then*

$$\|Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})\|_q \leq \rho_1^{-d_1} \rho_2^{-d_2} \|Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})\|_2.$$

Furthermore, when Q has no constant term, $Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})$ is $(2, q, \rho_1^{d_1} \rho_2^{d_2})$-hypercontractive.

Proof. By applying Theorem A.2, one obtains

$$\|Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})\|_q = \big\| \|Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})\|_{L^q(\boldsymbol{\mathcal{X}})} \big\|_{L^q(\boldsymbol{\mathcal{Y}})} \leq \rho_1^{-d_1} \big\| \|Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})\|_{L^2(\boldsymbol{\mathcal{X}})} \big\|_{L^q(\boldsymbol{\mathcal{Y}})}.$$

As a consequence of Minkowski's integral inequality, this can be upper bounded as

$$\|Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})\|_q \leq \rho_1^{-d_1} \big\| \|Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})\|_{L^q(\boldsymbol{\mathcal{Y}})} \big\|_{L^2(\boldsymbol{\mathcal{X}})}.$$

Another application of Theorem A.2 completes the proof. The claim about the hypercontractivity of $Q(\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}})$ when Q has no constant terms can be derived from Lemma A.3 using the same approach. □

Proof of Theorem 4.4. Note that we have

$$C^{\boldsymbol{F}_2 \circ \boldsymbol{F}_1}_{\chi_b, \chi_a} = \sum_{u \in \mathbb{F}_2^n} C^{\boldsymbol{F}_2}_{\chi_b, \chi_u} C^{\boldsymbol{F}_1}_{\chi_u, \chi_a}. \tag{A.1}$$

Let $\boldsymbol{\mathcal{X}} = \{C^{\boldsymbol{F}_1}_{\chi_u, \chi_a} \mid u \in \mathbb{F}_2^n\} \cup \{1\}$ and $\boldsymbol{\mathcal{Y}} = \{C^{\boldsymbol{F}_2}_{\chi_b, \chi_u} \mid u \in \mathbb{F}_2^n\} \cup \{1\}$. By the assumptions, $\boldsymbol{\mathcal{X}}$ is a $(2, q, \rho_1)$-hypercontractive orthogonal ensemble and $\boldsymbol{\mathcal{Y}}$ is $(2, q, \rho_2)$-hypercontractive. Note that (A.1) is a multilinear polynomial in $\boldsymbol{\mathcal{X}} \cup \boldsymbol{\mathcal{Y}}$ without a constant term. Hence, by Lemma A.4, the random variable (A.1) is $(2, q, \rho_1\rho_2)$-hypercontractive. □

Appendix B

Poster

This appendix includes the poster presented at the “masterproefbeurs” in April 2019. Its content is limited to Chapter 3.

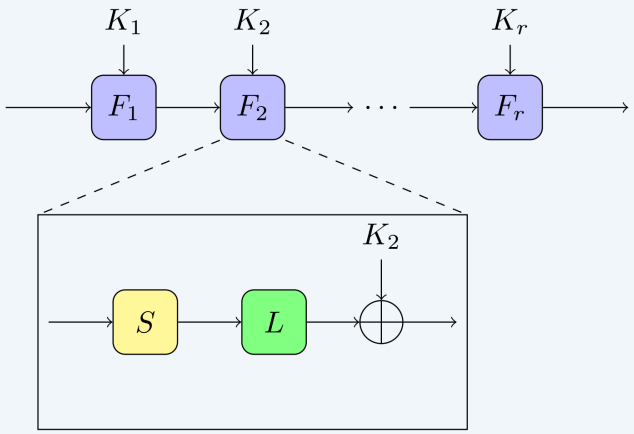
KU LEUVEN

Faculteit Ingenieurswetenschappen

Linear Cryptanalysis in the Weak-Key Setting

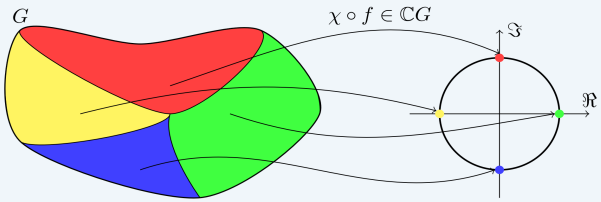
Iterated block ciphers & permutations

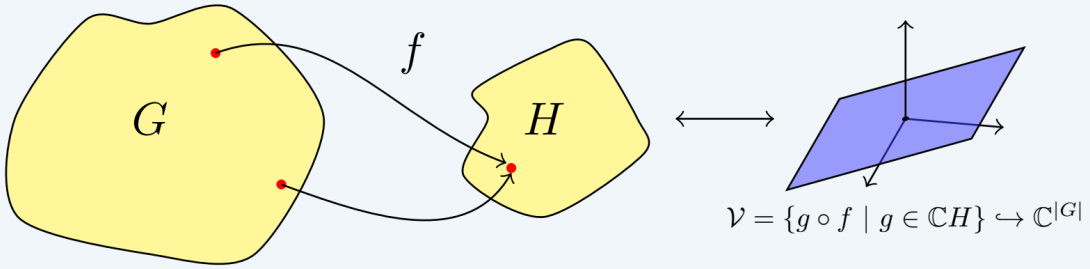
- Statistical cryptanalysis of primitives



Vector Space of Functions $G \to \mathbb{C}$

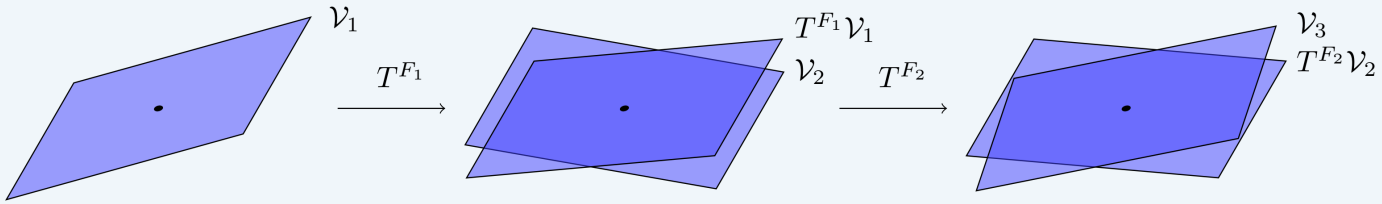
- Probability distributions
- Functions $f : G \to H$ (subspace of $\mathbb{C}G$)
- If G and H are abelian groups:
  role of Pontryagin dual $\hat{G}$, $\hat{H}$ (Fourier characters χ)





Geometric Interpretation of the Piling-Up Approximation

- Trail of vector spaces $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_r$
- Successive orthogonal projections
- Cosines of principal angles as a generalization of "correlation"



Conclusions

- Fixed-key understanding of linear cryptanalysis (in particular piling-up)
- Framework for nonlinear cryptanalysis generalizing and unifying several attacks
  i.a. resolution of open problems posed by Beierle et.al. (FSE 2019)
- Extends my work on block cipher invariants (ASIACRYPT 2018)
- Upper bounds on the correlation of linear approximations in the weak-key setting

Master in de ingenieurswetenschappen: wiskundige ingenieurstechnieken

Masterproef
Tim Beyne

Promotor
Prof. dr. ir. Vincent Rijmen

Academiejaar 2008-2009

COSIC

Bibliography

[1] Tomer Ashur. *Cryptanalysis of Symmetric-Key Primitives*. PhD thesis, Katholieke Universiteit Leuven, 2017.

[2] Tomer Ashur, Tim Beyne, and Vincent Rijmen. Revisiting the wrong-key-randomization hypothesis. Cryptology ePrint Archive, Report 2016/990, 2016. http://eprint.iacr.org/2016/990.

[3] Roberto Avanzi. The QARMA block cipher family. *IACR Trans. Symm. Cryptol.*, 2017(1):4–44, 2017.

[4] Thomas Baignères, Pascal Junod, and Serge Vaudenay. How far can we go beyond linear cryptanalysis? In Pil Joong Lee, editor, *ASIACRYPT 2004*, volume 3329 of *LNCS*, pages 432–450, Jeju Island, Korea, December 5–9, 2004. Springer, Heidelberg, Germany.

[5] Thomas Baignères, Jacques Stern, and Serge Vaudenay. Linear cryptanalysis of non binary ciphers. In Carlisle M. Adams, Ali Miri, and Michael J. Wiener, editors, *SAC 2007*, volume 4876 of *LNCS*, pages 184–211, Ottawa, Canada, August 16–17, 2007. Springer, Heidelberg, Germany.

[6] Subhadeep Banik, Andrey Bogdanov, Takanori Isobe, Kyoji Shibutani, Harunaga Hiwatari, Toru Akishita, and Francesco Regazzoni. Midori: A block cipher for low energy. In Tetsu Iwata and Jung Hee Cheon, editors, *ASIACRYPT 2015, Part II*, volume 9453 of *LNCS*, pages 411–436, Auckland, New Zealand, November 30 – December 3, 2015. Springer, Heidelberg, Germany.

[7] Subhadeep Banik, Sumit Kumar Pandey, Thomas Peyrin, Yu Sasaki, Siang Meng Sim, and Yosuke Todo. GIFT: A small present - towards reaching the limit of lightweight encryption. In Wieland Fischer and Naofumi Homma, editors, *CHES 2017*, volume 10529 of *LNCS*, pages 321–345, Taipei, Taiwan, September 25–28, 2017. Springer, Heidelberg, Germany.

[8] Lawrence Bassham, Çağdaş Çalık, Kerry McKay, and Meltem Sönmez Turan. NIST lightweight cryptography project. https://csrc.nist.gov/Projects/Lightweight-Cryptography.

[9] Christof Beierle, Anne Canteaut, and Gregor Leander. Nonlinear approximations in cryptanalysis revisited. *IACR Transactions on Symmetric Cryptology*, 2018(4):80–101, Dec. 2018.

[10] Christof Beierle, Jérémy Jean, Stefan Kölbl, Gregor Leander, Amir Moradi, Thomas Peyrin, Yu Sasaki, Pascal Sasdrich, and Siang Meng Sim. The SKINNY family of block ciphers and its low-latency variant MANTIS. In Matthew Robshaw and Jonathan Katz, editors, *CRYPTO 2016, Part II*, volume 9815 of *LNCS*, pages 123–153, Santa Barbara, CA, USA, August 14–18, 2016. Springer, Heidelberg, Germany.

[11] D. J. Bernstein. Competition for authenticated encryption: Security, applicability, and robustness. https://competitions.cr.yp.to/index.html.

[12] Daniel J. Bernstein and Tanja Lange. Non-uniform cracks in the concrete: The power of free precomputation. In Kazue Sako and Palash Sarkar, editors, *ASIACRYPT 2013, Part II*, volume 8270 of *LNCS*, pages 321–340, Bengalore, India, December 1–5, 2013. Springer, Heidelberg, Germany.

[13] Guido Bertoni, Joan Daemen, Michael Peeters, and Gilles Van Assche. On alignment in keccak. In *ECRYPT II Hash Workshop*, volume 51, page 122, 2011.

[14] Tim Beyne. Block cipher invariants as eigenvectors of correlation matrices. In Thomas Peyrin and Steven Galbraith, editors, *ASIACRYPT 2018, Part I*, volume 11272 of *LNCS*, pages 3–31, Brisbane, Queensland, Australia, December 2–6, 2018. Springer, Heidelberg, Germany.

[15] Tim Beyne and Begül Bilgin. Uniform first-order threshold implementations. In Roberto Avanzi and Howard M. Heys, editors, *SAC 2016*, volume 10532 of *LNCS*, pages 79–98, St. John's, NL, Canada, August 10–12, 2016. Springer, Heidelberg, Germany.

[16] Tim Beyne, Yu Long Chen, Christoph Dobraunig, and Bart Mennink. Dumbo, Jumbo, and Delirium: Parallel authenticated encryption for the lightweight circus. Submitted to ASIACRYPT, 2019.

[17] Tim Beyne, Yu Long Chen, Christoph Dobraunig, and Bart Mennink. Elephant v1. https://csrc.nist.gov/CSRC/media/Projects/Lightweight-Cryptography/documents/round-1/spec-doc/elephant-spec.pdf, 2019. Submission to the NIST lightweight cryptography project.

[18] Eli Biham. New types of cryptanalytic attacks using related keys. *Journal of Cryptology*, 7(4):229–246, December 1994.

[19] Eli Biham. On Matsui's linear cryptanalysis. In Alfredo De Santis, editor, *EUROCRYPT'94*, volume 950 of *LNCS*, pages 341–355, Perugia, Italy, May 9–12, 1995. Springer, Heidelberg, Germany.

[20] Eli Biham and Stav Perle. Conditional linear cryptanalysis – cryptanalysis of DES with less than 2^{42} complexity. *IACR Trans. Symm. Cryptol.*, 2018(3):215–264, 2018.

[21] Eli Biham and Adi Shamir. Differential cryptanalysis of DES-like cryptosystems. In Alfred J. Menezes and Scott A. Vanstone, editors, *CRYPTO'90*, volume 537 of *LNCS*, pages 2–21, Santa Barbara, CA, USA, August 11–15, 1991. Springer, Heidelberg, Germany.

[22] Eli Biham and Adi Shamir. Differential cryptanalysis of the full 16-round DES. In Ernest F. Brickell, editor, *CRYPTO'92*, volume 740 of *LNCS*, pages 487–496, Santa Barbara, CA, USA, August 16–20, 1993. Springer, Heidelberg, Germany.

[23] Alex Biryukov, Christophe De Cannière, and Michaël Quisquater. On multiple linear approximations. In Matthew Franklin, editor, *CRYPTO 2004*, volume 3152 of *LNCS*, pages 1–22, Santa Barbara, CA, USA, August 15–19, 2004. Springer, Heidelberg, Germany.

[24] Alex Biryukov and Leo Perrin. State of the art in lightweight symmetric cryptography. Cryptology ePrint Archive, Report 2017/511, 2017. http://eprint.iacr.org/2017/511.

[25] Céline Blondeau, Asli Bay, and Serge Vaudenay. Protecting against multidimensional linear and truncated differential cryptanalysis by decorrelation. In Gregor Leander, editor, *FSE 2015*, volume 9054 of *LNCS*, pages 73–91, Istanbul, Turkey, March 8–11, 2015. Springer, Heidelberg, Germany.

[26] Céline Blondeau and Kaisa Nyberg. Joint data and key distribution of simple, multiple, and multidimensional linear cryptanalysis test statistic and its impact to data complexity. *Designs, Codes and Cryptography*, 82(1):319–349, Jan 2017.

[27] Andrey Bogdanov, Miroslav Knežević, Gregor Leander, Deniz Toz, Kerem Varici, and Ingrid Verbauwhede. Spongent: A lightweight hash function. In Bart Preneel and Tsuyoshi Takagi, editors, *CHES 2011*, volume 6917 of *LNCS*, pages 312–325, Nara, Japan, September 28 – October 1, 2011. Springer, Heidelberg, Germany.

[28] Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J. B. Robshaw, Yannick Seurin, and C. Vikkelsoe. PRESENT: An ultra-lightweight block cipher. In Pascal Paillier and Ingrid Verbauwhede, editors, *CHES 2007*, volume 4727 of *LNCS*, pages 450–466, Vienna, Austria, September 10–13, 2007. Springer, Heidelberg, Germany.

[29] Andrey Bogdanov and Vincent Rijmen. Linear hulls with correlation zero and linear cryptanalysis of block ciphers. *Designs, Codes and Cryptography*, 70(3):369–383, Mar 2014.

[30] Andrey Bogdanov and Elmar Tischhauser. On the wrong key randomisation and key equivalence hypotheses in Matsui's algorithm 2. In Shiho Moriai, editor, *FSE 2013*, volume 8424 of *LNCS*, pages 19–38, Singapore, March 11–13, 2014. Springer, Heidelberg, Germany.

[31] Anne Canteaut, Eran Lambooij, Samuel Neves, Shahram Rasoolzadeh, Yu Sasaki, and Marc Stevens. Refined probability of differential characteristics including dependency between multiple rounds. *IACR Trans. Symm. Cryptol.*, 2017(2):203–227, 2017.

[32] Anne Canteaut and Joëlle Roué. On the behaviors of affine equivalent sboxes regarding differential and linear attacks. In Elisabeth Oswald and Marc Fischlin, editors, *EUROCRYPT 2015, Part I*, volume 9056 of *LNCS*, pages 45–74, Sofia, Bulgaria, April 26–30, 2015. Springer, Heidelberg, Germany.

[33] Claude Carlet, Pascale Charpin, and Victor Zinoviev. Codes, bent functions and permutations suitable for DES-like cryptosystems. *Designs, Codes and Cryptography*, 15(2):125–156, 1998.

[34] Tullio Ceccherini-Silberstein, Fabio Scarabotti, and Filippo Tolli. *Harmonic Analysis on Finite Groups*. Cambridge University Press, 2008.

[35] Dong Hyeon Cheon, Sangjin Lee, Jong In Lim, and Sung Jae Lee. New block cipher DONUT using pairwise perfect decorrelation. In Bimal K. Roy and Eiji Okamoto, editors, *INDOCRYPT 2000*, volume 1977 of *LNCS*, pages 262–270, Calcutta, India, December 10–13, 2000. Springer, Heidelberg, Germany.

[36] Baudoin Collard and François-Xavier Standaert. A statistical saturation attack against the block cipher PRESENT. In Marc Fischlin, editor, *CT-RSA 2009*, volume 5473 of *LNCS*, pages 195–210, San Francisco, CA, USA, April 20–24, 2009. Springer, Heidelberg, Germany.

[37] Baudoin Collard and François-Xavier Standaert. Multi-trail statistical saturation attacks. In Jianying Zhou and Moti Yung, editors, *ACNS 10*, volume 6123 of *LNCS*, pages 123–138, Beijing, China, June 22–25, 2010. Springer, Heidelberg, Germany.

[38] Baudoin Collard, François-Xavier Standaert, and Jean-Jacques Quisquater. Improving the time complexity of Matsui's linear cryptanalysis. In Kil-Hyun Nam and Gwangsoo Rhee, editors, *ICISC 07*, volume 4817 of *LNCS*, pages 77–88, Seoul, Korea, November 29–30, 2007. Springer, Heidelberg, Germany.

[39] Joan Daemen. *Cipher and hash function design*. PhD thesis, Katholieke Universiteit Leuven, 1995.

[40] Joan Daemen, René Govaerts, and Joos Vandewalle. Correlation matrices. In Bart Preneel, editor, *FSE'94*, volume 1008 of *LNCS*, pages 275–285, Leuven, Belgium, December 14–16, 1995. Springer, Heidelberg, Germany.

[41] Joan Daemen, Lars R. Knudsen, and Vincent Rijmen. The block cipher Square. In Eli Biham, editor, *FSE'97*, volume 1267 of *LNCS*, pages 149–165, Haifa, Israel, January 20–22, 1997. Springer, Heidelberg, Germany.

[42] Joan Daemen and Vincent Rijmen. AES proposal: Rijndael, 1999.

[43] Joan Daemen and Vincent Rijmen. The wide trail design strategy. In Bahram Honary, editor, *8th IMA International Conference on Cryptography and Coding*, volume 2260 of *LNCS*, pages 222–238, Cirencester, UK, December 17–19, 2001. Springer, Heidelberg, Germany.

[44] Joan Daemen and Vincent Rijmen. Plateau characteristics. *IET information security*, 1(1):11–17, 2007.

[45] Joan Daemen and Vincent Rijmen. Probability distributions of correlation and differentials in block ciphers. *J. Mathematical Cryptology*, 1(3):221–242, 2007.

[46] Joan Daemen and Vincent Rijmen. *The design of Rijndael: AES-the advanced encryption standard*. Springer Science & Business Media, 2013.

[47] Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988.

[48] W. Feller. *An Introduction to Probability Theory and Its Applicatons*, volume 2. John Wiley & Sons, 1971.

[49] Henri Gilbert, Marc Girault, Philippe Hoogvorst, Fabrice Noilhan, Thomas Pornin, Guillaume Poupard, Jacques Stern, and Serge Vaudenay. Decorrelated fast cipher: an aes candidate. In *Extended Abstract.) In Proceedings from the First Advanced Encryption Standard Candidate Conference, National Institute of Standards and Technology (NIST)*, 1998.

[50] Louis Granboulan, Éric Levieil, and Gilles Piret. Pseudorandom permutation families over Abelian groups. In Matthew J. B. Robshaw, editor, *FSE 2006*, volume 4047 of *LNCS*, pages 57–77, Graz, Austria, March 15–17, 2006. Springer, Heidelberg, Germany.

[51] Jian Guo, Jérémy Jean, Ivica Nikolic, Kexin Qiao, Yu Sasaki, and Siang Meng Sim. Invariant subspace attack against Midori64 and the resistance criteria for S-box designs. *IACR Trans. Symm. Cryptol.*, 2016(1):33–56, 2016. http://tosc.iacr.org/index.php/ToSC/article/view/534.

[52] Carlo Harpes, Gerhard G. Kramer, and James L. Massey. A generalization of linear cryptanalysis and the applicability of Matsui's piling-up lemma. In Louis C. Guillou and Jean-Jacques Quisquater, editors, *EUROCRYPT'95*, volume 921 of *LNCS*, pages 24–38, Saint-Malo, France, May 21–25, 1995. Springer, Heidelberg, Germany.

[53] Carlo Harpes and James L. Massey. Partitioning cryptanalysis. In Eli Biham, editor, *FSE'97*, volume 1267 of *LNCS*, pages 13–27, Haifa, Israel, January 20–22, 1997. Springer, Heidelberg, Germany.

[54] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional linear cryptanalysis of reduced round Serpent. In Yi Mu, Willy Susilo, and Jennifer Seberry, editors, *ACISP 08*, volume 5107 of *LNCS*, pages 203–215, Wollongong, Australia, July 7–9, 2008. Springer, Heidelberg, Germany.

[55] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional extension of Matsui's algorithm 2. In Orr Dunkelman, editor, *FSE 2009*, volume 5665 of *LNCS*, pages 209–227, Leuven, Belgium, February 22–25, 2009. Springer, Heidelberg, Germany.

[56] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pages 409–426. Springer, 1994.

[57] Seokhie Hong, Sangjin Lee, Jongin Lim, Jaechul Sung, Dong Hyeon Cheon, and Inho Cho. Provable security against differential and linear cryptanalysis for the SPN structure. In Bruce Schneier, editor, *FSE 2000*, volume 1978 of *LNCS*, pages 273–283, New York, NY, USA, April 10–12, 2001. Springer, Heidelberg, Germany.

[58] Pascal Junod. On the optimality of linear, differential, and sequential distinguishers. In Eli Biham, editor, *EUROCRYPT 2003*, volume 2656 of *LNCS*, pages 17–32, Warsaw, Poland, May 4–8, 2003. Springer, Heidelberg, Germany.

[59] Pascal Junod and Serge Vaudenay. Optimal key ranking procedures in a statistical cryptanalysis. In Thomas Johansson, editor, *FSE 2003*, volume 2887 of *LNCS*, pages 235–246, Lund, Sweden, February 24–26, 2003. Springer, Heidelberg, Germany.

[60] David Kahn. *The Codebreakers*. New York, McMillan, 1967.

[61] Burton S. Kaliski Jr. and Matthew J. B. Robshaw. Linear cryptanalysis using multiple approximations. In Yvo Desmedt, editor, *CRYPTO'94*, volume 839 of *LNCS*, pages 26–39, Santa Barbara, CA, USA, August 21–25, 1994. Springer, Heidelberg, Germany.

[62] Liam Keliher, Henk Meijer, and Stafford Tavares. Completion of computation of improved upper bound on the maximum average linear hull probability for Rijndael. Cryptology ePrint Archive, Report 2004/074, 2004. http://eprint.iacr.org/2004/074.

[63] Liam Keliher, Henk Meijer, and Stafford E. Tavares. Improving the upper bound on the maximum average linear hull probability for Rijndael. In Serge Vaudenay and Amr M. Youssef, editors, *SAC 2001*, volume 2259 of *LNCS*, pages 112–128, Toronto, Ontario, Canada, August 16–17, 2001. Springer, Heidelberg, Germany.

[64] Liam Keliher, Henk Meijer, and Stafford E. Tavares. New method for upper bounding the maximum average linear hull probability for SPNs. In Birgit Pfitzmann, editor, *EUROCRYPT 2001*, volume 2045 of *LNCS*, pages 420–436, Innsbruck, Austria, May 6–10, 2001. Springer, Heidelberg, Germany.

[65] Liam Keliher and Jiayuan Sui. Exact maximum expected differential and linear probability for 2-round advanced encryption standard (AES). Cryptology ePrint Archive, Report 2005/321, 2005. http://eprint.iacr.org/2005/321.

[66] Liam Keliher and Jiayuan Sui. Exact maximum expected differential and linear probability for two-round advanced encryption standard. *IET Information Security*, 1(2):53–57, 2007.

[67] Lars R. Knudsen and Vincent Rijmen. On the decorrelated fast cipher (dfc) and its theory. In Lars R. Knudsen, editor, *FSE'99*, volume 1636 of *LNCS*, pages 81–94, Rome, Italy, March 24–26, 1999. Springer, Heidelberg, Germany.

[68] Lars R. Knudsen and Vincent Rijmen. Known-key distinguishers for some block ciphers. In Kaoru Kurosawa, editor, *ASIACRYPT 2007*, volume 4833 of *LNCS*, pages 315–324, Kuching, Malaysia, December 2–6, 2007. Springer, Heidelberg, Germany.

[69] Lars R. Knudsen and Matthew J. B. Robshaw. Non-linear approximations in linear cryptanalysis. In Ueli M. Maurer, editor, *EUROCRYPT'96*, volume 1070 of *LNCS*, pages 224–236, Saragossa, Spain, May 12–16, 1996. Springer, Heidelberg, Germany.

[70] Lars R. Knudsen and David Wagner. Integral cryptanalysis. In Joan Daemen and Vincent Rijmen, editors, *FSE 2002*, volume 2365 of *LNCS*, pages 112–127, Leuven, Belgium, February 4–6, 2002. Springer, Heidelberg, Germany.

[71] Gregor Leander, Mohamed Ahmed Abdelraheem, Hoda AlKhzaimi, and Erik Zenner. A cryptanalysis of PRINTcipher: The invariant subspace attack. In Phillip Rogaway, editor, *CRYPTO 2011*, volume 6841 of *LNCS*, pages 206–221, Santa Barbara, CA, USA, August 14–18, 2011. Springer, Heidelberg, Germany.

[72] Yehuda Lindell and Jonathan Katz. *Introduction to modern cryptography*. Chapman and Hall/CRC, 2014.

[73] Yunwen Liu. *Techniques for Block Cipher Cryptanalysis*. PhD thesis, Katholieke Universiteit Leuven, 2018.

[74] Bao Luong. *Fourier Analysis on Finite Abelian Groups*. Springer Science & Business Media, 2009.

[75] Atul Luykx. *The Design and Analysis of Message Authentication and Authenticated Encryption Schemes*. PhD thesis, Katholieke Universiteit Leuven, 2016.

[76] Mitsuru Matsui. The first experimental cryptanalysis of the data encryption standard. In Yvo Desmedt, editor, *CRYPTO'94*, volume 839 of *LNCS*, pages 1–11, Santa Barbara, CA, USA, August 21–25, 1994. Springer, Heidelberg, Germany.

[77] Mitsuru Matsui. Linear cryptanalysis method for DES cipher. In Tor Helleseth, editor, *EUROCRYPT'93*, volume 765 of *LNCS*, pages 386–397, Lofthus, Norway, May 23–27, 1994. Springer, Heidelberg, Germany.

[78] Mitsuru Matsui and Atsuhiro Yamagishi. A new method for known plaintext attack of FEAL cipher. In Rainer A. Rueppel, editor, *EUROCRYPT'92*, volume 658 of *LNCS*, pages 81–91, Balatonfüred, Hungary, May 24–28, 1993. Springer, Heidelberg, Germany.

[79] Ueli M. Maurer, Renato Renner, and Clemens Holenstein. Indifferentiability, impossibility results on reductions, and applications to the random oracle methodology. In Moni Naor, editor, *TCC 2004*, volume 2951 of *LNCS*, pages 21–39, Cambridge, MA, USA, February 19–21, 2004. Springer, Heidelberg, Germany.

[80] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics. Second Series*, 171(1):295–341, 2010.

[81] Sean Murphy. The independence of linear approximations in symmetric cryptanalysis. *IEEE Transactions on Information Theory*, 52(12):5510–5518, 2006.

[82] Sean Murphy. The effectiveness of the linear hull effect. *J. Mathematical Cryptology*, 6(2):137–147, 2012.

[83] Mridul Nandi. The characterization of Luby-Rackoff and its optimum single-key variants. In Guang Gong and Kishan Chand Gupta, editors, *INDOCRYPT 2010*, volume 6498 of *LNCS*, pages 82–97, Hyderabad, India, December 12–15, 2010. Springer, Heidelberg, Germany.

[84] Kaisa Nyberg. Linear approximation of block ciphers (rump session). In Alfredo De Santis, editor, *EUROCRYPT'94*, volume 950 of *LNCS*, pages 439–444, Perugia, Italy, May 9–12, 1995. Springer, Heidelberg, Germany.

[85] Kaisa Nyberg. Affine linear cryptanalysis. *Cryptography and Communications*, pages 1–11, 2018.

[86] Ryan O'Donnell. Analysis of boolean functions, 2012.

[87] Sangwoo Park, Soo Hak Sung, Sangjin Lee, and Jongin Lim. Improving the upper bound on the maximum differential and the maximum linear Hull probability for SPN structures and AES. In Thomas Johansson, editor, *FSE 2003*, volume 2887 of *LNCS*, pages 247–260, Lund, Sweden, February 24–26, 2003. Springer, Heidelberg, Germany.

[88] Vincent Rijmen. *Cryptanalysis and design of iterated block ciphers*. PhD thesis, Katholieke Universiteit Leuven, 1997.

[89] Vincent Rijmen, Deniz Toz, and Kerem Varici. On the four-round AES characteristics. *Pre-proceedings of WCC*, pages 15–19, 2013.

[90] Ali Aydin Selçuk. On probability of success in linear and differential cryptanalysis. *Journal of Cryptology*, 21(1):131–147, January 2008.

[91] Ali Aydin Selçuk. On non-monotonicity of the success probability in linear cryptanalysis. *IACR Cryptology ePrint Archive*, 2018:478, 2018.

[92] Jean-Pierre Serre. *Linear representations of finite groups*, volume 42. Springer Science & Business Media, 2012.

[93] Anne Tardy-Corfdir and Henri Gilbert. A known plaintext attack of FEAL-4 and FEAL-6. In Joan Feigenbaum, editor, *CRYPTO'91*, volume 576 of *LNCS*, pages 172–181, Santa Barbara, CA, USA, August 11–15, 1992. Springer, Heidelberg, Germany.

[94] Yosuke Todo, Gregor Leander, and Yu Sasaki. Nonlinear invariant attack - practical attack on full SCREAM, iSCREAM, and Midori64. In Jung Hee Cheon and Tsuyoshi Takagi, editors, *ASIACRYPT 2016, Part II*, volume 10032 of *LNCS*, pages 3–33, Hanoi, Vietnam, December 4–8, 2016. Springer, Heidelberg, Germany.

[95] James Townsend, Niklas Koep, and Sebastian Weichwald. Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. *Journal of Machine Learning Research*, 17(137):1–5, 2016.

[96] Serge Vaudenay. An experiment on DES statistical cryptanalysis. In *ACM CCS 96*, pages 139–147, New Delhi, India, March 14–15, 1996. ACM Press.

[97] Serge Vaudenay. Adaptive-attack norm for decorrelation and super-pseudorandomness. In Howard M. Heys and Carlisle M. Adams, editors, *SAC 1999*, volume 1758 of *LNCS*, pages 49–61, Kingston, Ontario, Canada, August 9–10, 1999. Springer, Heidelberg, Germany.

[98] Serge Vaudenay. Feistel ciphers with L_2-decorrelation. In Stafford E. Tavares and Henk Meijer, editors, *SAC 1998*, volume 1556 of *LNCS*, pages 1–14, Kingston, Ontario, Canada, August 17–18, 1999. Springer, Heidelberg, Germany.

[99] Serge Vaudenay. Resistance against general iterated attacks. In Jacques Stern, editor, *EUROCRYPT'99*, volume 1592 of *LNCS*, pages 255–271, Prague, Czech Republic, May 2–6, 1999. Springer, Heidelberg, Germany.

[100] Serge Vaudenay. Decorrelation over infinite domains: The encrypted CBC-MAC case. In Douglas R. Stinson and Stafford E. Tavares, editors, *SAC 2000*, volume 2012 of *LNCS*, pages 189–201, Waterloo, Ontario, Canada, August 14–15, 2001. Springer, Heidelberg, Germany.

[101] Serge Vaudenay. Decorrelation: A theory for block cipher security. *Journal of Cryptology*, 16(4):249–286, September 2003.

List of Publications

1. Tim Beyne and Begül Bilgin. Uniform first-order threshold implementations. In Roberto Avanzi and Howard M. Heys, editors, *SAC 2016*, volume 10532 of *LNCS*, pages 79–98, St. John's, NL, Canada, August 10–12, 2016. Springer, Heidelberg, Germany

2. Tomer Ashur, Tim Beyne, and Vincent Rijmen. Revisiting the wrong-key-randomization hypothesis. Cryptology ePrint Archive, Report 2016/990, 2016. http://eprint.iacr.org/2016/990. Submitted to *Journal of Cryptology*

3. Tim Beyne. Block cipher invariants as eigenvectors of correlation matrices. In Thomas Peyrin and Steven Galbraith, editors, *ASIACRYPT 2018, Part I*, volume 11272 of *LNCS*, pages 3–31, Brisbane, Queensland, Australia, December 2–6, 2018. Springer, Heidelberg, Germany

4. Tim Beyne, Yu Long Chen, Christoph Dobraunig, and Bart Mennink. Elephant v1. https://csrc.nist.gov/CSRC/media/Projects/Lightweight-Cryptography/documents/round-1/spec-doc/elephant-spec.pdf, 2019. Submission to the NIST lightweight cryptography project

5. Tim Beyne, Yu Long Chen, Christoph Dobraunig, and Bart Mennink. Dumbo, Jumbo, and Delirium: Parallel authenticated encryption for the lightweight circus. Submitted to ASIACRYPT, 2019

Fiche masterproef

Student: Tim Beyne

Titel: Linear Cryptanalysis in the Weak Key Model

Nederlandse titel: Lineaire Cryptanalyse in het Zwakke Sleutel Model

UDC: 51-7

Korte inhoud:
De grondslagen van lineaire cryptanalyse worden herbekeken en veralgemeend in het licht van recente ontwikkelingen in de symmetrische-sleutel cryptografie. Afwegingen bij het ontwerp van *lichtgewicht cryptografische primitieven* hebben geleid tot nieuwe aanvallen zoals blokcijfer invarianten, en hebben de interesse voor klassieke problemen zoals het gebruik van niet-lineaire benaderingen in de cryptanalyse heropgewekt. Deze ontwikkelingen zijn intrinsiek verbonden met de beschrijving van lineaire cryptanalyse in het zwakke sleutel model. Bovendien wint permutatie-gebaseerde cryptografie – die gebruikmaakt van primitieven zonder sleutel – aan belang.

Als reactie op deze dwingende tendensen ontwikkelt de voorliggende thesis een doortastende veralgemening van lineaire cryptanalyse. De voorgestelde "meetkundige aanpak" laat een uniforme behandeling van een groot aantal varianten van de klassieke lineaire aanval toe en is geschikt voor gebruik in het sleutel-loze en zwakke sleutel model. Bovendien maakt het nieuwe raamwerk bijkomende uitbreidingen van lineaire cryptanalyse mogelijk en leidt het tot de oplossing van problemen gerelateerd aan het gebruik van niet-lineaire benaderingen. Ten slotte worden pogingen tot het bewijzen van veiligheid ten aanzien van lineaire cryptanalyse heroverwogen in de context van het zwakke sleutel model.

Thesis voorgedragen tot het behalen van de graad van Master of Science in de ingenieurswetenschappen: wiskundige ingenieurstechnieken

Promotor: Prof. dr. ir. V. Rijmen

Assessoren: Prof. dr. ir. L. De Lathauwer
Prof. dr. ir. F. Vercauteren

Begeleider: