Diss. ETH No. 13076

On the Validity of Certain Hypotheses Used in Linear Cryptanalysis

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Technical Sciences

presented by
ZSOLT KUKORELLY
dipl. Math. ETH
born April 23, 1970
citizen of Thônex GE

accepted on the recommendation of
Prof. Dr. James L. Massey, examiner
Prof. Dr. Ueli Maurer, co-examiner

1999

Acknowledgements

I wish to express my profound gratitude to Professor James L. Massey for supervising me as a doctoral student even though I asked him so as to say in the last minute. Jim taught me the beauty and efficiency of simplicity and precision. Not only was his door always open when I needed him, but he also succeeded in giving me the pleasant impression that I was the one who was explaining something to him and not the other way around.

I thank Professor Ueli Maurer for having accepted to act as the co-referee for this dissertation.

I am deeply indebted to Dr. Paul K. Wah, without whose advice to attend Jim's ADIT lecture I would probably never have come into contact with Information Theory.

It is much more easy to work intensively when the atmosphere is as pleasant and friendly as it was at ISI. Many thanks go for this to my friends and colleagues at the institute, especially to my last office-mate Jossy Sayir, who made me discover and appreciate good espressos and whose joy of living was only exceeded by his generosity, particularly at two in the morning; to Gerhard Krämer for his patience when listening to my entangled mathematical problems and for the varied discussions we had; to Beat Keusch and Richard De Moliner, who often stayed awake as long and for the same reason as I did; to 高群, who showed me how to drink green tea properly, taught me a little Chinese and told a lot about China; to Drahoslav Lím and Hanspeter Schmid for their support in LaTeX matters; to Ralf Kretzschmar, who helped making the abstract more understandable; to Dieter Arnold, Martin Hänggi, Markus Helfenstein, Markus Hofbauer, Felix Lustenberger, Pascal Vontobel and Sigi Wyrsch; and to the students whose projects I supervised, Stefan Moser and Kenneth De Spiegeleire, who forced me to explain the object of my research as clearly as possible.

Finally, I thank the members of my family for their support.

Abstract

Linear cryptanalysis and its generalisations are possible ways to attack an iterated block cipher. Their success relies on a certain number of assumptions made by the attacker. In this thesis, the validity of some of these assumptions is investigated.

According to Matsui's Piling-up Lemma, the imbalance of a sum of independent binary random variables is equal to the product of the imbalances of these random variables. One uses this fact in linear cryptanalysis to compute a lower bound on the probability of success of one's attack. It is shown that, on average, the imbalance of the sum is at least as large as the product of the imbalances and that for large sample spaces, both expressions are almost always approximately equal. It is deduced that, at least as an approximation, the Piling-up Lemma is applicable in a linear cryptanalysis attack to linked threefold sums even if they are not independent.

The validity of the hypothesis of fixed-key equivalence is investigated. The hypothesis asserts that for any effective input/output sum (I/O sum) virtually all key-dependent imbalances are approximately equal to their average, the average-key imbalance of the I/O sum. A counter-example is given. It is further proved that, for one round of encryption, the average and the variance of the key-dependent imbalances are approximately the same for virtually all I/O sums. Whether the key-dependent imbalances of an I/O sum can then be considered as "approximately equal" is subjective and therefore no conclusion about it is drawn. Finally, the average, over all I/O sums, of the average-key imbalances is computed for any number of rounds. Based on this result, a new quantitative definition of effective I/O sums is given.

The validity of the piling-up hypothesis is studied. This hypothesis is an m-ary analogue to Matsui's Piling-up Lemma. It says that (for certain imbalance measures) the imbalance of a product of independent m-ary random variables is in virtually all cases approximately equal to

the product of the imbalances of these random variables. The family of all imbalance measures that are convex-$\cup$ on the set of m-ary probability distributions, equal to 1 for a constant random variable and equal to 0 for a uniformly distributed random variable, is considered. It is argued that they are all equally appropriate for measuring the goodness of an expression used in the group generalisation of linear cryptanalysis attack. For the measure I_2^2, which belongs to this family, it is shown that the imbalance of a product of two random variables is on average equal to the product of the imbalances of the two random variables. It is inferred that the piling-up hypothesis holds for two random variables when m is large enough. By induction, it is shown that the hypothesis also holds for any number of random variables when m is large enough. Finally, it is argued that I_2^2 is an appropriate imbalance measure to use in the group generalisation of linear cryptanalysis.

Keywords. Iterated block cipher, linear cryptanalysis, imbalance, Piling-up Lemma, hypothesis of fixed-key equivalence, piling-up hypothesis.

Zusammenfassung

Die lineare Kryptoanalyse und ihre Verallgemeinerungen sind mögliche Methoden, um ein iteriertes Blockverschlüsselungsverfahren anzugreifen. Ihr Erfolg basiert auf einer Anzahl Annahmen, welche vom Angreifer (Kryptoanalysten) gemacht werden. In dieser Doktorarbeit wird die Gültigkeit einiger dieser Annahmen untersucht.

Nach Matsuis Auftürmlemma (Piling-up Lemma) ist die Unausgeglichenheit einer Summe von binären, unabhängigen Zufallsvariablen gleich dem Produkt der einzelnen Unausgeglichenheiten dieser Zufallsvariablen. Diese Tatsache kann in der linearen Kryptoanalyse benutzt werden, um eine untere Schranke der Erfolgswahrscheinlichkeit eines Angriffs zu berechnen. Für allgemeine Zufallsvariablen wird gezeigt, dass im Durchschnitt die Unausgeglichenheit der Summe mindestens so gross ist wie das Produkt der Unausgeglichenheiten, sowie dass für grosse Ergebnisräume beide Ausdrücke in der Regel gut übereinstimmen. Daraus wird geschlossen, dass in einem Angriff mit linearer Kryptoanalyse das Auftürmlemma auf zusammengekettete Dreifachsummen als Approximation anwendbar ist, selbst wenn diese Dreifachsummen nicht unabhängig sind.

Die Gültigkeit der Hypothese der Gleichartigkeit fester Schlüssel (hypothesis of fixed-key equivalence) wird erforscht. Die Hypothese behauptet, dass für jede wirksame Eingangs-/Ausgangssumme (E/A-Summe) die schlüsselabhängigen Unausgeglichenheiten fast alle ungefähr gleich ihrem Durchschnitt, der mittleren Unausgeglichenheit der E/A-Summe sind. Dazu wird ein Gegenbeispiel gegeben. Des weiteren wird gezeigt, dass für eine Verschlüsselungsrunde der Durchschnitt und die Varianz der schlüsselabhängigen Unausgeglichenheiten für fast alle E/A Summen beinahe gleich sind. Ob man die schlüsselabhängigen Unausgeglichenheiten einer E/A Summe dann als "ungefähr gleich" betrachten darf, ist subjektiv; aus diesem Grund wird auf einen solchen Schluss verzichtet. Schliesslich wird der Durchschnitt der mittleren Unausgeglichenheiten über alle E/A-Summen für eine beliebige Anzahl

von Runden berechnet. Beruhend auf diesem Resultat wird eine neue quantitative Definition einer wirksamen E/A-Summe gegeben.

Die Gültigkeit der Auftürmhypothese (piling-up hypothesis) wird studiert. Diese Hypothese ist ein Analogon zu Matsuis Auftürmlemma im m-wertigen Fall. Sie besagt, dass (für gewisse Unausgeglichenheitsmaße) die Unausgeglichenheit eines Produktes von m-wertigen, unabhängigen Zufallsvariablen in praktisch allen Fällen ungefähr gleich dem Produkt der Unausgeglichenheiten dieser Zufallsvariablen ist. Die Familie der Unausgeglichenheitsmaße, welche auf der Menge der m-wertigen Wahrscheinlichkeitsverteilungen konvex sind, sowie für konstante Zufallsvariablen den Wert 1 und für gleichverteilte Zufallsvariablen den Wert 0 haben, wird betrachtet. Es wird behauptet, dass sie alle für die Messung der Güte eines in einem Angriff mittels der Gruppenverallgemeinerung der linearen Kryptoanalyse gebrauchten Ausdrucks gleich gut sind. Für das Maß I_2^2 aus dieser Familie wird gezeigt, dass die Unausgeglichenheit des Produktes zweier Zufallsvariablen im Durchschnitt gleich dem Produkt der Unausgeglichenheiten beider Zufallsvariablen ist. Es wird bewiesen, dass die Auftürmhypothese für zwei Zufallsvariablen gültig ist, falls m groß genug ist. Mittels vollständiger Induktion wird dann gezeigt, dass bei genug großem m die Hypothese auch für eine beliebige Anzahl von Zufallsvariablen gültig ist. Schließlich wird dargelegt, dass I_2^2 in einem Angriff mittels Gruppenverallgemeinerung der linearen Kryptoanalyse ein geeignetes Unausgeglichenheitsmaß ist.

Schlüsselwörter. Iterierte Blockverschlüsselungsverfahren, lineare Kryptoanalyse, Unausgeglichenheit, Auftürmlemma, Hypothese der Gleichartigkeit fester Schlüssel, Auftürmhypothese.

Contents

Chapter 1

Introduction

1.1 'What is Cryptology?'

In the last few years, I have often been asked what I was researching. As I answered 'cryptology', most people's reaction was 'Cryptology? What is cryptology'? I was asked more than once whether my activity had something to do with the study of tombs and graves or of ancient languages. As I tried to explain that cryptology was 'the theory of secret codes', I often received the reply 'So you are a kind of spy'! A minority of my interlocutors could imagine that some 'large companies' might 'code' communications with their subsidiaries or protect their local networks with passwords; but virtually no one was aware that such coding was relevant to, or important for, the man in the street.

1.2 Terminology

The word "cryptology" comes from the ancient greek words "$\kappa\rho\acute{\upsilon}\pi\tau\varepsilon\iota\nu$", meaning "to hide", and "$\lambda\acute{o}\gamma o\varsigma$", meaning "word", "reason", or "explanation". Hence, cryptology is the teaching, or the theory, of hiding.

Recently, Simmons has characterized cryptology as the "science of information integrity" [50]. With time, cryptology has become a scientific field of its own and can be divided roughly into two closely related subfields: *cryptography* and *cryptanalysis*. The former is concerned with production or construction ("$\gamma\rho\acute{\alpha}\varphi\varepsilon\iota\nu$" means to write) and the latter one with dismantling. More precisely, cryptography is the design or implementation of techniques intended to aid the purposes of the *cryptographer*, namely achieving *secrecy* and/or *authenticity*. A technique provides secrecy if it

determines who can receive a message; it provides authenticity if it determines who can have sent a message [32]. In contrast to what one might assume, secrecy and authenticity are independent in the sense that a system that ensures the one property might not ensure the other. In the case of secrecy, there is usually a message, called the *plaintext* or *cleartext*, that one wants to be read only by a restricted group of individuals. To attain this goal, one *encrypts* the plaintext and obtains the *ciphertext*. This is done by means of a *cipher* (also called a *cryptographic system* or *cryptosystem*). A *secret key* controls the encryption to ensure that only the authorized people can *decrypt* the ciphertext and regain the original message.

Cryptanalysis is performed by *cryptanalysts* who try to thwart the cryptographer's plans, e.g., to read a message not intended for them, to find the secret key, or to impersonate someone else. One says that they launch an *attack* on the algorithm. When they succeed to an extent that they consider as being enough, one says that they have *broken* the system. (One can distinguish several stages in breaking a system [24], but we shall not struggle with these subtleties.)

Cryptosystems are divided into *secret-key* systems, also called *symmetric* systems, and *public-key* or *asymmetric* systems. In the former, the same secret-key is used both for encryption and decryption; in the latter, encryption and decryption are performed with two different keys, the *public key* and the *private key*, respectively. As the name suggests, only the private key need be kept secret by the decrypter; this is enough if the goal is to prevent anyone else from reading the message. The idea of public-key encryption is due to Diffie and Hellman in 1976 [12].

Cryptographic techniques have existed for a long time. In ancient times when one wanted to send a secret message to another party, one sometimes had it written on a slave's head, waited until enough hair had grown, sent the slave on his way, and, upon his arrival, had his hair cut. Of course, the message was not urgent. More technical methods were used, too. Caesar shifted all letters of the alphabet by three positions to encrypt his text, that is, he *permuted* the letters of the alphabet. During the following centuries, more and more complicated permutations were used; permutations are still the building blocks of many cryptographic techniques. The 19^{th} century saw the advent of mechanical encryption methods.

Until 1949, cryptology was mainly an empirical field. In this year Shannon's celebrated paper [49] made cryptology a science and not just an art. Shannon presented the model of a secret-key cryptosystem shown in Figure 1.1. In this model, the secret key Z is passed to both the en-

crypter and the decrypter. This key is used to encrypt the plaintext X so as to obtain the ciphertext Y, and then used to decrypt Y back to recover X. The ciphertext Y is transmitted over an insecure channel, i.e., anyone can intercept it. All three quantities X, Y and Z are modeled as random variables and X is considered to be statistically independent of Z. This is the model we will use. In his paper, Shannon described two basic principles on which practical ciphers should be built: *diffusion* and *confusion*. The principle of diffusion stipulates that the statistical structure of the plaintext should be "dissipated" into the statistical structure involving long combinations of digits in the ciphertext; this can be achieved when each digit of the plaintext and/or of the key influences many digits in the ciphertext. The principle of confusion states that "the relation between the simple statistics of Y and the simple description of Z be very complex", i.e., it must be complicated to tell from the ciphertext which key could have been used. These are still today the principles that prevail in the design of encryption algorithms.

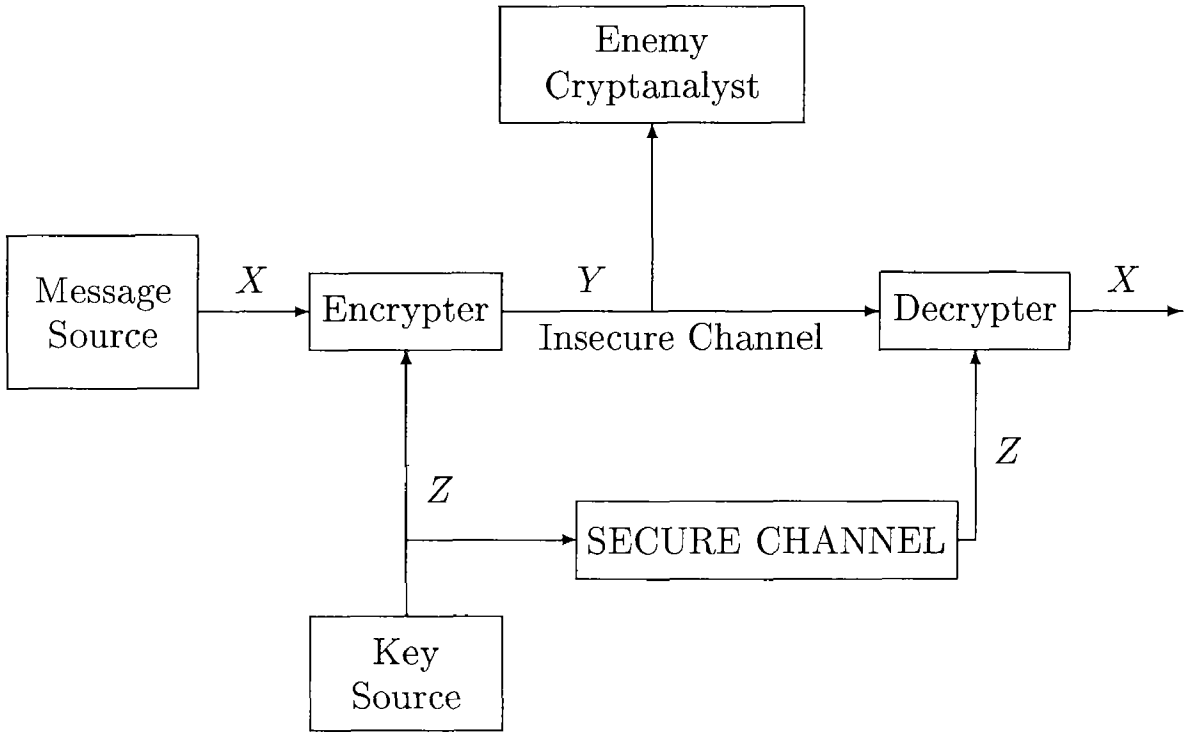


Figure 1.1: Shannon's Model of a Secret-Key Cryptosystem.

In [49], Shannon also gave a meaning to the concept of security. He defined a cryptosystem to be *theoretically secure* (in today's terminology: *unconditionally secure*) if the system is immune against a cryptanalyst who "has unlimited time and manpower available for the analysis" and

who knows only the ciphertext. A system is *practically secure* (today: *computationally secure*) against a cryptanalyst if the amount of time and/or computational power necessary to break the system exceeds the attacker's capabilities. Shannon showed that theoretically secure systems exist but that they require that the key be at least as long as the plaintext that it is used to encrypt.

An important principle in cipher design is *Kerckhoffs' assumption.* Kerckhoffs stated that users of a cryptosystem should always assume that the cryptanalyst knows everything about the process of encryption and decryption except the value of the secret key. This includes complete knowledge of the encryption algorithm. This assumption makes practical sense because everything in a practical cryptosystem except the key must usually remain fixed for a long time so that the cryptanalyst may well learn them by non-technical methods. Also, if one keeps one's algorithm secret, then one may never know whether it is secure or not. A published algorithm that has been analysed for years by many people and in which no one claims to have found a serious flaw is much more likely to be secure than an algorithm kept secret. Kerckhoffs' assumption is accepted by almost all cipher designers.

Secret-key ciphers are divided into two families: *stream ciphers* and *block ciphers.* In a stream cipher, the encryption process has internal memory whereas this is not the case in a block cipher. In a stream cipher, the plaintext is divided into small fragments, often one bit long, and each fragment is processed in a way that depends on the key and on the *cipher state.* In a block cipher, the plaintext is divided into large blocks (typically of 64 or 128 bits) and each block is encrypted in the same manner. This implies that two identical plaintext blocks yield two identical ciphertext blocks, i.e., patterns of the plaintext leak through. If the plaintext consists of text, this does not matter; it does matter, however, if it is an image. There are ways to prevent this [11]. With a stream cipher, this phenomenon does not occur in general.

1.3 Outline

This thesis treats of an attack on block ciphers called *linear cryptanalysis* and of some of its generalisations. In particular it examines the validity of three assumptions made by the cryptanalyst when applying different variants of linear cryptanalysis. These assumptions are in fact approximations. In each of these, the cryptanalyst approximates an expression that he would like to but is unable to compute, with another expression that he can calculate. It was not known previously whether the approximations

are allowable, but this ignorance has not prevented attackers from making these assumptions.

Chapter 2 is an introduction to iterated block ciphers and to the linear cryptanalysis attack. We begin by describing iterated block ciphers and mention possible ways to attack them. Then we explain the main idea behind linear cryptanalysis and how the attack itself is performed. We explain the problems in the calculation of the probability of success of the attack and a possible solution. We state Matsui's Piling-up Lemma. Finally, we introduce the imbalance of a binary-valued function and pave the way for Chapters 3 and 4.

In Chapter 3, we consider the application of the Piling-up Lemma to dependent random variables. We first prove two identites involving the imbalance of a sum *modulo* 2 of binary-valued random variables and the product of the imbalances of these random variables. The first identity is an inequality that narrows the set of values that both expressions can take on; it also shows that in particular cases, the imbalance of the sum can differ considerably from the product of the imbalances. The second identity is a generalisation of the Piling-up Lemma; unfortunately, it is not applicable in practical cases. In the remainder of the chapter, we show that, on average, the imbalance of the sum is at least as large as the product of the imbalances. We begin with two random variables and proceed to an arbitrary number of random variables. We show that, if the sample space on which the random variables are defined is large, then the imbalance of a sum of random variables is in virtually all cases approximately equal to the product of the imbalance of the random variables that compose the sum. We also show for any sample space on which the random variables are defined that, when enough random variables are involved, the imbalance of the sum is in virtually all cases larger than the product of the imbalances. We conclude that in linear cryptanalysis one can use an approximate version of the Piling-up Lemma for linked threefold sums even if they are not independent.

Chapter 4 is concerned with the hypothesis of fixed-key equivalence. We begin by recalling some definitions, after which we reformulate the hypothesis based on our newly defined *fixed-key equivalence condition.* We introduce a measure for the validity of the fixed-key equivalence condition – the larger the measure, the less the condition is satisfied. The measure is in fact the variance of the key-dependent imbalances. We continue with some algebraic considerations upon which we base the study of the hypothesis for one round of encryption. We fix two balanced functions and compute both the average and the variance of both the average-key imbalance and the validity measure over all one-round I/O sums that are

defined by some round function and by the two fixed balanced functions. We do this first as an example for ciphers where the text blocks and round keys have length three and then generalise to any length. We show that the moments calculated do not depend on the balanced functions chosen. From the results, we conclude that the average-key imbalance and the validity measure are approximately the same for almost all I/O sums. This has consequences on the distribution of the key-dependent imbalances of the I/O sums – we leave it to the reader to decide whether they can be considered as "almost equal". Finally, we take a brief look at the multi-round case. We show that the above average of the average-key imbalance does not depend on the number of rounds. This allows us to make precise the notion of an effective I/O sum.

The piling-up hypothesis is the topic of the last main chapter, Chapter 5. The piling-up hypothesis states that there is an m-ary analogue to the Piling-up Lemma, but with an approximation sign instead of an equality. Also, the sum modulo 2 is replaced by a group operation. We show that the piling-up hypothesis holds for a certain imbalance measure if m is large enough. We first recall the group generalisation of linear cryptanalysis and some possible m-ary imbalance measures. We choose the imbalance measure I_2^2 and examine the validity of the piling-up hypothesis with respect to I_2^2 for two independent random variables. We show that on average the imbalance of the product is equal to the product of the imbalances. From the properties of the variance, we deduce that the imbalance of the product is close to the product of the imbalances in virtually all cases if m is large enough. This means that the piling-up hypothesis holds for two random variables. Then it holds by induction for any number of random variables; one must only adapt the meaning of the approximation sign. Because of this and because it is relatively easy to compute, we assert that I_2^2 is the "right" imbalance measure to use in the group generalisation of linear cryptanalysis.

Chapter 2

The Linear Cryptanalysis Attack

In this chapter, we describe iterated block ciphers and some attacks on them. One of the attacks, linear cryptanalysis, is presented in detail. We mention two hypotheses made by the cryptanalyst when he uses this attack and recall Matsui's Piling-up Lemma, which is essential in computing the probability of success of the attack.

2.1 Iterated Block Ciphers

Definition 2.1.1
In an *iterated block cipher*, the plaintext is first divided into blocks of equal size and then each block is encrypted separately, but with the same key (see Figure 2.1). The plaintext block, denoted by X, undergoes a certain number of similar transformations called *rounds*. In round i, the input $Y(i-1)$ is modified by means of a function g, the *round function*, which is also a function of the *round key* Z_i, to produce the round output $Y(i)$. Thus, $Y(i) = g(Y(i-1), Z_i)$. The round keys $Z_1, \ldots, Z_r$ are produced from a *master key* Z by some *key schedule algorithm*. The number of rounds is denoted by r. After r rounds, one obtains the ciphertext Y.

The round function is chosen such that the function $g_z = g(\cdot, z)$ is invertible for every value z of the second argument. This is necessary to be able to decrypt the ciphertext and recover the plaintext. Decryption is made in the direction opposite to encryption, i.e., from right to left in Figure 2.1, by using the same round keys and the inverse of g.

Most practical ciphers are iterated ciphers. Examples are DES [11], IDEA [29, 36], the different variants of SAFER [33, 34, 35], FEAL [41, 42, 43], the various LOKI ciphers [8, 9], BlowFish [48], or the RC5 family [46].

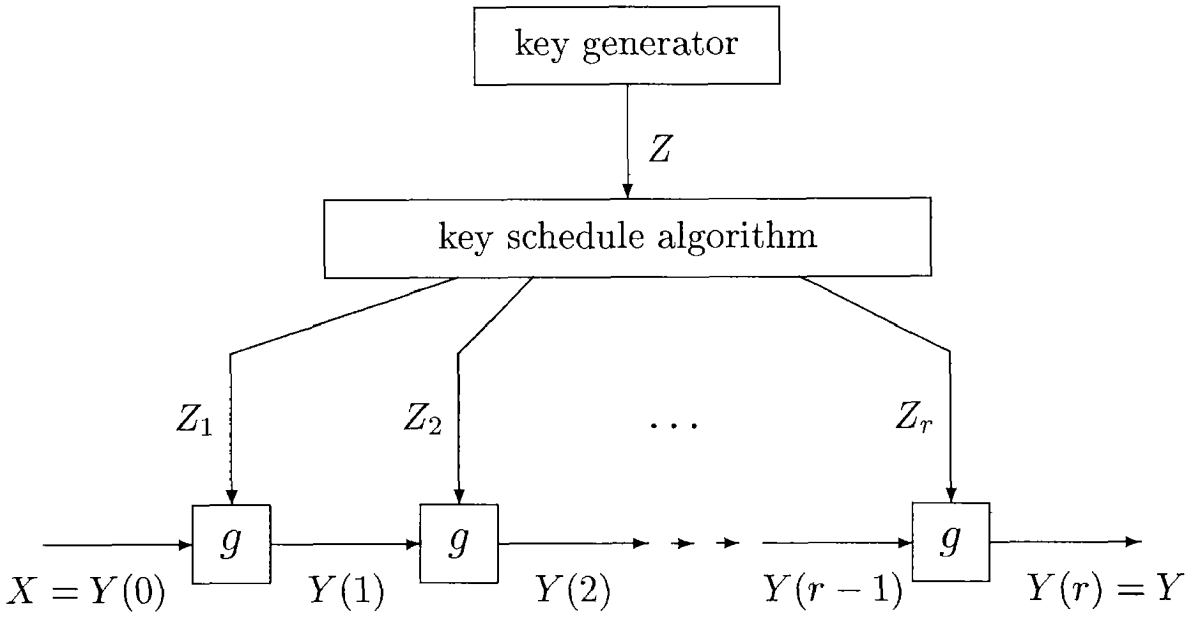


Figure 2.1: An Iterated Block Cipher.

Hereafter, by "block cipher", we mean an iterated block cipher. We denote the length of the text blocks (the cipher's *blocklength*) by n and the length of the round keys by k.

2.2 Attacks on an Iterated Block Cipher

2.2.1 Kinds of Attacks

Attacks on a cipher (and not only a block cipher) are generally divided into the following categories:

- **Ciphertext-only attack:** The cryptanalyst is assumed to have access only to a certain number of ciphertext blocks, which are all the result of an encryption using the same key. The goal is to use this knowledge to find 1) the corresponding plaintexts, or 2) (parts of) the secret key, or 3) a way to decrypt other messages encrypted with the same key. The goodness of the attack is measured in part by the number of ciphertext blocks necessary to reach a certain probability that the output of the attack is correct. (This probability increases with the number of ciphertext blocks.)

- **Known-plaintext attack:** The cryptanalyst is assumed to know a certain number of pairs (X, Y), where X is a plaintext block which, when encrypted with the actual secret key, yields the ciphertext block Y. The goal of the attack is to find (parts of) the secret key or a way to decrypt other messages encrypted with the same key. The goodness of the attack is measured in part by the number of pairs necessary to reach a certain probability that the output of the attack is correct.
- **Chosen-plaintext attack:** The cryptanalyst is assumed not only to have access to plaintext/ciphertext pairs, but also to be able to choose a certain number of plaintexts, to have them encrypted with the actual secret key and to get the corresponding ciphertexts. The goal is the same as in a known-plaintext attack and the goodness is measured in part by the number of plaintexts needed.
- **Other technical attacks:** These attacks are more seldom. Examples are the *chosen-ciphertext attack* or the *chosen-key attack* [48].
- **Non-technical attacks:** Blackmail, torture, theft, and thelike. These methods can be easier, cheaper, and faster than technical attacks.

2.2.2 Examples

Before we explain the linear cryptanalysis attack in the next section, we give here some examples of technical attacks on a block cipher.

An obvious attack is *exhaustive key search.* It consists in taking a few ciphertexts and trying out all possible keys until one has found the right one. That the key found is the right one is verified by decrypting the ciphertexts and determining whether the plaintexts obtained are valid, e.g., whether they are sequences of ASCII-characters that make sense in some language. Usually, two or three decrypted ciphertext blocks are enough to be sure that the key found is the right one [32]. Exhaustive key search can be a ciphertext-only attack. It can be also used as a known-plaintext attack if one knows some plaintext/ciphertext pairs. In that case, there is no need to test whether the plaintext is valid. On average, half of the possible values of the key have to be tested. This attack is how, in a recent contest, DES was broken in 22 hours 15 minutes [47]. This attack is impractical on most modern systems since the number of possible values of the key is too large.

Another often implemented attack is *differential cryptanalysis.* This chosen-plaintext attack, introduced in 1990 by Biham and Shamir [3], works as follows: one chooses at random a certain number N of plaintexts; then, for each plaintext X, one chooses another plaintext block

X^* such that the two plaintexts have a fixed difference $\Delta X = \alpha$. For each pair of plaintexts, the difference is the same. (The definition of difference can vary from cipher to cipher. For instance, in DES, it is the bitwise addition mod 2.) The $2N$ plaintexts are encrypted with the same key. Each pair produces a sequence of differences $\Delta X = \Delta Y(0), \Delta Y(1), \ldots, \Delta Y(r) = \Delta Y$. Differential cryptanalysis uses the fact that the round function is usually *cryptographically weak*, i.e., if one knows the triple $(\Delta Y(i-1), Y(i), Y^*(i))$ for a few pairs of plaintexts (X, X^*), then it is feasible to determine (part of) the subkey Z_i. This is achieved by *choosing* a pair (X, X^*) with a specified difference α such that $\Delta Y(r-1)$ takes a particular value β with high probability.

Differential cryptanalysis has been tried out on many ciphers. Lai, Massey and Murphy [36] identified a class of ciphers called Markov ciphers and showed that a Markov cipher is immune against differential cryptanalysis if, for all possible initial differences α between the plaintexts, the $(r-1)$-round differentials $\Delta Y(r-1)$ are virtually equally likely. FEAL, BlowFish, DES, LOKI, IDEA and SAFER are Markov ciphers. In their attack on DES with this method [5], Biham and Shamir needed 2^{47} chosen plaintexts in order for the probability that the found key is the right one to be acceptably large. FEAL-4 was broken with only 20 plaintexts by Murphy [44]. (This was actually the first published successful differential cryptanalysis attack on a real cipher.) Later, Biham and Shamir broke FEAL-4 with 8 plaintexts and FEAL-8 with 2000 plaintexts [4]. LOKI91 [8] was found secure against differential cryptanalysis by Knudsen [23]. According to Lai, IDEA is secure against differential cryptanalysis already after four rounds [29]. Massey showed [34] that the amount of work necessary to break SAFER K-64 with differential cryptanalysis is at least as large as exhaustive search for six or more rounds; for SAFER K-128, this probably happens for ten or more rounds. (This is because the key in SAFER K-128 is twice as long as in SAFER K-64, and does not mean that SAFER K-128 is less secure than SAFER K-64.)

Linear cryptanalysis is another attack applicable to any block cipher. We will present it more extensively in the next section.

In the last years, cryptanalysts have developed other attacks deriving from differential and linear cryptanalysis or generalising them, like:

- *truncated differential attack* [6, 25];
- *differential-linear cryptanalysis* [6, 18, 30]; this attack, which combines differential and linear cryptanalysis, seems very promising; it broke DES reduced to 8 rounds with only 768 chosen plaintext blocks [30]; it is to date the only attack that can break IDEA reduced to 4 rounds for an

arbitrary key [18]; however, it has not yet been applied successfully on a practical cipher with its full number of rounds.

- *linear cryptanalysis using multiple approximations* [21, 22];
- *linear cryptanalysis using non-linear approximations* [27];
- *binary generalisation of linear cryptanalysis* [16, 17];
- *group generalisation of linear cryptanalysis* [16];
- *partitioning cryptanalysis* [16];
- *correlation cryptanalysis* [19].

Also, it has been found that linear cryptanalysis and differential cryptanalysis are related [1, 10, 37].

2.3 Binary Generalisation of Linear Cryptanalysis

2.3.1 Philosophy Behind The Attack

Linear cryptanalysis is a known-plaintext attack that was first applied to FEAL [40] but first became known by its current name after Matsui's attack on DES [38]. In this section, we describe its binary generalisation, due to Harpes, Kramer and Massey [17]. The attack is based on the following construction (see Figure 2.2):

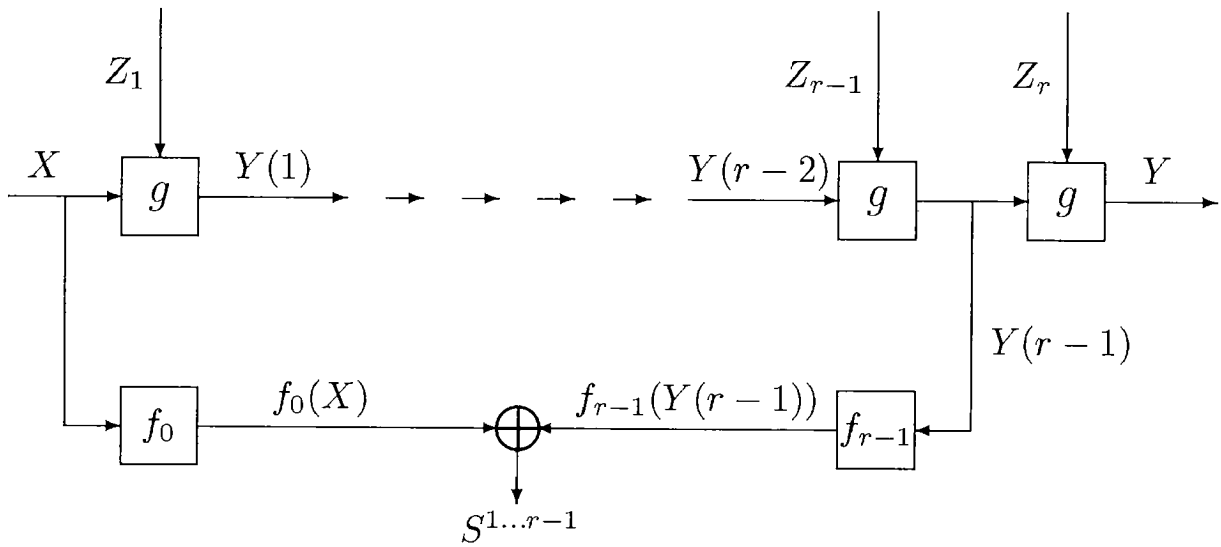


Figure 2.2: The Idea Behind Linear Cryptanalysis.

One applies a binary-valued function f_0 to the plaintext and a binary-valued function f_{r-1} to the output of the second last round; the mod 2 sum of the images is called an $(r-1)$-*round input/ouput sum (I/O sum)* and is denoted by $S^{1\ldots r-1}$. The cryptanalyst chooses f_0 and f_{r-1} to have the following properties:

- both functions are *balanced*, i.e., they take on each of the values 0 and 1 for half of their arguments;
- the expression $\left|P[S^{1\ldots r-1} = 0 | Z_1 = z_1, \ldots, Z_{r-1} = z_{r-1}] - \frac{1}{2}\right|$ is large for almost all values $z_1, \ldots, z_{r-1}$ of the round keys.

The cryptanalyst does not know the plaintext so he assumes that all possible plaintexts are equally likely. He models this with a uniformly distributed random variable. This is usual in cryptanalysis. Then, because for fixed round keys all round transformations are invertible, the possible values of $Y(r-1)$ are equally likely. If X and $Y(r-1)$ were independent random variables, then $f_0(X)$ and $f_{r-1}(Y(r-1))$ would be independent and equally likely to be zero or one. Then also $S^{1\ldots r-1}$ would be equally likely to take on the values 0 and 1 and one would have $\left|P[S^{1\ldots r-1} = 0 | Z_1 = z_1, \ldots, Z_{r-1} = z_{r-1}] - \frac{1}{2}\right| = 0$ for all balanced functions f_0 and f_{r-1}. But in a every cipher, X and $Y(r-1)$, when conditioned on $z_1, \ldots, z_{r-1}$, are not independent and there are balanced functions f_0 and f_{r-1} such that $\left|P[S^{1\ldots r-1} = 0 | Z_1 = z_1, \ldots, Z_{r-1} = z_{r-1}] - \frac{1}{2}\right| > 0$. This property can be exploited. The goal of the cryptanalyst is to find f_0 and f_{r-1} such that $\left|P[S^{1\ldots r-1} = 0 | Z_1 = z_1, \ldots, Z_{r-1} = z_{r-1}] - \frac{1}{2}\right|$ is as large as possible for as many round keys as possible.

Now we have seen the idea upon which linear cryptanalysis is based, we explain how it is realized.

2.3.2 How The Attack Works

The attack is made under the following assumptions:

- X is a uniformly distributed random variable on the set of all possible plaintexts;
- the round keys $Z_1, \ldots, Z_r$ are independent and uniformly distributed; (this is not actually true but is a good approximation to reality if the key schedule algorithm is good.)
- the cryptanalyst knows N plaintext/ciphertext pairs (*p/c pairs*) (X, Y) where the same key has been used.

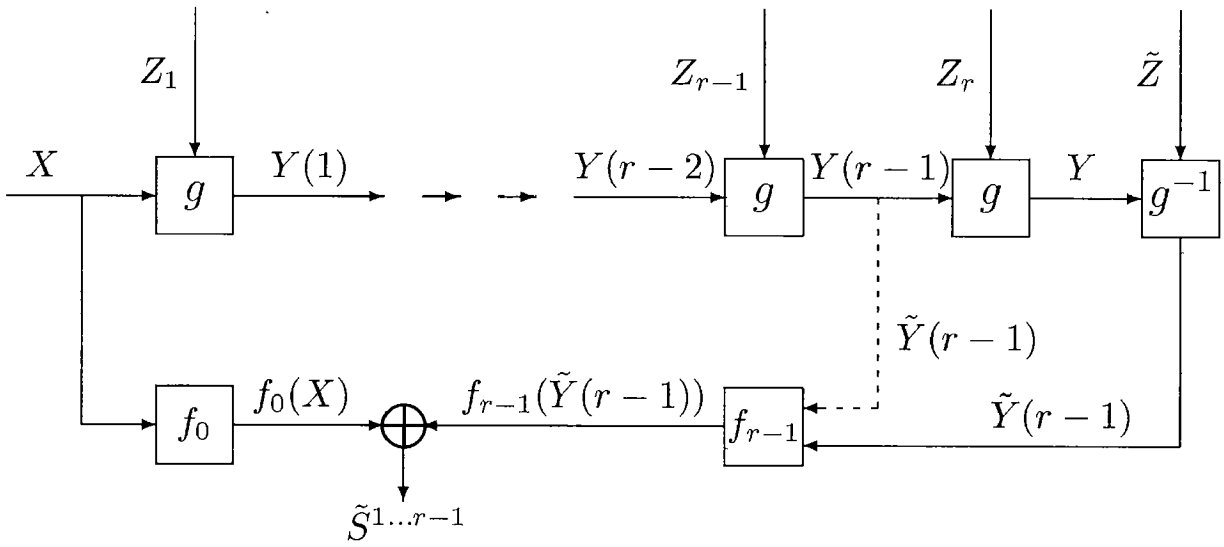


Figure 2.3: Principle of the Linear Cryptanalysis Attack.

The goal of the attack is to find as many bits as possible of the last round key. (There are refinements [39], but we will not discuss them in this thesis.) Basically, the attack itself goes as follows (see Figure 2.3):

```
For all possible estimates z̃ of the last round key do
  Set c(z̃) = 0;
  For all p/c pairs (X,Y) do
    Decrypt one round of the ciphertext by means of the key
    estimate z̃;
    Call the result Ỹ(r − 1);
    Set S̃^{1...r−1} := f_0(X) ⊕ f_{r−1}(Ỹ(r − 1));
    If S̃^{1...r−1} = 0, then increase c(z̃) by one
  end
end;
Output as estimate(s) of z_r the key(s) for which c(z̃) is
farthest from N/2, that is, set z̃_r := arg max_z̃ |c(z̃) − N/2|.
```

Notice that the outer loop might require many executions: if the round keys have length k, it requires 2^k executions. The algorithm can be speeded up by choosing f_0 and f_{r-1} such that $\tilde{S}^{1\ldots r-1}$ involves only $\tilde{k} < k$ bits of the last round key. Then only $2^{\tilde{k}}$ executions are required but one can estimate only those $\tilde{k}$ bits of z_r. More generally (see also [16, 17]), one can build equivalence classes of keys, where two keys z_r and z'_r are equivalent if and only if there is a $c \in \{0,1\}$ such that $f_{r-1} \circ g^{-1}(\cdot, z_r) = f_{r-1} \circ g^{-1}(\cdot, z'_r) \oplus c$. Then it is enough in the outer loop to examine one

representative of each class; on the other hand, the attack can only estimate the class in which the true key lies. One speaks then of the *right class* and of *wrong classes* and the representatives are called the *right key* and *wrong keys*, respectively [16, 17].

The reason why one outputs precisely the above estimate of z_r is the following assumption [17]:

Conjecture 2.3.1 (Hypothesis of Wrong-Key Randomization) *For any I/O sum $S^{1\ldots r-1}$ for which $\left|P[S^{1\ldots r-1}=0|Z_1=z_1,\ldots,Z_{r-1}=z_{r-1}]-\frac{1}{2}\right|$ is large for virtually all values $z_1,\ldots,z_{r-1}$ of the round keys, the following is true: for virtually all possible full keys $(z_1,\ldots,z_r)$ and for all estimates $\tilde{z}$ of the last round key,*

$$\frac{\left|P[\tilde{S}^{1\ldots r-1}=0|\tilde{Z}=z_r]-\frac{1}{2}\right|}{\left|P[\tilde{S}^{1\ldots r-1}=0|\tilde{Z}=\tilde{z}]-\frac{1}{2}\right|}\gg 1 \quad \text{for all } \tilde{z}\neq z_r.$$

This conjecture is plausible for the following reason: if one happens to choose the right key as the estimate, then the one-round decryption is in fact inverting the last encryption round, i.e., one follows actually the dashed path on Figure 2.3, whereas if one chooses a wrong key, one follows the detour through g^{-1} (solid line). But on the detour, one essentially performs two more "encryption" rounds than on the dashed path; therefore, one expects that, on the detour, $\tilde{Y}(r-1)$ is less dependent on X than is $\tilde{Y}(r-1)$ of the more direct way. By the argument of Subsection 2.3.1, $\left|P[\tilde{S}^{1\ldots r-1}=0|\tilde{Z}=z_r]-\frac{1}{2}\right|$ is (much) larger than $\left|P[\tilde{S}^{1\ldots r-1}=0|\tilde{Z}=\tilde{z}]-\frac{1}{2}\right|$.

Moreover, $c(\tilde{z})$ is a natural estimate of $P[\tilde{S}^{1\ldots r-1}=0|\tilde{Z}=\tilde{z}]$ [16]. Thus, $\frac{1}{N}\left|c(\tilde{z})-\frac{1}{2}\right|$ is an estimate of $\left|P[\tilde{S}^{1\ldots r-1}=0|\tilde{Z}=\tilde{z}]-\frac{1}{2}\right|$. According to the hypothesis of wrong-key randomization, the right key z_r should maximize $\frac{1}{N}\left|c(\tilde{z})-\frac{1}{2}\right|$ and hence also maximize $\left|c(\tilde{z})-\frac{N}{2}\right|$.

2.3.3 Probability of Success of the Attack

We have seen above that the distance between $1/2$ and the probability of some event is an important figure. The following definition, which is used throughout the thesis, keeps the expressions shorter [17].

Definition 2.3.2
Let X be a binary-valued random variable; then

$$I(X) := 2|P[X=0]-1/2| = |2P[X=0]-1| = |2P[X=1]-1|$$

is the *imbalance of* X.

(The factor of two in the definition of $I(X)$ is only for convenience.) One can also define imbalances based on conditional probabilities.

Definition 2.3.3
Let $S^{1\ldots i}$ be an i-round I/O sum. Then

- $I(S^{1\ldots i}|z_1,\ldots,z_i) := |2P[S^{1\ldots i}=0|(Z_1,\ldots,Z_i)=(z_1,\ldots,z_i)\,]-1|$ is the *key-dependent imbalance of* $S^{1\ldots i}$.
- The expectation of the key-dependent imbalance over all keys,

$$\overline{I}(S^{1\ldots i}) := E[I(S^{1\ldots i}|Z_1,\ldots,Z_i)\,] = \frac{1}{(2^k)^i}\sum_{z_1,\ldots,z_i} I(S^{1\ldots i}|z_1,\ldots,z_i),$$

  is called the *average-key imbalance of* $S^{1\ldots i}$.
- An I/O sum is called *effective* if $\overline{I}(S^{1\ldots i}) \approx 1$.

Example 2.3.4
Let $S^1 = f_0(X) \oplus f_1(Y(1)) = h(Z_1)$, where h is balanced. Then $I(S^1)=0$ but, since the I/O sum is the constant $h(z_1)$ when $Z_1=z_1$, $I(S^1|z_1)=1$ for all z_1 and therefore $\overline{I}(S^1)=1$.

We now come to the probability of success [17].

Definition 2.3.5
- The *probability of success of the attack*, p_{GLC}, is the probability of the event that the output list contains only the right class.
- The *conditional probability of success of the attack*, $p_{GLC|z_1,\ldots,z_r}$, is the probability of the same event given that $(Z_1,\ldots,Z_r)=(z_1,\ldots,z_r)$.

It is shown in [16] that, for fixed values of the round keys, if the hypothesis of wrong-key randomization holds, then the attack finds the true key as reliably as desired if enough p/c pairs are available.

The conditional probability of success is an increasing function of the square of the key-dependent imbalance of the I/O sum used; this suggests that the imbalance is a robust measure for the usefulness of such an I/O sum [17].

The conditional probability of success $p_{GLC|z_1,\ldots,z_r}$ is the true probability of success since in an actual encryption, a fixed key is used. However, the cryptanalyst does not know the key and can at most compute the overall probability of success p_{GLC}. To overcome this difficulty, he relies on the following hypothesis.

Conjecture 2.3.6 (Hypothesis of fixed-key equivalence)
For any effective I/O sum and for virtually all keys $(z_1, \ldots, z_{r-1})$, *the key-dependent imbalance* $I(S^{1\ldots r-1}|z_1, \ldots, z_{r-1})$ *is virtually independent of the value* $(z_1, \ldots, z_{r-1})$ *of the key, or equivalently,*

$$I(S^{1\ldots r-1}|z_1, \ldots, z_{r-1}) \approx \overline{I}(S^{1\ldots r-1}). \tag{2.1}$$

This hypothesis seems to hold in many cases and, if it holds, has the following consequences:

1. The conditional probability of success is approximately the same for all keys, i.e., the actual probability of success of the attack is approximately the same whatever key has been used in the encryption.

2. The overall probability of success, p_{GLC}, depends on the square of $\overline{I}(S^{1\ldots r-1})$ in approximately the same way as the conditional probability of success $p_{GLC|z_1,\ldots,z_r}$ depends on the square of $I(S^{1\ldots r-1}|z_1, \ldots, z_{r-1})$.

3. It allows the cryptanalyst to estimate the overall probability of success, p_{GLC}.

4. It is important for the cryptanalyst to find an I/O sum with an average-key imbalance as large as possible.

The validity of the hypothesis of fixed-key equivalence is the topic of Chapter 4.

2.3.4 The Piling-up Lemma

It is usually infeasible to compute the key-dependent imbalances of $S^{1\ldots r-1}$ and thus its average-key imbalance. An efficient way out of this dead-end can be found in [16, 17] and is briefly described here: one defines *threefold sums* T_i, $i = 1, \ldots, r$, by

$$T_i := f_{i-1}(Y(i-1)) \oplus g_{i-1}(Y(i)) \oplus h_i(Z_i), \tag{2.2}$$

where f_{i-1}, g_{i-1} and h_i are binary-valued functions the first two of which are balanced. Then $I(T_i) \leq \overline{I}(f_{i-1}(Y(i-1)) \oplus g_{i-1}(Y(i)))$. If $g_i = f_{i+1}$, $0 \leq i \leq r-2$, then the threefold sums are called *linked* and we have

$$\begin{aligned} T_1 \oplus \cdots \oplus T_{r-1} &= f_0(X) \oplus f_{r-1}(Y(r-1)) \oplus h_1(Z_1) \oplus \cdots \oplus h_{r-1}(Z_{r-1}) \\ &= S^{1\ldots r-1} \oplus h_1(Z_1) \oplus \cdots \oplus h_{r-1}(Z_{r-1}) \end{aligned}$$

and $I(T_1 \oplus \cdots \oplus T_{r-1}) \leq \overline{I}(S^{1\ldots r-1})$. Thus, finding linked threefold sums allows one at least to lower-bound the average-key imbalance of $S^{1\ldots r-1}$. But another problem arises: it is also infeasible in general to compute the imbalance of $T_1 \oplus \cdots \oplus T_{r-1}$. However, if $T_1, \ldots, T_{r-1}$ are independent, then one can use the following lemma.

Lemma 2.3.7 (Piling-up Lemma (Matsui))
Let $X_1, \ldots, X_s$ be independent random variables with values on $\mathbb{Z}_2$. Then

$$I\left(\bigoplus_{i=1}^{s} X_i\right) = \prod_{i=1}^{s} I(X_i). \tag{2.3}$$

Proof: A proof can be found in [16]. □

It is usually feasible to compute the imbalance of T_i as the threefold sum involves only the input, the output, and the round key of a single round. Then, if $T_1, \ldots, T_{r-1}$ are independent, we have

$$I(T_1 \oplus \cdots \oplus T_{r-1}) = I(T_1)I(T_2)\cdots I(T_{r-1}). \tag{2.4}$$

Still, finding linked, independent threefold sums is very difficult. What one often does is to assume that one's threefold sums are independent and then to apply the Piling-up Lemma. We call this the *piling-up approximation* and examine the consequences that it entails in Chapter 3. This is also what Matsui did is his attack on DES [38, 39]. There, he applied (2.4) with no concern for independence and used the result directly to compute the probability of success. It is shown in [16] that if T_i is independent of the round input $Y(i-1)$ for all i, then $T_1, \ldots, T_{r-1}$ are independent. A possible way to find threefold sums with this property is given in [16, 17] for a class of ciphers where, in each round, part of the key is combined with the round input by a group operation to form the input to the remainder of the round function in which only the remainder of the round key is used. We know of no other general way to construct independent threefold sums.

Remark 2.3.8
The name *linear cryptanalysis* comes from the fact that, in his original attack on DES, Matsui used linear and affine functions instead of general balanced ones. (Those functions are balanced, too.) On the other hand, this name is a little presomptuous since, with this "linear approximation", one cannot even represent all linear ciphers, as we show in the example below.

Example 2.3.9
Consider a one-round cipher where the ciphertext Y depends on the plaintext X and on the key Z in the manner $Y = A(Z)X$, where A(Z) is an invertible matrix depending on Z. Suppose there are two vectors α and β, not both the zero vector, such that $\alpha \bullet X = \beta \bullet Y$, where $a \bullet X = \alpha_1 X_1 \oplus \alpha_2 X_2 \oplus \cdots \oplus \alpha_n X_n$. Then we have

$$\begin{aligned} 0 &= \alpha \bullet X \oplus \beta \bullet Y = \alpha \bullet X \oplus \beta \bullet A(Z)X = \alpha \bullet X \oplus A(Z)^\top \beta \bullet X \\ &= (\alpha \oplus A(Z)^\top \beta) \bullet X. \end{aligned}$$

But in general, $c \bullet x = 0$ for all values of x if and only if $c = 0$. Thus, we must have $\alpha = A(z)^\top \beta$ for all values z of the key. Now take $A(z_1) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $A(z_2) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Let $\alpha = (\alpha_1, \alpha_2)^\top$ and $\beta = (\beta_1, \beta_2)^\top$. One gets the four equations $\alpha_1 = \beta_1 \oplus \beta_2$, $\alpha_2 = \beta_2$, $\alpha_1 = \beta_1$, and $\alpha_2 = \beta_1 \oplus \beta_2$. It follows that $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$, which we had excluded. By contradiction, we conclude that $\alpha \bullet X \neq \beta \bullet Y$.

2.3.5 Imbalance of Functions

In this subsection, we define the imbalance of a function and show some obvious, but useful properties of imbalances.

Definition 2.3.10
For any binary-valued function f on some set M, define the *imbalance of* f by $I(f) := I(f(X))$, where X is a uniformly distributed random variable on M. For instance, balanced functions have imbalance zero.

Lemma 2.3.11
Let g be an invertible function on M and $f : M \to \mathbb{Z}_2$ be any function. Then $I(f \circ g) = I(f)$. In particular, f is balanced if and only if $f \circ g$ is balanced.
Proof:
The function g is a permutation of M. Thus, f and $f \circ g$ are equal to zero for the same number of arguments. □

In a very similar way, we also have:

Lemma 2.3.12
Let X be a binary-valued random variable on a sample space Ω and π be a permutation on Ω. Then $I(X \circ \pi) = I(X)$.
Proof:
$X \circ \pi$ is a binary-valued random variable on Ω with the same probability distribution as X. □

Remark 2.3.13
In fact, for any binary-valued function f on a finite set M one can consider M as a sample space and f as a binary-valued random variable on M. Then the imbalance of f as a random variable coincides with the imbalance of f as a function. Thus, every property of the imbalance of a random variable has its equivalent as a property of the imbalance of a function. We shall use this duality several times. A first illustration of it was given by Lemmata 2.3.11 and 2.3.12; a second example is given by the two Lemmata below.

Lemma 2.3.14
Let Ω be a finite sample space, and let X be a binary-valued random variable on Ω. Then

1. *if $|\Omega|$ is even, i.e., $|\Omega| = 2\vartheta$, then $I(X)$ is of the form i/ϑ, $0 \leq i \leq \vartheta$;*
2. *if $|\Omega|$ is odd, i.e., $|\Omega| = 2\vartheta+1$, then $I(X)$ is of the form $\frac{2i+1}{2\vartheta+1}$, $0 \leq i \leq \vartheta$.*

Proof:

1. Let α be the number of times X takes on the value 0. Then $I(X) = |2\frac{\alpha}{2\vartheta} - 1| = \frac{1}{\vartheta}|\alpha - \vartheta|$, which is of the stated form.
2. The proof is similar. □

The dual statement for functions is:

Lemma 2.3.15
Let M be a finite set, and let f be a binary-valued function on M. Then

1. *if $|M|$ is even, i.e., $|M| = 2\vartheta$, then $I(f)$ is of the form i/ϑ, $0 \leq i \leq \vartheta$;*
2. *if $|M|$ is odd, i.e., $|M| = 2\vartheta+1$, then $I(f)$ is of the form $\frac{2i+1}{2\vartheta+1}$, $0 \leq i \leq \vartheta$.*

□

We end the chapter with the following remark.

Remark 2.3.16
When the round keys are fixed, the mapping $X \mapsto Y(r-1)$ is invertible; we denote this invertible function by $g_{z_1,\ldots,z_{r-1}}$. Then the $(r-1)$-round I/O sum $S^{1\ldots r-1}$ can be written as

$$S^{1\ldots r-1} = f_0(X) \oplus f_{r-1}(g_{Z_1,\ldots,Z_{r-1}}(X)) = (f_0 \oplus (f_{r-1} \circ g_{Z_1,\ldots,Z_{r-1}}))(X).$$

This emphasizes the dependence on X and shows the nature of the function applied to X. Moreover, since one assumes that the plaintext X

is a uniformly distributed random variable and by Definition 2.3.10, one can write the key-dependent imbalance of $S^{1\ldots r-1}$ as the imbalance of a function, namely

$$I(S^{1\ldots r-1}|z_1,\ldots,z_{r-1}) = I(f_0 \oplus f_{r-1} \circ g_{z_1,\ldots,z_{r-1}}). \tag{2.5}$$

We shall use this fact extensively in Chapter 4.

Chapter 3

The Imbalance of mod 2-Sums of Random Variables

3.1 Matsui's Piling-up Lemma

In the binary generalisation of linear cryptanalysis, one is interested in the imbalance of $T_1 \oplus \cdots \oplus T_{r-1}$, where $T_1, \ldots, T_{r-1}$ are linked threefold sums defined by (2.2). If $T_1, \ldots, T_{r-1}$ are independent, then one can apply Matsui's Piling-up Lemma, which we state here again.

Lemma 3.1.1 (Piling-up Lemma (Matsui))
Let $X_1, \ldots, X_r$ be independent random variables with values on $\mathbb{Z}_2$. Then

$$I\left(\bigoplus_{i=1}^{r} X_i\right) = \prod_{i=1}^{r} I(X_i). \tag{3.1}$$

Proof: A proof can be found in [16]. □

Then the imbalance of $T_1 \oplus \cdots \oplus T_{r-1}$ is equal to the product of the imbalances of $T_1, \ldots, T_{r-1}$. This is useful because the imbalance of the threefold sums is much more easy to compute than the imbalance of $T_1 \oplus \cdots \oplus T_{r-1}$. The problem is now to find independent threefold sums. This is very difficult in practice. What one does usually is to assume that one's threefold sums are independent and then to apply the Piling-up

Lemma. We call this the *piling-up approximation.* As we will see, this can be dangerous in particular cases but not on the average.

In this chapter, we find relations between the imbalance of a sum of (not necessarily independent) binary random variables and the product of their imbalances. In particular, we address the following question: given r random variables $X_1, \ldots, X_r$, for which we do not know their relationship, and given their imbalances $I(X_1), \ldots, I(X_r)$, what can we say about $I(X_1 \oplus \cdots \oplus X_r)$?

We will:

- see that $I(X_1 \oplus \cdots \oplus X_r)$ can differ considerably from $I(X_1) \cdots I(X_r)$;
- generalise the Piling-up Lemma to dependent random variables;
- examine how much one risks by assuming that $I(X_1 \oplus \cdots \oplus X_r) = I(X_1) \cdots I(X_r)$.

3.2 Two Identities Valid For All Binary Random Variables

In this section, we prove two relations between the imbalance of a mod 2-sum of binary-valued random variables and the product of the imbalances of these random variables; both relations are valid for both independent and dependent random variables.

3.2.1 An Inequality Between $I(X_1 \oplus \cdots \oplus X_r)$ and $I(X_1) \cdots I(X_r)$

The first relation is an inequality. Our first step to that inequality is the following lemma.

Lemma 3.2.1
For any random variables X_1 and X_2 with values in $\mathbb{Z}_2$, we have

$$I(X_1) + I(X_2) \leq 1 + I(X_1 \oplus X_2) \tag{3.2}$$

with equality if and only if one of the following holds:

- $P_{X_1X_2}(0,0) \geq 1/2, P_{X_1X_2}(1,1) = 0;$
- $P_{X_1X_2}(1,1) \geq 1/2, P_{X_1X_2}(0,0) = 0;$
- $P_{X_1X_2}(0,1) \geq 1/2, P_{X_1X_2}(1,0) = 0;$
- $P_{X_1X_2}(1,0) \geq 1/2, P_{X_1X_2}(0,1) = 0.$

Proof:

Let $p_{00} = P_{X_1X_2}(0,0), p_{01} = P_{X_1X_2}(0,1), p_{10} = P_{X_1X_2}(1,0)$, and $p_{11} = P_{X_1X_2}(1,1)$. Then $P_{X_1}(0) = p_{00}+p_{01}$, $P_{X_2}(0) = p_{00}+p_{10}$ and $P_{X_1\oplus X_2}(0) = p_{00}+p_{11} = 1-(p_{01}+p_{10})$. Hence, the imbalances of X_1, X_2, and $X_1 \oplus X_2$ are

$$\begin{aligned} I(X_1) &= |2(p_{00}+p_{01})-1| \\ I(X_2) &= |2(p_{00}+p_{10})-1| \\ I(X_1\oplus X_2) &= |2(p_{01}+p_{10})-1|. \end{aligned}$$

It remains to show that

$$f(p_{00},p_{01},p_{10}) := |2(p_{00}+p_{01})-1|+|2(p_{00}+p_{10})-1|-|2(p_{01}+p_{10})-1|$$

is upper-bounded by 1. An easy way to prove this is the following: the value of the expressions inside of each of the three absolute value signs can be either non-negative or negative. This gives eight cases denoted by $+++, ++-, \ldots, ---$. We examine two cases; the others are treated similarly.

$+++$ Here $f = 2(p_{00}+p_{01})-1+2(p_{00}+p_{10})-1-2(p_{01}+p_{10})+1 = 4p_{00}-1$. But $4p_{00}-1 \le 1$ because $2(p_{01}+p_{10})-1 \ge 0$ implies $p_{00} \le 1/2$. Moreover, $f = 1$ if and only if $p_{00} = 1/2$, which, because $p_{01}+p_{10} \ge 1/2$, also implies $p_{01}+p_{10} = 1/2$.

$++-$ Here $f = 2(p_{00}+p_{01})-1+2(p_{00}+p_{10})-1+2(p_{01}+p_{10})-1 = 4(p_{00}+p_{01}+p_{10})-3 \le 1$; equality holds if and only if $p_{11} = 1-(p_{00}+p_{01}+p_{10}) = 0$ and $p_{00} > 1/2$, because $p_{01}+p_{10} < 1/2$ implies $p_{00} > 1/2$. □

Figure 3.1 shows the regions of (p_{00},p_{01},p_{10}) where f is equal to 1. Now we note that $I(X_1)+I(X_2)+I(X_3) \le 1+I(X_1\oplus X_2)+I(X_3) \le 2+I(X_1\oplus X_2\oplus X_3)$. By a simple induction, it follows that for any random variables $X_1,\ldots,X_r$ with values in $\mathbb{Z}_2$,

$$I(X_1)+\cdots+I(X_r) \le (r-1)+I(X_1\oplus\cdots\oplus X_r). \tag{3.3}$$

Moreover, because $a_1a_2\cdots a_n \le (\frac{a_1+\cdots+a_n}{n})^n$ for any nonnegative numbers $a_1,\ldots,a_n$ with equality if and only if all a_i are equal, we have

$$\begin{aligned} I(X_1)\cdots I(X_r) &\le r^{-r}(I(X_1)+\cdots+I(X_r))^r \\ &\le r^{-r}((r-1)+I(X_1\oplus\cdots\oplus X_r))^r, \end{aligned} \tag{3.4}$$

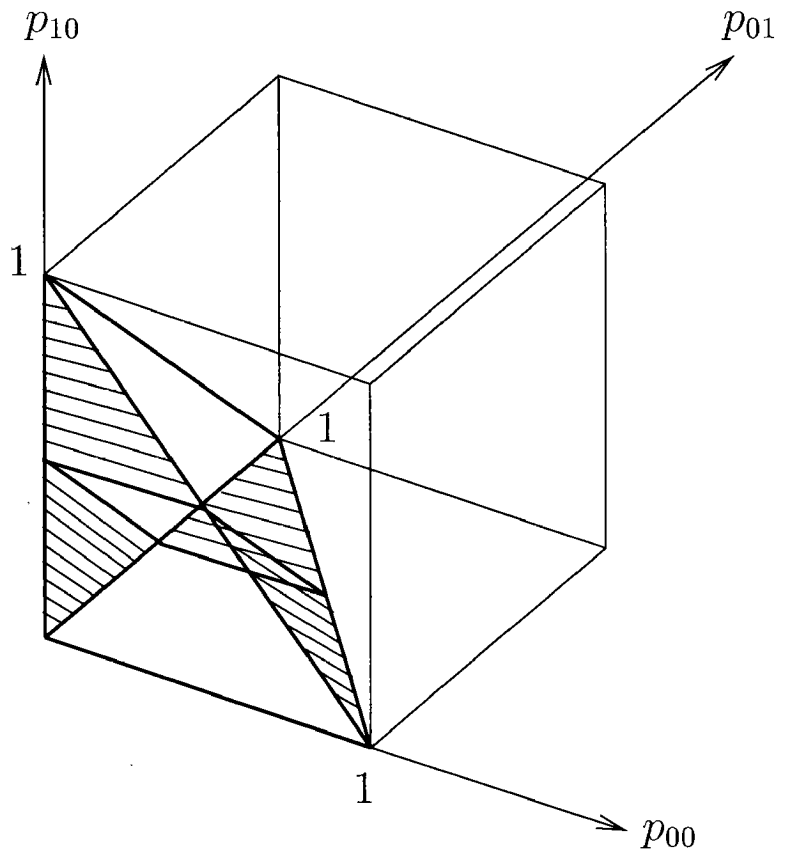


Figure 3.1: The regions where $f = 1$ (they include the boundaries).

with equality everywhere if and only if $I(X_i) = \frac{1}{r}((r-1)+I(X_1 \oplus \cdots \oplus X_r))$ for all i.

We show now that equality is indeed possible. Let a, b be real numbers with $0 \leq a, b \leq 1$, $a + b \leq 3/2$ and $a - b \geq 1/2$ (see Figure 3.2), and choose $p_{00} = a - b$, $p_{01} = b$, $p_{10} = 1 - a$ and $p_{11} = 0$; then $a \geq 1/2 \geq b$ so $I(X_1) = |2a - 1| = 2a - 1, I(X_2) = |2b - 1| = 1 - 2b \geq 1 - I(X_1)$ and $I(X_1 \oplus X_2) = |2(a - b) - 1| = 2a - 2b - 1$ so we have $I(X_1) + I(X_2) = 1 + I(X_1 \oplus X_2)$. Considering all possible such pairs (a, b), we see that the following holds:

> For any random variable X_1 and any real number I_2 such that $1 \geq I_2 \geq 1 - I(X_1)$, there is a random variable X_2 such that $I(X_2) = I_2$ and $I(X_1) + I(X_2) = 1 + I(X_1 \oplus X_2)$.

Let now $j \geq 2$ and suppose that $I(X_1) + \cdots + I(X_j) = (j - 1) + I(X_1 \oplus \cdots \oplus X_j)$. Choose I_{j+1} such that $1 \geq I_{j+1} \geq j - (I(X_1) + \cdots + I(X_j)) = 1 - I(X_1 \oplus \cdots \oplus X_j)$. Then there exists a random variable X_{j+1} such that $I(X_{j+1}) = I_{j+1}$ and $I(X_{j+1}) + I(X_1 \oplus \cdots \oplus X_j) = 1 + I(X_1 \oplus \cdots \oplus X_j \oplus X_{j+1})$. But then $I(X_1) + \cdots + I(X_{j+1}) = j + I(X_1 \oplus \cdots \oplus X_{j+1})$.

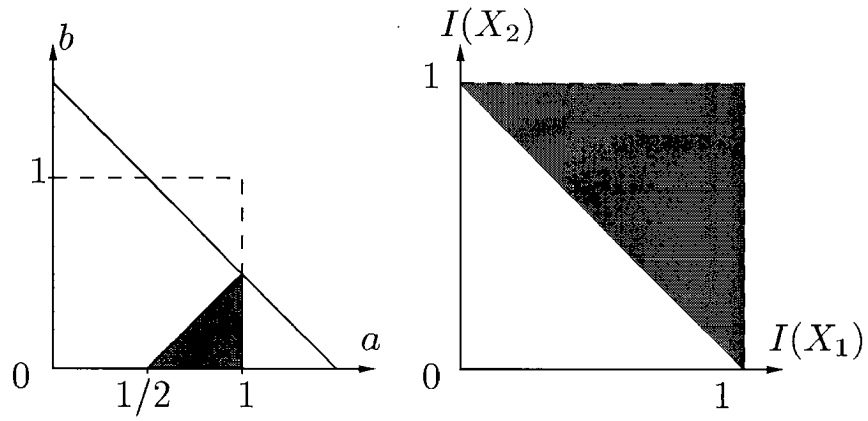


Figure 3.2: Possible choices of a and b and resulting imbalances.

So far we have shown only that equality is possible in (3.3). In order to show that equality everywhere is possible in (3.4), we still must show that this procedure can also yield the case $I(X_j) = \frac{1}{r}((r-1) + I(X_1 \oplus \cdots \oplus X_r))$, $j = 1, \ldots, r$. That is, we must show that once we have chosen $I(X_1) = \cdots = I(X_j) = \frac{1}{r}((r-1) + I(X_1 \oplus \cdots \oplus X_r))$, we can still choose $I(X_{j+1}) = I_{j+1} = \frac{1}{r}((r-1) + I(X_1 \oplus \cdots \oplus X_r)) \geq j - (I(X_1) + \cdots + I(X_j))$. The inequality is satisfied if and only if

$$\begin{aligned} \frac{1}{r}((r-1) + I(X_1 \oplus \cdots \oplus X_r)) &\geq j - \frac{j}{r}((r-1) + I(X_1 \oplus \cdots \oplus X_r)) \Leftrightarrow \\ I(X_1 \oplus \cdots \oplus X_r) &\geq 1 - \frac{r}{j+1}, \end{aligned}$$

which holds for all $j = 1, \ldots, r-1$ so it is indeed possible. The following proposition summarizes the results.

Proposition 3.2.2
For any r random variables $X_1 \ldots, X_r$ with values in $\mathbb{Z}_2$, we have:

1. $I(X_1) + \cdots + I(X_r) \leq (r-1) + I(X_1 \oplus \cdots \oplus X_r)$ *and*
2. $I(X_1) \cdots I(X_r) \leq r^{-r}((r-1) + I(X_1 \oplus \cdots \oplus X_r))^r$ *with equality if and only if* $I(X_i) = \frac{1}{r}((r-1) + I(X_1 \oplus \cdots \oplus X_r))$ *for all* i.

Moreover, equality can occur in both inequalities. □

Figure 3.3 visualises the second inequality of Proposition 3.2.2. The Piling-up Lemma says that if the random variables are independent, then we are always on the diagonal. The second inequality in Proposition 3.2.2 says that, for $r \geq 2$, all points on or above the solid line are possible. Hence, for dependent random variables, the product of the imbalances

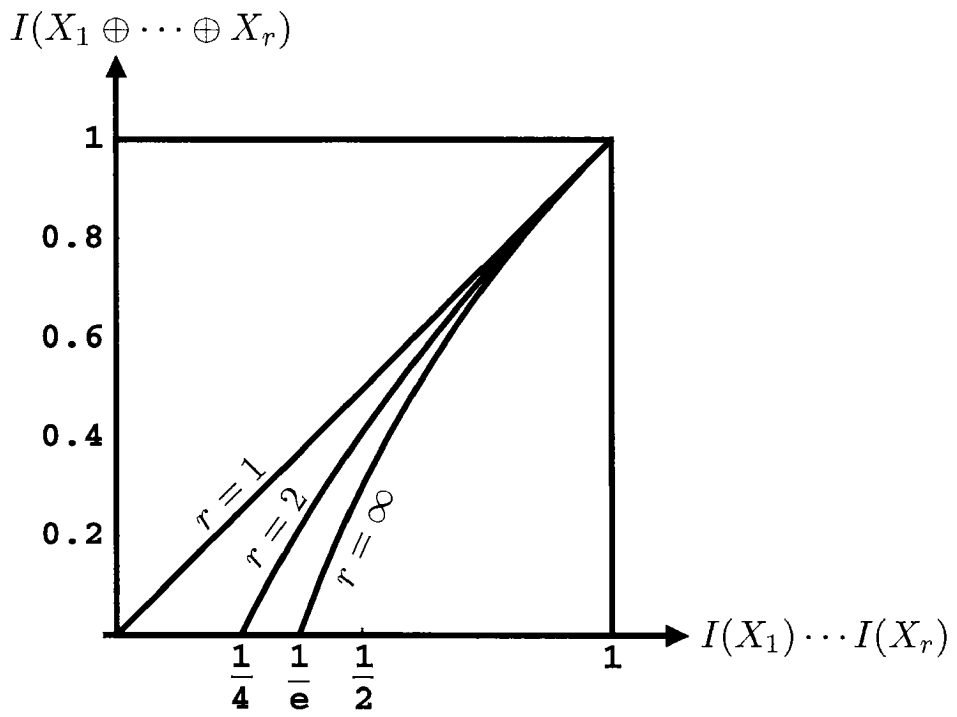


Figure 3.3: For $r \geq 2$, all points on or above the line are possible.

can differ considerably from the imbalance of the sum. However, in linear cryptanalysis, it is preferable to underestimate $I(X_1 \oplus \cdots \oplus X_r)$ as $I(X_1)\cdots I(X_r)$ rather than to overestimate it. Here most of the possible values of $I(X_1 \oplus \cdots \oplus X_r)$ are larger than or equal to $I(X_1)\cdots I(X_r)$ so one could say that there is nothing to worry about. But, as we will see later, for given $I(X_1),\ldots,I(X_r)$, small values of $I(X_1 \oplus \cdots \oplus X_r)$ occur more often than large ones.

3.2.2 Generalisation of Matsui's Piling-up Lemma

As next, we derive a formula involving both $I(X_1)\cdots I(X_r)$ and $I(X_1 \oplus \cdots \oplus X_r)$ that, when applied to independent random variables, reduces to the Piling-up Lemma. Again we require a preliminary result.

Lemma 3.2.3
For any integer $r \geq 2$ and any real numbers $x_1,\ldots,x_r$ in $\{0,1\}$, we have

$$x_1 \oplus \cdots \oplus x_r = \sum_{k=1}^{r} (-2)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq r} x_{i_1} \cdots x_{i_k}. \qquad (3.5)$$

Proof:
A proof can be found on page 120 of [31]. □

For instance, we can write:

- $x_1 \oplus x_2 = (x_1 + x_2) - 2x_1x_2$;
- $x_1 \oplus x_2 \oplus x_3 = (x_1 + x_2 + x_3) - 2(x_1x_2 + x_1x_3 + x_2x_3) + 4x_1x_2x_3$.

We can can consider $x_1, \ldots, x_r$ in Lemma 3.2.3 as realizations of binary random variables; taking expectations on both sides of (3.5) yields a formula for the expected value of $X_1 \oplus \cdots \oplus X_r$.

Corollary 3.2.4
For any real-valued random variables $X_1, \ldots, X_r$ with values in $\{0, 1\}$,

$$E\Big[\bigoplus_{i=1}^{r} X_i\Big] = \sum_{k=1}^{r} (-2)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le r} E[X_{i_1} \cdots X_{i_k}]. \qquad (3.6)$$

□

Definition 3.2.5
For any real-valued random variables $X_1, \ldots, X_r$ and any $r \ge 1$, define $\lambda(X_1, \ldots, X_r) := E[X_1 \cdots X_r] - E[X_1] \cdots E[X_r]$. For $r = 2$, $\lambda(X_1, X_2)$ is the covariance of X_1 and X_2.

Lemma 3.2.6
If $X_1, \ldots, X_r$ are real-valued random variables with values in $\{0, 1\}$, then they are independent if and only if $\lambda(X_{i_1}, \ldots, X_{i_k}) = 0$ for all k, $2 \le k \le r$, and all $1 \le i_1 < \cdots < i_k \le r$.
Proof:
The condition is obviously necessary. That it is sufficient follows from the fact that $E[X_{i_1}, \ldots, X_{i_k}] = P_{X_{i_1} \cdots X_{i_k}}(1, \ldots, 1)$ for all k, $2 \le k \le r$, and all $1 \le i_1 < \cdots < i_k \le r$, and that all the random variables considered have only two possible values, namely 0 and 1. □

This definition allows us to generalise the Piling-up Lemma. But first we need the following lemma:

Lemma 3.2.7
For any integer $r \ge 1$ and any real numbers $b_1, \ldots, b_r$, the following identity holds:

$$\sum_{k=1}^{r} (-1)^{r-k} \sum_{1 \le i_1 < \cdots < i_k \le r} \prod_{l=1}^{k} (1 + b_{i_l}) = b_1 b_2 \cdots b_r + (-1)^{r-1}. \qquad (3.7)$$

Proof:

For any real numbers $c_1, \ldots, c_r$, we have

$$\prod_{j=1}^{r}(c_j - 1) = (-1)^r + \sum_{k=1}^{r}(-1)^{r-k} \sum_{1 \le i_1 < \cdots < i_k \le r} \prod_{l=1}^{k} c_{i_l}.$$

Now set $c_j = b_j + 1$ and put the term $(-1)^r$ on the other side of the equation. □

For instance, we have:

- $1 + b_1 = b_1 + 1$;
- $(1 + b_1)(1 + b_2) - [(1 + b_1) + (1 + b_2)] = b_1 b_2 - 1$;
- $(1 + b_1)(1 + b_2)(1 + b_3) - \big[(1 + b_1)(1 + b_2) + (1 + b_1)(1 + b_3) + (1 + b_2)(1 + b_3)\big] + \big[(1 + b_1) + (1 + b_2) + (1 + b_3)\big] = b_1 b_2 b_3 + 1$.

We can now formulate our second general relationship between $I(X_1 \oplus \cdots \oplus X_r)$ and $I(X_1) \cdots I(X_r)$. Recall that $\mathrm{sgn}(b)$, the sign of a real number b, is equal to 1 if b is positive, to -1 if b is negative, and to 0 if $b = 0$.

Theorem 3.2.8
For any real-valued random variables $X_1, \ldots, X_r$ taking values in $\{0, 1\}$,

$$I(X_1 \oplus \cdots \oplus X_r) = \left| 2\left(\sum_{k=2}^{r}(-2)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le r} \lambda(X_{i_1} \cdots X_{i_k})\right) + (-1)^{r-1} \prod_{i=1}^{r} \varepsilon_i I(X_i) \right|$$

where $\varepsilon_i = sgn(2E[X_i] - 1)$.

Proof:

By (3.6), we have

$$\begin{aligned} E\Big[\bigoplus_{i=1}^{r} X_i\Big] &= \sum_{k=1}^{r}(-2)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le r} E[X_{i_1} \cdots X_{i_k}] \\ &= \sum_{k=1}^{r}(-2)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le r} \Big(\lambda(X_{i_1} \cdots X_{i_k}) + \prod_{l=1}^{k} E[X_{i_l}]\Big) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{r} (-2)^{k-1} \sum_{1\le i_1<\cdots<i_k\le r} \lambda(X_{i_1}\cdots X_{i_k}) \\
&\quad + \sum_{k=1}^{r} (-2)^{k-1} \sum_{1\le i_1<\cdots<i_k\le r} \prod_{l=1}^{k} E[X_{i_l}] \\
&= \sum_{k=2}^{r} (-2)^{k-1} \sum_{1\le i_1<\cdots<i_k\le r} \lambda(X_{i_1}\cdots X_{i_k}) \\
&\quad + \frac{1}{2}(-1)^{r-1} \sum_{k=1}^{r} (-1)^{r-k} \sum_{1\le i_1<\cdots<i_k\le r} \prod_{l=1}^{k} (1+\varepsilon_{i_l} I(X_{i_l})).
\end{aligned}
$$

where, in the last equality, we have used the fact that $E[X_i] = \frac{1}{2}(1+\varepsilon_i I(X_i))$ for all i. By Lemma 3.2.7, the second term is equal to $\frac{1}{2}((-1)^{r-1}\prod_{i=1}^{r}\varepsilon_i I(X_i)+1)$. This implies that

$$
\begin{aligned}
I(X_1\oplus\cdots\oplus X_r) &= \Big|2E\Big[\bigoplus_{i=1}^{r} X_i\Big] - 1\Big| \\
&= \Big|2\Big(\sum_{k=2}^{r}(-2)^{k-1} \sum_{1\le i_1<\cdots<i_k\le r} \lambda(X_{i_1}\cdots X_{i_k})\Big) + (-1)^{r-1}\prod_{i=1}^{r}\varepsilon_i I(X_i)\Big|.
\end{aligned}
$$

□

In our opinion, this result is more of theoretical than of practical importance since it is infeasible to compute $\lambda(X_{i_1}\cdots X_{i_k})$ in practical cases as it requires the knowledge of the joint distribution of $X_{i_1},\ldots,X_{i_k}$. But if we know the joint distribution of $X_1,\ldots,X_k$, $k=2,\ldots,r$, then we can also compute $I(X_1\oplus\cdots\oplus X_r)$ without the above formula.

In the remainder of this chapter, we will be concerned with the distribution of the values of $I(X_1\oplus\cdots\oplus X_r)$ over all binary-valued random variables $X_1,\ldots,X_r$ that have some given imbalance $I(X_1),\ldots,I(X_r)$, and with the corresponding average and variance. The results allow us to tell whether it is advisable to approximate $I(X_1\oplus\cdots\oplus X_r)$ by $I(X_1)\cdots I(X_r)$ in general. We begin by considering the case $r=2$.

3.3 Two Random Variables

This section deals with the average and the variance of the values of $I(X_1\oplus X_2)$ over all random variables X_1 and X_2 that have the same imbalances $I(X_1)$ and $I(X_2)$, respectively. We begin with an example.

Example 3.3.1
Let X_1 and X_2 be two random variables on a sample space Ω, with $|\Omega| = 4$. By Lemma 2.3.14, a random variable on Ω can have imbalance 0, 1/2, or 1. Let first $I(X_1) = I(X_2) = 0$, i.e., let X_1 and X_2 take on each of the values zero and one for two arguments. There are $\binom{4}{2}^2 = 36$ pairs of random variables having this property. Computing now $I(X_1 \oplus X_2)$ for the 36 possible pairs (X_1, X_2), we find 24 times that $I(X_1 \oplus X_2) = 0$ and 12 times that $I(X_1 \oplus X_2) = 1$, and that $I(X_1 \oplus X_2) = 1/2$ never occurs. We denote this by the triple $(24, 0, 12)$. On the average, $I(X_1 \oplus X_2) = 1/3$. One can do this for any value of $I(X_1)$ and $I(X_2)$. Table 3.1 shows the corresponding triples and, below each triple, the corresponding average. The results are of course symmetrical in $I(X_1)$ and $I(X_2)$.

An alternative view is proposed in Figure 3.4: considering all 256 pairs of random variables (X_1, X_2) on Ω, this figure indicates for how many pairs (X_1, X_2) the possible pair of values $(I(X_1)I(X_2), I(X_1 \oplus X_2))$ occurs. Empty circles mean that the first entry in the pair is a possible value of $I(X_1)I(X_2)$ and that the second entry is a possible value of $I(X_1 \oplus X_2)$, but that the pair $(I(X_1)I(X_2), I(X_1 \oplus X_2))$ never occurs. The full line shows the lower bound $I(X_1 \oplus X_2) \geq 2\sqrt{I(X_1)I(X_2)} - 1$ that we proved in Proposition 3.2.2.

		$I(X_2)$		
		0	1/2	1
$I(X_1)$	0	$(24, 0, 12)$ 1/3	$(0, 48, 0)$ 1/2	$(12, 0, 0)$ 0
	1/2	$(0, 48, 0)$ 1/2	$(48, 0, 16)$ 1/4	$(0, 16, 0)$ 1/2
	1	$(12, 0, 0)$ 0	$(0, 16, 0)$ 1/2	$(0, 0, 4)$ 1

Table 3.1: The Distribution of the Possible Values of $I(X_1 \oplus X_2)$ and their Average.

One sees that, in most cases, $I(X_1 \oplus X_2) \geq I(X_1)I(X_2)$ and (from Table 3.1) that, for $I(X_1)$ and $I(X_2)$ fixed, the average of $I(X_1 \oplus X_2)$ is always at least as large as $I(X_1)I(X_2)$. We want now to show this for any sample space Ω with an even number of elements and arbitrary imbalances $I(X_1)$ and $I(X_2)$. For this purpose, we could calculate and use the probability distribution of $I(X_1 \oplus X_2)$ given $I(X_1)$ and $I(X_2)$. However, it is easier to handle the average in a direct way as we do below. We will be concerned with the probability distribution when we take up the question of more than two random variables.

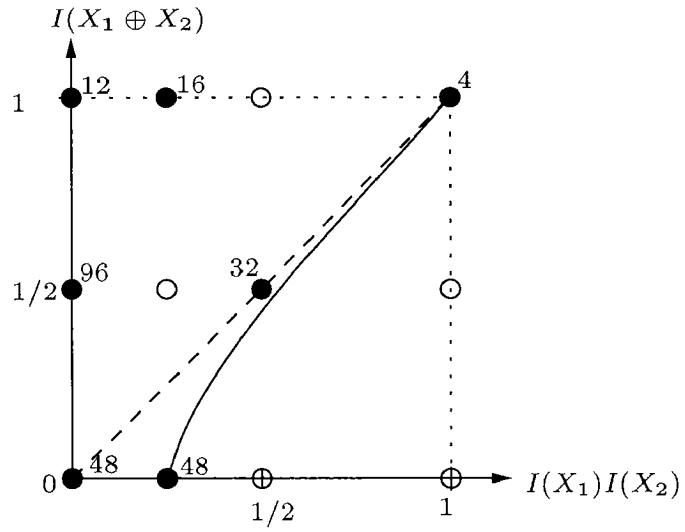


Figure 3.4: Comparison between $I(X_1)I(X_2)$ and $I(X_1 \oplus X_2)$ for $|\Omega| = 4$, with frequency of occurrence for each pair $(I(X_1)I(X_2), I(X_1 \oplus X_2))$.

Here is how we proceed: we take some sample space Ω with an even number 2ϑ of elements. Then we consider two integers $0 \leq i_1, i_2 \leq \vartheta$ and compute $I(X_1 \oplus X_2)$ for all random variables X_1 (resp. X_2) with imbalance $I(X_1) = i_1/\vartheta$ (resp. $I(X_2) = i_2/\vartheta$). The random variables X_1 and X_2 are chosen independently and uniformly at random from the set of all random variables with imbalance $I(X_1) = i_1/\vartheta$ and $I(X_2) = i_2/\vartheta$, respectively. The distribution of the values obtained has a certain average and a certain variance. We denote these as $E_\vartheta[I(X_1 \oplus X_2)|I(X_1) = \frac{i_1}{\vartheta}, I(X_2) = \frac{i_2}{\vartheta}]$ and $Var_\vartheta(I(X_1 \oplus X_2)|I(X_1) = \frac{i_1}{\vartheta}, I(X_2) = \frac{i_2}{\vartheta})$, respectively.

Remark 3.3.2
We choose $|\Omega|$ even because in this case the imbalance of a random variable defined on Ω is of the form i/ϑ for some integer i, $0 \leq i \leq \vartheta$ (Lemma 2.3.14). This covers all cases of interest because in "real" cases, threefold sums are functions of a uniformly distributed random vector, consisting of the plaintext and the key, that takes values in a set with an even number of elements, in fact a power of two. By the duality explained in Subsection 2.3.5, we can take that set as the sample space and the threefold sums as random variables on that sample space.

Remark 3.3.3
We allowed ourselves above to write an expected value and a variance for the following reason: we can consider a random variable X (resp. Y) whose values are random variables and that is uniformly distributed on

the set of all random variables having imbalance i_1/ϑ (resp. i_2/ϑ). That is, to write $X = x$ (resp. $Y = y$) means that x (resp. y) is a random variable with imbalance i_1/ϑ (resp. i_2/ϑ); thus, an expression like $I(x \oplus y)$ has a precise meaning and then $E_\vartheta[I(X_1 \oplus X_2)|I(X_1) = \frac{i_1}{\vartheta}, I(X_2) = \frac{i_2}{\vartheta}]$ is nothing but the expected value of $I(X \oplus Y)$ taken over the independent random variables X and Y.

Lemma 3.3.4
Let X_1 and X_1' be random variables defined on a sample space Ω and $I(X_1) = I(X_1')$. Then there is a permutation π on Ω such that $X_1' = X_1 \circ \pi$ or $X_1' = X_1 \circ \pi \oplus 1$ as functions on Ω.
Proof:

Let $A = \{\omega \in \Omega | X_1(\omega) = 0\}$, $B = \{\omega \in \Omega | X_1(\omega) = 1\}$, and $A' = \{\omega \in \Omega | X_1'(\omega) = 0\}$. Since $I(X_1) = I(X_1')$, either $|A'| = |A|$ or $|A'| = |B|$. In the former case, there is a bijection $\pi : \Omega \to \Omega$ that maps A' into A. Then $X_1' = X_1 \circ \pi$. In the latter case, there is a bijection $\pi : \Omega \to \Omega$ that maps A' into B. Then $X_1' = X_1 \circ \pi \oplus 1$. □

We will make use of the following Lemma several times.

Lemma 3.3.5
The multiset $\{I(X_1 \oplus X_2) | X_2$ random variable with $I(X_2) = i_2/\vartheta\}$ is the same for all random variables X_1 with imbalance $I(X_1) = i_1/\vartheta$.
Proof:

Let X_1 and X_1' be two random variables with imbalance i_1/ϑ. Then, by Lemma 3.3.4, there is a permutation π on Ω such that $X_1' = X_1 \circ \pi$ or $X_1' = X_1 \circ \pi \oplus 1$ as functions on Ω. Let $X_2' = X_2 \circ \pi$ or $X_2' = X_2 \circ \pi \oplus 1$, respectively. Then $I(X_1' \oplus X_2') = I((X_1 \oplus X_2) \circ \pi) = I(X_1 \oplus X_2)$ by Lemma 2.3.12. The proof is completed by noting that as X_2 runs over the set of all random variables with imbalance i_2/ϑ, $X_2 \circ \pi$ and $X_2 \circ \pi \oplus 1$ run over the same set. □

A consequence of this Lemma is that the average over all pairs (X_1, X_2) with constant imbalance i_1/ϑ and i_2/ϑ is the same as if we considered only one particular random variable X_1 with imbalance i_1/ϑ and then took the average over all random variables X_2 with $I(X_2) = i_2/\vartheta$ constant. Because $I(X_2 \oplus 1) = I(X_2)$ and $I(X_1 \oplus X_2 \oplus 1) = I(X_1 \oplus X_2)$, it is even enough to consider only all random variables X_2 for which $P[X_2 = 1]$ is some constant greater than or equal to $1/2$.

A simple choice for X_1 is the random variable that is 1 on the first $\vartheta + i_1$ elements of Ω, and equal to 0 on the remaining ones, where the elements of Ω have been numbered in some way and $i_1 \geq 0$. The imbalance of X_1 is equal to i_1/ϑ. Let then $P[X_2 = 1] = (\vartheta + i_2)/2\vartheta$ with $i_2 \geq 0$, which

gives $I(X_2) = i_2/\vartheta$. There are $\binom{2\vartheta}{\vartheta+i_2}$ such random variables X_2. Now if $P_{X_1,X_2}(1,1) = m/2\vartheta$ for some m, then $P_{X_1,X_2}(1,0) = (\vartheta + i_1 - m)/2\vartheta$, $P_{X_1,X_2}(0,1) = (\vartheta + i_2 - m)/2\vartheta$ and $P_{X_1,X_2}(0,0) = (m - i_1 - i_2)/2\vartheta$, that is, $P[X_1 \oplus X_2 = 0] = (2m - i_1 - i_2)/2\vartheta$, which implies $I(X_1 \oplus X_2) = |2m - i_1 - i_2 - \vartheta|/\vartheta$. For any m, there are $\binom{\vartheta+i_1}{m}\binom{\vartheta-i_1}{\vartheta+i_2-m} = \binom{\vartheta+i_1}{m}\binom{\vartheta-i_1}{m-i_1-i_2}$ random variables X_2 with $P_{X_2}(1) = (\vartheta+i_2)/2\vartheta$ and $P_{X_1,X_2}(1,1) = m/2\vartheta$.

Which values can m take on? Obviously, $0 \le m \le \vartheta + i_1$; it must also satisfy $0 \le m - i_1 - i_2 \le \vartheta - i_1$, i.e., $i_1 + i_2 \le m \le \vartheta + i_2$. Altogether, we must have $i_1 + i_2 \le m \le \vartheta + \min(i_1, i_2)$. The average we want to compute is symmetrical in i_1 and i_2 so we can assume $i_1 \le i_2$, which implies $i_1 + i_2 \le m \le \vartheta + i_1$. Thus, we have proved the following lemma.

Lemma 3.3.6
Let ϑ, i_1 and i_2 be integers with $\vartheta \ge 1$ and $0 \le i_1, i_2 \le \vartheta$. Then the average of $I(X_1 \oplus X_2)$ over all pairs of random variables (X_1, X_2) such that $I(X_1) = i_1/\vartheta$ and $I(X_2) = i_2/\vartheta$ is

$$E_\vartheta\left[I(X_1 \oplus X_2)\middle| I(X_1) = \frac{i_1}{\vartheta}, I(X_2) = \frac{i_2}{\vartheta}\right] = f(\vartheta, i_1, i_2), \tag{3.8}$$

where

$$f(\vartheta, i, j) := \frac{1}{\vartheta\binom{2\vartheta}{\vartheta+j}} \sum_{m=i+j}^{\vartheta+i} \left|2m - \vartheta - i - j\right| \binom{\vartheta+i}{m}\binom{\vartheta-i}{m-i-j} \tag{3.9}$$

if $0 \le i \le j \le \vartheta$, and $f(\vartheta, i, j) := f(\vartheta, j, i)$ if $i > j$. □

Next, we will be concerned with the properties of the function f. In general, there is no simple form for $f(\vartheta, i, j)$ because of the absolute value in its definition. But if i or j is zero, we can show the following.

Lemma 3.3.7
Let $\vartheta \ge 1$ and $k \ge 0$ be integers. Then

1. For ϑ even, $\vartheta = 2\alpha$:

$$f(2\alpha, 0, 2k) = \frac{(\alpha+k)(\alpha-k)}{2\alpha^2} \frac{\binom{2\alpha}{\alpha+k}^2}{\binom{4\alpha}{2\alpha+2k}};$$

$$f(2\alpha, 0, 2k+1) = \frac{(\alpha+k+1)(\alpha-k)}{2\alpha^2} \frac{\binom{2\alpha}{\alpha+k+1}\binom{2\alpha}{\alpha+k}}{\binom{4\alpha}{2\alpha+2k+1}}.$$

2. *For ϑ odd, $\vartheta = 2\alpha + 1$:*

$$f(2\alpha+1,0,2k) = \frac{2(\alpha+k+1)(\alpha-k+1)}{(2\alpha+1)^2} \frac{\binom{2\alpha+1}{\alpha+k+1}\binom{2\alpha+1}{\alpha+k}}{\binom{4\alpha+2}{2\alpha+2k+1}};$$

$$f(2\alpha+1,0,2k+1) = \frac{2(\alpha+k+1)(\alpha-k)}{(2\alpha+1)^2} \frac{\binom{2\alpha+1}{\alpha+k+1}^2}{\binom{4\alpha+2}{2\alpha+2k+2}}.$$

Proof:

1. In this case, for $0 \le j \le 2\alpha$,

$$f(2\alpha,0,j) = \frac{1}{2\alpha\binom{4\alpha}{2\alpha+j}} \sum_{m=j}^{2\alpha} \left|2m-2\alpha-j\right| \binom{2\alpha}{m}\binom{2\alpha}{m-j}.$$

We use the fact that the summand is the same for m as for $2\alpha+j-m$. This implies that

$$f(2\alpha,0,j) = \frac{1}{\alpha\binom{4\alpha}{2\alpha+j}} \sum_{m=j}^{A} \left(2\alpha+j-2m\right) \binom{2\alpha}{m}\binom{2\alpha}{m-j},$$

where A stands for $\alpha + j/2 - 1$ if j is even, and for $\alpha + (j-1)/2$ if j is odd. We have gotten rid of the unpleasant absolute value. Now, by Gosper's method [14], we find $T(m)$ such that the term in the new sum is equal to $T(m+1) - T(m)$, namely

$$T(m) = \frac{m(m-j)}{2\alpha} \binom{2\alpha}{m}\binom{2\alpha}{m-j}.$$

Then the sum is equal to $\sum_{m=j}^{A}(T(m+1)-T(m)) = T(A+1)-T(j) = T(A+1)$, since $T(j) = 0$. Dividing by $\alpha\binom{4\alpha}{2\alpha+j}$ completes the proof.

2. The proof is very much the same as in 1. Here, for all $0 \le j \le 2\alpha+1$,

$$f(2\alpha+1,0,j) = \frac{1}{(2\alpha+1)\binom{4\alpha+2}{2\alpha+1+j}} \sum_{m=j}^{2\alpha+1} \left|2m-2\alpha-j-1\right| \binom{2\alpha+1}{m}\binom{2\alpha+1}{m-j}.$$

We use the fact that the summand is the same for m as for $2\alpha+1+j-m$. Thus, we have

$$f(2\alpha+1,0,j) = \frac{2}{(2\alpha+1)\binom{4\alpha+2}{2\alpha+1+j}} \sum_{m=j}^{A} \left(2\alpha+1+j-2m\right) \binom{2\alpha+1}{m}\binom{2\alpha+1}{m-j},$$

where A stands for $\alpha + j/2$ if j is even, and for $\alpha + (j-1)/2$ if j is odd. Now, again by Gosper's method, we find $T(m)$ such that the expression inside the sum is equal to $T(m+1) - T(m)$, namely

$$T(m) = \frac{m(m-j)}{2\alpha+1}\binom{2\alpha+1}{m}\binom{2\alpha+1}{m-j}.$$

Then the sum is equal to $T(A+1) - T(j) = T(A+1)$, since $T(j) = 0$. Dividing by $(2\alpha+1)\binom{4\alpha+2}{2\alpha+1+j}$ and multiplying by two completes the proof. □

Remark 3.3.8
One could argue that, in linear cryptanalysis, one avoids at all costs having balanced threefold sums since then the piling-up approximation yields $I(T_1 \oplus \cdots \oplus T_{r-1}) = I(T_1) \cdots I(T_{r-1}) = 0$, that is, one lower-bounds the average-key imbalance of the $(r-1)$-round I/O sum $S^{1\ldots r-1}$ with the trivial lower bound 0. Thus, it is pointless to consider the case $i_1 = 0$ (or $i_2 = 0$), because this corresponds to $I(X_1) = i_1/\vartheta = 0$ (or to $I(X_2) = i_2/\vartheta = 0$). Nonetheless, Lemma 3.3.7 is useful as we will see later (Proposition 3.3.14).

We next prove that $f(\vartheta, i, j) \geq ij/\vartheta^2$ and that, for large ϑ, $f(\vartheta, i, j) \leq ij/\vartheta^2 + c/\sqrt{\vartheta}$ for some constant c. We need the following definition.

Definition 3.3.9
For all integers ϑ, i, j for which it exists, define

$$g(\vartheta, i, j) := \begin{cases} \frac{1}{\vartheta\binom{2\vartheta}{\vartheta+j}} \sum_{m=i+j}^{\vartheta+i} (2m - i - j - \vartheta)\binom{\vartheta+i}{m}\binom{\vartheta-i}{m-i-j}, & i \leq j \\ g(\vartheta, j, i), & i > j. \end{cases} \tag{3.10}$$

Lemma 3.3.10
For all integers ϑ, i, j such that $\vartheta \geq 1$ and $0 \leq i, j \leq \vartheta$, we have $g(\vartheta, i, j) = ij/\vartheta^2$.
Proof:

Because both $g(\vartheta, i, j)$ and ij/ϑ^2 are symmetrical in i and j, it is enough to prove the lemma for $i \leq j$. By (3.10), for $i \leq j$, $g(\vartheta, i, j)$ is the expected value of $(2M - i - j - \vartheta)/\vartheta$ where M is a hypergeometrically distributed random variable, that is,

$$P_M(m) = \binom{\vartheta+i}{m}\binom{\vartheta-i}{m-i-j} \Big/ \binom{2\vartheta}{\vartheta+j}.$$

But then $E[M] = (\vartheta+i)(\vartheta+j)/2\vartheta$ [7] and the lemma follows immediately.□

Now we have $f(\vartheta, i_1, i_2) \geq g(\vartheta, i_1, i_2) = i_1 i_2/\vartheta^2 = I(X_1)I(X_2)$ for $i_1 \leq i_2$. Because f and g obey the same symmetry rule in i_1, i_2, this property extends to all $0 \leq i_1, i_2 \leq \vartheta$ so we have proved:

Proposition 3.3.11
Let ϑ, i_1 and i_2 be integers with $\vartheta \geq 1$ and $0 \leq i_1, i_2 \leq \vartheta$. Then the average of $I(X_1 \oplus X_2)$ over all pairs of random variables (X_1, X_2) such that $I(X_1) = i_1/\vartheta$ and $I(X_2) = i_2/\vartheta$ is lower-bounded by $i_1 i_2/\vartheta^2$. □

Now we want to find out by how much, on the average, $I(X_1 \oplus X_2)$ is greater than $I(X_1)I(X_2)$. To this end, we make the following definition:

Definition 3.3.12
For all integers ϑ, i, j such that $\vartheta \geq 1$ and $0 \leq i, j \leq \vartheta$, define $\tilde{f}(\vartheta, i, j) := f(\vartheta, i, j) - g(\vartheta, i, j) = f(\vartheta, i, j) - ij/\vartheta^2$.

This function has the following properties:

Lemma 3.3.13
Let ϑ, i, j be integers. If $\vartheta \geq 1$ and $0 \leq i \leq j \leq \vartheta$, then $\tilde{f}(\vartheta, i, j)$ can be written as:

1. $\tilde{f}(\vartheta, i, j) = \frac{2}{\vartheta\binom{2\vartheta}{\vartheta+j}} \sum_{m=j}^{\lfloor \frac{\vartheta-i+j-1}{2} \rfloor} (\vartheta - i + j - 2m)\binom{\vartheta+i}{m+i}\binom{\vartheta-i}{m-j}$.

Moreover, if $\vartheta \geq 1$ and $0 \leq i, j \leq \vartheta$, then the following are true:

2. *If $i + j \geq \vartheta$, then $\tilde{f}(\vartheta, i, j) = 0$, otherwise $\tilde{f}(\vartheta, i, j) > 0$.*
3. *If $i + j < \vartheta$, then $\tilde{f}(\vartheta, i+1, j+1) < \tilde{f}(\vartheta, i, j)$.*
4. *If $i + j < \vartheta$ then $\tilde{f}(\vartheta, i, j+2) < \tilde{f}(\vartheta, i, j)$ and $\tilde{f}(\vartheta, i+2, j) < \tilde{f}(\vartheta, i, j)$.*

Proof:

1. We have $\tilde{f}(\vartheta, i, j) = f(\vartheta, i, j) - g(\vartheta, i, j) =$

$$\frac{1}{\vartheta\binom{2\vartheta}{\vartheta+j}} \sum_{m=i+j}^{\vartheta+i} \Big[\big|2m-i-j-\vartheta\big| - (2m-i-j-\vartheta)\Big] \binom{\vartheta+i}{m} \binom{\vartheta-i}{m-i-j}.$$

For any real number a, $|a| - a$ is equal to $2(-a)$ if $a < 0$, and to 0 if $a \geq 0$. This explains the factor 2 above. But $2m - i - j - \vartheta$ is negative only for $m \leq (\vartheta + i + j - 2)/2$ if $\vartheta + i + j$ is even, and for $m \leq (\vartheta + i + j - 1)/2$ if $\vartheta + i + j$ is odd. Both cases are covered by the inequality $m \leq \lfloor (\vartheta + i + j - 1)/2 \rfloor$. Finally, we changed the index of summation by means of the substitution $m \to m + i$. (That is, we wrote everywhere $m + i$ instead of m.)

2. Both $\tilde{f}$ and the statement to be proved are symmetrical in i and j so it is enough to perform the proof for $i \leq j$.

   For $i+j \geq \vartheta$, we have $\lfloor(\vartheta-i+j-1)/2\rfloor \leq \lfloor(i+j-i+j-1)/2\rfloor = j-1$ so the sum's upper limit is smaller than its lower limit. The sum computes to 0.

   If $i+j < \vartheta$, say, $\vartheta = i+j+\alpha$ for some $\alpha \geq 1$, then the sum's upper limit is equal to $\lfloor(2j+\alpha-1)/2\rfloor$, which is larger or equal to j; thus, the summation is not empty. It remains to show that it always contains at least one non-zero summand. One shows easily that the binomial coefficients never vanish. They are multiplied by $\vartheta - i + j - 2m = 2j - 2m + \alpha$. For $\alpha = 1$ and $\alpha = 2$, the sum's upper limit is j so the sum consists of one summand with $m = j$, but then $2j-2m+\alpha = \alpha \neq 0$. For larger α, there is more than one summand and hence there is at least one m for which $2j - 2m + \alpha \neq 0$.

3. For the same reason as in 2., it is enough to prove it for $i \leq j$. If $i + j = \vartheta - 1$ or if $i + j = \vartheta - 2$, then the statement follows from 2. Otherwise, let $\vartheta > i + j + 2$ be fixed. Define $b(i,j,m) := (\vartheta - i + j - 2m)\binom{\vartheta+i}{m+i}\binom{\vartheta-i}{m-j}/\binom{2\vartheta}{\vartheta+j}$. Then, for $i \leq j$,

$$\tilde{f}(\vartheta,i,j) = \frac{2}{\vartheta}\sum_{m=j}^{\lfloor\frac{\vartheta-i+j-1}{2}\rfloor} b(i,j,m) \qquad \text{and}$$

$$\tilde{f}(\vartheta,i+1,j+1) = \frac{2}{\vartheta}\sum_{m=j+1}^{\lfloor\frac{\vartheta-i+j-1}{2}\rfloor} b(i+1,j+1,m).$$

   We show that $b(i+1,j+1,m) < b(i,j,m)$ for all integers m in the range $j+1 \leq m \leq (\vartheta - i + j - 1)/2$, for $\vartheta + i + j$ odd and even. This is obviously enough for our purpose. Now

$$\frac{b(i+1,j+1,m)}{b(i,j,m)} = \frac{(m-j)(\vartheta+i+1)(\vartheta+j+1)}{(m+i+1)(\vartheta-i)(\vartheta-j)} < 1 \quad \Leftrightarrow \quad (3.11)$$

$$(m-j)(\vartheta+i+1)(\vartheta+j+1) < (m+i+1)(\vartheta-i)(\vartheta-j). \quad (3.12)$$

   When we increase m, then the left side of (3.12) grows faster than the right side; it is therefore enough to show (3.12) for the largest possible value of m, i.e., for $m = (\vartheta - i + j - 1)/2$ (even if this might not be an integer). In that case, (3.12) reduces to

$$(\vartheta - i - j - 1)(\vartheta + i + 1)(\vartheta + j + 1) < (\vartheta + i + j + 1)(\vartheta - i)(\vartheta - j).$$

   This is true if and only if $\vartheta+i+j+2ij+1 > 0$, which obviously holds.

4. If $i+j=\vartheta-1$ or if $i+j=\vartheta-2$, then both statements follow from 2. Let $i+j<\vartheta-2$. Notice that because $\tilde{f}$ is symmetrical in i and j, both inequalities are equivalent. However, we will prove the first one for $i \le j$ and the second one for $i \le j-1$; then both inequalities hold for all i,j because:

   - $i>j \Rightarrow j \le i-1$ and $\tilde{f}(\vartheta,i,j+2)=\tilde{f}(\vartheta,j+2,i)<\tilde{f}(\vartheta,j,i)=\tilde{f}(\vartheta,i,j)$.
   - $i>j-1 \Rightarrow j \le i$ and $\tilde{f}(\vartheta,i+2,j)=\tilde{f}(\vartheta,j,i+2)<\tilde{f}(\vartheta,j,i)=\tilde{f}(\vartheta,i,j)$;

   Let $b(i,j,m)$ be defined as in the proof of 3. We have for $i \le j$

$$\tilde{f}(\vartheta,i,j) = \frac{2}{\vartheta} \sum_{m=j}^{\lfloor\frac{\vartheta-i+j-1}{2}\rfloor} b(i,j,m)\,;$$

$$\tilde{f}(\vartheta,i,j+2) = \frac{2}{\vartheta} \sum_{m=j+2}^{\lfloor\frac{\vartheta-i+j+1}{2}\rfloor} b(i,j+2,m) = \frac{2}{\vartheta} \sum_{m=j+1}^{\lfloor\frac{\vartheta-i+j-1}{2}\rfloor} b(i,j+2,m+1)\,;$$

$$\tilde{f}(\vartheta,i+2,j) = \frac{2}{\vartheta} \sum_{m=j}^{\lfloor\frac{\vartheta-i+j-3}{2}\rfloor} b(i+2,j,m) = \frac{2}{\vartheta} \sum_{m=j+1}^{\lfloor\frac{\vartheta-i+j-1}{2}\rfloor} b(i+2,j,m-1).$$

   Let $i \le j$. In order to prove the first property, we show that $b(i,j+2,m+1) < b(i,j,m)$ for $j+1 \le m \le (\vartheta-i+j+1)/2$. This holds because

$$\frac{b(i,j+2,m+1)}{b(i,j,m)} = \frac{(\vartheta-m)(m-j)(\vartheta-j)(\vartheta-j-1)}{(\vartheta-m-i+j+1)(m+i+1)(\vartheta+j+1)(\vartheta+j+2)},$$

   which is smaller than 1 since the i^{th} term in the numerator is smaller than the i^{th} term in the denominator and all terms are positive. (Remember, $i \le j$; furthermore, $\vartheta-m-i+j+1>0$ for all m in the above range.)

   We now turn to the second property. Let $i \le j-1$. We show that $b(i+2,j,m-1) < b(i,j,m)$ for $j+1 \le m \le (\vartheta-i+j+1)/2$. We have

$$\frac{b(i+2,j,m-1)}{b(i,j,m)} = \frac{(\vartheta+i+1)(\vartheta+i+2)(m-j)(\vartheta-i-m+j)}{(\vartheta-i)(\vartheta-i-1)(m+i+1)(\vartheta-m+1)}.$$

   For $0 \le i < \vartheta-2$ and $j+1 \le m \le (\vartheta-i+j+1)/2$, the denominator is positive. Now consider m as a real number. Then the derivative of the numerator with respect to m is $(\vartheta+i+1)(\vartheta+i+2)(\vartheta-i+2j-2m)$,

which is larger than $(\vartheta - i)(\vartheta - i - 1)(\vartheta - i - 2m)$, the derivative of the denominator. Thus, it suffices to show that the quotient is smaller than 1 for the largest value of m that we allow, that is, for $m = (\vartheta - i + j + 1)/2$. This happens if and only if

$$\vartheta^2(-2i + 4j - 2) + 4j\vartheta(i + j + 1) + (2i^3 + 4i^2 j + 6i^2 + 2ij^2 + 8ij + 6i + 2j^2 + 4j + 2) > 0.$$

This holds because we assumed that $i \leq j - 1$ so $-2i + 4j - 2 \geq 2i + 2 > 0$. □

We can thus say:

Proposition 3.3.14
Let $\tilde{f}$ be defined by Definition 3.3.12. Then $\tilde{f}$ is nonnegative and takes its maximum at $(\vartheta, 0, 1)$ and at $(\vartheta, 1, 0)$ for even ϑ, and at $(\vartheta, 0, 0)$ for odd ϑ. Moreover, for large ϑ, both maxima are approximately equal to $1/\sqrt{\pi\vartheta}$.

Proof:

Because of Lemma 3.3.13, $\tilde{f}$ is non-negative and the points where $\tilde{f}$ reaches its maximum can be only $(i, j) = (0, 0)$, $(i, j) = (0, 1)$ or $(i, j) = (1, 0)$. Because of symmetry, it suffices to compare $\tilde{f}(\vartheta, 0, 0)$ and $\tilde{f}(\vartheta, 0, 1)$. At these points, $\tilde{f}$ is equal to f so we can apply Lemma 3.3.7. If $\vartheta = 1$, then $f(1, 0, 1) = 0$ while $f(1, 0, 0) = 1$. If $\vartheta > 1$, then $\vartheta = 2\alpha$ or $\vartheta = 2\alpha + 1$ for some $\alpha \geq 1$. From Lemma 3.3.7, it follows that

$$\frac{f(2\alpha, 0, 1)}{f(2\alpha, 0, 0)} = \frac{f(2\alpha + 1, 0, 0)}{f(2\alpha + 1, 0, 1)} = \frac{2\alpha + 1}{2\alpha} > 1$$

for all $\alpha \geq 1$. Thus, the points where $\tilde{f}$ reaches its maximum are the ones mentioned. There,

$$\tilde{f}(2\alpha, 0, 1) = \frac{(\alpha + 1)\alpha}{2\alpha^2} \frac{\binom{2\alpha}{\alpha+1}\binom{2\alpha}{\alpha}}{\binom{4\alpha}{2\alpha+1}} = \frac{2\alpha + 1}{4\alpha} \frac{\binom{2\alpha}{\alpha}^2}{\binom{4\alpha}{2\alpha}} \quad \text{and}$$

$$\tilde{f}(2\alpha + 1, 0, 0) = \frac{2(\alpha + 1)^2}{(2\alpha + 1)^2} \frac{\binom{2\alpha+1}{\alpha+1}\binom{2\alpha+1}{\alpha+1}}{\binom{4\alpha+2}{2\alpha+1}} = \frac{2\alpha + 1}{4\alpha + 1} \frac{\binom{2\alpha}{\alpha}^2}{\binom{4\alpha}{2\alpha}}.$$

With Stirling's approximation for the factorial [7, 15], we have $\binom{2n}{n}^2 / \binom{4n}{2n} \approx \sqrt{2/\pi n}$ for large n. Thus, the maximum of $\tilde{f}$ for large ϑ is approximately equal to $1/\sqrt{2\pi\alpha} \approx 1/\sqrt{\pi\vartheta}$. □

Figure 3.5 shows the behaviour of $\tilde{f}(32, i, j)$. We summarize the above results and translate them into the language of imbalances and expectations and obtain:

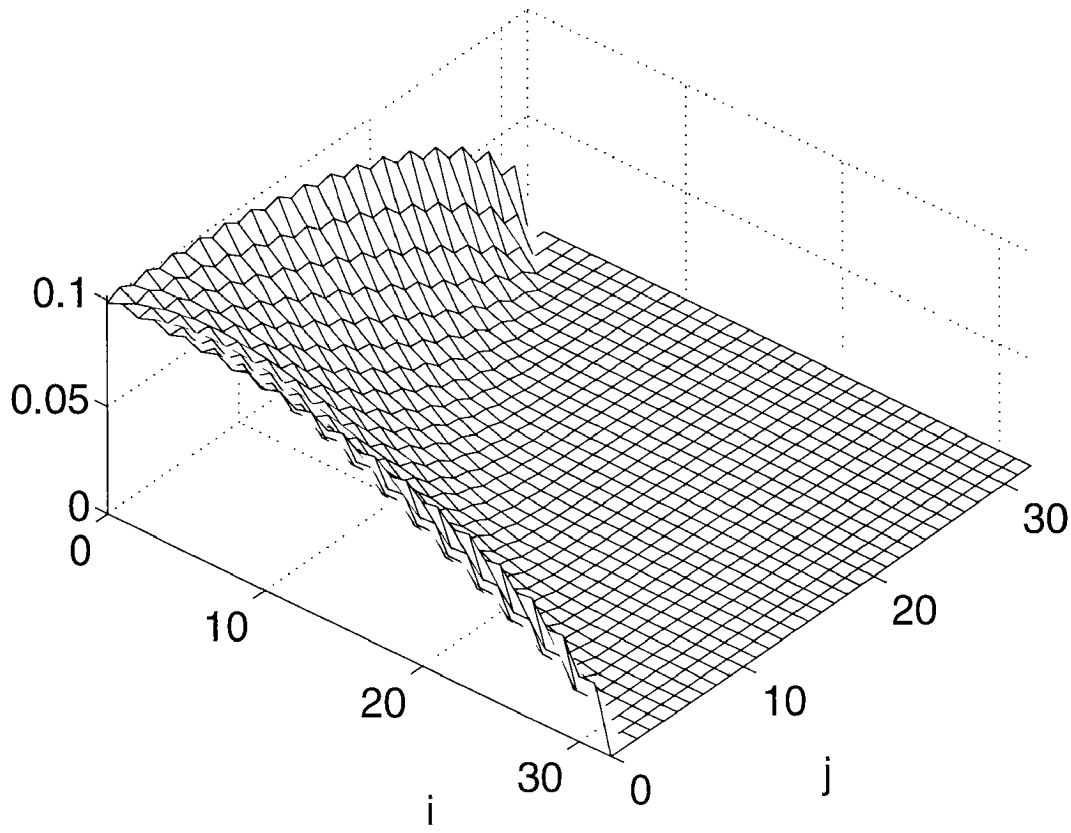


Figure 3.5: $\tilde{f}(\vartheta, i, j)$ for $\vartheta = 32$.

Theorem 3.3.15
Let Ω be some sample space with 2ϑ elements and i_1, i_2 be integers, $0 \leq i_1, i_2 \leq \vartheta$. Then the average of $I(X_1 \oplus X_2)$ over all pairs of random variables (X_1, X_2) such that $I(X_1) = i_1/\vartheta$ and $I(X_2) = i_2/\vartheta$ is lower-bounded by $i_1 i_2/\vartheta^2$ and upper-bounded by $i_1 i_2/\vartheta^2 + (1 + \varepsilon(\vartheta))/\sqrt{\pi\vartheta}$, where $\lim_{\vartheta\to\infty} \varepsilon(\vartheta) = 0$. □

Notice that because of the properties of $\tilde{f}$, the difference between $I(X_1)I(X_2)$ and the average of $I(X_1 \oplus X_2)$ tends to decrease as $I(X_1)$ or $I(X_2)$ increases and not only when ϑ does. In linear cryptanalysis, one is interested in large imbalances. Thus, from this point of view, the approximation of $I(X_1 \oplus X_2)$ by $I(X_1)I(X_2)$ is, on the average, a pessimistic one. However, the spread of the different values of $I(X_1 \oplus X_2)$ around the average might be wide, in which case the approximation is risky in the sense that in many cases $I(X_1 \oplus X_2)$ might be much smaller than $I(X_1)I(X_2)$. In order to remove this doubt, we now investigate the variance of $I(X_1 \oplus X_2)$ over all above mentioned pairs (X_1, X_2).

For that purpose, we first compute the average of $I^2(X_1 \oplus X_2)$. By the same argument as for the average of $I(X_1 \oplus X_2)$, it suffices to hold X_1 fixed and to let X_2 run over all random variables with $P_{X_2}(1) = (\vartheta + i_2)/2\vartheta$. The average is given by the following function.

Lemma 3.3.16
Let ϑ, i_1 and i_2 be integers with $\vartheta \geq 1$ and $0 \leq i_1, i_2 \leq \vartheta$. Then the average of $I^2(X_1 \oplus X_2)$ over all pairs of random variables (X_1, X_2) such that $I(X_1) = i_1/\vartheta$ and $I(X_2) = i_2/\vartheta$ is

$$E_\vartheta\left[I^2(X_1 \oplus X_2)\middle| I(X_1) = \frac{i_1}{\vartheta}, I(X_2) = \frac{i_2}{\vartheta}\right] = h(\vartheta, i_1, i_2), \tag{3.13}$$

where

$$h(\vartheta, i, j) := \frac{1}{\vartheta^2 \binom{2\vartheta}{\vartheta+j}} \sum_{m=i+j}^{\vartheta+i} \left(2m - \vartheta - i - j\right)^2 \binom{\vartheta+i}{m} \binom{\vartheta-i}{m-i-j} \tag{3.14}$$

if $0 \leq i \leq j \leq \vartheta$, and $h(\vartheta, i, j) := h(\vartheta, j, i)$ if $i > j$. □

Fortunately, $h(\vartheta, i, j)$ can be written in a simple form.

Lemma 3.3.17
For all integers ϑ, i, j such that $\vartheta \geq 1$ and $0 \leq i, j \leq \vartheta$,

$$\begin{aligned} h(\vartheta, i, j) &= \frac{1}{2\vartheta - 1}\left(1 + \frac{2i^2j^2 - \vartheta(i^2 + j^2)}{\vartheta^3}\right) \\ &= \frac{1}{2\vartheta - 1} + \frac{2\vartheta}{2\vartheta - 1}\frac{i^2}{\vartheta^2}\frac{j^2}{\vartheta^2} - \frac{1}{2\vartheta - 1}\left(\frac{i^2}{\vartheta^2} + \frac{j^2}{\vartheta^2}\right). \end{aligned} \tag{3.15}$$

Proof:

By the definition of h, it is enough to prove the lemma for $i \leq j$. Similarly to Lemma 3.3.10, we note that for $i \leq j$, $h(\vartheta, i, j)$ is the expected value of $(2M - i - j - \vartheta)^2/\vartheta^2$, where M is hypergeometrically distributed. We have [7]

$$E[M] = \frac{(\vartheta+i)(\vartheta+j)}{2\vartheta} \quad \text{and} \quad Var(M) = \frac{(\vartheta+i)(\vartheta+j)}{2\vartheta}\left(1 - \frac{\vartheta+i}{2\vartheta}\right)\frac{\vartheta-i}{2\vartheta-1}.$$

Then $h(\vartheta, i, j) = \frac{1}{\vartheta^2}\left[4E[M^2] - 4(\vartheta + i + j)E[M] + (\vartheta + i + j)^2\right]$ and the lemma follows. □

Now the variance of $I(X_1 \oplus X_2)$ is given by $V(\vartheta, i_1, i_2) := h(\vartheta, i_1, i_2) - f^2(\vartheta, i_1, i_2)$. Because $f(\vartheta, i_1, i_2) \geq i_1 i_2/\vartheta^2$, we can upper-bound the variance by

$$\begin{aligned} V(\vartheta, i_1, i_2) &\leq \frac{1}{2\vartheta - 1}\Big(1 + \frac{2i_1^2 i_2^2 - \vartheta(i_1^2 + i_2^2)}{\vartheta^3}\Big) - i_1^2 i_2^2/\vartheta^4 \\ &= \frac{(\vartheta^2 - i_1^2)(\vartheta^2 - i_2^2)}{\vartheta^4(2\vartheta - 1)} =: A(\vartheta, i_1, i_2). \end{aligned} \tag{3.16}$$

For $0 \leq i_1, i_2 \leq \vartheta$, A is minimal for $i_1 = \vartheta$ or $i_2 = \vartheta$, where $A = 0$, and maximal for $i_1 = i_2 = 0$ with $A(\vartheta, 0, 0) = \frac{1}{2\vartheta - 1}$. Hence, the expected value of $I(X_1 \oplus X_2) - I(X_1)I(X_2)$ lies between 0 and roughly $1/\sqrt{\pi\vartheta}$ and its standard deviation is upper-bounded by something of the order of $1/\sqrt{2\vartheta}$.

The bound (3.16) is rather loose. Numerical results show that $V(\vartheta, i_1, i_2)$ is much smaller than that. The *normalized variance*, defined for any random variable X by the variance of $X/E[X]$, and by infinity if $E[X] = 0$, is equal to $h/f^2 - 1$. For instance, if $i_1 = i_2 = 0$ and ϑ is large, this is approximately equal to $\pi/2 - 1 \approx 0.57$. The normalized variance seems to reach its maximum at $i_1 = 1, i_2 = \vartheta - 1$ (where it can be easily computed and is equal to $(\vartheta + 1)/(\vartheta - 1)$). With (3.16), the normalized variance is upper-bounded by $A\vartheta^4/i_1^2 i_2^2 - 1 = \frac{(\vartheta^2 - i_1^2)(\vartheta^2 - i_2^2)}{(2\vartheta - 1) i_1^2 i_2^2}$, which is not bounded in the range $0 \leq i_1, i_2 \leq \vartheta$. Figure 3.6 shows the behaviour of the normalized variance for $\vartheta = 32$.

3.3.1 Implication for The Piling-up Approximation for Two Random Variables

In the piling-up approximation, one approximates $I(X_1 \oplus X_2)$ by $I(X_1)I(X_2)$. If $I(X_1 \oplus X_2) > I(X_1)I(X_2)$, then the approximation is pessimistic and we are on the safe side. But it can happen that $I(X_1 \oplus X_2)$ is much smaller than $I(X_1)I(X_2)$.

If ϑ is "small" and $I(X_1), I(X_2)$ are small, too, then the average of $I(X_1 \oplus X_2)$ is approximately $1/\sqrt{\pi\vartheta}$, which is much larger than $I(X_1)I(X_2)$; because in this case the normalized variance lies around 0.57, most of the values of $I(X_1 \oplus X_2)$ are larger than $I(X_1)I(X_2)$ and it is safe to approximate $I(X_1 \oplus X_2)$ by $I(X_1)I(X_2)$.

Let $I(X_1)$ and $I(X_2)$ be fixed, and let ϑ be "large". Then, by Theorem 3.3.15, we can say that the average of $I(X_1 \oplus X_2)$ is between $I(X_1)I(X_2)$ and $I(X_1)I(X_2) + 1/\sqrt{\pi\vartheta}$; thus, it is close to $I(X_1)I(X_2)$. Moreover, for large ϑ, the average of $I^2(X_1 \oplus X_2)$ is by Lemma 3.3.17 equal to $h(\vartheta, i_1, i_2) = \frac{1}{2\vartheta - 1} + \frac{2\vartheta}{2\vartheta - 1} I^2(X_1)I^2(X_2) - \frac{1}{2\vartheta - 1}(I^2(X_1) + I^2(X_2)) \approx$

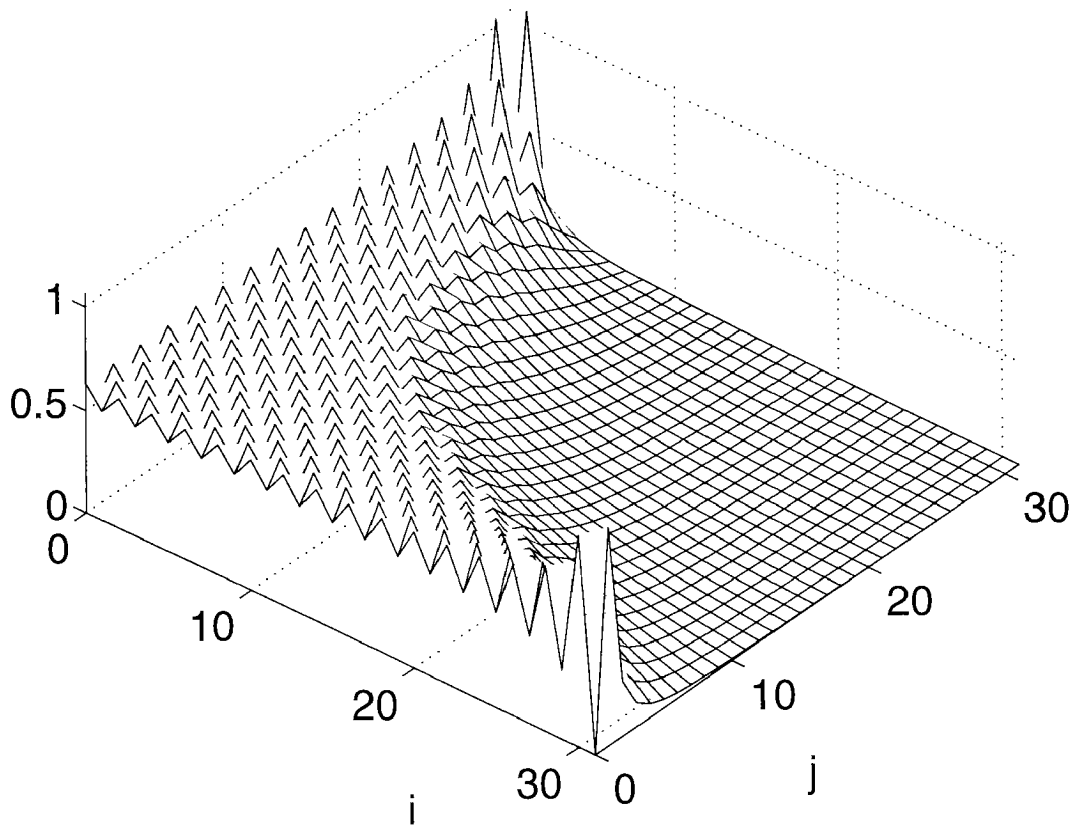


Figure 3.6: Normalized Variance $\frac{h(\vartheta,i,j)}{f^2(\vartheta,i,j)} - 1$ for $\vartheta = 32$.

$I^2(X_1)I^2(X_2)$. This means that most values of $I(X_1 \oplus X_2)$ are close to $I(X_1)I(X_2)$. Thus, if ϑ is large, one may use the piling-up approximation even for dependent random variables without risking too much.

3.4 Any Fixed Number of Random Variables

We now turn to the study of any number of random variables. We take some given integers $i_1, \ldots, i_r$ such that $0 \leq i_k \leq \vartheta$ for all k and look at the average of $I(X_1 \oplus \cdots \oplus X_r)$ and of $I^2(X_1 \oplus \cdots \oplus X_r)$ over all r-tuples of random variables $(X_1, \ldots, X_r)$ such that $I(X_k) = i_k/\vartheta$, $k = 1, \ldots, r$. If we wanted to compute the averages directly, e.g., by counting arguments as on Page 33, we would have to deal with sums over $2^{r-1} - 1$ indices which contain complicated multinomial coefficients and are therefore infeasible to compute. We will see that, fortunately, the averages can be computed recursively using properties of the probability distribution of $I(X_1 \oplus \cdots \oplus X_r)$ given that $I(X_1) = i_1/\vartheta, \ldots, I(X_r) = i_r/\vartheta$. We begin by calculating this probability distribution for $r = 2$.

3.4.1 The Conditional Probability Distribution of $I(X_1 \oplus X_2)$ given $I(X_1)$ and $I(X_2)$

Let X_1 be a random variable with $I(X_1) = i_1/\vartheta$. We have seen in Section 3.3 that for $0 \le i_1 \le i_2 \le \vartheta$, of the $\binom{2\vartheta}{\vartheta+i_2}$ random variables X_2 with $I(X_2) = i_2/\vartheta$, there were $\binom{\vartheta+i_1}{m}\binom{\vartheta-i_1}{m-i_1-i_2}$ that gave $I(X_1 \oplus X_2) = |2m - i_1 - i_2 - \vartheta|/\vartheta$, where $i_1 + i_2 \le m \le \vartheta + i_1$. Sometimes, there is an $m' \ne m$ such that $|2m' - i_1 - i_2 - \vartheta| = |2m - i_1 - i_2 - \vartheta|$. Thus, if we want to count the number of X_2 such that $I(X_1 \oplus X_2)$ is equal to, say, ℓ/ϑ, we must know for which ℓ there are two different values m and m' such that $|2m - i_1 - i_2 - \vartheta| = |2m' - i_1 - i_2 - \vartheta| = \ell$.

Now $|2m - i_1 - i_2 - \vartheta| = |2m' - i_1 - i_2 - \vartheta|$ for different m and m' only if $2m - i_1 - i_2 - \vartheta = \vartheta + i_1 + i_2 - 2m'$, i.e., only if $m' = \vartheta + i_1 + i_2 - m$. Such m and m' must satisfy $i_1 + i_2 \le m, m' \le \vartheta + i_1$; now $m' = \vartheta + i_1 + i_2 - m$ and $i_1 + i_2 \le m, m' \le \vartheta + i_1$ both hold if and only if $i_1 + i_2 \le m \le \vartheta$ and $m' = \vartheta + i_1 + i_2 - m$. Thus, if $m > \vartheta$, then m has no counterpart m' since this implies $m' < i_1 + i_2$. Finally, there is one value of m for which one must be careful: if $i_1 + i_2 \le \vartheta$ and $m = (\vartheta + i_1 + i_2)/2$ (which lies between $i_1 + i_2$ and ϑ so m' exists), then $m' = m$ so that in fact m has no counterpart. This corresponds to the case $\ell = 0$ and can happen only if $\vartheta + i_1 + i_2$ is even.

Let first $i_1 + i_2 > \vartheta$. Then $m \ge i_1 + i_2 > \vartheta$, so m has no counterpart m'. Moreover, $|2m - i_1 - i_2 - \vartheta| = 2m - i_1 - i_2 - \vartheta$. Thus, there are $\binom{\vartheta+i_1}{m}\binom{\vartheta-i_1}{m-i_1-i_2}$ random variables X_2 such that $I(X_1 \oplus X_2) = (2m - i_1 - i_2 - \vartheta)/\vartheta$. Since $(2m - i_1 - i_2 - \vartheta) = \ell$ if and only if $m = (\vartheta + i_1 + i_2 + \ell)/2$, there are $\binom{\vartheta+i_1}{\frac{1}{2}(\vartheta+i_1+i_2+\ell)}\binom{\vartheta-i_1}{\frac{1}{2}(\vartheta-i_1-i_2+\ell)}$ random variables X_2 such that $I(X_1 \oplus X_2) = \ell/\vartheta$. Furthermore, since $\ell = 2m - i_1 - i_2 - \vartheta$ and $i_1 + i_2 \le m \le \vartheta + i_1$, the probability is non-zero if and only if $\ell \equiv \vartheta + i_1 + i_2 \mod 2$ and $i_1 + i_2 - \vartheta \le \ell \le \vartheta + i_1 - i_2$.

Next, let $i_1 + i_2 = \vartheta$. Then $|2m - i_1 - i_2 - \vartheta| = 2|m - i_1 - i_2| = 2(m - i_1 - i_2) = 2m - i_1 - i_2 - \vartheta$. If $m > i_1 + i_2$, then $m > \vartheta$ and m has no counterpart m'. If $m = i_1 + i_2$, then $m = (\vartheta + i_1 + i_2)/2$ and again, as we have seen in the second paragraph, m has no counterpart because $m' = m$. (The latter case corresponds to $\ell = 0$.) Thus, we can draw the same conclusions as for $i_1 + i_2 > \vartheta$.

Now let $i_1 + i_2 < \vartheta < m$. Again m has no counterpart and $2m - i_1 - i_2 - \vartheta$ is positive so we can draw the same conclusions as for $i_1 + i_2 > \vartheta$.

Since $2m - i_1 - i_2 - \vartheta > \vartheta - i_1 - i_2$, the ℓ's that correspond to this case are the ones with $\ell > \vartheta - (i_1 + i_2)$.

Next, we look at the case $i_1 + i_2 < \vartheta$, $m \leq \vartheta$, $m \neq (\vartheta + i_1 + i_2)/2$. Here m' exists and $m' = \vartheta + i_1 + i_2 - m$. Thus, there are $\binom{\vartheta+i_1}{m}\binom{\vartheta-i_1}{m-i_1-i_2} + \binom{\vartheta+i_1}{m'}\binom{\vartheta-i_1}{m'-i_1-i_2} = \binom{\vartheta+i_1}{m}\binom{\vartheta-i_1}{m-i_1-i_2} + \binom{\vartheta+i_1}{m-i_2}\binom{\vartheta-i_1}{m-i_1}$ random variables X_2 such that $I(X_1 \oplus X_2) = |2m - i_1 - i_2 - \vartheta|/\vartheta$. Without losing generality, we can choose $m < m'$. (Otherwise, exchange m and m'.) Then $m < (\vartheta + i_1 + i_2)/2$ and therefore $|2m - i_1 - i_2 - \vartheta| = \vartheta + i_1 + i_2 - 2m$. Moreover, $\vartheta + i_1 + i_2 - 2m = \ell$ if and only if $m = (\vartheta + i_1 + i_2 - \ell)/2$ so there are $\binom{\vartheta+i_1}{(\vartheta+i_1+i_2-\ell)/2}\binom{\vartheta-i_1}{(\vartheta-i_1-i_2-\ell)/2} + \binom{\vartheta+i_1}{(\vartheta+i_1+i_2+\ell)/2}\binom{\vartheta-i_1}{(\vartheta-i_1-i_2+\ell)/2}$ random variables X_2 such that $I(X_1 \oplus X_2) = \ell/\vartheta$. Those ℓ corresponding to this case are the ones with $0 < \ell \leq \vartheta - (i_1 + i_2)$.

Finally, we consider $\ell = 0$, which occurs for $i_1 + i_2 \leq \vartheta$, $m = (\vartheta + i_1 + i_2)/2$ and $\vartheta + i_1 + i_2$ even. Similarly, we conclude that there are $\binom{\vartheta+i_1}{(\vartheta+i_1+i_2)/2}\binom{\vartheta-i_1}{(\vartheta-i_1-i_2)/2}$ random variables X_2 such that $I(X_1 \oplus X_2) = 0$.

If we divide everywhere by $\binom{2\vartheta}{\vartheta+i_2}$, the number of random variables X_2 with $I(X_2) = i_2/\vartheta$, we get the conditional probability distribution of $I(X_1 \oplus X_2)$ given $I(X_1)$ and $I(X_2)$. Thus, we have proved:

Proposition 3.4.1
Let X_1 and X_2 be random variables on a sample space Ω with 2ϑ elements, and let i_1, i_2 be integers, $0 \leq i_1 \leq i_2 \leq \vartheta$. Then the conditional probability $P := P_{I(X_1 \oplus X_2)|I(X_1)I(X_2)}(\frac{\ell}{\vartheta}|\frac{i_1}{\vartheta}, \frac{i_2}{\vartheta})$ that $I(X_1 \oplus X_2) = \ell/\vartheta$ given that $I(X_1) = i_1/\vartheta$ and $I(X_2) = i_2/\vartheta$ is equal to the following:

$$\begin{aligned}
i_1 + i_2 \geq \vartheta &\Rightarrow P = S(\ell), \\
i_1 + i_2 < \vartheta,\ \ell > \vartheta - (i_1 + i_2) &\Rightarrow P = S(\ell), \\
i_1 + i_2 < \vartheta,\ 0 < \ell \leq \vartheta - (i_1 + i_2) &\Rightarrow P = S(\ell) + S(-\ell), \quad \text{and} \\
\ell = 0 &\Rightarrow P = S(0) \quad \text{if } \vartheta + i_1 + i_2 \text{ is even,} \\
&\qquad \text{and } P = 0 \text{ otherwise,}
\end{aligned}$$

where $$S(\ell) = \binom{\vartheta + i_1}{\frac{1}{2}(\vartheta + i_1 + i_2 + \ell)}\binom{\vartheta - i_1}{\frac{1}{2}(\vartheta - i_1 - i_2 + \ell)} \Big/ \binom{2\vartheta}{\vartheta + i_2}.$$

If $i_1 > i_2$, then the above formulas hold after having exchanged i_1 and i_2. □

Remark 3.4.2
Because the lower index of the binomial coefficients must everywhere be an integer, we must have $\ell \equiv \vartheta + i_1 + i_2 \mod 2$. Otherwise the above probabilities are zero.

Hereafter in this chapter, unless stated otherwise, we write a conditional probability distribution without subscripts and without explicitely naming its arguments, e.g., $P(\ell|i_1,\ldots,i_r)$ will mean the probability that $I(X_1 \oplus \cdots \oplus X_r) = \ell/\vartheta$ given that $I(X_1) = i_1/\vartheta, \ldots, I(X_r) = i_r/\vartheta$.

Next, we derive the exact conditions for $P(\ell|i_1,i_2) = 0$.

Lemma 3.4.3
Let $0 \le i_1, i_2 \le \vartheta$. If $\ell \ge 0$ and $\ell \equiv \vartheta + i_1 + i_2 \mod 2$, then $P(\ell|i_1,i_2) = 0$ if and only if $\ell > \vartheta - |i_1 - i_2|$ or $\ell < i_1 + i_2 - \vartheta$. Otherwise, $P(\ell|i_1,i_2) = 0$ always holds.

Proof:

The last statement in the lemma follows immediately from the fact that $P(\ell|i_1,i_2) = 0$ for $\ell < 0$ and from Remark 3.4.2. Let now $\ell \ge 0$, $\ell \equiv \vartheta + i_1 + i_2 \mod 2$. By Proposition 3.4.1, we can write $P(\ell|i_1,i_2)$ as $P(\ell|i_1,i_2) = S(\ell) + \delta S(-\ell)$, where

$$S(\ell) = \binom{\vartheta + i_1}{\frac{1}{2}(\vartheta + i_1 + i_2 + \ell)} \binom{\vartheta - i_1}{\frac{1}{2}(\vartheta - i_1 - i_2 + \ell)} \Big/ \binom{2\vartheta}{\vartheta + i_2} \quad \text{and}$$

$$\delta = \begin{cases} 1, & i_1 + i_2 < \vartheta \quad \text{and} \quad 0 < \ell \le \vartheta - (i_1 + i_2), \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to check that $S(\ell) = 0$ if and only if $\ell > \vartheta - |i_1 - i_2|$ or $\ell < i_1 + i_2 - \vartheta$. Moreover, $\delta = 1$ only if $i_1 + i_2 < \vartheta$ and $0 < \ell \le \vartheta - (i_1 + i_2)$, but in that case neither $\ell > \vartheta - |i_1 - i_2|$ nor $\ell < i_1 + i_2 - \vartheta$ are satisfied so $S(\ell) > 0$ and thus $P > 0$. Thus, we have successively:

- $P = 0$, $\delta = 0$, $P = S(\ell)$, $S(\ell) = 0$, $\ell > \vartheta - |i_1 - i_2|$ or $\ell < i_1 + i_2 - \vartheta$;
- $\ell > \vartheta - |i_1 - i_2|$ or $\ell < i_1 + i_2 - \vartheta$, $S(\ell) = 0$ and $\delta = 0$, $P = S(\ell)$, $P = 0$.

□

Now the expected value of $I(X_1 \oplus X_2)$ and of $I^2(X_1 \oplus X_2)$ given $I(X_1)$ and $I(X_2)$ can be written as

$$f(\vartheta, i_1, i_2) = \sum_{\ell} \frac{\ell}{\vartheta} P(\ell|i_1, i_2), \tag{3.17}$$

$$h(\vartheta, i_1, i_2) = \sum_{\ell} \frac{\ell^2}{\vartheta^2} P(\ell|i_1, i_2), \tag{3.18}$$

where we can let the sums run over all the non-negative integers since only a finite number of terms are different from zero. From here on in this chapter, all sums without precise limits will be sums over all the non-negative integers. This spares us unnecessary perspiration when dealing with different parities of ϑ and of the i_k's.

3.4.2 Only One Random Variable

Before we turn to the study of more than two random variables, we take a brief look at the case of only one random variable. The statements here are all trivial but important for what follows. Let $0 \leq i_1 \leq \vartheta$. Then:

- the probability that $I(X_1) = \ell/\vartheta$ given that $I(X_1) = i_1/\vartheta$ is $P(\ell|i_1) = \delta_{\ell,i_1}$;
- the average of $I(X_1)$ over all random variables with $I(X_1) = i_1/\vartheta$ is $f(i_1) = i_1/\vartheta$;
- the average of $I^2(X_1)$ over all random variables with $I(X_1) = i_1/\vartheta$ is $h(i_1) = i_1^2/\vartheta^2$.

3.4.3 Various Properties For More Than Two Random Variables

We now consider expressions for more than two random variables. In the following, we show properties of the probability distribution of $I(X_1 \oplus \cdots \oplus X_r)$ given $I(X_1), \ldots, I(X_r)$, of its expected value and of the expected value of its square. These properties hold for any $r \geq 2$ unless stated otherwise. In particular, we show that the probability distribution of $I(X_1 \oplus \cdots \oplus X_r)$ given $I(X_1), \ldots, I(X_r)$ can be written with terms involving only the probability distribution of $I(X_1 \oplus X_2)$ given $I(X_1)$ and $I(X_2)$, and we deduce from this properties of the expected value of $I(X_1 \oplus \cdots \oplus X_r)$ and of $I^2(X_1 \oplus \cdots \oplus X_r)$ conditioned on particular values of $I(X_1), \ldots, I(X_r)$.

Proposition 3.4.4 (Recursion for P)
For any integers ϑ and $i_1, \ldots, i_r$ such that $\vartheta \geq 1$ and $0 \leq i_1, \ldots, i_r \leq \vartheta$, the probability $P(\ell|i_1, \ldots, i_r)$ that $I(X_1 \oplus \cdots \oplus X_r) = \ell/\vartheta$ given that $I(X_1) = i_1/\vartheta, \ldots, I(X_r) = i_r/\vartheta$, can be written recursively as

$$P(\ell|i_1, \ldots, i_r) = \sum_k P(\ell|k, i_r) P(k|i_1, \ldots, i_{r-1}). \tag{3.19}$$

Proof:

Here we write P explicitely with indices. We abbreviate $I(X_k)$ as I_k and $I(X_1 \oplus \cdots \oplus X_k)$ as $I_{1,\ldots,k}$, $k = 1, \ldots, r$. By the observations of Subsection 3.4.2, (3.19) holds for $r = 2$. Let now $r > 2$. We have

$$\begin{aligned}
P_{I_{1,\ldots,r}|I_1 \cdots I_r}(\ell|i_1, \ldots, i_r) &= \sum_k P_{I_{1,\ldots,r} I_{1,\ldots,r-1}|I_1 \cdots I_r}(\ell, k|i_1, \ldots, i_r) \\
&= \sum_k P_{I_{1,\ldots,r}|I_{1,\cdots,r-1} I_1 \ldots I_r}(\ell|k, i_1, \ldots, i_r) P_{I_{1,\ldots,r-1}|I_1 \cdots I_r}(k|i_1, \ldots, i_r) \\
&= \sum_k P_{I_{1,\ldots,r}|I_{1,\ldots,r-1} I_r}(\ell|k, i_r) P_{I_{1,\ldots,r-1}|I_1 \cdots I_{r-1}}(k|i_1, \ldots, i_{r-1}) \\
&= \sum_k P_{I_{1,2}|I_1 I_2}(\ell|k, i_r) P_{I_{1,\ldots,r-1}|I_1 \cdots I_{r-1}}(k|i_1, \ldots, i_{r-1}),
\end{aligned}$$

where in the penultimate equality we have used the facts that, if $I(X_1 \oplus \cdots \oplus X_{r-1})$ and $I(X_r)$ are given, then the probability distribution of $I(X_1 \oplus \cdots \oplus X_r)$ does not depend further on $I(X_1), \ldots, I(X_{r-1})$ and that $I(X_1 \oplus \cdots \oplus X_{r-1})$ does not depend on $I(X_r)$. To obtain the last equality we made the substitutions $X_1 \oplus \cdots \oplus X_{r-1} \to X_1$ and $X_r \to X_2$ in the first probability term. Notice that the above sequence of equalities also holds for $r = 2$. □

The next corollary follows now by induction.

Corollary 3.4.5
For any integers $\vartheta \geq 1$, $r \geq 3$ and $i_1, \ldots, i_r$ such that $0 \leq i_1, \ldots, i_r \leq \vartheta$,

$$P(\ell|i_1, \ldots, i_r) = \sum_{k_r, k_{r-1}, \ldots, k_3} P(\ell|k_r, i_r) \times P(k_r|k_{r-1}, i_{r-1}) \times \cdots \times P(k_4|k_3, i_3) \times P(k_3|i_1, i_2).$$

□

In analogy to the two-variable case, we introduce functions f and h for the expected values of $I(X_1 \oplus \cdots \oplus X_r)$ and $I^2(X_1 \oplus \cdots \oplus X_r)$, respectively, given that $I(X_1) = i_1/\vartheta, \ldots, I(X_r) = i_r/\vartheta$.

Definition 3.4.6
Let ϑ and $i_1, \ldots, i_r$ be integers such that $\vartheta \geq 1$ and $0 \leq i_1, \ldots, i_r \leq \vartheta$. Denote by $f(i_1, \ldots, i_r)$ (resp. $h(i_1, \ldots, i_r)$) the average of $I(X_1 \oplus \cdots \oplus X_r)$ (resp. of $I^2(X_1 \oplus \cdots \oplus X_r)$) over all r-tuples of random variables $(X_1, \ldots, X_r)$ such that $I(X_1) = i_1/\vartheta, \ldots, I(X_r) = i_r/\vartheta$, that is,

$$f(i_1, \ldots, i_r) := \sum_{\ell} \frac{\ell}{\vartheta} P(\ell | i_1, \ldots, i_r), \tag{3.20}$$

$$= E_\vartheta \Big[I(X_1 \oplus \cdots \oplus X_r) \Big| I(X_1) = \frac{i_1}{\vartheta}, \ldots, I(X_r) = \frac{i_r}{\vartheta} \Big],$$

$$h(i_1, \ldots, i_r) := \sum_{\ell} \Big(\frac{\ell}{\vartheta}\Big)^2 P(\ell | i_1, \ldots, i_r) \tag{3.21}$$

$$= E_\vartheta \Big[I^2(X_1 \oplus \cdots \oplus X_r) \Big| I(X_1) = \frac{i_1}{\vartheta}, \ldots, I(X_r) = \frac{i_r}{\vartheta} \Big].$$

(We cease writing the argument ϑ in f and h.)

We now look at some properties of f and h. We show that both functions can be defined recursively. We find lower and upper bounds for f and necessary and sufficient conditions for the lower bound to be achieved. We show that the bounds are close to each other when ϑ is large. Finally, we show that the value of f and h converges as r increases.

Lemma 3.4.7 (Recursion for f)
For any integers $\vartheta \geq 1$ and $0 \leq i_1, \ldots, i_r \leq \vartheta$, $f(i_1, \ldots, i_r) = \sum_k f(k, i_r) P(k | i_1, \ldots, i_{r-1})$.
Proof:

For $r = 2$, the equation follows from the observations of Subsection 3.4.2. For $r > 2$, we use Proposition 3.4.4 and equation (3.17).

$$f(i_1, \ldots, i_r) = \sum_{\ell} \frac{\ell}{\vartheta} P(\ell | i_1, \ldots, i_r) = \sum_{\ell,k} \frac{\ell}{\vartheta} P(\ell | k, i_r) P(k | i_1, \ldots, i_{r-1})$$

$$= \sum_k \Big(\sum_\ell \frac{\ell}{\vartheta} P(\ell | k, i_r) \Big) P(k | i_1, \ldots, i_{r-1}) = \sum_k f(k, i_r) P(k | i_1, \ldots, i_{r-1}).$$

Notice again that, thanks to the observations of Subsection 3.4.2, we could also perform the proof for $r = 2$ by the above sequence of equalities. □

This recursion formula for f allows us to generalise the properties of f from 2 to r random variables. As the first such property, we show that the expected value of $I(X_1 \oplus \cdots \oplus X_r)$ is lower-bounded by the product of imbalances $I(X_1) \cdots I(X_r)$.

Corollary 3.4.8
For any integers $\vartheta \geq 1$, $r \geq 1$, *and* $0 \leq i_1, \ldots, i_r \leq \vartheta$, $f(i_1, \ldots, i_r) \geq \frac{i_1}{\vartheta} \times \cdots \times \frac{i_r}{\vartheta}$.
Proof:

By the observations of Subsection 3.4.2, the corollary holds for $r = 1$. We showed in the preceding section that the corollary holds for $r = 2$. For $r > 2$, assume that the corollary holds up to $r - 1$. Then

$$\begin{aligned} f(i_1, \ldots, i_r) &= \sum_k f(k, i_r) P(k|i_1, \ldots, i_{r-1}) \\ &\geq \sum_k \frac{i_r k}{\vartheta^2} P(k|i_1, \ldots, i_{r-1}) = \frac{i_r}{\vartheta} \sum_k \frac{k}{\vartheta} P(k|i_1, \ldots, i_{r-1}) \\ &= \frac{i_r}{\vartheta} f(i_1, \ldots, i_{r-1}) \geq \frac{i_1}{\vartheta} \times \cdots \times \frac{i_r}{\vartheta}. \end{aligned}$$

□

Next, we give necessary and sufficient conditions for the lower bound of Corollary 3.4.8 to be achieved. For that, we use the following lemma.

Lemma 3.4.9
For any integers $\vartheta \geq 1$, $r \geq 1$, *and* $0 \leq i_1, \ldots, i_r \leq \vartheta$, *if* $\ell < i_1 + \cdots + i_r - (r-1)\vartheta$, *then* $P(\ell|i_1, \ldots, i_r) = 0$.
Proof:

By the observations of Subsection 3.4.2, the lemma holds for $r = 1$. By Lemma 3.4.3, the assertion holds also for $r = 2$. Let $r > 2$. Suppose that $P(k|i_1, \ldots, i_{r-1}) = 0$ for all k such that $k < i_1 + \cdots + i_{r-1} - (r-2)\vartheta$ and let $\ell < i_1 + \cdots + i_r - (r-1)\vartheta$. Then

$$\begin{aligned} P(\ell|i_1, \ldots, i_r) &= \sum_k P(\ell|k, i_r) P(k|i_1, \ldots, i_{r-1}) \\ &= \sum_{k \geq i_1 + \cdots + i_{r-1} - (r-2)\vartheta} P(\ell|k, i_r) P(k|i_1, \ldots, i_{r-1}). \end{aligned}$$

If $k \geq i_1 + \cdots + i_{r-1} - (r-2)\vartheta$, then $\ell < i_1 + \cdots + i_r - (r-1)\vartheta \leq k + i_r - \vartheta$. But then $P(\ell|k, i_r) = 0$ according to Lemma 3.4.3. Thus the last sum above is zero and the lemma follows. □

Notice that the statement of this lemma also follows immediately from inequality (3.3). However, the following corollary cannot be so derived.

Corollary 3.4.10
For any integer $r \geq 1$, the function f is equal to its lower bound in Corollary 3.4.8, i.e., $f(i_1, \ldots, i_r) = \frac{i_1}{\vartheta} \times \cdots \times \frac{i_1}{\vartheta}$, if and only if $i_1 + \cdots + i_r \geq (r-1)\vartheta$.

Proof:

By the observations of Subsection 3.4.2, the assertion of the corollary holds for $r = 1$. By Lemma 3.3.13, the assertion holds for $r = 2$. Let $r > 2$ and suppose that the corollary holds up to $r - 1$.

Let $i_1 + \cdots + i_r \geq (r-1)\vartheta$. Because $i_r \leq \vartheta$, we must have $i_1 + \cdots + i_{r-1} \geq (r-2)\vartheta$. Then $f(i_1, \ldots, i_{r-1}) = \frac{i_1}{\vartheta} \times \cdots \times \frac{i_{r-1}}{\vartheta}$ and

$$\begin{aligned}
f(i_1, \ldots, i_r) &= \sum_k f(k, i_r) P(k | i_1, \ldots, i_{r-1}) \\
&= \sum_{k \geq \vartheta - i_r} f(k, i_r) P(k | i_1, \ldots, i_{r-1}) \\
&= \frac{i_r}{\vartheta} \sum_{k \geq \vartheta - i_r} \frac{k}{\vartheta} P(k | i_1, \ldots, i_{r-1}) \\
&= \frac{i_r}{\vartheta} \sum_k \frac{k}{\vartheta} P(k | i_1, \ldots, i_{r-1}) \\
&= \frac{i_r}{\vartheta} f(i_1, \ldots, i_{r-1}) \quad = \quad \frac{i_1}{\vartheta} \times \cdots \times \frac{i_r}{\vartheta},
\end{aligned}$$

where in the third equality we used the fact that the assertion of the corollary holds for $r = 2$, and in the second and the fourth equalities the fact that $P(k | i_1, \ldots, i_{r-1}) = 0$ for $k < \vartheta - i_r$ as then $k < (r-1)\vartheta - (r-2)\vartheta - i_r \leq i_1 + \cdots + i_r - (r-2)\vartheta - i_r = i_1 + \cdots + i_{r-1} - (r-2)\vartheta$, which implies $P(k | i_1, \ldots, i_{r-1}) = 0$ by Lemma 3.4.9.

Let now $i_1 + \cdots + i_r < (r-1)\vartheta$. We examine the cases $i_1 + \cdots + i_{r-1} \geq (r-2)\vartheta$ and $i_1 + \cdots + i_{r-1} < (r-2)\vartheta$ separately.

Let $i_1 + \cdots + i_{r-1} \geq (r-2)\vartheta$. Then $f(i_1, \ldots, i_{r-1}) = \frac{i_1}{\vartheta} \times \cdots \times \frac{i_{r-1}}{\vartheta}$. If $k = i_1 + \cdots + i_{r-1} - (r-2)\vartheta$, then $k + i_r = i_1 + \cdots + i_r - (r-2)\vartheta < \vartheta$ so $f(k, i_r) > i_r k / \vartheta^2$. Thus, among all k's such that $k \geq i_1 + \cdots + i_{r-1} - (r-2)\vartheta$, there is at least one k with $f(k, i_r) > i_r k / \vartheta^2$. Hence,

$$\begin{aligned} f(i_1,\ldots,i_r) &= \sum_k f(k,i_r)P(k|i_1,\ldots,i_{r-1}) \\ &= \sum_{k\geq i_1+\cdots+i_{r-1}-(r-2)\vartheta} f(k,i_r)P(k|i_1,\ldots,i_{r-1}) \\ &> \frac{i_r}{\vartheta}\sum_{k\geq i_1+\cdots+i_{r-1}-(r-2)\vartheta} \frac{k}{\vartheta}\,P(k|i_1,\ldots,i_{r-1}) \\ &= \frac{i_r}{\vartheta}\sum_k \frac{k}{\vartheta}P(k|i_1,\ldots,i_{r-1}) \\ &= \frac{i_r}{\vartheta}f(i_1,\ldots,i_{r-1}) = \frac{i_1}{\vartheta}\times\cdots\times\frac{i_r}{\vartheta}, \end{aligned}$$

where to get the second and the fourth equalities we used Lemma 3.4.9.

If $i_1+\cdots+i_{r-1} < (r-2)\vartheta$, then $f(i_1,\ldots,i_{r-1}) > \frac{i_1}{\vartheta}\times\cdots\times\frac{i_{r-1}}{\vartheta}$. Thus,

$$\begin{aligned} f(i_1,\ldots,i_r) &= \sum_k f(k,i_r)P(k|i_1,\ldots,i_{r-1}) \\ &\geq \frac{i_r}{\vartheta}\sum_k \frac{k}{\vartheta}P(k|i_1,\ldots,i_{r-1}) \\ &= \frac{i_r}{\vartheta}f(i_1,\ldots,i_{r-1}) > \frac{i_1}{\vartheta}\times\cdots\times\frac{i_r}{\vartheta}. \end{aligned}$$

□

We translate our conclusions into the language of imbalances and expectations to obtain:

Proposition 3.4.11
Let Ω be some sample space with 2ϑ elements, let $r\in\mathbb{N}$, and let $i_1,\ldots,i_r$ be integers, $0\leq i_1,\ldots,i_r\leq\vartheta$. Then the average of $I(X_1\oplus\cdots\oplus X_r)$ over all r-tuples of random variables $(X_1,\ldots,X_r)$ such that $I(X_1)=i_1/\vartheta,\ldots,I(X_r)=i_r/\vartheta$ is lower-bounded by $i_1\times\cdots\times i_r/\vartheta^r$. The lower bound is attained if and only if $i_1+\cdots+i_r\geq(r-1)\vartheta$. □

Next, we show that the expected value of $I^2(X_1\oplus\cdots\oplus X_r)$ can be written recursively.

Lemma 3.4.12 (Recursion for h)
For any integers $\vartheta \geq 1$ and $0 \leq i_1, \ldots, i_r \leq \vartheta$,

$$h(i_1, \ldots, i_r) = \tag{3.22}$$
$$= \frac{1}{2\vartheta - 1}\left(1 + \frac{1}{\vartheta^3}\left(2i_r^2\vartheta^2 h(i_1, \ldots, i_{r-1}) - \vartheta(\vartheta^2 h(i_1, \ldots, i_{r-1}) + i_r^2)\right)\right)$$
$$= \frac{1}{2\vartheta - 1} + \frac{2\vartheta}{2\vartheta - 1}\frac{i_r^2}{\vartheta^2} h(i_1, \ldots, i_{r-1}) - \frac{1}{2\vartheta - 1}\left(h(i_1, \ldots, i_{r-1}) + \frac{i_r^2}{\vartheta^2}\right).$$

Proof:

For $r = 2$, the proof is made by noticing that (3.22) reduces to (3.15) if we apply the identity $h(i_1) = i_1^2/\vartheta^2$ of Subsection 3.4.2. For $r > 2$, we assume that the corollary holds up to $r - 1$ and use Proposition 3.4.4 and equation (3.18) to obtain

$$\begin{aligned}
h(i_1, \ldots, i_r) &= \sum_{\ell}\left(\frac{\ell}{\vartheta}\right)^2 P(\ell | i_1, \ldots, i_r) \\
&= \sum_{\ell,k}\left(\frac{\ell}{\vartheta}\right)^2 P(\ell | k, i_r) P(k | i_1, \ldots, i_{r-1}) \\
&= \sum_{k}\left(\sum_{\ell}\left(\frac{\ell}{\vartheta}\right)^2 P(\ell | k, i_r)\right) P(k | i_1, \ldots, i_{r-1}) \\
&= \sum_{k} h(k, i_r) P(k | i_1, \ldots, i_{r-1}) \\
&= \sum_{k} \frac{1}{2\vartheta - 1}\left(1 + \frac{1}{\vartheta^3}\left(2k^2 i_r^2 - \vartheta(k^2 + i_r^2)\right)\right) P(k | i_1, \ldots, i_{r-1}) \\
&= \frac{1}{2\vartheta - 1}\left(1 + \frac{1}{\vartheta^3}\left(2i_r^2\vartheta^2 h(i_1, \ldots, i_{r-1}) - \vartheta(\vartheta^2 h(i_1, \ldots, i_{r-1}) + i_r^2)\right)\right).
\end{aligned}$$

□

We now briefly investigate the value of f for small $i_1, \ldots, i_r$.

Approximation 3.4.13
If $i_1, \ldots, i_r$ are small and ℓ is large, then $P(\ell | i_1, \ldots, i_r) \approx 0$.
Motivation:

For $r = 2$ this follows from Proposition 3.4.1: i_1, i_2 small implies $i_1 + i_2 < \vartheta$; moreover, $\binom{\vartheta + i_1}{\frac{1}{2}(\vartheta + i_1 + i_2 + \ell)}$, $\binom{\vartheta - i_1}{\frac{1}{2}(\vartheta - i_1 - i_2 + \ell)}$, $\binom{\vartheta + i_1}{\frac{1}{2}(\vartheta + i_1 + i_2 - \ell)}$ and $\binom{\vartheta - i_1}{\frac{1}{2}(\vartheta - i_1 - i_2 - \ell)}$ are small compared to $\binom{2\vartheta}{\vartheta + i_2}$; thus, $P(\ell | i_1, i_2)$ is small. For $r > 2$, we suppose that the approximation holds up to $r - 1$ and use Proposition 3.4.4:

$$\begin{aligned} P(\ell|i_1,\ldots,i_r) &= \sum_k P(\ell|k,i_r)P(k|i_1,\ldots,i_{r-1}) \\ &\approx \sum_{k \text{ small}} P(\ell|k,i_r)P(k|i_1,\ldots,i_{r-1}) \approx 0, \end{aligned}$$

where both approximations hold by induction: the first because $P(k|i_1,\ldots,i_{r-1})$ is negligible for large k, and the second because $P(\ell|k,i_r)$ is negligible for small k. □

Approximation 3.4.14
If $i_1,\ldots,i_r$ are small, then $f(i_1,\ldots,i_r) \approx 1/\sqrt{\pi\vartheta}$ and $\tilde{f}(i_1,\ldots,i_r) \approx 1/\sqrt{\pi\vartheta}$.
Motivation:

Consider first $r = 2$. From the proof of points 3 and 4 in Lemma 3.3.13, we have

$$\tilde{f}(i_1,i_2) = \frac{2}{\vartheta} \sum_{m=i_2}^{\lfloor \frac{\vartheta - i_1 + i_2 - 1}{2} \rfloor} b(i_1,i_2,m)$$

$$\text{and} \quad \tilde{f}(i_1+1,i_2+1) = \frac{2}{\vartheta} \sum_{m=i_2+1}^{\lfloor \frac{\vartheta - i_1 + i_2 - 1}{2} \rfloor} b(i_1+1,i_2+1,m).$$

There is one term more in $\tilde{f}(i_1,i_2)$ than in $\tilde{f}(i_1+1,i_2+1)$, namely $\frac{2}{\vartheta}b(i_1,i_2,i_2)$, which is equal to $\frac{2}{\vartheta}(\vartheta - i_1 - i_2)\binom{\vartheta+i_1}{i_1+i_2}/\binom{2\vartheta}{\vartheta+i_2}$. For small i_1,i_2, by Stirling's approximation of the factorial [7], this is approximately equal to $4^{-\vartheta}(\frac{\vartheta}{i_1+i_2})^{i_1+i_2}\sqrt{2/\vartheta}$, which is small compared to $1/\sqrt{\pi\vartheta}$. The terms that appear in both sums can be compared by means of the ratio $b(i_1+1,i_2+1,m)/b(i_1,i_2,m)$ (equation (3.11)). For small m, this is small compared to 1, but $b(i_1,i_2,m)$ is also small so the terms for small m do not differ much. The terms that really influence the value of $\tilde{f}$ are those with m large; but for these $b(i_1+1,i_2+1,m)$ and $b(i_1,i_2,m)$ are almost equal. Thus, $\tilde{f}(i_1,i_2)$ and $\tilde{f}(i_1+1,i_2+1)$ are almost equal for small i_1,i_2.

Similar reasoning applies for $\tilde{f}(i_1,i_2+2)$ and $\tilde{f}(i_1+2,i_2)$. Hence, for small i_1,i_2, modifying i_1 and i_2 does not change the value of $\tilde{f}$ much. But, by Lemma 3.3.7 and Stirling's approximation, $\tilde{f}(0,0), \tilde{f}(0,1) \approx 1/\sqrt{\pi\vartheta}$. Thus, $\tilde{f}(i_1,i_2) \approx 1/\sqrt{\pi\vartheta}$ for small i_1,i_2. Now, because $\tilde{f}(i_1,i_2) - f(i_1,i_2) = i_1 i_2/\vartheta^2 \ll 1/\sqrt{\pi\vartheta}$, we also have that $f(i_1,i_2) \approx 1/\sqrt{\pi\vartheta}$ for small i_1,i_2.

For $r > 2$, suppose that the approximation is valid up to $r-1$. Then, by Proposition 3.4.7,

$$\begin{aligned} f(i_1,\ldots,i_r) &= \sum_k f(k,i_r)P(k|i_1,\ldots,i_{r-1}) \\ &\approx \sum_{k \text{ small}} f(k,i_r)P(k|i_1,\ldots,i_{r-1}) \approx \frac{1}{\sqrt{\pi\vartheta}} \sum_{k \text{ small}} P(k|i_1,\ldots,i_{r-1}) \\ &\approx \frac{1}{\sqrt{\pi\vartheta}} \sum_k P(k|i_1,\ldots,i_{r-1}) = \frac{1}{\sqrt{\pi\vartheta}}, \end{aligned}$$

where the first and the last approximation follow from Approximation 3.4.13 and the second approximation from the fact that the approximation to be proved holds for $r = 2$. Finally, $\tilde{f}(i_1,\ldots,i_r) \approx 1/\sqrt{\pi\vartheta}$ because $\tilde{f}(i_1,\ldots,i_r) - f(i_1,\ldots,i_r) = \frac{i_1}{\vartheta} \times \cdots \times \frac{i_r}{\vartheta} \ll 1/\sqrt{\pi\vartheta}$. □

In Approximations 3.4.13 and 3.4.14, we approximated sums over all k by sums over only the small k's; one could worry that this contradicts Lemma 3.4.9. But in the above approximations, we supposed that the i's were small, which implies that $i_1 + \cdots + i_{r-1} - (r-2)\vartheta$ is negative and so Lemma 3.4.9 does not apply. ($P(\ell|i_1,\ldots,i_r)$ is zero for negative ℓ.)

We translate this result into the language of imbalances and expectations to obtain:

Approximation 3.4.15
Let Ω be some sample space with 2ϑ elements, let $r \geq 2$ and let $i_1,\ldots,i_r$ be integers, $0 \leq i_1,\ldots,i_r \ll \vartheta$. Then the average of $I(X_1 \oplus \cdots \oplus X_r)$ over all r-tuples of random variables $(X_1,\ldots,X_r)$ such that $I(X_1) = i_1/\vartheta,\ldots,I(X_r) = i_r/\vartheta$ is approximately $1/\sqrt{\pi\vartheta}$. □

3.4.4 Large Sample Spaces

In a linear cryptanalysis attack, one often encounters threefold sums with a fairly large imbalance, typically 1/4 or 1/2, and ϑ equal to a very large power of two (see Remark 3.3.2). In that case, the lower bound $\frac{i_1}{\vartheta} \times \cdots \times \frac{i_r}{\vartheta}$ on the average of $I(X_1 \oplus \cdots \oplus X_r)$ might very well be (much) larger than $1/\sqrt{\pi\vartheta}$. Let r be fixed. Since ϑ is large, we can write $f(i_1,i_2) \leq i_1 i_2/\vartheta^2 + 1/\sqrt{\pi\vartheta}$ for all i_1, i_2. Then, by the recursion formula for f (Proposition 3.4.7), a simple induction shows that

$$f(i_1,\ldots,i_r) \le \frac{i_1}{\vartheta}\cdots\frac{i_r}{\vartheta} + \frac{1}{\sqrt{\pi\vartheta}}\Big(1+\frac{i_r}{\vartheta}\Big(1+\frac{i_{r-1}}{\vartheta}\Big(\cdots\frac{i_4}{\vartheta}\Big(1+\frac{i_3}{\vartheta}\Big)\cdots\Big)\Big)\Big)$$
$$\le \frac{i_1}{\vartheta}\cdots\frac{i_r}{\vartheta} + \frac{r-1}{\sqrt{\pi\vartheta}}.$$

Thus, if we let ϑ go to infinity while keeping $\frac{i_k}{\vartheta}$ constant, the product $\frac{i_1}{\vartheta}\cdots\frac{i_r}{\vartheta}$ dominates $\frac{r-1}{\sqrt{\pi\vartheta}}$ and we can conclude that $f(i_1,\ldots,i_r) \approx \frac{i_1}{\vartheta}\cdots\frac{i_r}{\vartheta}$. Before giving an example, we state this result, which is important for our work.

Approximation 3.4.16
Let Ω be some sample space with 2ϑ elements, let $r \ge 2$ and let $i_1,\ldots,i_r$ be integers, $0 \le i_1,\ldots,i_r \le \vartheta$. If ϑ is large enough so that $\frac{i_1}{\vartheta}\cdots\frac{i_r}{\vartheta}$ is much larger than $\frac{r-1}{\sqrt{\pi\vartheta}}$, then the average of $I(X_1\oplus\cdots\oplus X_r)$ over all r-tuples of random variables $(X_1,\ldots,X_r)$ such that $I(X_1)=i_1/\vartheta,\ldots,I(X_r)=i_r/\vartheta$ is approximately $\frac{i_1}{\vartheta}\cdots\frac{i_r}{\vartheta}$. □

Example 3.4.17
We cannot generally compute $f(i_1,\ldots,i_r)$ for large ϑ in practice, but we can show for small values of ϑ what happens as ϑ increases. Let $r=16$ and $(I(X_1),I(X_2),\ldots,I(X_r)) = \frac{1}{32}(3,4,5,7,9,11,12,12,14,16,17,21,23,28)$. The product of the imbalances is approximately $4.29*10^{-8}$. For $\vartheta=32$, the average of $I(X_1\oplus\cdots\oplus X_r)$ is approximately $3.13/32$, for $\vartheta=64$, it is approximately $4.47/64$, for $\vartheta=128$, it is approximately $6.35/128$, and for $\vartheta=256$, it is approximately $9.00/256$. One sees that as ϑ increases, the average decreases. As ϑ goes to infinity, the average converges to approximately $4.29*10^{-8}$.

And what happens to $h(i_1,\ldots,i_r)$, the average of $I^2(X_1\oplus\cdots\oplus X_r)$, as ϑ increases? By (3.15), we have $h(i_1,i_2) \approx \frac{i_1^2}{\vartheta^2}\frac{i_2^2}{\vartheta^2} = I^2(X_1)I^2(X_2)$ for large ϑ. The recursion formula (3.22) for h implies that $h(i_1,\ldots,i_r) \approx \frac{i_1^2}{\vartheta^2}\cdots\frac{i_r^2}{\vartheta^2} = I^2(X_1)\cdots I^2(X_r)$ for large ϑ.

Implication for the piling-up approximation: In this subsection, we have seen that, for large ϑ, the average of $I(X_1\oplus\cdots\oplus X_r)$ is approximately equal to $I(X_1)\cdots I(X_r)$ and that the average of $I^2(X_1\oplus\cdots\oplus X_r)$ is approximately equal to $I^2(X_1)\cdots I^2(X_r)$. Although this assertion is not

precise and does not give a quantitative value for the variance, it allows us to say that for large ϑ, $I(X_1 \oplus \cdots \oplus X_r)$ is approximately equal to $I(X_1) \cdots I(X_r)$ in most cases, even when the random variables in question are not independent.

3.5 Letting The Number of Random Variables Go To Infinity

In this last section of the chapter, we investigate the behaviour as r goes to infinity of the probability distribution of $I(X_1 \oplus \cdots \oplus X_r)$, given that $I(X_1) = \frac{i_1}{\vartheta}, \ldots, I(X_r) = \frac{i_r}{\vartheta}$. We begin with an example.

3.5.1 An Example

Example 3.5.1

Table 3.2 gives some examples of the value of $f(i_1, \ldots, i_r)$ and $h(i_1, \ldots, i_r)$ for $1 \leq r \leq 16$. We abbreviate the vector $(i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8, i_9, i_{10}, i_{11}, i_{12}, i_{13}, i_{14}, i_{15}, i_{16})$ as $\underline{i}$. The i-sequences are chosen increasing for convenience, but we recall that f and h are symmetrical in all their arguments.

(a) $\vartheta = 32$, $\underline{i} = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$;

(b) $\vartheta = 32$, $\underline{i} = (2,2,4,6,6,6,8,8,8,8,12,12,12,12,12,12)$;

(c) $\vartheta = 32$, $\underline{i} = (3,3,3,3,5,5,6,7,7,7,7,7,8,8,8,8)$;

(d) $\vartheta = 32$, $\underline{i} = (20,22,22,22,28,28,28,29,29,29,30,30,30,32,32,32)$;

(e) $\vartheta = 133$, $\underline{i} = (11,11,13,14,17,21,22,34,34,34,40,56,59,64,78,90)$.

The results suggest that the sequences $\big(f(i_1, \ldots, i_r)\big)_r$ and $\big(h(i_1, \ldots, i_r)\big)_r$ converge and that, if the arguments i_k are not too close to ϑ, then the convergence is very fast. How can this be explained?

The convergence of the values of h can be explained by the recursion formula (3.22). Let $(i_r)_{r \in \mathbb{N}}$ be some sequence of integers, $0 \leq i_r \leq \vartheta$. Define

$$F(x, i) := \frac{1}{2\vartheta - 1}\Big(1 - \frac{i^2}{\vartheta^2} + \Big(\frac{2i^2}{\vartheta} - 1\Big)x\Big).$$

Then, by (3.22), $h(i_1, \ldots, i_r) = F(h(i_1, \ldots, i_{r-1}), i_r)$, $r \geq 2$. Now

	(a)		(b)		(c)		(d)		(e)	
r	$\sqrt{\pi\vartheta}\cdot f$	$2\vartheta\cdot h$	$\sqrt{\pi\vartheta}\cdot f$	$2\vartheta\cdot h$	$\sqrt{\pi\vartheta}\cdot f$	$2\vartheta\cdot h$	$\sqrt{\pi\vartheta}\cdot f$	$2\vartheta\cdot h$	$\sqrt{\pi\vartheta}\cdot f$	$2\vartheta\cdot h$
1	0.00000	0.00000	0.62666	0.25000	0.93999	0.56250	6.26657	25.0000	1.69060	1.81955
2	0.98835	1.01587	0.98485	1.00893	0.98185	1.00304	4.30827	12.1429	1.00408	1.00253
3	0.98016	0.99975	0.98029	1	1.01191	0.99998	2.96193	6.17347	1.00284	1.00001
4	0.98029	1.00000	0.98029	1	0.98029	1.00000	2.05694	3.40197	0.99529	1.00000
5	0.98029	1.00000	0.98029	1	1.01191	1.00000	1.82871	2.83007	0.99529	1.00000
6	0.98029	1.00000	0.98029	1	0.98029	1.00000	1.64355	2.39434	0.99529	1.00000
7	0.98029	1.00000	0.98029	1	0.98029	1.00000	1.49558	2.06235	1.00283	1.00000
8	0.98029	1.00000	0.98029	1	1.01191	1.00000	1.42594	1.86949	0.99529	1.00000
9	0.98029	1.00000	0.98029	1	0.98029	1.00000	1.33254	1.71163	1.00283	1.00000
10	0.98029	1.00000	0.98029	1	1.01191	1.00000	1.29354	1.58244	0.99529	1.00000
11	0.98029	1.00000	0.98029	1	0.98029	1.00000	1.25983	1.51079	1.00283	1.00000
12	0.98029	1.00000	0.98029	1	1.01191	1.00000	1.23004	1.44795	0.99529	1.00000
13	0.98029	1.00000	0.98029	1	1.01191	1.00000	1.20375	1.39285	0.99529	1.00000
14	0.98029	1.00000	0.98029	1	1.01191	1.00000	1.20375	1.39285	1.00283	1.00000
15	0.98029	1.00000	0.98029	1	1.01191	1.00000	1.20375	1.39285	0.99529	1.00000
16	0.98029	1.00000	0.98029	1	1.01191	1.00000	1.20375	1.39285	1.00283	1.00000

Table 3.2: Value of $\sqrt{\pi\vartheta}\cdot f$ and $2\vartheta\cdot h$.

$$\Big|F(x,i) - F(y,i)\Big| = \frac{2i^2 - \vartheta}{\vartheta(2\vartheta - 1)}\Big|x - y\Big| \quad \text{and} \quad F(x,\vartheta) = x, \quad \text{all } x, y \text{ in } \mathbb{R}.$$

The absolute value of $\frac{2i^2-\vartheta}{\vartheta(2\vartheta-1)}$ is smaller than 1 for $0 \leq i < \vartheta$, and equal to 1 for $i = \vartheta$. The Banach Fixed-Point Theorem [28] tells us that, for $i < \vartheta$, $F(\cdot, i)$ has exactly one fixed point and that the iteration $x_{n+1} = F(x_n, i)$ converges to that fixed point for any starting value. The fixed point is easily shown to be $1/2\vartheta$ for all i; thus, the sequence $\big(h(i_1, \ldots, i_r)\big)_r$ converges to $1/2\vartheta$ if infinitely many i_k are smaller than ϑ. (The speed of convergence might vary from sequence to sequence.) Since $h(i_1, \ldots, i_{r-1}, \vartheta) = h(i_1, \ldots, i_{r-1})$, the convergence halts for one step each time $i_k = \vartheta$ occurs.

It is more complicated to show the convergence of f because we have no recursion formula similar to that for h. We begin our investigation of this matter by going back to the recursion formula for $P(\ell | i_1, \ldots, i_r)$, the conditional probability that $I(X_1 \oplus \cdots \oplus X_r) = \ell/\vartheta$ given that $I(X_1) = i_1/\vartheta, \ldots, I(X_r) = i_r/\vartheta$ (Proposition 3.4.4).

3.5.2 The Convergence of The Probability Distribution Explained With Markov Chains

In what follows, we consider a sequence $(i_r)_{r \in \mathbb{N}}$ of integers, $0 \leq i_r \leq \vartheta$, specified in advance, that has infinitely many terms different from ϑ. For each $r \geq 1$, let $W(r)$ be a random variable with probability distribution $P_{W(r)}(\ell) = P(\ell | i_1, \ldots, i_r)$. Note that $W(r)$ takes values in $\{0, 1, \ldots, \vartheta\}$. The sequence $(W(r))_r$ is a discrete-time stochastic process. Because

$$P(\ell | i_1, \ldots, i_r) = \sum_k P(\ell | k, i_r) P(k | i_1, \ldots, i_{r-1}), \tag{3.23}$$

we have

$$P_{W(r)}(\ell) = \sum_k P(\ell | k, i_r) P_{W(r-1)}(k). \tag{3.24}$$

In general, $(W(r))_r$ is not a Markov chain. But we can imagine a non-homogeneous Markov chain $(V(r))_r$ with values in $\{0, 1, \ldots, \vartheta\}$ and transition probabilities $P_{V(r)|V(r-1)}(\ell | k) = A_{k\ell}(i_r)$, where $A_{k\ell}(i_r) = P(\ell | k, i_r)$. Then we have $P_{V(r)}(\ell) = \sum_k P(\ell | k, i_r) P_{V(r-1)}(k)$. This implies that, if $W(0)$ and $V(0)$ have the same probability distribution, then $W(r)$ and

$V(r)$ also do for all $r \geq 1$. Our goal is to show that, if infinitely many of the i_r are smaller than ϑ, then $(V(r))_r$ has two steady-state distributions and to calculate these. This will explain the convergence of $f(i_1, \ldots, i_r)$ observed above. We denote the states of V by $0, \ldots, \vartheta$, where "being in state i" means that $V(r) = i$. Then $A_{k\ell}(i_r) = P(\ell|k, i_r)$ is the probability of going from state k to state ℓ. We do all our calculations for $V(r)$ and pass the conclusions allowed on to $W(r)$. We first show that the possible states at a particular "time instant" all have the same parity.

Lemma 3.5.2
If ℓ and $\vartheta + i_r + k$ do not have the same parity, then $A_{k\ell}(i_r) = 0$.
Proof:

This is a restatement of Remark 3.4.2. □

Because $P_{W(0)}(\ell) = P(\ell|i_1) = \delta_{\ell,i_1}$, $W(0)$ is a constant. Thus, at all time instants r, $W(r)$ has parity determined by $i_1, \ldots, i_r$. If we take $V(0) = i_1$, the same holds for $V(r)$.

To facilitate the study of the behaviour of $V(r)$, we consider the states with odd and even parity separately. We rewrite the transition matrix $A(i_r)$ with the rows (resp. columns) corresponding to the even-parity states in the upper (resp. left) half and the rows (resp. columns) corresponding to the odd-parity states in the lower (resp. right) half. We obtain in this manner a transition matrix $E(i_r)$ subdivided into four matrices $E_{ee}(i_r)$, $E_{eo}(i_r)$, $E_{oe}(i_r)$, and $E_{oo}(i_r)$, where $E_{ee}(i_r)$ is the submatrix that governs the transition from the even-parity states to the even-parity states as time increases from $r-1$ to r, $E_{eo}(i_r)$ is the submatrix governing the transition from the even-parity states to the odd-parity states, and so on, i.e.,

$$E(i_r) = \left(\begin{array}{c|c} E_{ee}(i_r) & E_{eo}(i_r) \\ \hline E_{oe}(i_r) & E_{oo}(i_r) \end{array} \right). \tag{3.25}$$

We choose to begin the indices of the submatrices at 0. For instance, the i,j-entry of $E_{eo}(i_r)$ is $P(2j+1|2i, i_r)$. The fact that $A_{k\ell}(i_r) = 0$ if ℓ and $\vartheta + i_r + k$ have different parity implies that, for all r, $E_{ee}(i_r) = E_{oo}(i_r) = 0$ if $i_r \equiv \vartheta + 1 \mod 2$ and $E_{eo}(i_r) = E_{oe}(i_r) = 0$ if $i_r \equiv \vartheta \mod 2$. We also wish to differentiate between ϑ even and ϑ odd because the transition matrices are not the same. We reserve the notation $E(i_r)$ only for ϑ even and write $O(i_r)$ for the transition matrix when ϑ is odd.

We now concentrate on the matrices $E_{ee}(i_r), E_{eo}(i_r), \ldots, O_{oo}(i_r)$. Because $0 \leq k, \ell \leq \vartheta$ in $P(\ell|k, i_r)$, their indices run from 0 to some maximum which is either $\vartheta/2$, $(\vartheta-1)/2$ or $\vartheta/2-1$. The form of these matrices, which depends only on i_r and on the parity of ϑ, will explain the convergence phenomenon encountered. We show in the appendix to this chapter that, for any ϑ and any $i_{r-1}, i_r < \vartheta$, each of the matrices $E_{ee}(i_r)$, $E_{oo}(i_r)$, $O_{ee}(i_r)$, $O_{oo}(i_r)$, $E_{oe}(i_{r-1})E_{eo}(i_r)$, $E_{eo}(i_{r-1})E_{oe}(i_r)$, $O_{oe}(i_{r-1})O_{eo}(i_r)$ and $O_{eo}(i_{r-1})O_{oe}(i_r)$, if it is not the zero matrix because of the parity of i_{r-1}, i_r and ϑ, induces a homogeneous, irreducible, recurrent and aperiodic Markov chain (Proposition 3.A.1). We now calculate the corresponding unique stationary distribution, which is at the same time their steady-state distribution [13]. We first make the following definition.

Definition 3.5.3

- For ϑ even, let

$$\boldsymbol{e_1} := \frac{1}{2^{2\vartheta-2}}\left(\frac{1}{2}\binom{2\vartheta}{\vartheta}, \binom{2\vartheta}{\vartheta-2}, \ldots, \binom{2\vartheta}{2}, \binom{2\vartheta}{0}\right) \quad \text{and}$$

$$\boldsymbol{e_2} := \frac{1}{2^{2\vartheta-2}}\left(\binom{2\vartheta}{\vartheta-1}, \binom{2\vartheta}{\vartheta-3}, \ldots, \binom{2\vartheta}{3}, \binom{2\vartheta}{1}\right).$$

- For ϑ odd, let

$$\boldsymbol{o_1} := \frac{1}{2^{2\vartheta-2}}\left(\binom{2\vartheta}{\vartheta-1}, \binom{2\vartheta}{\vartheta-3}, \ldots, \binom{2\vartheta}{2}, \binom{2\vartheta}{0}\right) \quad \text{and}$$

$$\boldsymbol{o_2} := \frac{1}{2^{2\vartheta-2}}\left(\frac{1}{2}\binom{2\vartheta}{\vartheta}, \binom{2\vartheta}{\vartheta-2}, \ldots, \binom{2\vartheta}{3}, \binom{2\vartheta}{1}\right).$$

These vectors are *probability vectors*, i.e., their components are non-negative and sum up to one. They have the following property.

Lemma 3.5.4

For any ϑ and any i_r, we have, if the matrices below are non-zero:

1. $\boldsymbol{e_1}E_{ee}(i_r) = \boldsymbol{e_1}$, $\boldsymbol{e_2}E_{oe}(i_r) = \boldsymbol{e_1}$, $\boldsymbol{e_1}E_{eo}(i_r) = \boldsymbol{e_2}$, $\boldsymbol{e_2}E_{oo}(i_r) = \boldsymbol{e_2}$ *if ϑ is even;*

2. $\boldsymbol{o_1}O_{oo}(i_r) = \boldsymbol{o_1}$, $\boldsymbol{o_2}O_{eo}(i_r) = \boldsymbol{o_1}$, $\boldsymbol{o_1}O_{oe}(i_r) = \boldsymbol{o_2}$, $\boldsymbol{o_2}O_{ee}(i_r) = \boldsymbol{o_2}$ *if ϑ is odd.*

Proof:

We show that $\boldsymbol{e_2}E_{oo}(i_r) = \boldsymbol{e_2}$ and that $\boldsymbol{e_1}E_{ee}(i_r) = \boldsymbol{e_1}$. The proof of the other equalities is similar. These matrices are non-zero if and only if ϑ and i_r are even.

The easier proof is that of the equality $\boldsymbol{e_2}E_{oo}(i_r) = \boldsymbol{e_2}$. We have to show that $\sum_{m=0}^{\vartheta/2-1} P(2s+1|2m+1,i_r)e_{2,m} = e_{2,s}$ for $0 \le s \le \vartheta/2-1$. Let

$$S_m(s) := \binom{\vartheta+2m+1}{\frac{1}{2}(\vartheta+2m+1+i_r+2s+1)}\binom{\vartheta-2m-1}{\frac{1}{2}(\vartheta-2m-1-i_r+2s+1)},$$

$$T_m(s) := \binom{\vartheta+2m+1}{\frac{1}{2}(\vartheta+2m+1+i_r-2s-1)}\binom{\vartheta-2m-1}{\frac{1}{2}(\vartheta-2m-1-i_r-2s-1)}.$$

Then, according to Proposition 3.4.1,

$$\binom{2\vartheta}{\vartheta+i_r}P(2s+1|2m+1,i_r) =$$

$$\begin{cases} S_m(s)+T_m(s), & 2m+1+i_r<\vartheta \text{ and } 0<2s+1\le\vartheta-2m-1-i_r\,; \\ S_m(s), & \text{otherwise.} \end{cases}$$

We note also that $T_m(s) = S_{-(m+1)}(s)$. Let $t_n = \frac{1}{2^{2\vartheta-2}}\binom{2\vartheta}{\vartheta-2n-1}$. Then $e_{2,m} = t_m$ for $m = 0,1,\ldots,\vartheta/2-1$ and $t_m = t_{-(m+1)}$ for all integers m. Let first $\frac{\vartheta-i_r}{2} - s \ge 1$. Then

$$\begin{aligned}
&\binom{2\vartheta}{\vartheta+i_r}\sum_{m=0}^{\vartheta/2-1} P(2s+1|2m+1,i_r)e_{2,m} \\
&= \sum_{m=0}^{\frac{\vartheta-i_r}{2}-1-s}\big(S_m(s)+T_m(s)\big)t_m + \sum_{m=\frac{\vartheta-i_r}{2}-s}^{\frac{\vartheta}{2}-1} S_m(s)t_m \\
&= \sum_{m=0}^{\frac{\vartheta-i_r}{2}-1-s}\big(S_m(s)t_m + S_{-(m+1)}(s)t_{-(m+1)}\big) + \sum_{m=\frac{\vartheta-i_r}{2}-s}^{\frac{\vartheta}{2}-1} S_m(s)t_m \\
&= \sum_{m=\frac{i_r-\vartheta}{2}+s}^{\frac{\vartheta}{2}-1} S_m(s)t_m.
\end{aligned}$$

Let now $\frac{\vartheta-i_r}{2} - s \le 0$. Then $\binom{2\vartheta}{\vartheta+i_r}P(2s+1|2m+1,i_r)e_{2,m} = S_m(s)$ for $0 \le m \le \frac{\vartheta}{2}-1$ so that

$$\binom{2\vartheta}{\vartheta+i_r}\sum_{m=0}^{\frac{\vartheta}{2}-1}P(2s+1|2m+1,i_r)e_{2,m}=\sum_{m=0}^{\frac{\vartheta}{2}-1}S_m(s)t_m=\sum_{m=\frac{i_r-\vartheta}{2}+s}^{\frac{\vartheta}{2}-1}S_m(s)t_m,$$

where we used the fact that $S_m(s)=0$ for $m<\frac{i_r-\vartheta}{2}+s$. Because of this and because $t_m=0$ for $m\geq\frac{\vartheta}{2}$, we have

$$\begin{aligned}
&\sum_{m=\frac{i_r-\vartheta}{2}+s}^{\frac{\vartheta}{2}-1}S_m(s)t_m=\sum_{m\in\mathbb{Z}}S_m(s)t_m\\
&=\frac{1}{2^{2\vartheta-2}}\sum_{m\in\mathbb{Z}}\binom{\vartheta+2m+1}{\frac{1}{2}(\vartheta+2m+1+i_r+2s+1)}\binom{\vartheta-2m-1}{\frac{1}{2}(\vartheta-2m-1-i_r+2s+1)}\binom{2\vartheta}{\vartheta-2m-1}\\
&=\frac{1}{2^{2\vartheta-2}}\sum_{m\in\mathbb{Z}}\binom{2\vartheta}{\frac{1}{2}(\vartheta+2m+i_r+2s+2),\frac{1}{2}(\vartheta+2m-i_r-2s),\frac{1}{2}(\vartheta-2m-i_r+2s),\frac{1}{2}(\vartheta-2m-i_r-2s-2)}\\
&=\frac{1}{2^{2\vartheta-2}}\binom{2\vartheta}{\vartheta-2s-1}\binom{2\vartheta}{\vartheta+i_r}=t_s\binom{2\vartheta}{\vartheta+i_r}=e_{2,s}\binom{2\vartheta}{\vartheta+i_r},
\end{aligned}$$

where in the third equality we used the fact that some factorials cancel each other and where in the fourth equality we used Proposition 3.A.2 with $a=2\vartheta$, $b=\frac{\vartheta+i_r}{2}+s+1$, $c=\frac{\vartheta-i_r}{2}-s$, $d=\frac{\vartheta-i_r}{2}+s$. Notice that up to the last equality actually the equations hold for any integer s.

The proof of $\boldsymbol{e_1}E_{ee}(i_r)=\boldsymbol{e_1}$ is a little more complicated, although similar. We have to show that $\sum_{m=0}^{\vartheta/2}P(2s|2m,i_r)e_{1,m}=e_{1,s}$ for $0\leq s\leq\frac{\vartheta}{2}$. Let

$$\begin{aligned}
S_m(s)&:=\binom{\vartheta+2m}{\frac{1}{2}(\vartheta+2m+i_r+2s)}\binom{\vartheta-2m}{\frac{1}{2}(\vartheta-2m-i_r+2s)}\qquad\text{and}\\
T_m(s)&:=\binom{\vartheta+2m}{\frac{1}{2}(\vartheta+2m+i_r-2s)}\binom{\vartheta-2m}{\frac{1}{2}(\vartheta-2m-i_r-2s)}.
\end{aligned}$$

Then, according to Proposition 3.4.1,

$$\binom{2\vartheta}{\vartheta+i_r}P(2s|2m,i_r)=$$

$$\begin{cases}S_m(s)+T_m(s), & 2m+i_r<\vartheta,\ 0<2s\leq\vartheta-2m-i_r;\\ S_m(s), & \text{otherwise.}\end{cases}$$

We note also that $T_m(s)=S_{-m}(s)$. Let $t_n=\frac{1}{2^{2\vartheta-2}}\binom{2\vartheta}{\vartheta-2n}$ for $n\neq 0$ and $t_0=\frac{1}{2}\frac{1}{2^{2\vartheta-2}}\binom{2\vartheta}{\vartheta}$. Then $e_{1,m}=t_m$ for $m=0,1,\ldots,\frac{\vartheta}{2}$ and $t_m=t_{-m}$ for all integers m. Let now $s>0$ and $\frac{\vartheta-i_r}{2}-s\geq 0$. Then

$$\binom{2\vartheta}{\vartheta+i_r}\sum_{m=0}^{\vartheta/2}P(2s|2m,i_r)e_{1,m}$$

$$= \sum_{m=0}^{\frac{\vartheta-i_r}{2}-s}\big(S_m(s)+T_m(s)\big)t_m + \sum_{m=\frac{\vartheta-i_r}{2}-s+1}^{\vartheta/2} S_m(s)t_m$$

$$= \sum_{m=0}^{\frac{\vartheta-i_r}{2}-s}\big(S_m(s)t_m + S_{-m}(s)t_{-m}\big) + \sum_{m=\frac{\vartheta-i_r}{2}-s+1}^{\vartheta/2} S_m(s)t_m$$

$$= \sum_{m=\frac{i_r-\vartheta}{2}+s}^{\vartheta/2} S_m(s)t_m = \frac{1}{2^{2\vartheta-2}}\sum_{m=\frac{i_r-\vartheta}{2}+s}^{\vartheta/2} S_m(s)\binom{2\vartheta}{\vartheta-2m},$$

where in the last equality we used the fact that, on its left, $S_0(s)t_0$ is counted twice but t_0 is the only one of the t_i's that has a factor $1/2$. Suppose now that $\frac{\vartheta-i_r}{2}-s<0$. Then $\binom{2\vartheta}{\vartheta+i_r}P(2s|2m,i_r)=S_m(s)$ for $0\le m\le \vartheta/2$ so that

$$\binom{2\vartheta}{\vartheta+i_r}\sum_{m=0}^{\vartheta/2}P(2s|2m,i_r)e_{1,m} = \sum_{m=0}^{\vartheta/2}S_m(s)t_m$$

$$= \sum_{m=\frac{i_r-\vartheta}{2}+s}^{\vartheta/2} S_m(s)t_m = \frac{1}{2^{2\vartheta-2}}\sum_{m=\frac{i_r-\vartheta}{2}+s}^{\vartheta/2} S_m(s)\binom{2\vartheta}{\vartheta-2m},$$

where we used the fact that $S_m(s)=0$ for $m<\frac{i_r-\vartheta}{2}+s$. Because of this and because $\binom{2\vartheta}{\vartheta-2m}=0$ for $m>\frac{\vartheta}{2}$, we have

$$\frac{1}{2^{2\vartheta-2}}\sum_{m=\frac{i_r-\vartheta}{2}+s}^{\vartheta/2} S_m(s)\binom{2\vartheta}{\vartheta-2m} = \frac{1}{2^{2\vartheta-2}}\sum_{m\in\mathbb{Z}} S_m(s)\binom{2\vartheta}{\vartheta-2m}$$

$$= \frac{1}{2^{2\vartheta-2}}\binom{2\vartheta}{\vartheta-2s}\binom{2\vartheta}{\vartheta+i_r} = t_s\binom{2\vartheta}{\vartheta+i_r} = e_{1,s}\binom{2\vartheta}{\vartheta+i_r},$$

where in the third last equality we used Proposition 3.A.2 again. For $s=0$, we have, by Proposition 3.4.1, $P(0|2m,i_r)=S_m(0)$ so that

$$\binom{2\vartheta}{\vartheta+i_r}\sum_{m=0}^{\vartheta/2}P(0|2m,i_r)e_{1,m} = \sum_{m=0}^{\vartheta/2}S_m(0)e_{1,m} = \sum_{m=0}^{\vartheta/2}S_m(0)t_m$$

$$= \frac{1}{2}\sum_{m=-\vartheta/2}^{\vartheta/2}S_m(0)\frac{1}{2^{2\vartheta-2}}\binom{2\vartheta}{\vartheta-2m} = \frac{1}{2}\frac{1}{2^{2\vartheta-2}}\sum_{m\in\mathbb{Z}}S_m(0)\binom{2\vartheta}{\vartheta-2m}$$

$$= \frac{1}{2}\frac{1}{2^{2\vartheta-2}}\binom{2\vartheta}{\vartheta}\binom{2\vartheta}{\vartheta+i_r} = t_0\binom{2\vartheta}{\vartheta+i_r} = e_{1,0}\binom{2\vartheta}{\vartheta+i_r}. \quad \square$$

In particular, we have $\boldsymbol{e_1}E_{eo}(i_{r-1})E_{oe}(i_r) = \boldsymbol{e_1}$, $\boldsymbol{e_2}E_{oe}(i_{r-1})E_{eo}(i_r) = \boldsymbol{e_2}$, $\boldsymbol{o_1}O_{oe}(i_{r-1})O_{eo}(i_r) = \boldsymbol{o_1}$, and $\boldsymbol{o_2}O_{eo}(i_{r-1})O_{oe}(i_r) = \boldsymbol{o_2}$. We have found the stationary distribution, and thus the steady-state distribution when $i_{r-1}, i_r < \vartheta$, of the eight implicitely defined homogeneous Markov chains mentioned above.

3.5.3 The Convergence of The Average

By Lemma 3.5.4 and because $E(i_r)$ and $O(i_r)$ are either of the form $\left(\begin{array}{c|c} * & 0 \\ \hline 0 & * \end{array}\right)$ or of the form $\left(\begin{array}{c|c} 0 & * \\ \hline * & 0 \end{array}\right)$ for different parities of ϑ and i_r, we have:

Corollary 3.5.5
Let $\vartheta \geq 1$ and i_r be integers, $0 \leq i_r \leq \vartheta$. Then

- *For ϑ even, i_r even, $(\boldsymbol{e_1}, \boldsymbol{0})E(i_r) = (\boldsymbol{e_1}, \boldsymbol{0})$ and $(\boldsymbol{0}, \boldsymbol{e_2})E(i_r) = (\boldsymbol{0}, \boldsymbol{e_2})$.*
- *For ϑ even, i_r odd, $(\boldsymbol{e_1}, \boldsymbol{0})E(i_r) = (\boldsymbol{0}, \boldsymbol{e_2})$ and $(\boldsymbol{0}, \boldsymbol{e_2})E(i_r) = (\boldsymbol{e_1}, \boldsymbol{0})$.*
- *For ϑ odd, i_r even, $(\boldsymbol{o_2}, \boldsymbol{0})O(i_r) = (\boldsymbol{0}, \boldsymbol{o_1})$ and $(\boldsymbol{0}, \boldsymbol{o_1})O(i_r) = (\boldsymbol{o_2}, \boldsymbol{0})$.*
- *For ϑ odd, i_r odd, $(\boldsymbol{o_2}, \boldsymbol{0})O(i_r) = (\boldsymbol{o_2}, \boldsymbol{0})$ and $(\boldsymbol{0}, \boldsymbol{o_1})O(i_r) = (\boldsymbol{0}, \boldsymbol{o_1})$,*

where $\boldsymbol{0}$ denotes a sequence of sufficiently many zeros to make the multiplication well defined. $\square$

Now let ϑ be even. The conclusion for ϑ odd are similar. For any $r \geq 1$, the product $E(i_1)\cdots E(i_r)$ is either of the form $\left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right)$ or of the form $\left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right)\left(\begin{array}{c|c} 0 & E_{eo}(i_r) \\ \hline E_{oe}(i_r) & 0 \end{array}\right)$, where A (resp. B) is a product of matrices that are either the identity matrix or matrices of the form $E_{ee}(i)$ or $E_{eo}(i)E_{oe}(j)$ (resp. $E_{oo}(i)$ or $E_{oe}(i)E_{eo}(j)$). By Corollary 3.A.10,

as r increases, $\boldsymbol{p}A$ (resp. $\boldsymbol{p}B$) converges to $\boldsymbol{e_1}$ (resp. $\boldsymbol{e_2}$) for any starting probability distribution $\boldsymbol{p}$. This implies that $(\boldsymbol{p}, \boldsymbol{0})\left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right)$ (resp. $(\boldsymbol{0}, \boldsymbol{p})\left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right)$) converges to $(\boldsymbol{e_1}, \boldsymbol{0})$ (resp. $(\boldsymbol{0}, \boldsymbol{e_2})$) for any probability distribution $\boldsymbol{p}$. Moreover,

$$(\boldsymbol{e_1}, \boldsymbol{0})\left(\begin{array}{c|c} 0 & E_{eo}(i_r) \\ \hline E_{oe}(i_r) & 0 \end{array}\right) = (\boldsymbol{0}, \boldsymbol{e_2}) \quad \text{and}$$

$$(\boldsymbol{0}, \boldsymbol{e_2})\left(\begin{array}{c|c} 0 & E_{eo}(i_r) \\ \hline E_{oe}(i_r) & 0 \end{array}\right) = (\boldsymbol{e_1}, \boldsymbol{0}).$$

It follows that, if $\boldsymbol{p}$ is the probability distribution of a constant random variable, any sequence $(\boldsymbol{p}E(i_1)\cdots E(i_r))_r$ such that infinitely many i_r are smaller than ϑ has the two accumulation points $(\boldsymbol{e_1}, \boldsymbol{0})$ and $(\boldsymbol{0}, \boldsymbol{e_2})$.

Thus, for any sequence $(i_r)_r$ that contains infinitely many terms smaller than ϑ, the sequence of probability distributions $\left(P_{I(X_1\oplus\cdots\oplus X_r)|I(X_1)I(X_2)\cdots I(X_r)}(\cdot|\frac{i_1}{\vartheta}, \frac{i_2}{\vartheta}, \ldots, \frac{i_r}{\vartheta})\right)_r$ has the two accumulation points $(e_{10}, 0, e_{11}, 0, e_{12}, 0, \ldots)$ and $(0, e_{20}, 0, e_{21}, 0, e_{22}, \ldots)$. (This is after rearrangement. Recall that $\boldsymbol{e_1}$ is a probability distribution on the even values beginning with 0, and that $\boldsymbol{e_2}$ is a probability distribution on the odd values.) Hence, the sequence of averages $(f(i_1, \ldots, i_r))_r$ has two accumulation points, namely the expected values of these two distributions. The expected values are computed below, based on the following Lemma.

Lemma 3.5.6
For any integer $\vartheta \geq 1$,

1. $\sum_{k=1}^{\vartheta} k\binom{2\vartheta}{\vartheta-k} = \frac{\vartheta}{2}\binom{2\vartheta}{\vartheta}$;

For any $\vartheta \geq 2$ even,

2. $\sum_{k=1}^{\vartheta/2} 2k\binom{2\vartheta}{\vartheta-2k} = \vartheta\binom{2\vartheta-2}{\vartheta-2}$;

3. $\sum_{k=0}^{\vartheta/2-1}(2k+1)\binom{2\vartheta}{\vartheta-(2k+1)} = \frac{\vartheta^2}{\vartheta-1}\binom{2\vartheta-2}{\vartheta-2}$, *which is equal to* $\vartheta\binom{2\vartheta-2}{\vartheta-1}$.

For any $\vartheta \geq 3$ odd,

4. $\sum_{k=0}^{(\vartheta-1)/2}(2k+1)\binom{2\vartheta}{\vartheta-(2k+1)} = \vartheta\binom{2\vartheta-2}{\vartheta-1}$;

5. $\sum_{k=1}^{(\vartheta-1)/2}(2k)\binom{2\vartheta}{\vartheta-2k} = (\vartheta-1)\binom{2\vartheta-2}{\vartheta-1}$, *which is equal to* $\vartheta\binom{2\vartheta-2}{\vartheta-2}$.

Proof:

1. With Gosper's method [14], we find $T(k)$ such that $T(k+1) - T(k) = k\binom{2\vartheta}{\vartheta-k}$, namely $T(k) = -\frac{1}{2}(\vartheta + k)\binom{2\vartheta}{\vartheta-k}$. Then $\sum_{k=1}^{\vartheta} k\binom{2\vartheta}{\vartheta-k} = \sum_{k=0}^{\vartheta} k\binom{2\vartheta}{\vartheta-k} = T(\vartheta+1) - T(0) = -T(0) = \frac{\vartheta}{2}\binom{2\vartheta}{\vartheta}$, since $T(\vartheta+1) = 0$.

2. With Gosper's method, one finds $T(k)$ such that $T(k+1) - T(k) = 2k\binom{2\vartheta}{\vartheta-2k}$, namely $T(k) = -\vartheta\binom{2\vartheta-2}{\vartheta+2k-2}$. Then $T(\frac{\vartheta}{2}+1) = -\vartheta\binom{2\vartheta-2}{2\vartheta} = 0$ and the sum is equal to $T(\frac{\vartheta}{2}+1) - T(0) = \vartheta\binom{2\vartheta-2}{\vartheta-2}$.

3. The left side of the equation is the left side of 2. substracted from the left side of 1.

4. Again with Gosper's method, one finds $T(k)$ such that $T(k+1) - T(k) = (2k+1)\binom{2\vartheta}{\vartheta-(2k+1)}$, namely $T(k) = -\vartheta\binom{2\vartheta-2}{\vartheta+2k-1}$. Now $T(\frac{\vartheta+1}{2}) = -\vartheta\binom{2\vartheta-2}{2\vartheta} = 0$ and the sum is equal to $T(\frac{\vartheta+1}{2}) - T(0) = \vartheta\binom{2\vartheta-2}{\vartheta-1}$.

5. The left side of the equation is the left side of 4. substracted from the left side of 1. □

Corollary 3.5.7

1. *The expected value of* $(e_{10}, 0, e_{11}, 0, e_{12}, 0, \ldots)$ *is* $\binom{2\vartheta-2}{\vartheta-2} \Big/ 2^{2\vartheta-2}$;

2. *The expected value of* $(0, e_{20}, 0, e_{21}, 0, e_{22}, \ldots)$ *is* $\binom{2\vartheta-2}{\vartheta-1} \Big/ 2^{2\vartheta-2}$;

3. *The expected value of* $(0, o_{10}, 0, o_{11}, 0, o_{12}, \ldots)$ *is* $\binom{2\vartheta-2}{\vartheta-1} \Big/ 2^{2\vartheta-2}$;

4. *The expected value of* $(o_{20}, 0, o_{21}, 0, o_{22}, 0, \ldots)$ *is* $\binom{2\vartheta-2}{\vartheta-2} \Big/ 2^{2\vartheta-2}$;

Proof:

Divide the last four identities of Lemma 3.5.6 by $\vartheta 2^{2\vartheta-2}$. □

With Stirling's approximation to the factorial, one shows that both $\binom{2\vartheta-2}{\vartheta-2} \Big/ 2^{2\vartheta-2}$ and $\binom{2\vartheta-2}{\vartheta-1} \Big/ 2^{2\vartheta-2}$ behave like $1/\sqrt{\pi\vartheta}$ as ϑ becomes large. In particular, for $\vartheta = 32$, we have $\binom{2\vartheta-2}{\vartheta-2} \Big/ 2^{2\vartheta-2} \approx 0.98029/\sqrt{\pi\vartheta}$ and $\binom{2\vartheta-2}{\vartheta-1} \Big/ 2^{2\vartheta-2} \approx 1.01191/\sqrt{\pi\vartheta}$. One obtains the data given in Table 3.2.

3.5.4 Implication For The Piling-up Approximation

We saw at the end of Section 3.4 that the average of $I^2(X_1 \oplus \cdots \oplus X_r)$ given $I(X_1) = i_1/\vartheta, \ldots, I(X_r) = i_r/\vartheta$ converges, as r increases, to $1/2\vartheta$. Here we have seen that the average of $I(X_1 \oplus \cdots \oplus X_r)$ converges to $1/\sqrt{\pi\vartheta}$.

Thus, the variance of $I(X_1 \oplus \cdots \oplus X_r)$ converges to $\frac{1}{\vartheta}(\frac{1}{2} - \frac{1}{\pi})$ and the normalized variance, defined as the variance of $X/E[X]$, to $\pi/2-1 \approx 0.57$.

Thus, for ϑ fixed, if r is large enough and sufficiently many of the numbers $i_1, \ldots, i_r$ are smaller than ϑ, then for random variables X_k with $I(X_k) = i_k/\vartheta$ for all k, $I(X_1 \oplus \cdots \oplus X_r)$ is about $1/\sqrt{\pi\vartheta}$. Because the normalized variance is approximately 0.57 and because $I(X_1) \cdots I(X_r)$ is much smaller than $1/\sqrt{\pi\vartheta}$ if r is large enough, in almost all cases $I(X_1 \oplus \cdots \oplus X_r) > I(X_1) \cdots I(X_r)$. For this reason, it is not too risky to approximate $I(X_1 \oplus \cdots \oplus X_r)$ with $I(X_1) \cdots I(X_r)$ if r is large.

3.6 Conclusions

We saw in Subsection 3.4.4 that, when the sample space on which the random variables are defined is large, then $I(X_1 \oplus \cdots \oplus X_r)$ is approximately equal to $I(X_1) \cdots I(X_r)$ in most cases, even if the random variables in question are not independent.

Moreover, we concluded in Subsection 3.5.4 that, for any sample space, if r is large enough, then the approximation $I(X_1 \oplus \cdots \oplus X_r) \approx I(X_1) \cdots I(X_r)$, although it is not actually valid, errs on the safe side from the linear cryptanalysis point of view.

Thus, in linear cryptanalysis, by Remark 3.3.2 and Subsection 3.5.4, one can safely approximate the imbalance of $T_1 \oplus \cdots \oplus T_{r-1}$ by the product of the imbalances of $T_1, \ldots, T_{r-1}$, even if the threefold sums are not independent.

3.A Proofs

Proposition 3.A.1
For any integers $i_{r-1}, i_r < \vartheta$, each of the matrices $E_{ee}(i_r)$, $E_{oo}(i_r)$, $O_{ee}(i_r)$, $O_{oo}(i_r)$, $E_{oe}(i_{r-1})E_{eo}(i_r)$, $E_{eo}(i_{r-1})E_{oe}(i_r)$, $O_{oe}(i_{r-1})O_{eo}(i_r)$ and $O_{eo}(i_{r-1})O_{oe}(i_r)$, provided it is not the zero matrix because of the parity of i_{r-1}, i_r and ϑ, is the transition matrix of a homogeneous, irreducible, recurrent and aperiodic Markov chain.

Proof:

In all cases, we use from Lemma 3.4.3 the fact that $P(\ell|i_1, i_2) = 0$ if and only if $\ell > \vartheta - |i_1 - i_2|$ or $\ell < i_1 + i_2 - \vartheta$.

ϑ even, i_r even

Here $E_{eo}(i_r) = E_{oe}(i_r) = 0$ so that $E(i_r) = \left(\begin{array}{c|c} E_{ee}(i_r) & 0 \\ \hline 0 & E_{oo}(i_r) \end{array}\right)$.

- $\boldsymbol{E_{ee}(i_r)}$: $\big(E_{ee}(i_r)\big)_{ms} = P(2s|2m, i_r)$, $m, s = 0, \ldots, \vartheta/2$ and

$$\big(E_{ee}(i_r)\big)_{ms} \neq 0 \Leftrightarrow \begin{cases} 2s \leq \vartheta - 2m + i_r \\ 2s \leq \vartheta - i_r + 2m \\ 2s \geq 2m + i_r - \vartheta \end{cases} \Leftrightarrow \begin{array}{l} m + s \leq \frac{\vartheta + i_r}{2} \\ s - m \leq \frac{\vartheta - i_r}{2} \\ m - s \leq \frac{\vartheta - i_r}{2} \end{array}$$

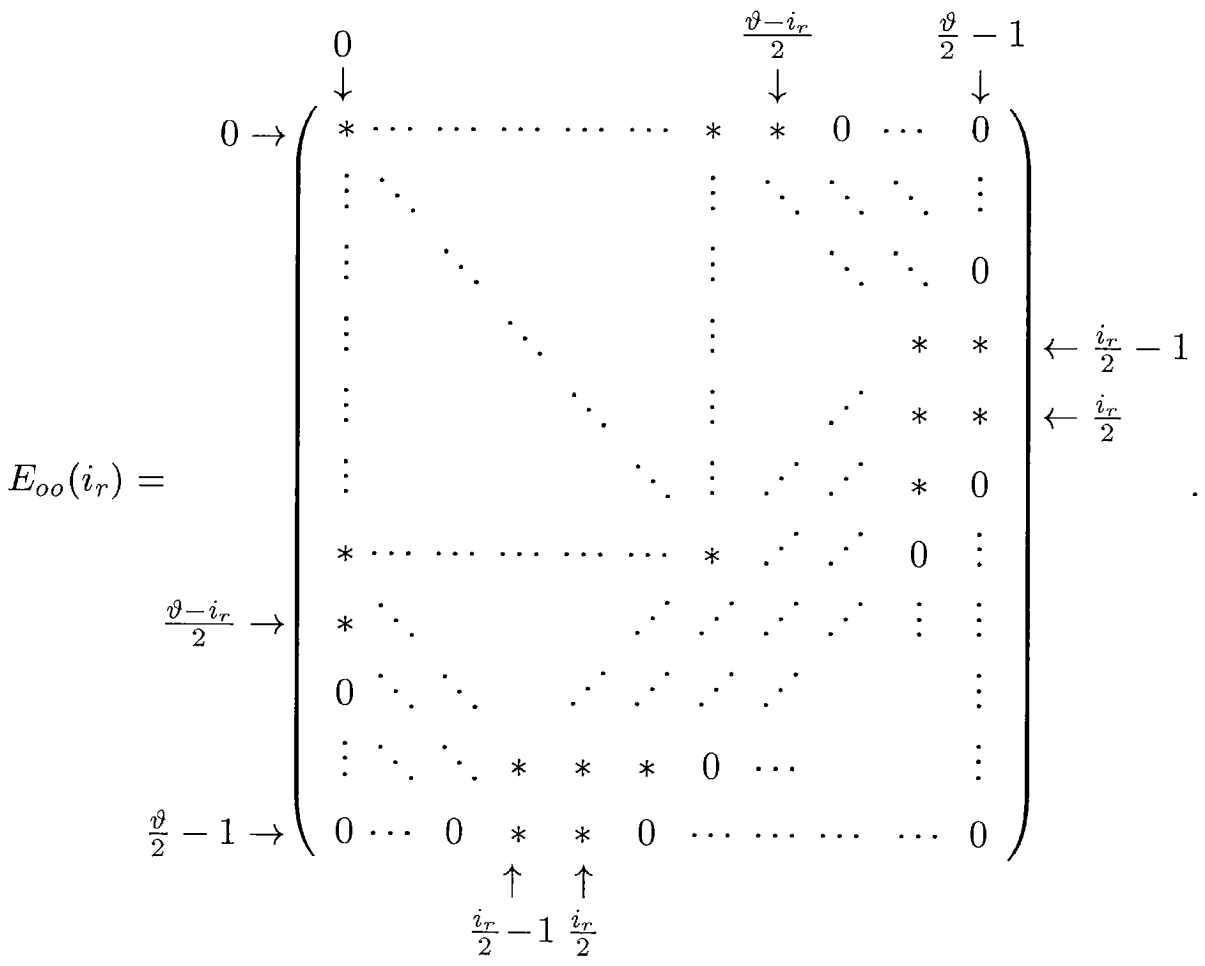
so that $E_{ee}(i_r)$ has the following form (stars stand for non-zero entries):

$$E_{ee}(i_r) = \begin{array}{c} \\ 0 \to \\ \\ \\ \\ \\ \\ \\ \frac{\vartheta - i_r}{2} \to \\ \\ \\ \\ \frac{\vartheta}{2} \to \end{array} \begin{array}{c} \begin{array}{ccccccccccc} & 0 & & & & & \frac{\vartheta - i_r}{2} & & & & \frac{\vartheta}{2} \\ & \downarrow & & & & & \downarrow & & & & \downarrow \end{array} \\ \begin{pmatrix} * & \cdots & \cdots & \cdots & \cdots & \cdots & * & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & & & & & \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & & & & \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & & & \vdots & & & * & 0 \\ \vdots & & & & \ddots & & \vdots & & \iddots & * & * \\ \vdots & & & & & \ddots & \vdots & \iddots & \iddots & * & 0 \\ * & \cdots & \cdots & \cdots & \cdots & \cdots & * & \iddots & \iddots & 0 & \vdots \\ 0 & \ddots & & & & \iddots & \iddots & \iddots & \iddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & & \iddots & \iddots & \iddots & \iddots & & & \vdots \\ \vdots & & \ddots & * & * & * & 0 & \cdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & * & 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \\ \begin{array}{ccccc} & & & \uparrow & \\ & & & \frac{i_r}{2} & \end{array} \end{array} \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \leftarrow \frac{i_r}{2} \\ \\ \\ \\ \\ \end{array} .$$

- $\boldsymbol{E_{oo}(i_r)}$: $\big(E_{oo}(i_r)\big)_{ms} = P(2s+1|2m+1, i_r)$, $m, s = 0, \ldots, \vartheta/2 - 1$ and

$$\big(E_{oo}(i_r)\big)_{ms} \neq 0 \Leftrightarrow \begin{cases} 2s + 1 \leq \vartheta - 2m - 1 + i_r \\ 2s + 1 \leq \vartheta - i_r + 2m + 1 \\ 2s + 1 \geq 2m + 1 + i_r - \vartheta \end{cases} \Leftrightarrow \begin{array}{l} m + s \leq \frac{\vartheta + i_r}{2} - 1 \\ s - m \leq \frac{\vartheta - i_r}{2} \\ m - s \leq \frac{\vartheta - i_r}{2} \end{array}$$

so that $E_{oo}(i_r)$ has the following form:



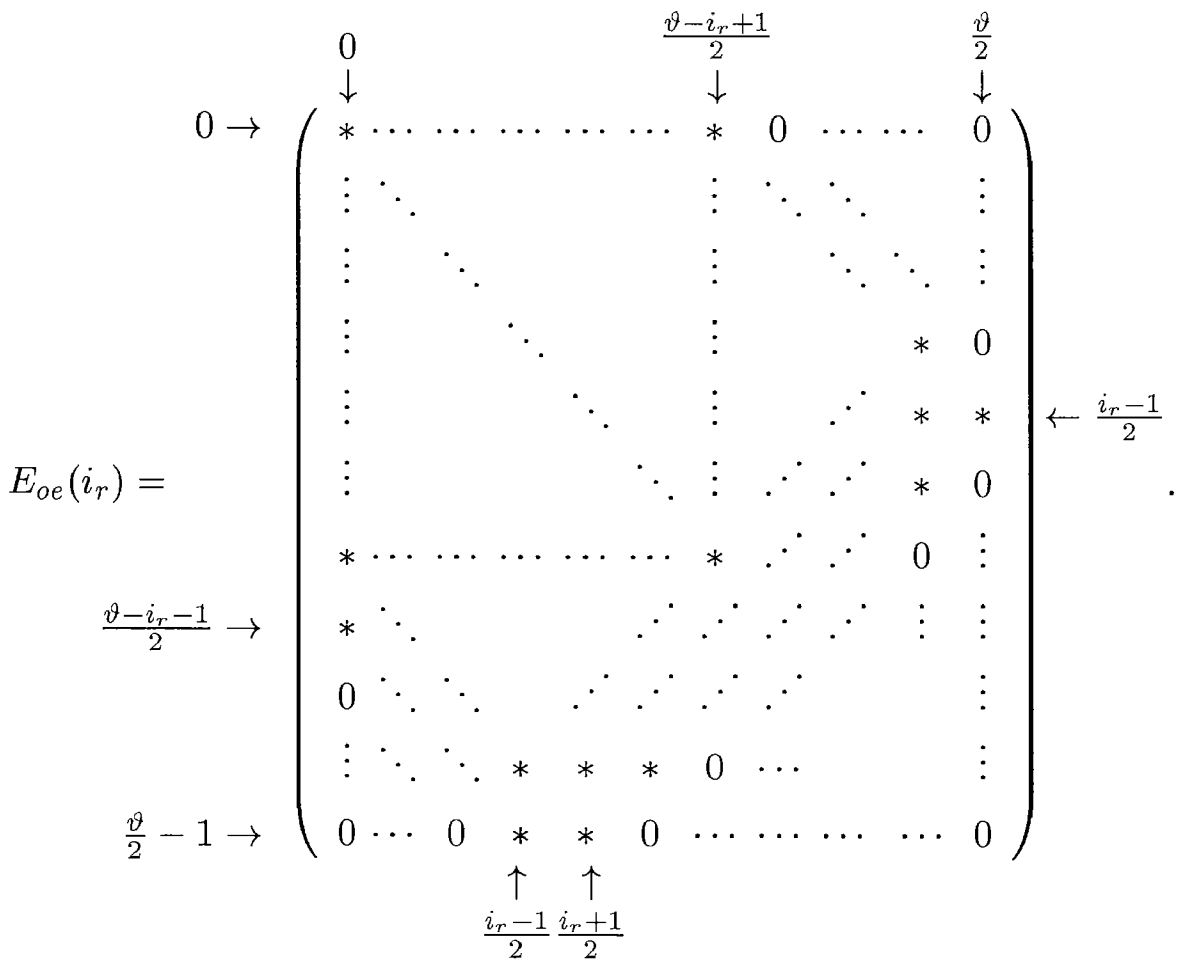
ϑ even, i_r odd

Here $E_{ee}(i_r) = E_{oo}(i_r) = 0$ so that $E(i_r) = \left(\begin{array}{c|c} 0 & E_{eo}(i_r) \\ \hline E_{oe}(i_r) & 0 \end{array}\right)$.

- $\boldsymbol{E_{oe}(i_r)}$: $\left(E_{oe}(i_r)\right)_{ms} = P(2s|2m+1, i_r)$, $m = 0, \ldots, \vartheta/2 - 1$, $s = 0, \ldots, \vartheta/2$ and

$$\left(E_{oe}(i_r)\right)_{ms} \neq 0 \Leftrightarrow \begin{cases} 2s \leq \vartheta - 2m - 1 + i_r \\ 2s \leq \vartheta - i_r + 2m + 1 \\ 2s \geq 2m + 1 + i_r - \vartheta \end{cases} \Leftrightarrow \begin{array}{l} m + s \leq \frac{\vartheta + i_r - 1}{2} \\ s - m \leq \frac{\vartheta - i_r + 1}{2} \\ m - s \leq \frac{\vartheta - i_r - 1}{2} \end{array}$$

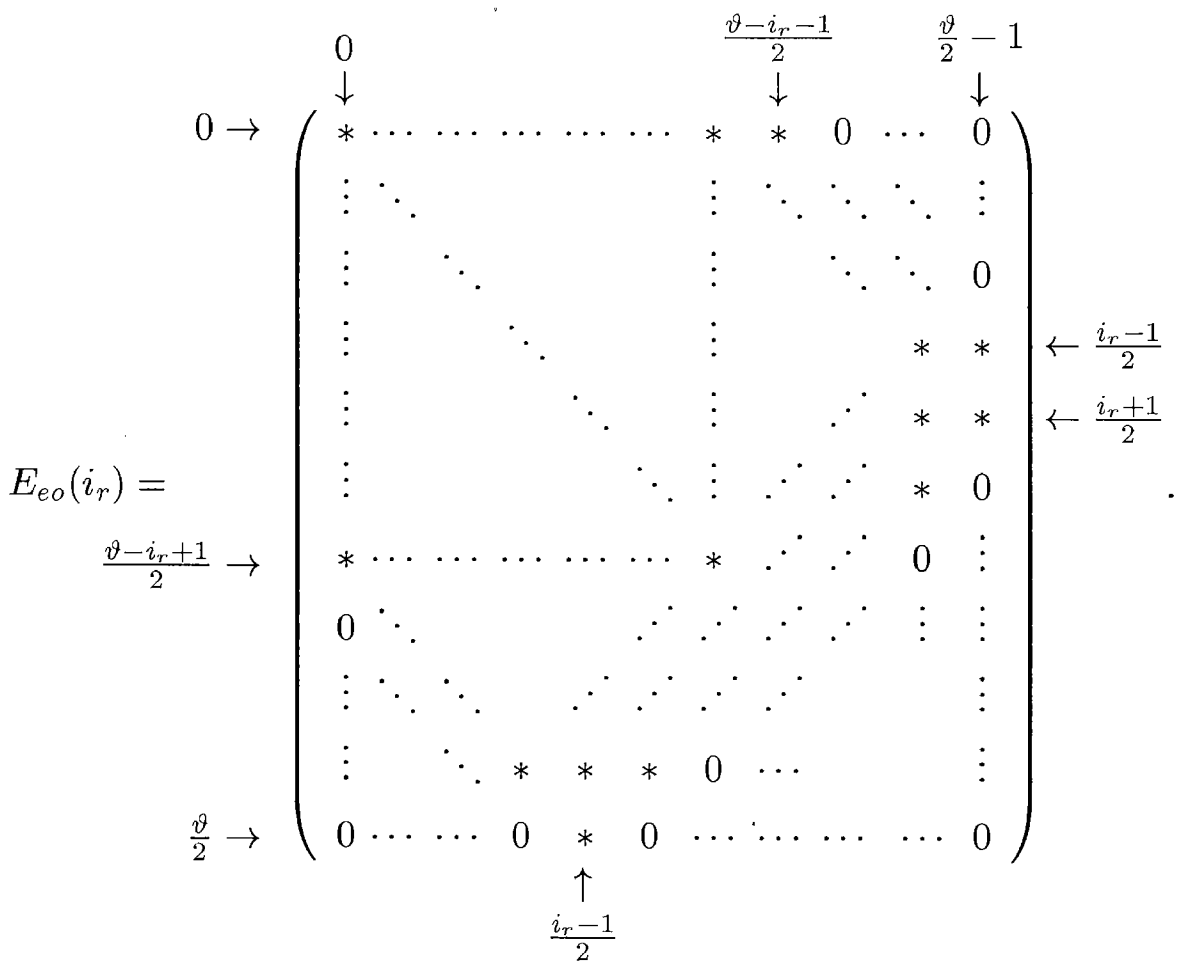
so that $E_{oe}(i_r)$ has the following form:



- $\boldsymbol{E_{eo}(i_r)}$: $\left(E_{eo}(i_r)\right)_{ms} = P(2s+1|2m, i_r)$, $m = 0, \ldots, \vartheta/2$, $s = 0, \ldots, \vartheta/2 - 1$ and

$$\left(E_{eo}(i_r)\right)_{ms} \neq 0 \Leftrightarrow \begin{cases} 2s+1 \leq \vartheta - 2m + i_r \\ 2s+1 \leq \vartheta - i_r + 2m \\ 2s+1 \geq 2m + i_r - \vartheta \end{cases} \Leftrightarrow \begin{array}{l} m + s \leq \frac{\vartheta + i_r - 1}{2} \\ s - m \leq \frac{\vartheta - i_r - 1}{2} \\ m - s \leq \frac{\vartheta - i_r + 1}{2} \end{array}$$

so that $E_{eo}(i_r)$ has the following form:



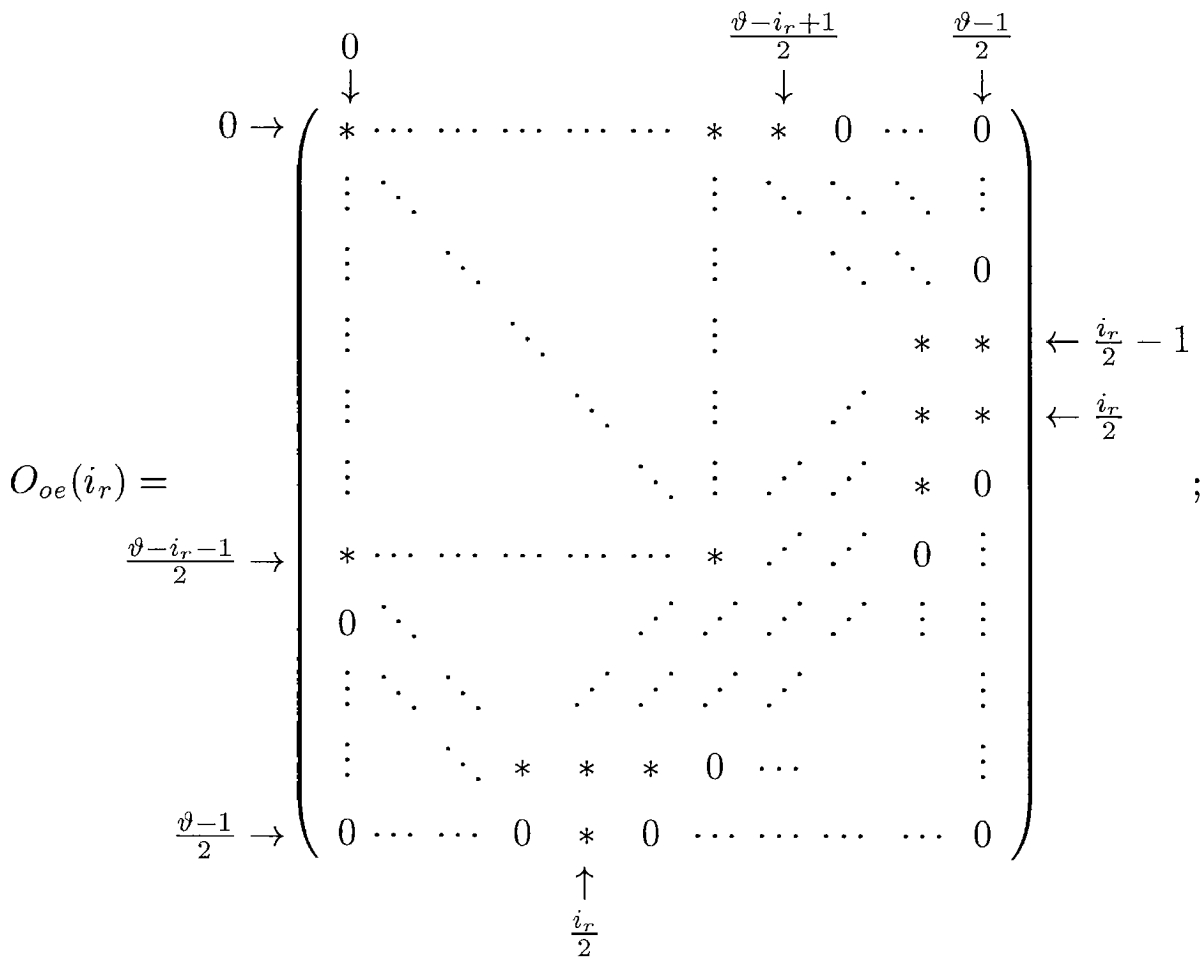
ϑ odd, i_r even

Here $O_{ee}(i_r) = O_{oo}(i_r) = 0$ so that $O(i_r) = \left(\begin{array}{c|c} 0 & O_{eo}(i_r) \\ \hline O_{oe}(i_r) & 0 \end{array}\right)$.

- $\boldsymbol{O_{oe}(i_r)}$ **:** $\left(O_{oe}(i_r)\right)_{ms} = P(2s|2m+1, i_r)$, $m, s = 0, \ldots, (\vartheta - 1)/2$ and

$$\left(O_{oe}(i_r)\right)_{ms} \neq 0 \Leftrightarrow \begin{cases} 2s \leq \vartheta - 2m - 1 + i_r \\ 2s \leq \vartheta - i_r + 2m + 1 \\ 2s \geq 2m + 1 + i_r - \vartheta \end{cases} \Leftrightarrow \begin{array}{l} m + s \leq \frac{\vartheta + i_r - 1}{2} \\ s - m \leq \frac{\vartheta - i_r + 1}{2} \\ m - s \leq \frac{\vartheta - i_r - 1}{2} \end{array}$$

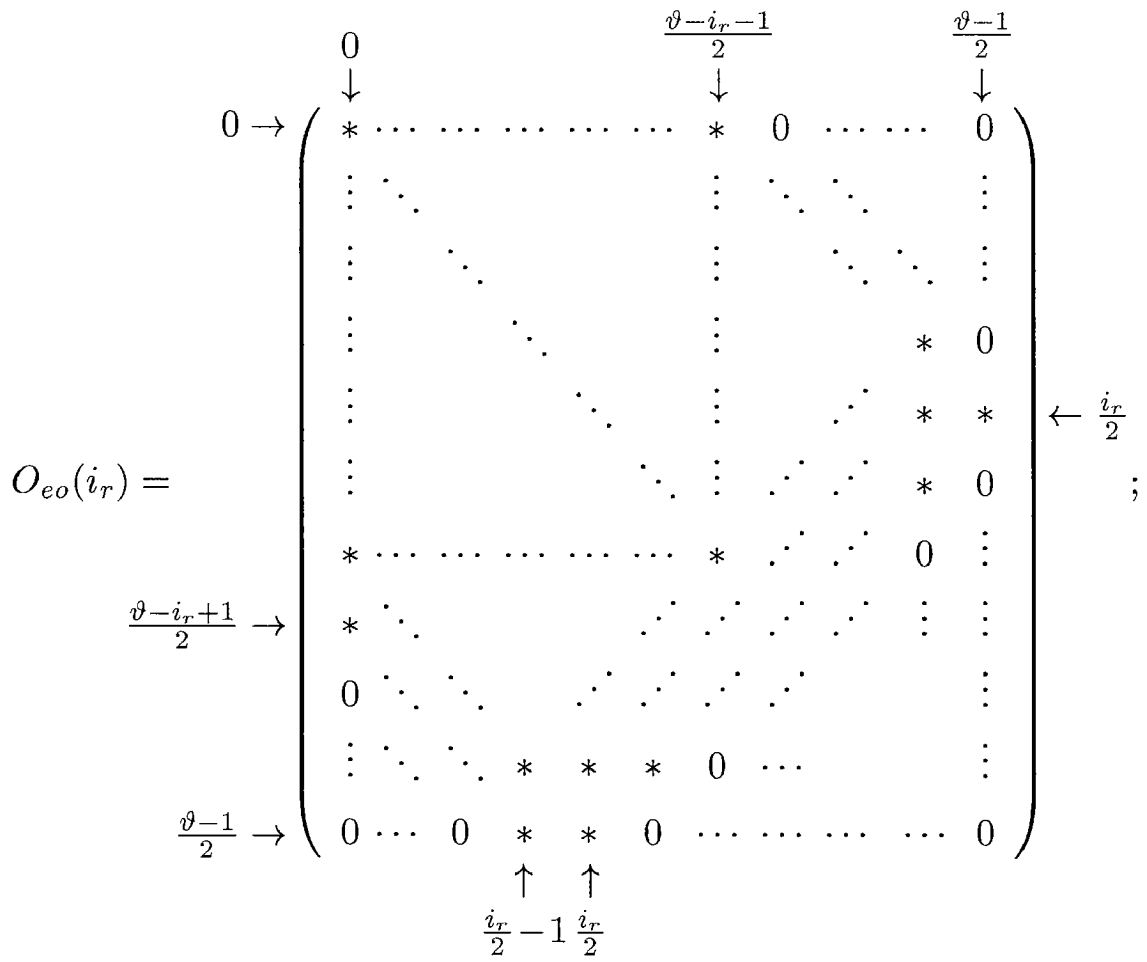
so that $O_{oe}(i_r)$ has the following form:



- $\boldsymbol{O_{eo}(i_r)}$: $\left(O_{eo}(i_r)\right)_{ms} = P(2s+1|2m, i_r)$, $m, s = 0, \ldots, (\vartheta-1)/2$ and

$$\left(O_{eo}(i_r)\right)_{ms} \neq 0 \Leftrightarrow \begin{cases} 2s+1 \leq \vartheta - 2m + i_r \\ 2s+1 \leq \vartheta - i_r + 2m \\ 2s+1 \geq 2m + i_r - \vartheta \end{cases} \Leftrightarrow \begin{array}{l} m+s \leq \frac{\vartheta+i_r-1}{2} \\ s-m \leq \frac{\vartheta-i_r-1}{2} \\ m-s \leq \frac{\vartheta-i_r+1}{2} \end{array}$$

so that $O_{eo}(i_r)$ has the following form:



ϑ odd, i_r odd

Here $O_{eo}(i_r) = O_{oe}(i_r) = 0$ so that $O(i_r) = \left(\begin{array}{c|c} O_{ee}(i_r) & 0 \\ \hline 0 & O_{oo}(i_r) \end{array}\right)$.

- $\boldsymbol{O_{ee}(i_r)}$: $\big(O_{ee}(i_r)\big)_{ms} = P(2s|2m, i_r)$, $m, s = 0, \ldots, (\vartheta-1)/2$ and

$$\big(O_{ee}(i_r)\big)_{ms} \neq 0 \Leftrightarrow \begin{cases} 2s \leq \vartheta - 2m + i_r \\ 2s \leq \vartheta - i_r + 2m \\ 2s \geq 2m + i_r - \vartheta \end{cases} \Leftrightarrow \begin{array}{l} m + s \leq \frac{\vartheta+i_r}{2} \\ s - m \leq \frac{\vartheta-i_r}{2} \\ m - s \leq \frac{\vartheta-i_r}{2} \end{array}$$

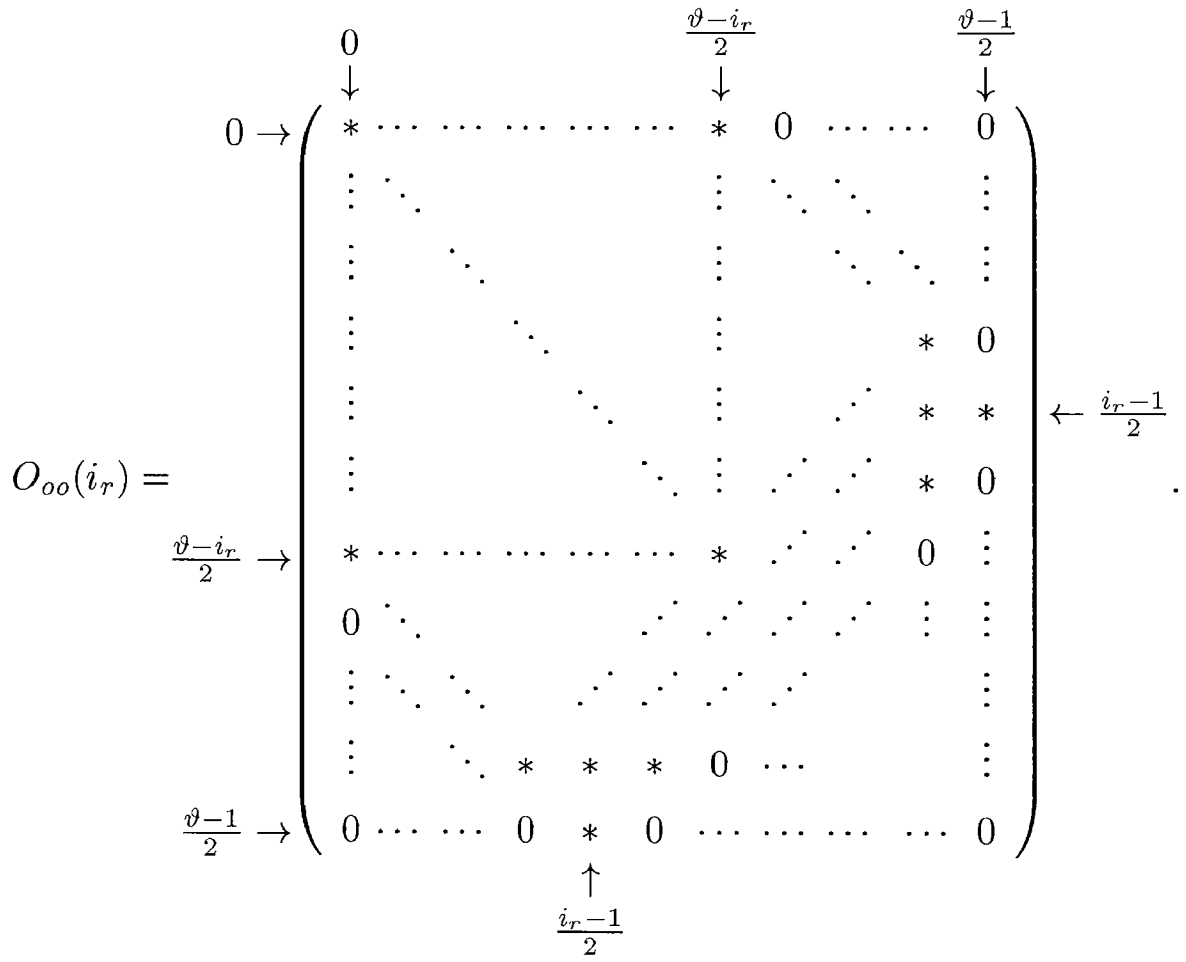
so that $O_{ee}(i_r)$ has the following form:

$$O_{ee}(i_r) = \begin{array}{c} \\ 0 \to \\ \\ \\ \\ \\ \\ \\ \frac{\vartheta-i_r}{2} \to \\ \\ \\ \frac{\vartheta-1}{2} \to \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccccccccc} & 0 & & & & & & & \frac{\vartheta-i_r}{2} & & & \frac{\vartheta-1}{2} & \\ & \downarrow & & & & & & & \downarrow & & & \downarrow & \\ * & \cdots & \cdots & \cdots & \cdots & \cdots & * & * & 0 & \cdots & & 0 & \\ \vdots & \ddots & & & & & \vdots & \ddots & \ddots & \ddots & & \vdots & \\ \vdots & & \ddots & & & & \vdots & & \ddots & \ddots & & 0 & \\ \vdots & & & \ddots & & & \vdots & & & * & & * & \\ \vdots & & & & \ddots & & \vdots & & \iddots & * & & * & \\ \vdots & & & & & \ddots & \vdots & \iddots & \iddots & * & & 0 & \\ * & \cdots & \cdots & \cdots & \cdots & \cdots & * & \iddots & \iddots & 0 & & \vdots & \\ * & \ddots & & & & \iddots & \iddots & \iddots & \iddots & \iddots & & \vdots & \\ 0 & \ddots & \ddots & & \iddots & \iddots & \iddots & \iddots & & & & \vdots & \\ \vdots & \ddots & \ddots & * & * & * & 0 & \cdots & & & & \vdots & \\ 0 & \cdots & 0 & * & * & 0 & \cdots & \cdots & \cdots & \cdots & & 0 & \end{array} \\ \begin{array}{ccccccccccccc} & & & \uparrow & \uparrow & & & & & & & & \\ & & & \frac{i_r-1}{2} & \frac{i_r+1}{2} & & & & & & & & \end{array} \end{array} \begin{array}{l} \\ \\ \\ \\ \leftarrow \frac{i_r-1}{2} \\ \leftarrow \frac{i_r+1}{2} \\ \\ \\ \\ \\ \\ \\ \\ \end{array}.$$

- $\boldsymbol{O_{oo}(i_r)}$ **:** $\left(O_{oo}(i_r)\right)_{ms} = P(2s+1|2m+1, i_r)$, $m, s = 0, \ldots, (\vartheta-1)/2$ and

$$\left(O_{oo}(i_r)\right)_{ms} \neq 0 \Leftrightarrow \begin{cases} 2s+1 \leq \vartheta - 2m - 1 + i_r \\ 2s+1 \leq \vartheta - i_r + 2m + 1 \\ 2s+1 \geq 2m + 1 + i_r - \vartheta \end{cases} \Leftrightarrow \begin{array}{l} m+s \leq \frac{\vartheta+i_r}{2} - 1 \\ s-m \leq \frac{\vartheta-i_r}{2} \\ m-s \leq \frac{\vartheta-i_r}{2} \end{array}$$

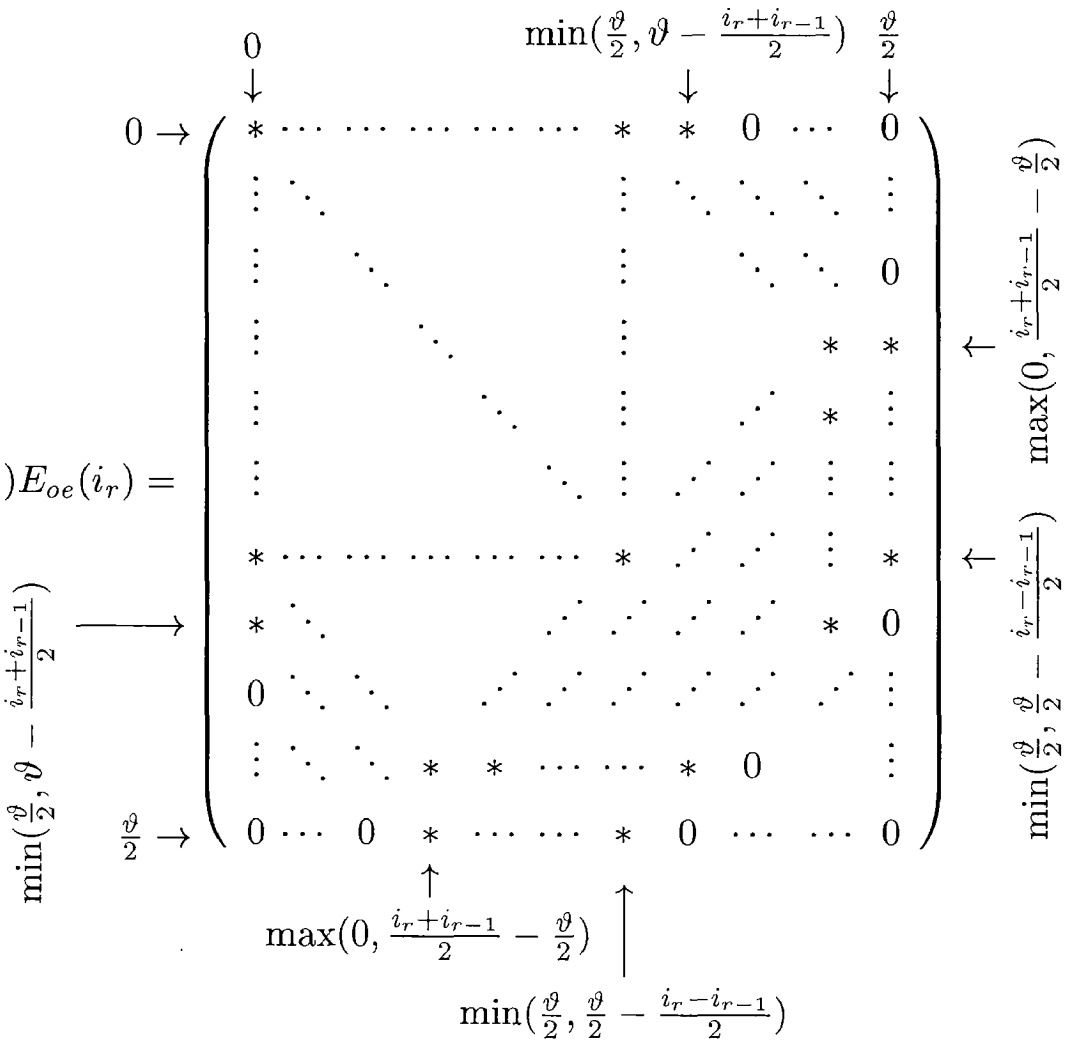
so that $O_{oo}(i_r)$ has the following form:



The matrices E_{ee} to O_{oo} depend only on ϑ and i_r. Each of the four matrices $E_{ee}(i_r), E_{oo}(i_r), O_{ee}(i_r)$ and $O_{oo}(i_r)$, when ϑ and i_r have the same parity, can be viewed as transition matrices of a homogeneous Markov chain on the even-parity states and the odd-parity states, respectively (each of these matrices *defines implicitely* a Markov chain). From the form of these matrices, it can be easily seen that, if $i_r < \vartheta$, this Markov chain is irreducible, recurrent and aperiodic. (Irreducibility is obvious. The Markov chain is recurrent because any two states are reachable from one another, even if the largest index with a non-zero element in the first column of the matrix is smaller than the smallest index with a non-zero element in the last column. Aperiodicity follows from the fact that any two states are reachable from one another and that there is at least one state that has a self-loop, i.e., it is reachable from itself.) The other four matrices $E_{oe}(i_r), E_{eo}(i_r), O_{eo}(i_r)$ and $O_{oe}(i_r)$, which are non-zero when ϑ and i_r have different parities, are the ones that make us leave the odd-parity states for the even-parity states or vice-versa. Because of this and because of dimensional considerations, we cannot consider them as transition matrices of a homogeneous Markov chain. However, we can do this

for the products $E_{oe}(i_{r-1})E_{eo}(i_r)$, $E_{eo}(i_{r-1})E_{oe}(i_r)$, $O_{oe}(i_{r-1})O_{eo}(i_r)$ and $O_{eo}(i_{r-1})O_{oe}(i_r)$ (which are non-zero only when $i_{r-1} \equiv i_r \equiv \vartheta+1 \bmod 2$), since they yield quadratic matrices that leave the state parity invariant. For instance, the form of $E_{eo}E_{oe}$ is as follows.

$E_{eo}(i_{r-1})E_{oe}(i_r) =$



The other three products are similar. Thus, they implicitely define a homogeneous, irreducible, recurrent and aperiodic Markov chain, either on the even-parity states, or on the odd-parity states. □

Proposition 3.A.2
For any positive integer a and any integers $b, c, d \geq 0$,

$$\sum_{m\in\mathbb{Z}} \binom{a}{b+m,\, c+m,\, d-m,\, a-b-c-d-m} = \binom{a}{a-c-d}\binom{a}{a-b-d}. \tag{3.26}$$

Proof:

The lemma holds if $b+d > a$ or if $c+d > a$ because then both sides are zero. Otherwise, (3.26) holds for $a = 1$: it suffices to verify (3.26) for the cases $(b,c,d) = (0,0,0)$, $(0,1,0)$, $(1,0,0)$, $(1,1,0)$, $(0,0,1)$; in these cases, the computation is easy. Now denote the left side of (3.26) by $S(a)$ and the multinomial coefficient by $t(a,m)$. We show that $S(a+1)/S(a) = (a+1)^2/(a+1-b-d)(a+1-c-d)$. This is enough because the right side of (3.26) has the same quotient for two successive values of a. To this end, we use Zeilberger's method [52], a generalisation of Gosper's, to find $T(m)$ such that $T(m+1) - T(m) = t(a+1,m) - \frac{(a+1)^2}{(a+1-b-d)(a+1-c-d)}\, t(a,m)$, namely

$$\begin{aligned} T(m) &= \frac{(m+b)(m+c)(a+1)}{(a+1-b-d)(a+1-c-d)(m-a-1+b+c+d)} \\ &\quad \times \binom{a}{b+m,\ c+m,\ d-m,\ a-b-c-d-m}. \end{aligned}$$

Now because $T(m) = 0$ for m outside of a certain range, we have

$$\begin{aligned} 0 &= \sum_{m\in\mathbb{Z}} \big(T(m+1) - T(m)\big) \\ &= \sum_{m\in\mathbb{Z}} \Big[t(a+1,m) - \frac{(a+1)^2}{(a+1-b-d)(a+1-c-d)} t(a,m)\Big] \\ &= \sum_{m\in\mathbb{Z}} t(a+1,m) - \frac{(a+1)^2}{(a+1-b-d)(a+1-c-d)} \sum_{m\in\mathbb{Z}} t(a,m) \\ &= S(a+1) - \frac{(a+1)^2}{(a+1-b-d)(a+1-c-d)}\, S(a). \end{aligned}$$

□

We now prove the convergence property mentioned on Page 66. In the following, indices of matrices and vectors begin with zero.

Definition 3.A.3

Let a, b and N be nonnegative integers with $a \le N$ and $b \ge 2N - a$. A stochastic matrix A is *of type* (a,b,N) if it is an $(N+1) \times (N+1)$ matrix such that $A_{ij} \neq 0$ if and only if $|i-j| \le a$ and $i+j \le b$.

Lemma 3.A.4

Let A and $\tilde{A}$ be stochastic matrices of type (a,b,N) and $(\tilde{a},\tilde{b},N)$, respectively. Then the stochastic matrix $A\tilde{A}$ is of type $(\min(N, a+\tilde{a}), \min(2N, a+\tilde{b}, b+\tilde{a}), N)$.

Proof:

We have $(A\tilde{A})_{k\ell} = \sum_m A_{km}\tilde{A}_{m\ell}$, which is positive if and only there exists m such that all the following inequalities are satisfied:

$$\begin{cases} A_{km} > 0 \Leftrightarrow \begin{cases} k-m \le a \\ m-k \le a \\ k+m \le b \end{cases} \\ \tilde{A}_{m\ell} > 0 \Leftrightarrow \begin{cases} m-\ell \le \tilde{a} \\ \ell-m \le \tilde{a} \\ m+\ell \le \tilde{b} \end{cases} \end{cases} \Leftrightarrow \begin{cases} m \ge k-a \\ m \le k+a \\ m \le b-k \\ m \ge \ell-\tilde{a} \\ m \le \ell+\tilde{a} \\ m \le \tilde{b}-\ell \end{cases} \Leftrightarrow \begin{cases} m \ge \max(k-a, \ell-\tilde{a}) \\ m \le \min(k+a, \ell+\tilde{a}). \\ m \le \min(b-k, \tilde{b}-\ell) \end{cases} \tag{3.27}$$

From the first set of six inequalities follows that

$$k-\ell \le a+\tilde{a},\ \ell-k \le a+\tilde{a},\ k+\ell \le a+\tilde{b},\ \text{and}\ \ k+\ell \le \tilde{a}+b. \tag{3.28}$$

On the other hand, from the inequalities (3.28) follows that

$$k-a \le \ell+\tilde{a},\ \ell-\tilde{a} \le k+a,\ \ell-\tilde{a} \le b-k,\ \text{and}\ \ k-a \le \tilde{b}-\ell. \tag{3.29}$$

Moreover, because $a+b, \tilde{a}+\tilde{b} \ge 2N$ and $k, \ell \le N$, we also have $\ell-\tilde{a} \le \tilde{b}-\ell$ and $k-a \le b-k$. From this and with (3.29), we have

$$\begin{aligned} \max(k-a, \ell-\tilde{a}) &\le \min(k+a, \ell+\tilde{a}) \quad \text{and} \\ \max(k-a, \ell-\tilde{a}) &\le \min(b-k, \tilde{b}-\ell), \end{aligned}$$

and there is at least one m that satisfies the last three inequalities in (3.27). Hence, $(A\tilde{A})_{kl} \neq 0$ if and only if (3.28) is satisfied, i.e., if and only if $|k-\ell| \le a+\tilde{a}$, $k+\ell \le a+\tilde{b}$, and $k+\ell \le \tilde{a}+b$. But $|k-\ell|$ (resp. $k+\ell$) cannot be larger than N (resp. $2N$) so $A\tilde{A}$ is of type $(\min(N, a+\tilde{a}), \min(2N, a+\tilde{b}, b+\tilde{a}), N)$. □

Lemma 3.A.5

Let $a, \tilde{a}, N, b, \tilde{b}$ be positive integers with $a, \tilde{a} \le N$, $b \ge 2N-a$ and $\tilde{b} \ge 2N-\tilde{a}$. Consider the recursion

$$a_{r+1} = \min(N, a_r + a|\tilde{a}), \qquad b_{r+1} = \min(2N, a_r + b|\tilde{b}, b_r + a|\tilde{a}) \tag{3.30}$$

with the initial conditions

$$a_1 = a|\tilde{a}, \qquad b_1 = b|\tilde{b}, \tag{3.31}$$

where in both equations $a|\tilde{a}$ (resp. $b|\tilde{b}$) means that either a or $\tilde{a}$ (resp. either b or $\tilde{b}$) can occur; more precisely, at each step, either

$$a_{r+1} = \min(N, a_r + a) \quad \text{and} \quad b_{r+1} = \min(2N, a_r + b, b_r + a) \quad \text{or} \tag{3.32}$$

$$a_{r+1} = \min(N, a_r + \tilde{a}) \quad \text{and} \quad b_{r+1} = \min(2N, a_r + \tilde{b}, b_r + \tilde{a}). \tag{3.33}$$

Let $k_0 := \max(\lceil \frac{N}{a} \rceil, \lceil \frac{N}{\tilde{a}} \rceil)$. Then $a_{2k_0} = N$ and $b_{2k_0} = 2N$.

Proof:

The proof consists of the following steps:

1. for all $r \geq 1$, $a_r < N \Rightarrow a_{r+1} > a_r$;
2. for all $r \geq 1$, $a_r = N \Rightarrow a_{r+1} = N$;
3. $a_{3k_0} = N$ and $b_{3k_0} = 2N$.

Proof of the three parts:

1. If $a_r < N$, then either $a_{r+1} = N > a_r$ or $a_{r+1} = a_r + a|\tilde{a} > a_r$.
2. If $a_r = N$, then $a_{r+1} = \min(N, N + a|\tilde{a}) = N$.
3. Without losing generality, let $a \leq \tilde{a}$. Then $k_0 = \lceil \frac{N}{a} \rceil$. We have $a_1 \geq a$, and by a simple induction one shows that $a_n \geq an$ for $n < k_0$. Then $a_{k_0} = \min(N, a_{k_0-1} + a|\tilde{a})$. But $a_{k_0-1} + a|\tilde{a} \geq a(k_0 - 1) + a|\tilde{a} \geq a(k_0 - 1) + a = ak_0 \geq N$. Thus, $a_{k_0} = N$, and, by 2., $a_{3k_0} = N$.

   Furthermore, we have $b_{k_0+n} \geq \min(2N, b_{k_0} + an)$ for $n \geq 0$: this is true for $n = 0$ since $b_{k_0} \leq 2N$ and hence $b_{k_0} \geq \min(2N, b_{k_0})$. Let it be true for some n; then

$$\begin{aligned}
b_{k_0+n+1} &= \min(2N, a_{k_0+n} + b|\tilde{b}, b_{k_0+n} + a|\tilde{a}) \\
&= \min(2N, N + b|\tilde{b}, b_{k_0+n} + a|\tilde{a}) \\
&\quad \text{(from above and by 2. of Lemma 3.A.5)} \\
&= \min(2N, b_{k_0+n} + a|\tilde{a}) \qquad \text{(since } b, \tilde{b} \geq N\text{)} \\
&\geq \min(2N, b_{k_0+n} + a) \quad \geq \quad \min(2N, \min(2N, b_{k_0} + an) + a) \\
&= \min(2N, 2N + a, b_{k_0} + a(n+1)) \quad = \quad \min(2N, b_{k_0} + a(n+1)).
\end{aligned}$$

   Then $b_{3k_0} \geq \min(2N, b_{k_0} + 2ak_0)$. Because $2ak_0 \geq 2N$, we have $b_{3k_0} = 2N$. □

Corollary 3.A.6

Let A and $\tilde{A}$ be stochastic matrices of type (a, b, N) and $(\tilde{a}, \tilde{b}, N)$, respectively. Then there exists $m \in \mathbb{N}$ such that any product of m matrices containing only A's and $\tilde{A}$'s gives a matrix with no zero entry.

Proof:

By Lemma 3.A.4, any partial product of A's and $\tilde{A}$'s involving n multiplicands yields a matrix of type (a_n, b_n, N), where a_n and b_n are defined by (3.30) and (3.31). [When the $(r+1)$-st matrix in the product is A, then (3.32) is applied, otherwise (3.33) is applied.] But then, by Lemma 3.A.5, $a_{3k_0} = N$ and $b_{3k_0} = 2N$, where $k_0 = \max(\lceil \frac{N}{a} \rceil, \lceil \frac{N}{\tilde{a}} \rceil)$. Thus, the product of $3k_0$ matrices yields a matrix of type $(N, 2N, N)$, which has no zero entry. Hence, it is enough to set $m = 3k_0$. □

Lemma 3.A.7

Let A be a stochastic matrix with no zero entry. Denote by $||\cdot||_1$ the vector norm defined by $||\boldsymbol{x}||_1 = \sum_i |x_i|$. Then there exists $0 \leq \lambda < 1$ such that $||\boldsymbol{x}A||_1 \leq \lambda \cdot ||\boldsymbol{x}||_1$ for all vectors $\boldsymbol{x}$ with $\sum_i x_i = 0$.

Proof[1]**:**

If $\boldsymbol{x}A = \boldsymbol{0}$, then the inequality is satisfied for any λ. Let now $\boldsymbol{x}A \neq \boldsymbol{0}$. Let $A_{min} := \min_{i,j} A_{ij}$. This is positive and upper-bounded by $1/n$. Let H be the matrix all entries of which are equal to A_{min}, and let $G = A - H$. All entries of G are nonnegative. Then $\boldsymbol{x}H = \boldsymbol{0}$ and $\boldsymbol{x}A = \boldsymbol{x}G$. Further, we have $||\boldsymbol{x}M||_1 \leq ||\boldsymbol{x}||_1 \cdot ||M||_1$ for any matrix M of suitable dimension, where $||M||_1 = \max_i \sum_j |M_{ij}|$ [28]. But then

$$||G||_1 = \max_i \sum_j |G_{ij}| = \max_i \sum_j G_{ij} = \max_i \sum_j (A_{ij} - A_{min}) = 1 - NA_{min}$$

which implies $||\boldsymbol{x}A||_1 = ||\boldsymbol{x}G||_1 \leq ||\boldsymbol{x}||_1 \cdot ||G||_1 = (1 - NA_{min})||\boldsymbol{x}||_1$. □

Proposition 3.A.8

Let A and $\tilde{A}$ be stochastic matrices of type (a, b, N) and $(\tilde{a}, \tilde{b}, N)$, respectively, that are the transition probability matrices of two Markov chains (X_n) and (Y_n) that have the same steady-state distribution $\boldsymbol{\pi}$. Define the sequence of matrices (B_n) as follows: $B_1 = A$ or $B_1 = \tilde{A}$; then, at each step, either $B_{n+1} = B_n A$ or $B_{n+1} = B_n \tilde{A}$. (That is, we consider any product containing only A's and $\tilde{A}$'s.) Then $\lim_{n\to\infty} \boldsymbol{p}B_n = \boldsymbol{\pi}$ for any starting probability vector $\boldsymbol{p}$.

Proof:

By Corollary 3.A.6, there exists $m \in \mathbb{N}$ such that any product B of m matrices containing only A's and $\tilde{A}$'s gives a matrix with no zero entry. By Lemma 3.A.7, for every such matrix B, there exists $0 \leq \lambda(B) < 1$ such that $||\boldsymbol{x}B||_1 \leq \lambda(B) \cdot ||\boldsymbol{x}||_1$ for all vectors $\boldsymbol{x}$ with $\sum_i x_i = 0$. Since there is only a finite number of such products, there exists λ_* such that

[1] I thank Pascal Vontobel for making the proof clearer and shorter.

$||\boldsymbol{x}B||_1 \leq \lambda_* \cdot ||\boldsymbol{x}||_1$ for all vectors $\boldsymbol{x}$ with $\sum_i x_i = 0$ and for all such products B. In particular, $||\boldsymbol{x}B_m||_1 \leq \lambda_* \cdot ||\boldsymbol{x}||_1$ since B_m is such a product. Then, by induction, $\lim_{k\to\infty} \boldsymbol{x}B_{km} = \mathbf{0}$ and thus $\lim_{n\to\infty} \boldsymbol{x}B_n = \mathbf{0}$. (The latter holds because $||\boldsymbol{x}A||_1 \leq ||\boldsymbol{x}||_1$ for any vector $\boldsymbol{x}$ and any stochastic matrix A, and because B_n is stochastic for any n.)

Now the steady-state distribution $\boldsymbol{\pi}$ is also a stationary distribution of the Markov chains (X_n) and (Y_n) so $\boldsymbol{\pi}A = \boldsymbol{\pi}\tilde{A} = \boldsymbol{\pi}$ and thus $\boldsymbol{\pi}B_n = \boldsymbol{\pi}$ for all $n \geq 1$. Hence,

$$\lim_{n\to\infty} \boldsymbol{p}B_n = \lim_{n\to\infty} (\boldsymbol{p} - \boldsymbol{\pi})B_n + \lim_{n\to\infty} \boldsymbol{\pi}B_n = \mathbf{0} + \boldsymbol{\pi} = \boldsymbol{\pi}. \qquad \square$$

Remark 3.A.9
Instead of only two stochastic matrices A and $\tilde{A}$, one can also consider any finite number of stochastic matrices different from the identity matrix. The conclusions are the same.

Corollary 3.A.10
Let the matrices $E_{ee}(k), E_{eo}(k), E_{oe}(k)$, and $E_{oo}(k)$ be as in Proposition 3.A.1, and let $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$ be defined by Definition 3.5.3. Let $\boldsymbol{p}$ be a probability vector and let B_0 be the identity matrix. Then:

1. *Define $(B_n)_{n\in\mathbb{N}}$ such that at each step, $B_{n+1} = B_n$ or $B_{n+1} = B_n E_{ee}(i)$ or $B_{n+1} = B_n E_{eo}(i)E_{oe}(j)$ for some $i, j < \vartheta$, and such that among the second and the third recursion, at least one is applied an infinite number of times. Then $\lim_{n\to\infty} \boldsymbol{p}B_n = \boldsymbol{e_1}$.*

2. *Define $(B_n)_{n\in\mathbb{N}}$ such that at each step, $B_{n+1} = B_n$ or $B_{n+1} = B_n E_{oo}(i)$ or $B_{n+1} = B_n E_{oe}(i)E_{eo}(j)$ for some $i, j < \vartheta$, and such that among the second and the third recursion, at least one is applied an infinite number of times. Then $\lim_{n\to\infty} \boldsymbol{p}B_n = \boldsymbol{e_2}$.*

Proof:
We prove the first statement. The proof of the second statement is similar. The matrices $E_{ee}(i)$ and $E_{eo}(i)E_{oe}(j)$ are stochastic and of type $(\frac{\vartheta-i}{2}, \frac{\vartheta+i}{2}, \frac{\vartheta}{2})$ and $(\min(\frac{\vartheta}{2}, \vartheta - \frac{i+j}{2}), \min(\vartheta, \vartheta - \frac{j-i}{2}), \frac{\vartheta}{2})$, respectively. Moreover, all have $\boldsymbol{e_1}$ as steady-state distribution. The statement holds then by Proposition 3.A.8 and Remark 3.A.9. $\square$

Chapter 4

The Hypothesis of Fixed-Key Equivalence

This chapter treats of the hypothesis of fixed-key equivalence [16, 17]. Recall that the cryptanalyst wishes for the following reasons that this hypothesis holds:

- It assures him that the probability of success of his linear cryptanalysis attack is roughly the same regardless of which key has been used in the encryption.
- It allows him to calculate, or at least to estimate, his probability of success.

We begin by recalling some definitions, among others the statement of the hypothesis itself. Then we introduce a measure that allows us to quantify the validity of the hypothesis, i.e., to give a meaning to the approximation sign. We proceed then by presenting some properties of balanced functions and of invertible functions. After this, we come to the heart of the problem and study the validity of the hypothesis of fixed-key equivalence for one round. We give the approximation sign '$\approx$' a quantitative meaning, but we leave to the reader the decision as to the "validity" of the hypothesis. Finally, we take a brief look at the multiple-round case.

4.1 Reminder and Definitions

4.1.1 Modification of The Statement of The Hypothesis

Recall that the i-round input/output sum (I/O sum) is defined as $S^{1\ldots i} = f_0(X) \oplus f_i(Y(i))$, where the functions $f_0, f_i : \mathbb{Z}_2^n \to \mathbb{Z}_2$ are balanced, X is

the plaintext and $Y(i)$ is the output of the i^{th} round of the cipher. When the round keys for rounds 1 to i are fixed, say, $Z_1 = z_1, \ldots, Z_i = z_i$, then the transformation from X to $Y(i)$ is performed by an invertible function that we denote by $g_{z_1,\ldots,z_i}$. Then the i-round I/O sum can be written as

$$f_0(X) \oplus f_i(g_{z_1,\ldots,z_i}(X)) = (f_0 \oplus f_i \circ g_{z_1,\ldots,z_i})(X). \tag{4.1}$$

This emphasizes the dependence on X and shows the form of the function applied to X. This is the notation we shall use hereafter. Recall also that the key-dependent imbalance of $S^{1\ldots i}$ given $Z_1 = z_1, \ldots, Z_i = z_i$ is defined by

$$I(S^{1\ldots i}|z_1, \ldots, z_i) := \left|2P[S^{1\ldots i} = 0|(Z_1, \ldots, Z_i) = (z_1, \ldots, z_i)\,] - 1\right|. \tag{4.2}$$

The average-key imbalance of $S^{1\ldots i}$ is the expected value of the key-dependent imbalance over all keys, i.e.,

$$\overline{I}(S^{1\ldots i}) := E[I(S^{1\ldots i}|Z_1, \ldots, Z_i)\,] = \frac{1}{(2^k)^i} \sum_{z_1,\ldots,z_i} I(S^{1\ldots i}|z_1, \ldots, z_i), \tag{4.3}$$

since the round keys are considered to be independent random variables uniformly distributed on $\mathbb{Z}_2^k$. The I/O sum $S^{1\ldots i}$ is called effective if $\overline{I}(S^{1\ldots i}) \approx 1$. We come now to the hypothesis of fixed-key equivalence as stated in [16] and [17].

Conjecture 4.1.1 (Hypothesis of Fixed-Key Equivalence)
For any effective i-round I/O sum $S^{1\ldots i}$ and for virtually all keys $z_1, \ldots, z_i$, the key-dependent imbalance $I(S^{1\ldots i}|z_1, \ldots, z_i)$ is virtually independent of the value $(z_1, \ldots, z_i)$ of the key, i.e.,

$$I(S^{1\ldots i}|z_1, \ldots, z_i) \approx \overline{I}(S^{1\ldots i}). \tag{4.4}$$

(Note that we state the hypothesis for i rounds and not just for $i = r - 1$ rounds as in (2.1); this makes no difference since r and i can take on any value.)

In the statement of this hypothesis, there is much room for interpretation: what does the approximation sign mean? What is to be understood by "virtually all keys"? When is an I/O sum to be considered as "effective", i.e., when can one consider its average-key imbalance to be approximately equal to 1? When an average of numbers between 0 and 1 is almost 1, then it is natural to conclude that virtually all of these numbers are close to 1 and therefore close to the average. But in linear cryptanalysis, an I/O sum with an average-key imbalance of 1/2 is already to be considered as

good and should therefore fall within the category of "effective" I/O sums. We find in the following example "effective" I/O sums for which (4.4) is not fulfilled.

Example 4.1.2
Consider a cipher with text blocklength 3 and round key length 3 defined by the round function of Table 4.1.

X \ Z	(0,0,0)	(0,0,1)	(0,1,0)	(0,1,1)	(1,0,0)	(1,0,1)	(1,1,0)	(1,1,1)
(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,0)	(0,0,1)
(0,0,1)	(1,0,0)	(0,0,1)	(1,0,0)	(1,1,0)	(0,0,1)	(0,1,1)	(0,1,0)	(1,1,1)
(0,1,0)	(1,1,0)	(1,0,1)	(0,0,1)	(1,1,1)	(1,0,0)	(1,0,0)	(0,1,1)	(1,1,0)
(0,1,1)	(0,1,0)	(1,0,0)	(0,1,1)	(0,0,1)	(1,0,1)	(1,0,1)	(1,0,0)	(0,0,0)
(1,0,0)	(0,0,1)	(1,1,1)	(1,0,1)	(1,0,0)	(1,1,0)	(1,1,1)	(1,0,1)	(0,1,1)
(1,0,1)	(1,1,1)	(1,1,0)	(1,1,0)	(1,0,1)	(0,1,0)	(0,0,1)	(1,1,0)	(1,0,1)
(1,1,0)	(1,0,1)	(0,1,0)	(0,1,0)	(0,1,0)	(1,1,1)	(0,1,0)	(0,1,1)	(0,1,0)
(1,1,1)	(0,1,1)	(0,1,1)	(1,1,1)	(0,1,1)	(0,1,1)	(1,1,0)	(1,1,1)	(1,0,0)

Table 4.1: Example of a round function.

As we will see below, for any round function, the variance of the key-dependent imbalances can never exceed 1/4. For all 1-round I/O sums defined by a pair of balanced functions (f_0, f_1), we have computed the average-key imbalance and the variance of the key-dependent imbalances. The largest average-key imbalance is 17/32 and it is reached for five pairs (f_0, f_1). For these pairs, the variance can take on different values: for $f_0(X) = X_1 \oplus X_2 \oplus X_3$ and $f_1(X) = X_1X_2 \oplus X_2 \oplus X_3$, it is only 7/1024. For $f_0(X) = X_1 \oplus X_2 \oplus X_3$ and $f_1(X) = X_1X_3 \oplus X_2X_3 \oplus X_1$, it is 167/1024. In the other three cases, it is 71/1024. Obviously, the first pair (f_0, f_1) is the best from the cryptanalyst's point of view. However, if one searches for only the pair of balanced functions that produces the largest, or at least one of the largest, average-key imbalance, one might very well find the second pair, which has a fairly large variance of the key-dependent imbalances and for which (4.4) does not hold.

For this reason, we regard the hypothesis of fixed-key equivalence as stated above as inadequately formulated. We modify it as follows.

Definition 4.1.3
An i-round I/O sum $S^{1\ldots i}$ *satisfies the fixed-key equivalence condition*, or *the fixed-key equivalence condition is valid for* $S^{1\ldots i}$, if $I(S^{1\ldots i}|z_1,\ldots,z_i) \approx \overline{I}(S^{1\ldots i})$ for virtually all keys $z_1,\ldots,z_i$.

Conjecture 4.1.4 (Hypothesis of Fixed-Key Equivalence)
For given blocklength n, round-key length k and integer i, almost all effective i-round I/O sums satisfy the fixed-key equivalence condition.

This version, too, contains some imprecision. Our goal is to give this a quantitative meaning. We will not end by concluding whether the hypothesis of fixed-key equivalence holds or not; what we will do is to quantify *how well* the approximation (4.4) holds. The reader can then decide whether this is well enough in his opinion to say that the hypothesis of fixed-key equivalence is valid.

Throughout this chapter, we will sometimes refer to the hypothesis of fixed-key equivalence simply as "the hypothesis".

4.1.2 A Measure of The Validity of The Fixed-Key Equivalence Condition

The statement that "the key-dependent imbalances are approximately equal to the average-key imbalance" contains the expression "approximately equal", which leaves much room for interpretation. To give it a quantitative meaning, we define a measure of the validity of the approximation (4.4) for an I/O sum. It will be equal to 0 if and only if the approximation is in fact an equality for all keys (in which case one says that the I/O sum satisfies the fixed-key equivalence condition *exactly*), it will be "small" if and only if the I/O sum satisfies the fixed-key equivalence condition to a certain extent, and it will be bounded by a constant for all I/O sums. Moreover, it will allow us to compare the validity of the fixed-key equivalence condition for different I/O sums.

Definition 4.1.5
For any i-round I/O sum $S^{1\ldots i}$, define

$$V(S^{1\ldots i}) := \frac{1}{(2^k)^i} \sum_{z_1,\ldots,z_i} (I(S^{1\ldots i}|z_1,\ldots,z_i) - \overline{I}(S^{1\ldots i}))^2. \qquad (4.5)$$

This is in fact nothing else than the variance of the key-dependent imbalances and depends only on the two balanced functions and the round function that define the I/O sum.

We will soon see some properties of V, based on the following Lemma.

Lemma 4.1.6
Let $a_1, \ldots, a_n$ be real numbers in the interval $[0,1]$ with average $\overline{a} = \frac{1}{n}\sum_{i=1}^{n} a_i$. Denote by $Var(a)$ the variance of the numbers $a_1, \ldots, a_n$ and by $\max(Var(a))$ the greatest variance that the numbers $a_1, \ldots, a_n$ can have under the condition that their average be $\overline{a}$. Then

1. *If $\sum_{i=1}^{n} a_i$ is an integer, then $\max(Var(a)) = \overline{a}(1-\overline{a})$.*
2. *If $\sum_{i=1}^{n} a_i$ is not an integer, then $\overline{a}(1-\overline{a}) - \frac{1}{4n} \leq \max(Var(a)) < \overline{a}(1-\overline{a})$.*

Proof:
Imagine the numbers $a_1, \ldots, a_n$ as points in the interval $[0,1]$, e.g., as in Figure 4.1. Now take two points that are not endpoints of the interval

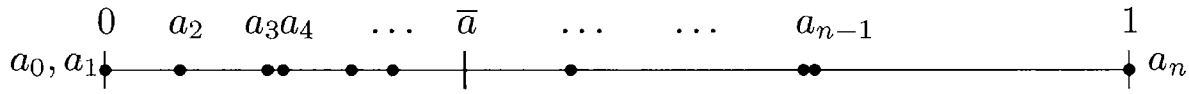


Figure 4.1: A Possible Distribution of Numbers on The Interval $[0,1]$.

and move the left one to the left and the right one to the right by the same amount until one of the points reaches the end of the interval. This does not change the average but increases the variance: if the left point is to the left of $\overline{a}$ and the right point to the right of $\overline{a}$, then the squared distance of both points to $\overline{a}$ increases; if both points are on the left of $\overline{a}$, then by this procedure the squared distance between the left point and $\overline{a}$ is larger than the squared distance between the right point and $\overline{a}$, increases faster than the latter decreases, and after a while the points will possibly be on opposite sides of $\overline{a}$; if both points are to the right of $\overline{a}$, a similar thing happens.

Continue with this procedure until you cannot move anymore points. At the end, there is at most one point left in the interior of the interval, and all other points sit either at 0 or at 1. (There cannot be two points in the interior because one could still move them as above.) Any starting configuration ends up with $\lfloor \sum_{i=1}^{n} a_i \rfloor$ points at 1, one point at $\sum_{i=1}^{n} a_i - \lfloor \sum_{i=1}^{n} a_i \rfloor$ (if $\sum_{i=1}^{n} a_i$ is not an integer), and the other points at 0. Thus, this procedure leads to the configuration that maximizes $Var(a)$ under the constraint of a given average $\overline{a}$.

1. If $\overline{a}n = \sum_{i=1}^{n} a_i$ is an integer, then no point remains in the interior, since otherwise the numbers would not add up to an integer. Thus, we

end up with $\overline{a}n$ points at 1 and the others at 0. The corresponding variance is $\big(\overline{a}n(1-\overline{a})^2 + (n-\overline{a}n)\overline{a}^2\big)/n = \overline{a}(1-\overline{a})$.

2. If $\overline{a}n = \sum_{i=1}^{n} a_i$ is not an integer, then a point remains in the interior of the interval at position $\beta = \sum_{i=1}^{n} a_i - \lfloor \sum_{i=1}^{n} a_i \rfloor$ $(0 < \beta < 1)$. Then there are $\lfloor \overline{a}n \rfloor = \overline{a}n - \beta$ points at 1 and $n - \overline{a}n + \beta - 1$ points at 0. The corresponding variance is $\big((\overline{a}n-\beta)(1-\overline{a})^2 + (n-\overline{a}n+\beta-1)\overline{a}^2 + (\beta-\overline{a})^2\big)/n = \overline{a}(1-\overline{a}) - \beta(1-\beta)/n$. The statement of the Lemma follows now from the fact that $0 < \beta(1-\beta) \leq 1/4$. □

Corollary 4.1.7
For any i-round I/O sum $S^{1\ldots i}$:

1. $0 \leq V(S^{1\ldots i}) \leq 1/4$;
2. $V(S^{1\ldots i}) = 0$ *if and only if* $S^{1\ldots i}$ *satisfies the fixed-key equivalence condition exactly;*
3. $V(S^{1\ldots i}) = 1/4$ *if and only if half of the key-dependent imbalances of* $S^{1\ldots i}$ *are zero and the remaining key-dependent imbalances are 1.*

Proof:

1. The lower bound holds by definition. Since $V(S^{1\ldots i})$ is the variance of the key-dependent imbalances, we have, by Lemma 4.1.6, $V(S^{1\ldots i}) \leq \overline{I}(S^{1\ldots i})\big(1-\overline{I}(S^{1\ldots i})\big) \leq 1/4$.
2. Is obvious.
3. Let $V(S^{1\ldots i}) = 1/4$, i.e., we have the largest value V can take on. This means that there is a value $\overline{I}$ of the average-key imbalance such that $V = \overline{I}(1-\overline{I})$. This implies that $\overline{I} = 1/2$; this situation can be reached from a starting configuration with average $\overline{I} = 1/2$ by applying the procedure described in the above Lemma. Because we end up with equality in $V = \overline{I}(1-\overline{I})$, we are in situation 1. of the Lemma and at the end there is no point, that is, no key-dependent imbalance, in the interior of the interval. Thus, half of the key-dependent imbalances must be 1 and the others 0. The proof of the converse is trivial. □

The hypothesis of fixed-key equivalence holds if and only if $V(S^{1\ldots i}) \approx 0$ for almost all effective I/O sums. Again, what is meant by the approximation sign is subjective. Since $V(S^{1\ldots i})$ is a variance, it would be nice if it were (much) smaller than the square of the average $\overline{I}(S^{1\ldots i})$ for almost all effective I/O sums, since in this case we could in good conscience consider $I(S^{1\ldots i}|z_1,\ldots,z_i)$ as "almost the same" for all $z_1,\ldots,z_i$. Before proceeding further, we give an example.

Example 4.1.8
Let $n = 3$, $f_0(X) = X_1X_2 \oplus X_3$, $f_1(X) = X_1 \oplus X_2 \oplus X_3$ for the balanced functions, and $g(X, Z) = (X_3, X_2, X_2X_3Z_1 \oplus X_1 \oplus X_2Z_2 \oplus X_3Z_3)$ for the round function. (Here, exceptionally, the Z_i's denote the components of the first-round key.) Then

$$S^1 = X_1X_2 \oplus X_2X_3Z_1 \oplus X_1 \oplus X_2(Z_2 \oplus 1) \oplus X_3Z_3.$$

Table 4.2 shows the value of S^1 as a function of X and Z. In this example, $\overline{I}(S^1) = 1/4$ and $V(S^1) = 1/16$.

			0	1	0	1	0	1	0	1
		Z	0	0	1	1	0	0	1	1
	X		0	0	0	0	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	1	0	0	0	1
0	1	0	1	1	0	0	1	0	0	0
0	1	1	1	0	0	1	0	1	1	0
1	0	0	1	1	1	1	1	1	1	1
1	0	1	1	0	1	0	1	1	1	0
1	1	0	1	1	0	0	1	0	0	0
1	1	1	1	0	0	1	0	1	1	0
$I(S^1\|Z) \rightarrow$			1/2	0	1/2	0	0	1/2	0	1/2

Table 4.2: Value of S^1 in function of X and Z.

Remark 4.1.9
Hereafter in this chapter, we use the following notation: f_1 and f_2 stand for *any* balanced functions, not necessarily the balanced functions applied to the output of the first and the second round, respectively. For instance, we might say that an i-round I/O sum is of the form $f_1(X) \oplus f_2(Y(i))$ for some balanced functions f_1 and f_2.

4.2 Balanced Functions

In this section, we study some properties of balanced functions, in particular sums of two balanced functions. These observations are not only necessary for our further argumentation, but they also give some insight into the behaviour of balanced functions.

4.2.1 Playing With Balanced Functions

Recall that the imbalance of a function f is defined as the imbalance of the random variable $f(X)$, where X is a uniformly distributed random variable and that, if g is an invertible function, then $I(f \circ g) = I(f)$ (see Subsection 2.3.5).

Lemma 4.2.1
For any integer $n \geq 2$, we have:

1. *For any function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, $I(f) \in \{0, \frac{1}{2^{n-1}}, \ldots, 1\}$.*
2. *If $f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{Z}_2$ are balanced, then $I(f_1 \oplus f_2) \in \{0, \frac{1}{2^{n-2}}, \ldots, 1\}$.*
3. *For any positive integer i, any fixed keys $z_1, \ldots, z_i$ and any i-round I/O sum $S^{1 \ldots i}$, the key-dependent imbalance $I(S^{1 \ldots i} | z_1, \ldots, z_i)$ must take on one of the values $0, \frac{1}{2^{n-2}}, \frac{2}{2^{n-2}}, \ldots, 1$, where n is the text block length.*

Proof:

1. This follows immediately from Lemma 2.3.15.
2. Call the subset of $\mathbb{Z}_2^n$ on which f_1 is equal to 0 "the first half of $\mathbb{Z}_2^n$" and the rest of $\mathbb{Z}_2^n$ "the second half of $\mathbb{Z}_2^n$". In the first half, f_2 has value 0 α times and value 1 $2^{n-1} - \alpha$ times for some $0 \leq \alpha \leq 2^{n-1}$. The same holds for $f_1 \oplus f_2$. In the second half, f_2 has value 0 $2^{n-1} - \alpha$ times and value 1 $2^{n-1} - (2^{n-1} - \alpha) = \alpha$ times, while $f_1 \oplus f_2$ takes on the value 1 $2^{n-1} - \alpha$ times and the value 0 $2^{n-1} - (2^{n-1} - \alpha) = \alpha$ times. Hence, on the whole of $\mathbb{Z}_2^n$, $f_1 \oplus f_2$ is equal to zero 2α times. Thus, $I(f_1 \oplus f_2) = |2P[(f_1 \oplus f_2)(X) = 0] - 1| = |2\frac{2\alpha}{2^n} - 1| = |\frac{\alpha}{2^{n-2}} - 1|$, which is of the stated form.
3. By Remark 2.3.16, the key-dependent imbalance of $S^{1 \ldots i}$ is equal to $I(f_1 \oplus f_2 \circ g_{z_1, \ldots, z_i})$, where $g_{z_1, \ldots, z_i}$ is some invertible function and f_1 and f_2 are some balanced functions. By Lemma 2.3.11, $f_2 \circ g_{z_1, \ldots, z_i}$ is balanced. By point 2. of this Lemma, $I(f_1 \oplus f_2 \circ g_{z_1, \ldots, z_i}) \in \{0, \frac{1}{2^{n-2}}, \ldots, 1\}$.

□

Corollary 4.2.2
Let $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ and $0 \leq i \leq 2^{n-1}$ with $I(f) = i/2^{n-1}$. Then f takes on the value zero either $2^{n-1} + i$ or $2^{n-1} - i$ times.

Proof:

Let X be a uniformly distributed random variable on $\mathbb{Z}_2^n$ and $\alpha = P[f(X) = 0]$. Then

$$\frac{i}{2^{n-1}} = I(f) = |2P[f(X) = 0] - 1| = |2\alpha - 1| \quad \Leftrightarrow$$

$$\left(2\alpha - 1 = \frac{i}{2^{n-1}} \quad \text{or} \quad 1 - 2\alpha = \frac{i}{2^{n-1}}\right) \quad \Leftrightarrow$$

$$\left(\alpha = \frac{2^{n-1} + i}{2^n} \quad \text{or} \quad \alpha = \frac{2^{n-1} - i}{2^n}\right). \qquad \square$$

Proposition 4.2.3
Let $f_1 : \mathbb{Z}_2^n \to \mathbb{Z}_2$ be balanced. Then the set $\{f_1 \oplus f_2 | \, f_2$ balanced$\}$ contains $2\binom{2^{n-1}}{2^{n-2}-i}^2$ elements with $I(f_1 \oplus f_2) = i/2^{n-2}$, $i = 1, \ldots, 2^{n-2}$ and $\binom{2^{n-1}}{2^{n-2}}^2$ elements with $I(f_1 \oplus f_2) = 0$.

Proof:

Let again $\{x \in \mathbb{Z}_2^n | \, f_1(x) = 0\}$ be "the first half of $\mathbb{Z}_2^n$" and the rest of $\mathbb{Z}_2^n$ be "the second half of $\mathbb{Z}_2^n$". We have seen in the proof of Lemma 4.2.1 that if f_2 takes on the value 0 α times in the first half, then $f_1 \oplus f_2$ will be equal to 0 2α times on $\mathbb{Z}_2^n$. Moreover, there are $\binom{2^{n-1}}{\alpha}^2$ functions f_2 with this property. (One must choose α elements in the first half of $\mathbb{Z}_2^n$ for which f_2 is zero and α elements in the second half for which f_2 is one.)

Let $i \neq 0$. In order to have $I(f_1 \oplus f_2) = i/2^{n-2} = 2i/2^{n-1}$, $f_1 \oplus f_2$ must have value 0 either $2^{n-1} + 2i$ times or $2^{n-1} - 2i$ times (by Lemma 4.2.2). So necessarily $2\alpha = 2^{n-1} \pm 2i$, i.e., $\alpha = 2^{n-2} \pm i$. The total number of possibilities for f_2 is then $\binom{2^{n-1}}{2^{n-2}+i}^2 + \binom{2^{n-1}}{2^{n-2}-i}^2 = 2\binom{2^{n-1}}{2^{n-2}-i}^2$.

Now in order to have $I(f_1 \oplus f_2) = 0 = 0/2^{n-2}$, $f_1 \oplus f_2$ must have value 0 2^{n-1} times so necessarily $2\alpha = 2^{n-1}$, i.e., $\alpha = 2^{n-2}$. Thus, the total number of possibilities for f_2 is $\binom{2^{n-1}}{2^{n-2}}^2$. $\square$

Example 4.2.4
Table 4.3 shows these numbers for $n = 3$.

Value of $I(f_1 \oplus f_2)$	number of times	
	combinatorial	numerical
1	$2\binom{4}{0}^2$	2
1/2	$2\binom{4}{1}^2$	32
0	$\binom{4}{2}^2$	36

Table 4.3: Frequency of occurrence of $I(f_1 \oplus f_2)$ for $n = 3$ for a fixed balanced function f_1.

4.2.2 Algebraic Considerations

In this subsection, G will always be the set of invertible functions on $\mathbb{Z}_2^n$. We recall that G with the operation $\circ$, the concatenation of functions, is a group.

Lemma 4.2.5
For any balanced function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, the set $H(f) = \{g \in G | f \circ g = f\}$ is a subgroup of G.
Proof: $H(f) \subseteq G$ is obvious. Let g_1 and g_2 be elements of $H(f)$. Then $f \circ (g_1 \circ g_2) = (f \circ g_1) \circ g_2 = f \circ g_2 = f$ and $f \circ g_1^{-1} = (f \circ g_1) \circ g_1^{-1} = f \circ (g_1 \circ g_1^{-1}) = f$. Thus, $g_1 \circ g_2 \in H(f)$ and $g_1^{-1} \in H(f)$. □

Lemma 4.2.6
For any balanced function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ and any $g \in G$, the right coset $H(f) \circ g$ is characterised by

$$H(f) \circ g = \{h \in G | f \circ h = f \circ g\}.$$

Proof:
If $h \in H(f) \circ g$ then there is $h' \in H(f)$ with $h = h' \circ g$. But then $f \circ h' = f$ and $f \circ h = f \circ h' \circ g = f \circ g$.
If $f \circ h = f \circ g$, then $f \circ h \circ g^{-1} = f$, i.e., $h \circ g^{-1} \in H(f)$, from which it follows that $h \in H(f) \circ g$. □

Corollary 4.2.7
For any balanced $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ and any $g \in G$, there are $|H(f)|$ functions h in G such that $f \circ h = f \circ g$. □

Corollary 4.2.8
For any balanced $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, the multiset $\{f \circ g | g \in G\}$ contains each balanced function $|H(f)|$ times.
Proof:
Let e be the neutral element of $\langle G, \circ \rangle$. By Corollary 4.2.7, for each balanced function $\tilde{f}$ there are $|H(f)|$ elements $h \in G$ with $f \circ h = \tilde{f} \circ e = \tilde{f}$, and all functions of the form $f \circ h$, $h \in G$, are balanced. □

Lemma 4.2.9
Let $f, f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{Z}_2$ be balanced, let $g_1, g_2 \in G$ with $f \circ g_1 = f_1$ and $f \circ g_2 = f_2$, and let $f_1 \neq f_2$. (By Corollary 4.2.8, g_1 and g_2 exist.) Then $H(f) \circ g_1 \cap H(f) \circ g_2 = \varnothing$.
Proof:
If there were an element g in $H(f) \circ g_1 \cap H(f) \circ g_2$, then there would exist h_1 and h_2 in $H(f)$ such that $g = h_1 \circ g_1 = h_2 \circ g_2$. But then

$f_1 = f \circ g_1 = f \circ h_1 \circ g_1 = f \circ g = f \circ h_2 \circ g_2 = f \circ g_2 = f_2$, which is a contradiction. □

Corollary 4.2.10
For any balanced $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, $|H(f)| = (2^n)!/\binom{2^n}{2^{n-1}}$.
Proof:

By Lemma 4.2.9, there are as many cosets of $H(f)$ as there are balanced functions, namely, $\binom{2^n}{2^{n-1}}$. Thus, $(2^n)! = |G| = \binom{2^n}{2^{n-1}}|H(f)|$. □

Proposition 4.2.11
For any balanced functions $f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{Z}_2$, *the multiset* $\{f_1 \oplus f_2 \circ g \,|\, g \in G\}$ *has* $(2^n)!$ *elements, of which*

$$\binom{2^{n-1}}{2^{n-2}}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}} \quad \textit{have imbalance 0 and}$$

$$2\binom{2^{n-1}}{2^{n-2}-i}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}} \quad \textit{have imbalance } \frac{i}{2^{n-2}}, \quad 1 \le i \le 2^{n-2}.$$

Proof:

By Corollary 4.2.8, the multiset $\{f_2 \circ g | \, g \in G\}$ contains each balanced function $|H(f_2)| = (2^n)!/\binom{2^n}{2^{n-1}}$ times. By Proposition 4.2.3, the set $\{f_1 \oplus f | \, f : \mathbb{Z}_2^n \to \mathbb{Z}_2 \text{ balanced}\}$ contains $\binom{2^n}{2^{n-1}}$ elements, of which $\binom{2^{n-1}}{2^{n-2}}^2$ have imbalance 0 and $2\binom{2^{n-1}}{2^{n-2}-i}^2$ have imbalance $i/2^{n-2}$, for $i = 1, \ldots, 2^{n-2}$. Combining these two facts completes the proof. □

4.3 Validity of The Hypothesis of Fixed-Key Equivalence for One Round

In this long section, we study the validity of the hypothesis of fixed-key equivalence for one round, that is, with $i = 1$ in (4.4). We use an averaging argument: we consider a fixed pair (f_1, f_2) of balanced functions; then both the average-key imbalance $\overline{I}(S^1)$ and the validity measure $V(S^1)$ depend on the round function only. We compute the proportion of round functions that yield a particular value of $\overline{I}$ or of V; then we can calculate the average and the variance of $\overline{I}$ and V over all round functions. These moments depend only on the text length n and the key length k. This means that we are not considering only effective I/O sums. Nevertheless, we shall be able to say how well (4.4) holds in general for 1-round I/O

sums. More importantly, the results will also allow us to say for any number of rounds what average-key imbalance can be expected from an I/O sum; this will help in making the notion of an effective I/O sum more precise.

We begin by considering the cases $k = n = 1$ and $k = n = 3$.

4.3.1 Text and Key of Length 1

In this case, the balanced functions are $f_1(X) = X \oplus \alpha_1$ and $f_2(X) = X \oplus \alpha_2$ for some α_1 and α_2 in $\mathbb{Z}_2$ and the round function is $g(X, z) = X \oplus \beta(z)$ for some function $\beta : \mathbb{Z}_2 \to \mathbb{Z}_2$. Then $S^1 = f_1(X) \oplus f_2(g(X, Z)) = \alpha_1 \oplus \alpha_2 \oplus \beta(Z)$; thus, $I(S^1|z) = 1$ for all z and $\overline{I}(S^1) = 1$ so all one-round I/O sums satisfy the fixed-key equivalence condition exactly and the hypothesis of fixed-key equivalence holds exactly.

4.3.2 Text and Key of Length 3

We skip to the case where both the text and the key have length 3 because this case is still simple enough to keep a general view but already complicated enough to give us an idea about how to tackle the problem for general text and key lengths.

By Proposition 4.2.11, we know that if G is the set of all invertible functions on $\mathbb{Z}_2^3$, then for any fixed balanced functions f_1 and f_2, the multiset $\{f_1 \oplus f_2 \circ g | g \in G\}$ contains $2\binom{4}{0}^2 \cdot 8!/\binom{8}{4} = 1152$ elements with imbalance 1, $2\binom{4}{1}^2 \cdot 8!/\binom{8}{4} = 18432$ elements with imbalance $1/2$ and $\binom{4}{2}^2 \cdot 8!/\binom{8}{4} = 20736$ elements with imbalance 0.

By Lemma 4.2.1, the key-dependent imbalance of S^1 can take on the values $0, 1/2$ or 1. For any round function g, let α_g be the number of first-round keys z for which $I(S^1|z) = 0$, β_g the number of z for which $I(S^1|z) = 1/2$ and γ_g the number of z for which $I(S^1|z) = 1$. (α_g, β_g, γ_g sum up to 8.) Then

$$\overline{I}(S^1) = \frac{\beta_g + 2\gamma_g}{16} \quad \text{and} \quad V(S^1) = \frac{\beta_g + 4\gamma_g}{32} - (\overline{I}(S^1))^2.$$

We continue to consider a fixed pair (f_1, f_2) of balanced functions. In order to determine how many round functions give a particular value of $\overline{I}(S^1)$ or of $V(S^1)$, we calculate, for each triple of non-negative integers (α, β, γ) with $\alpha + \beta + \gamma = 8$, the number of round functions g such that $(\alpha_g, \beta_g, \gamma_g) = (\alpha, \beta, \gamma)$. We use the fact that choosing a round function is equivalent to choosing eight invertible functions, one for each possible value of the first-round key. What must we do in order to get (α, β, γ)?

- Among the 8 first-round keys, we must choose γ for which $I(S^1|z) = 1$; for each of these γ keys, we have a choice among 1152 possible invertible functions for g_z.
- Among the remaining $8 - \gamma$ keys, we must then choose β for which $I(S^1|z) = 1/2$; for each of these keys, there are 18432 possible invertible functions g_z.
- For each of the remaining $8 - \gamma - \beta$ keys, there are 20736 possible invertible functions.

Thus, there are $\binom{8}{\gamma}\binom{8-\gamma}{\beta} \times 1152^{\gamma} \times 18432^{\beta} \times 20736^{8-\gamma-\beta}$ round functions giving (α, β, γ). The probability, taken over all round functions, of obtaining (α, β, γ) is then

$$\binom{8}{\gamma}\binom{8-\gamma}{\beta} \times 1152^{\gamma} \times 18432^{\beta} \times 20736^{8-\gamma-\beta} \times 40320^{-8} = \binom{8}{\gamma}\binom{8-\gamma}{\beta} \times 1^{\gamma} \times 16^{\beta} \times 18^{8-\gamma-\beta} \times 35^{-8}.$$

It should be noticed that different (α, β, γ) can yield the same value of $\overline{I}(S^1)$ or of $V(S^1)$; by adding the corresponding numbers we get the distribution of the values of $\overline{I}$ and of V over all round functions; this distribution is shown in Tables 4.4 and 4.5.

The interpretation of the tables is the following: when the text and the key both have length three, then for any balanced functions $f_1, f_2 : \mathbb{Z}_2^3 \to \mathbb{Z}_2$, approximately 25% of the round functions for instance will give $\overline{I}(S^1) = 1/4$ and approximately 4% of the round functions will give $V(S^1) = 1/8$. One sees that for most round functions, $V(S^1)$ is "small", which could indicate that (4.4) holds for most one-round I/O sums. We also notice that, for most round functions, the average-key imbalance is also "small". But if an average of non-negative numbers that are upper-bounded by 1 is small, then most of those numbers must be small and so must also be their variance. Thus, it is natural that $V(S^1)$ be small in most cases. Our hope is to prove that, in most cases, V is sufficiently small to allow us to say that (4.4) holds.

We can consider the round functions g as realizations of a discrete, uniformly distributed random variable and $V(S^1)$ resp. $\overline{I}(S^1)$ as functions of it. Then what is shown in the tables is the probability distribution of $V(S^1)$ and of $\overline{I}(S^1)$. The corresponding expected values and variances are:

$$E[\overline{I}] = 0.2571 = 9/35;$$
$$Var(\overline{I}) = 9.59184 * 10^{-3} = 47/4900;$$

$\overline{I}$	$16 * \overline{I}$	proportion
0	0	$4.89 * 10^{-3}$
0.063	1	$3.48 * 10^{-2}$
0.125	2	$1.10 * 10^{-1}$
0.188	3	$2.06 * 10^{-1}$
0.250	4	$2.50 * 10^{-1}$
0.313	5	$2.08 * 10^{-1}$
0.375	6	$1.20 * 10^{-1}$
0.438	7	$4.87 * 10^{-2}$
0.500	8	$1.38 * 10^{-2}$
0.563	9	$2.70 * 10^{-3}$
0.625	10	$3.71 * 10^{-4}$
0.688	11	$3.56 * 10^{-5}$
0.750	12	$2.39 * 10^{-6}$
0.813	13	$1.09 * 10^{-7}$
0.888	14	$3.25 * 10^{-9}$
0.938	15	$3.25 * 10^{-11}$
1	16	$4.44 * 10^{-13}$

Table 4.4: Distribution of the values of $\overline{I}(S^1)$.

V	$256 * V$	proportion
0	0	$6.80 * 10^{-3}$
0.027	7	$5.29 * 10^{-2}$
0.047	12	$1.76 * 10^{-1}$
0.059	15	$3.45 * 10^{-1}$
0.063	16	$2.21 * 10^{-1}$
0.090	23	$2.68 * 10^{-2}$
0.109	28	$4.98 * 10^{-2}$
0.121	31	$6.70 * 10^{-2}$
0.125	32	$4.00 * 10^{-2}$
0.152	39	$6.27 * 10^{-3}$
0.172	44	$5.03 * 10^{-3}$
0.184	47	$2.26 * 10^{-3}$
0.188	48	$7.94 * 10^{-4}$
0.215	55	$2.20 * 10^{-4}$
0.234	60	$4.71 * 10^{-5}$
0.250	64	$3.26 * 10^{-6}$

Table 4.5: Distribution of the values of $V(S^1)$.

$$E[V] = 0.0671 = 47/700;$$
$$Var(V) = 7.97549 * 10^{-4} = 87539/109'760'000.$$

The exact fractional expressions will be proved later (Corollary 4.3.6 and end of this section).

4.3.3 Any Text and Key Lengths

One can follow the same procedure for any text length $n \geq 2$ and any key length $k \geq 1$. However, even for $n = 4$, the tables of the distribution of the values of $\overline{I}(S^1)$ and of $V(S^1)$ become too large to overview. Thus, we concentrate on the probability distribution of $\overline{I}$ and V over all round functions and on their average and variance, for which we will find simple forms. We will prove that, if n and k are sufficiently large, then the variance of $\overline{I}$ and of V is so much smaller than their average that we can consider $\overline{I}$ and V as having virtually constant value. Then we shall be able to make quantitative statements about the validity of (4.4), because $\overline{I}$ and V are the average and the variance, respectively, of the key-dependent imbalances.

In the general case, there are 2^k possible keys for the first round. Thus, choosing a round function is the same as choosing 2^k invertible functions. We again consider two fixed balanced functions $f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{Z}_2$. We have seen in Lemma 4.2.1 that the key-dependent imbalance $I(S^1|z)$ must take on one of the values $0, \frac{1}{2^{n-2}}, \frac{2}{2^{n-2}}, \ldots, 1$. We make the following definitions.

Definition 4.3.1

- For any round function g and any integer i, $0 \leq i \leq 2^{n-2}$, let $\alpha_i(g)$ be the number of keys z such that $I(S^1|z) = i/2^{n-2}$.
- Let $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_{2^{n-2}})$ and denote by $\overline{I}(\boldsymbol{\alpha})$ (resp. $V(\boldsymbol{\alpha})$) the average-key imbalance (resp. the value of the validity measure) resulting from the distribution $\boldsymbol{\alpha}$ of the key-dependent imbalances.

Knowing $\boldsymbol{\alpha}$, the values of $\overline{I}(\boldsymbol{\alpha})$ and $V(\boldsymbol{\alpha})$ are easy to calculate:

Lemma 4.3.2
If $n \geq 2$ and $k \geq 1$, then

$$\begin{aligned} \overline{I}(\boldsymbol{\alpha}) &= \frac{\alpha_1 + 2\alpha_2 + \cdots + 2^{n-2}\alpha_{2^{n-2}}}{2^{n-2}2^k} \quad \textit{and} \\ V(\boldsymbol{\alpha}) &= \frac{\alpha_1 + 4\alpha_2 + \cdots + 2^{2(n-2)}\alpha_{2^{n-2}}}{2^{2(n-2)}2^k} - \left(\overline{I}(\boldsymbol{\alpha})\right)^2. \end{aligned}$$

Proof:
The right side of the first equation is

$$\frac{1}{2^k} \sum_{i=0}^{2^{n-2}} \left(\frac{i}{2^{n-2}} \times \left| \left\{ z \,\middle|\, I(S^1|z) = \frac{i}{2^{n-2}} \right\} \right| \right),$$

which is the average of the key-dependent imbalance $I(S^1|z)$ over all keys. The proof of the second equality is similar: one considers the variance of the key-dependent imbalance over all keys, which is

$$\frac{1}{2^k} \sum_{i=0}^{2^{n-2}} \left(\left(\frac{i}{2^{n-2}} \right)^2 \times \left| \left\{ z \,\middle|\, I(S^1|z) = \frac{i}{2^{n-2}} \right\} \right| \right) - \left(\overline{I}(\boldsymbol{\alpha})\right)^2.$$

□

Next, we compute, for each $\boldsymbol{\alpha}$, the number and the proportion of the round functions g such that $\boldsymbol{\alpha}(g) = \boldsymbol{\alpha}$.

Proposition 4.3.3
Let $P(\boldsymbol{\alpha})$ denote the probability, taken over all round functions g, that $\boldsymbol{\alpha}(g) = \boldsymbol{\alpha}$. Then, for any $n \geq 2$ and any $k \geq 1$,

$$P(\boldsymbol{\alpha}) = \binom{2^k}{\alpha_{2^{n-2}}} \times \binom{2^k - \alpha_{2^{n-2}}}{\alpha_{2^{n-2}-1}} \times \cdots \times \binom{2^k - \alpha_{2^{n-2}} - \cdots - \alpha_2}{\alpha_1} \times$$

$$\left(2\left(\frac{\binom{2^{n-1}}{0}}{\binom{2^{n-1}}{2^{n-2}}}\right)^2\right)^{\alpha_{2^{n-2}}} \times \left(2\left(\frac{\binom{2^{n-1}}{1}}{\binom{2^{n-1}}{2^{n-2}}}\right)^2\right)^{\alpha_{2^{n-2}-1}} \times \cdots \times \left(2\left(\frac{\binom{2^{n-1}}{2^{n-2}-1}}{\binom{2^{n-1}}{2^{n-2}}}\right)^2\right)^{\alpha_1}$$

$$\times \left(\frac{\binom{2^{n-1}}{2^{n-2}}^2}{\binom{2^n}{2^{n-1}}}\right)^{2^k}. \qquad (4.6)$$

(Purposely, we have not converted the binomial coefficients on the first line into a multinomial coefficient.)

Proof:

There are $\binom{2^k}{\alpha_{2^{n-2}}} \times \binom{2^k - \alpha_{2^{n-2}}}{\alpha_{2^{n-2}-1}} \times \cdots \times \binom{2^k - \alpha_{2^{n-2}} - \cdots - \alpha_2}{\alpha_1}$ ways to split a set of 2^k elements (the set of all keys for the first round) into $2^{n-2}+1$ sets with $\alpha_{2^{n-2}}, \alpha_{2^{n-2}-1}, \ldots, \alpha_1, \alpha_0 = 2^k - \alpha_{2^{n-2}} - \cdots - \alpha_1$ elements, respectively.

Moreover, we know from Proposition 4.2.11 that if G is the set of invertible functions on $\mathbb{Z}_2^n$, then for any balanced functions $f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{Z}_2$, the multiset $\{f_1 \oplus f_2 \circ g | g \in G\}$ contains

$$\binom{2^{n-1}}{2^{n-2}}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}} \quad \text{elements with imbalance 0 and}$$

$$2\binom{2^{n-1}}{2^{n-2}-i}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}} \quad \text{elements with imbalance } \frac{i}{2^{n-2}}, \quad 1 \leq i \leq 2^{n-2}.$$

Thus, once α_i keys have been chosen, $1 \leq i \leq 2^{n-2}$, there are

$$\left(2\binom{2^{n-1}}{2^{n-2}-i}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}}\right)^{\alpha_i}$$

ways to choose α_i invertible functions g_z such that $I(S^1|z) = i/2^{n-2}$, $1 \leq i \leq 2^{n-2}$, and

$$\left(\binom{2^{n-1}}{2^{n-2}}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}}\right)^{2^k - \alpha_{2^{n-2}} - \cdots - \alpha_1}$$

ways to choose the remaining $2^k - \alpha_{2^{n-2}} - \cdots - \alpha_1$ invertible functions g_z such that $I(S^1|z) = 0$. Hence, the number of round functions such that $\boldsymbol{\alpha}(g) = \boldsymbol{\alpha}$ is

$$\binom{2^k}{\alpha_{2^{n-2}}} \times \binom{2^k - \alpha_{2^{n-2}}}{\alpha_{2^{n-2}-1}} \times \cdots \times \binom{2^k - \alpha_{2^{n-2}} - \cdots - \alpha_2}{\alpha_1} \times$$

$$\left(2\binom{2^{n-1}}{0}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}}\right)^{\alpha_{2^{n-2}}} \times \left(2\binom{2^{n-1}}{1}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}}\right)^{\alpha_{2^{n-2}-1}} \times \cdots$$

$$\times \left(2\binom{2^{n-1}}{2^{n-2}-1}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}}\right)^{\alpha_1} \times \left(\binom{2^{n-1}}{2^{n-2}}^2 \frac{(2^n)!}{\binom{2^n}{2^{n-1}}}\right)^{2^k - \alpha_{2^{n-2}} - \cdots - \alpha_1}.$$

Finally, dividing by $(2^n)!^{(2^k)}$, the number of possible round functions, and simplifying the expression yields the result (4.6). □

This allows us now to write $E[\overline{I}]$, $Var(\overline{I})$, $E[V]$ and $Var(V)$, the average and the variance of $\overline{I}$ and of V over all round functions, where f_1, f_2 are two fixed balanced functions, as functions of n and k only. We begin with $E[\overline{I}]$.

Corollary 4.3.4
For any integers $n \geq 2$ and $k \geq 1$, we have $E[\overline{I}] = \sum_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha})\overline{I}(\boldsymbol{\alpha}) =$

$$\left(\frac{\binom{2^{n-1}}{2^{n-2}}^2}{\binom{2^n}{2^{n-1}}}\right)^{2^k} \sum_{\alpha_{2^{n-2}}=0}^{2^k} \sum_{\alpha_{2^{n-2}-1}=0}^{2^k - \alpha_{2^{n-2}}} \cdots \sum_{\alpha_1=0}^{2^k - \alpha_{2^{n-2}} - \cdots - \alpha_2} \Bigg\{$$

$$\binom{2^k}{\alpha_{2^{n-2}}} \times \binom{2^k - \alpha_{2^{n-2}}}{\alpha_{2^{n-2}-1}} \times \cdots \times \binom{2^k - \alpha_{2^{n-2}} - \cdots - \alpha_2}{\alpha_1} \times$$

$$\left(2\left(\frac{\binom{2^{n-1}}{0}}{\binom{2^{n-1}}{2^{n-2}}}\right)^2\right)^{\alpha_{2^{n-2}}} \times \left(2\left(\frac{\binom{2^{n-1}}{1}}{\binom{2^{n-1}}{2^{n-2}}}\right)^2\right)^{\alpha_{2^{n-2}-1}} \times \cdots \times \left(2\left(\frac{\binom{2^{n-1}}{2^{n-2}-1}}{\binom{2^{n-1}}{2^{n-2}}}\right)^2\right)^{\alpha_1}$$

$$\times \frac{\alpha_1 + 2\alpha_2 + \cdots + 2^{n-2}\alpha_{2^{n-2}}}{2^{n-2}2^k} \Bigg\}.$$

□

Fortunately, this can be written in a simple form. The following definition will be a big help.

Definition 4.3.5
For any non-negative integers i and j, we define

$$f(i,j) := 2\left(\frac{\binom{2j}{j-i}}{\binom{2j}{j}}\right)^2 .$$

Proposition 4.3.6
For any integers $n \geq 2$ and $k \geq 1$, we have $E[\bar{I}] = \binom{2^{n-1}}{2^{n-2}}^2 / 2\binom{2^n}{2^{n-1}}$.

Proof:

In order to simplify the notation, we introduce $m := 2^{n-2}$ and $p := 2^k$. We take the long expression in Corollary 4.3.4 and equate it to $\binom{2^{n-1}}{2^{n-2}}^2 / 2\binom{2^n}{2^{n-1}}$. Then we move to the other side of the equation obtained the term before the sums and the denominator of the last fraction to get the equation

$$\sum_{\alpha_m=0}^{p} \sum_{\alpha_{m-1}=0}^{p-\alpha_m} \cdots \sum_{\alpha_1=0}^{p-\alpha_m-\cdots-\alpha_2} \left\{ \binom{p}{\alpha_m} \times \binom{p-\alpha_m}{\alpha_{m-1}} \times \cdots \times \binom{p-\alpha_m-\cdots-\alpha_2}{\alpha_1} \right.$$
$$\times \left(2\left(\frac{\binom{2m}{0}}{\binom{2m}{m}}\right)^2\right)^{\alpha_m} \times \left(2\left(\frac{\binom{2m}{1}}{\binom{2m}{m}}\right)^2\right)^{\alpha_{m-1}} \times \cdots \times \left(2\left(\frac{\binom{2m}{m-1}}{\binom{2m}{m}}\right)^2\right)^{\alpha_1}$$
$$\left. \times (\alpha_1 + 2\alpha_2 + \cdots + m\,\alpha_m) \right\} = \frac{pm}{2}\left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^{p-1} .$$

Now we take out of each sum everything that does not depend on its index and use the definition of f. The equation becomes

$$\sum_{\alpha_m=0}^{p} \binom{p}{\alpha_m} f(m,m)^{\alpha_m} \sum_{\alpha_{m-1}=0}^{p-\alpha_m} \binom{p-\alpha_m}{\alpha_{m-1}} f(m-1,m)^{\alpha_{m-1}} \cdots$$
$$\sum_{\alpha_1=0}^{p-\alpha_m-\cdots-\alpha_2} \binom{p-\alpha_m-\cdots-\alpha_2}{\alpha_1} f(1,m)^{\alpha_1} (\alpha_1 + 2\alpha_2 + \cdots + m\alpha_m)$$
$$= \frac{pm}{2}\left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^{p-1} . \tag{4.7}$$

It appears as if we could define the left side of the equation recursively. In fact, we can: for any integers $n \geq 0$, $k \geq 1$ and $a \geq 0$ and any complex number b, we define recursively

$$H_1(n,k,a,b) := \sum_{\gamma=0}^{n} \binom{n}{\gamma} f(k,k+a)^{\gamma} H_1(n-\gamma, k-1, a+1, b+k\gamma) \quad (4.8)$$

with the initial condition $H_1(n,0,a,b) := b$. As the reader can easily check, the left side of (4.7) is $H_1(p,m,0,0)$. In the appendix to this chapter (Proposition 4.A.5), we prove that, for all $k \geq 0$,

$$H_1(n,k,a,b) = \quad (4.9)$$
$$\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-1} \times \left\{ b\left[1+\sum_{j=1}^{k} f(j,k+a)\right] + n\sum_{j=1}^{k} j f(j,k+a) \right\}$$

where empty sums are treated as zero. (Since $1+\sum_{j=1}^{\ell} f(j,k+a) \neq 0$ for all ℓ, neither (4.8) nor (4.9) gives problems when the first argument of H_1 is zero.) Thus,

$$H_1(p,m,0,0) = \left(1+\sum_{j=1}^{m} f(j,m)\right)^{p-1} \left\{ p\sum_{j=1}^{m} j f(j,m) + 0 \right\}.$$

Finally, we use Lemmata 4.A.1 and 4.A.2 by multiplying their identities by the appropriate constant and by changing the index of summation:

- Statement 1) of Lemma 4.A.1 implies $1+\sum_{j=1}^{m} f(j,m) = \binom{4m}{2m} / \binom{2m}{m}^2$.
- Statement 1) of Lemma 4.A.2 implies $\sum_{j=1}^{m} j f(j,m) = m/2$.

Then $H_1(p,m,0,0) = \frac{pm}{2}\left(\binom{4m}{2m} / \binom{2m}{m}^2\right)^{p-1}$. □

The other moments $Var(\bar{I})$, $E[V]$ and $Var(V)$ can also be rewritten with the help of Proposition 4.3.3:

Lemma 4.3.7

1. $Var(\bar{I}) + (E[\bar{I}])^2 = E\left[\bar{I}^2\right] =$

$$\sum_{\alpha_{2^n-2}=0}^{2^k} \cdots \sum_{\alpha_1=0}^{2^k-\alpha_{2^n-2}-\cdots-\alpha_2} P(\boldsymbol{\alpha}) \left(\frac{\alpha_1 + 2\alpha_2 + \cdots + 2^{n-2}\alpha_{2^n-2}}{2^{n-2}2^k}\right)^2;$$

2. $$E[V] = \sum_{\alpha_{2^{n-2}}=0}^{2^k} \cdots \sum_{\alpha_1=0}^{2^k-\alpha_{2^{n-2}}-\cdots-\alpha_2} P(\boldsymbol{\alpha}) \times$$
$$\left[\frac{\alpha_1 + 4\alpha_2 + \cdots + 2^{2(n-2)}\alpha_{2^{n-2}}}{2^{2(n-2)}2^k} - \left(\frac{\alpha_1 + 2\alpha_2 + \cdots + 2^{n-2}\alpha_{2^{n-2}}}{2^{n-2}2^k}\right)^2\right]$$
$$= \sum_{\alpha_{2^{n-2}}=0}^{2^k} \cdots \sum_{\alpha_1=0}^{2^k-\alpha_{2^{n-2}}-\cdots-\alpha_2} P(\boldsymbol{\alpha}) \frac{\alpha_1 + 4\alpha_2 + \cdots + 2^{2(n-2)}\alpha_{2^{n-2}}}{2^{2(n-2)}2^k}$$
$$- \left[Var(\bar{I}) + (E[\bar{I}])^2\right];$$

3. $$Var(V) + (E[V])^2 = E\left[V^2\right] =$$
$$= \sum_{\alpha_{2^{n-2}}=0}^{2^k} \cdots \sum_{\alpha_1=0}^{2^k-\alpha_{2^{n-2}}-\cdots-\alpha_2} P(\boldsymbol{\alpha}) \times$$
$$\left[\frac{\alpha_1 + 4\alpha_2 + \cdots + 2^{2(n-2)}\alpha_{2^{n-2}}}{2^{2(n-2)}2^k} - \left(\frac{\alpha_1 + 2\alpha_2 + \cdots + 2^{n-2}\alpha_{2^{n-2}}}{2^{n-2}2^k}\right)^2\right]^2$$
$$= \sum_{\alpha_{2^{n-2}}=0}^{2^k} \cdots \sum_{\alpha_1=0}^{2^k-\alpha_{2^{n-2}}-\cdots-\alpha_2} P(\boldsymbol{\alpha}) \left[\left(\frac{\alpha_1 + 4\alpha_2 + \cdots + 2^{2(n-2)}\alpha_{2^{n-2}}}{2^{2(n-2)}2^k}\right)^2\right.$$
$$- 2\frac{\alpha_1 + 4\alpha_2 + \cdots + 2^{2(n-2)}\alpha_{2^{n-2}}}{2^{2(n-2)}2^k} \cdot \left(\frac{\alpha_1 + 2\alpha_2 + \cdots + 2^{n-2}\alpha_{2^{n-2}}}{2^{n-2}2^k}\right)^2$$
$$\left. + \left(\frac{\alpha_1 + 2\alpha_2 + \cdots + 2^{n-2}\alpha_{2^{n-2}}}{2^{n-2}2^k}\right)^4\right].$$

□

There are simple forms for these expressions, too. The way to find them is similar to that used in the proof of Proposition 4.3.6. We need the following functions.

Definition 4.3.8
For any integers $n \geq 0$, $k \geq 1$ and $a \geq 0$ and any complex numbers b and c (b and c subject to the conditions mentioned below), we define recursively

$$H_2(n, k, a, b^2) := \sum_{\gamma=0}^{n} \binom{n}{\gamma} f(k, k+a)^\gamma H_2(n-\gamma, k-1, a+1, (b+k\gamma)^2);$$
$$(-\pi/2 < \arg(b) \leq \pi/2 \quad \text{or} \quad b = 0) \tag{4.10}$$

$$H_3(n,k,a,b) := \sum_{\gamma=0}^{n} \binom{n}{\gamma} f(k,k+a)^\gamma H_3(n-\gamma,k-1,a+1,b+k^2\gamma);$$

$$\text{(no condition)} \tag{4.11}$$

$$H_4(n,k,a,b^2) := \sum_{\gamma=0}^{n} \binom{n}{\gamma} f(k,k+a)^\gamma H_4(n-\gamma,k-1,a+1,(b+k^2\gamma)^2);$$

$$(-\pi/2 < \arg(b) \le \pi/2 \quad \text{or} \quad b=0) \tag{4.12}$$

$$H_5(n,k,a,b^4) := \sum_{\gamma=0}^{n} \binom{n}{\gamma} f(k,k+a)^\gamma H_5(n-\gamma,k-1,a+1,(b+k\gamma)^4);$$

$$(-\pi/4 < \arg(b) \le \pi/4 \quad \text{or} \quad b=0) \tag{4.13}$$

$$H_6(n,k,a,b,c^2) := \sum_{\gamma=0}^{n} \binom{n}{\gamma} f(k,k+a)^\gamma H_6(n-\gamma,k-1,a+1,b+k^2\gamma,(c+k\gamma)^2)$$

$$(-\pi/2 < \arg(c) \le \pi/2 \quad \text{or} \quad c=0) \tag{4.14}$$

(H_6 has one argument more than the other functions) with the initial conditions $H_2(n,0,a,b^2) = H_4(n,0,a,b^2) = b^2$, $H_3(n,0,a,b) = b$, $H_5(n,0,a,b^4) = b^4$ and $H_6(n,0,a,b,c^2) = bc^2$.

We have then (still with $m = 2^{n-2}$ and $p = 2^k$):

$$Var(\bar{I}) + (E[\bar{I}])^2 = \frac{1}{m^2p^2}\left(\frac{\binom{2m}{m}^2}{\binom{4m}{2m}}\right)^p H_2(p,m,0,0)\,; \tag{4.15}$$

$$E[V] = \frac{1}{m^2p^2}\left(\frac{\binom{2m}{m}^2}{\binom{4m}{2m}}\right)^p p\,H_3(p,m,0,0) - \left[Var(\bar{I}) + (E[\bar{I}])^2\right]; \tag{4.16}$$

$$Var(V) + (E[V])^2 = \tag{4.17}$$

$$\frac{1}{m^4p^4}\left(\frac{\binom{2m}{m}^2}{\binom{4m}{2m}}\right)^p \Big[p^2\,H_4(p,m,0,0) - 2p\,H_6(p,m,0,0,0) + H_5(p,m,0,0)\Big].$$

Similarly to the result (4.9), we find the following formulas for $H_2(n,k,a,b^2),\ldots,H_6(n,k,a,b,c^2)$.

Proposition 4.3.9
For any integers $n \geq 0$, $k \geq 0$ and $a \geq 0$ and any complex numbers b and c (b and c subject to the conditions of Definition 4.3.8), we have

$$H_2(n,k,a,b^2) = \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-2}\Bigg\{$$

$$b^2\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^2 + 2\binom{n}{2}\left(\sum_{j=1}^{k} j f(j,k+a)\right)^2$$

$$+\binom{n}{1}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)\left[\sum_{j=1}^{k} j^2 f(j,k+a) + 2b\sum_{j=1}^{k} j f(j,k+a)\right]\Bigg\};$$

$$H_3(n,k,a,b) = \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-1}\Bigg\{$$

$$b\left(1+\sum_{j=1}^{k} f(j,k+a)\right) + n\sum_{j=1}^{k} j^2 f(j,k+a)\Bigg\};$$

$$H_4(n,k,a,b^2) = \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-2}\Bigg\{$$

$$b^2\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^2 + 2\binom{n}{2}\left(\sum_{j=1}^{k} j^2 f(j,k+a)\right)^2$$

$$+\binom{n}{1}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)\left[\sum_{j=1}^{k} j^4 f(j,k+a) + 2b\sum_{j=1}^{k} j^2 f(j,k+a)\right]\Bigg\};$$

$$H_5(n,k,a,b^4) = \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-4}\Bigg\{\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^4\left[b^4\right]$$

$$+ \binom{n}{1}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^3\Bigg[4b^3\sum_{j=1}^{k} j f(j,k+a)$$

$$+ 6b^2\sum_{j=1}^{k} j^2 f(j,k+a) + 4b\sum_{j=1}^{k} j^3 f(j,k+a) + \sum_{j=1}^{k} j^4 f(j,k+a)\Bigg]$$

$$\ldots$$

(continued on next page)

(continued from previous page) ...

$$
\begin{aligned}
&+ \quad 2\binom{n}{2}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^2\left[6b^2\left(\sum_{j=1}^{k} jf(j,k+a)\right)^2\right.\\
&\qquad +12b\sum_{j=1}^{k} jf(j,k+a)\sum_{j=1}^{k} j^2 f(j,k+a)\\
&\qquad \left.+3\left(\sum_{j=1}^{k} j^2 f(j,k+a)\right)^2+4\sum_{j=1}^{k} jf(j,k+a)\sum_{j=1}^{k} j^3 f(j,k+a)\right]\\
&+ \quad 6\binom{n}{3}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)\left[4b\left(\sum_{j=1}^{k} jf(j,k+a)\right)^3\right.\\
&\qquad \left.+6\left(\sum_{j=1}^{k} jf(j,k+a)\right)^2\sum_{j=1}^{k} j^2 f(j,k+a)\right]\\
&+ \quad 24\binom{n}{4}\left(\sum_{j=1}^{k} jf(j,k+a)\right)^4\Bigg\};
\end{aligned}
$$

$$
\begin{aligned}
H_6(n,k,a,b,c^2) &= \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-3}\left\{\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^3\left[bc^2\right]\right.\\
&+ \quad \binom{n}{1}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^2\left[b\sum_{j=1}^{k} j^2 f(j,k+a)+2bc\sum_{j=1}^{k} jf(j,k+a)\right.\\
&\qquad \left.+2c\sum_{j=1}^{k} j^3 f(j,k+a)+c^2\sum_{j=1}^{k} j^2 f(j,k+a)+\sum_{j=1}^{k} j^4 f(j,k+a)\right]\\
&+ \quad 2\binom{n}{2}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)\left[b\left(\sum_{j=1}^{k} jf(j,k+a)\right)^2\right.\\
&\qquad +2c\sum_{j=1}^{k} jf(j,k+a)\sum_{j=1}^{k} j^2 f(j,k+a)\\
&\qquad \left.+\left(\sum_{j=1}^{k} j^2 f(j,k+a)\right)^2+2\sum_{j=1}^{k} jf(j,k+a)\sum_{j=1}^{k} j^3 f(j,k+a)\right]\\
&+ \quad 6\binom{n}{3}\left(\sum_{j=1}^{k} jf(j,k+a)\right)^2\sum_{j=1}^{k} j^2 f(j,k+a)\Bigg\}.
\end{aligned}
$$

Proof:
Most of the appendix to this chapter is dedicated to the proof of these identities. □

Corollary 4.3.10
(Recall that $m = 2^{n-2}$ and $p = 2^k$.)

$$H_2(p,m,0,0) = \left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^p p\frac{m^2}{4m-1} + \left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^{p-2} p(p-1)\frac{m^2}{4};$$

$$H_3(p,m,0,0) = \left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^p p\frac{m^2}{4m-1};$$

$$H_4(p,m,0,0) = \left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^p \left[p\frac{m^3(16m^2-5m-2)}{(4m-1)(4m-3)} + p(p-1)\frac{m^4}{(4m-1)^2}\right];$$

$$\begin{aligned} H_5(p,m,0,0) = {} & \left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^p \left[p\frac{m^3(16m^2-5m-2)}{(4m-1)(4m-3)} + p(p-1)\frac{3m^2}{(4m-1)^2}\right] \\ & + \left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^{p-2} \left[p(p-1)\frac{m^4}{2m-1} + p(p-1)(p-2)\frac{3}{2}\frac{m^4}{4m-1}\right] \\ & + \left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^{p-4} p(p-1)(p-2)(p-3)\frac{m^4}{16}; \end{aligned}$$

$$\begin{aligned} H_6(p,m,0,0,0) = {} & \left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^p \left[p\frac{m^3(16m^2-5m-2)}{(4m-1)(4m-3)} + p(p-1)\frac{m^4}{(4m-1)^2}\right] \\ & + \left(\frac{\binom{4m}{2m}}{\binom{2m}{m}^2}\right)^{p-2} \left[p(p-1)\frac{m^4}{2(2m-1)} + p(p-1)(p-2)\frac{m^4}{4(4m-1)}\right]. \end{aligned}$$

Proof:
From Proposition 4.3.9, we have

$$\begin{aligned} H_2(p,m,0,0) = {} & \left(1+\sum_{j=1}^{m} f(j,m)\right)^{p-2} \left\{2\binom{p}{2}\left(\sum_{j=1}^{m} jf(j,m)\right)^2 \right. \\ & \left. + \binom{p}{1}\left(1+\sum_{j=1}^{m} f(j,m)\right)\left[\sum_{j=1}^{m} j^2 f(j,m)\right]\right\}; \end{aligned}$$

$$H_3(p,m,0,0) = \left(1+\sum_{j=1}^{m} f(j,m)\right)^{p-1}\left\{p\sum_{j=1}^{m} j^2 f(j,m)\right\};$$

$$H_4(p,m,0,0) = \left(1+\sum_{j=1}^{m} f(j,m)\right)^{p-2}\left\{2\binom{p}{2}\left(\sum_{j=1}^{m} j^2 f(j,m)\right)^2 + \binom{p}{1}\left(1+\sum_{j=1}^{m} f(j,m)\right)\left[\sum_{j=1}^{m} j^4 f(j,m)\right]\right\};$$

$$H_5(p,m,0,0) = \left(1+\sum_{j=1}^{m} f(j,m)\right)^{p-4}\Bigg\{ \binom{p}{1}\left(1+\sum_{j=1}^{m} f(j,m)\right)^3\left[\sum_{j=1}^{m} j^4 f(j,m)\right] + 2\binom{p}{2}\left(1+\sum_{j=1}^{m} f(j,m)\right)^2\Bigg[7\left(\sum_{j=1}^{m} j^2 f(j,m)\right)^2 + 4\left(\sum_{j=1}^{m} j f(j,m)\sum_{j=1}^{m} j^3 f(j,m) - \left(\sum_{j=1}^{m} j^2 f(j,m)\right)^2\right)\Bigg] + 6\binom{p}{3}\left(1+\sum_{j=1}^{m} f(j,m)\right)\left[6\left(\sum_{j=1}^{m} j f(j,m)\right)^2\sum_{j=1}^{m} j^2 f(j,m)\right] + 24\binom{p}{4}\left(\sum_{j=1}^{m} j f(j,m)\right)^4\Bigg\};$$

$$H_6(p,m,0,0,0) = \left(1+\sum_{j=1}^{m} f(j,m)\right)^{p-3}\Bigg\{ \binom{p}{1}\left(1+\sum_{j=1}^{m} f(j,m)\right)^2\left[\sum_{j=1}^{m} j^4 f(j,m)\right] + 2\binom{p}{2}\left(1+\sum_{j=1}^{m} f(j,m)\right)\Bigg[3\left(\sum_{j=1}^{m} j^2 f(j,m)\right)^2 + 2\left(\sum_{j=1}^{m} j f(j,m)\sum_{j=1}^{m} j^3 f(j,m) - \left(\sum_{j=1}^{m} j^2 f(j,m)\right)^2\right)\Bigg] + 6\binom{p}{3}\left(\sum_{j=1}^{m} j f(j,m)\right)^2\sum_{j=1}^{m} j^2 f(j,m)\Bigg\}.$$

Now we use Lemmata 4.A.1 and 4.A.2. By multiplying with the appropriate constant and changing the index of summation, we get:

- Statement 1) of Lemma 4.A.1 implies $1+\sum_{j=1}^{m} f(j,m) = \frac{\binom{4m}{2m}}{\binom{2m}{m}^2}$;
- the statements of Lemma 4.A.2 imply

(a) $\sum_{j=1}^{m} jf(j,m) = \frac{m}{2}$;

(b) $\sum_{j=1}^{m} j^2 f(j,m) = \frac{m^2}{4m-1} \frac{\binom{4m}{2m}}{\binom{2m}{m}^2}$;

(c) $\sum_{j=1}^{m} j^3 f(j,m) = \frac{m^3}{2(2m-1)}$;

(d) $\sum_{j=1}^{m} j^4 f(j,m) =$

$$\frac{1}{\binom{2m}{m}^2}\left(-3m^4\binom{4m}{2m} + 8m^4\binom{4m-2}{2m-1} + 4m^2(2m-1)^2\binom{4m-4}{2m-2}\right)$$

$$= \frac{m^3(16m^2-5m-2)}{(4m-1)(4m-3)} \frac{\binom{4m}{2m}}{\binom{2m}{m}^2}.$$

The proof is completed by substituting these identities into the expressions for $H_2(p,m,0,0),\ldots,H_6(p,m,0,0,0)$ and making some factorisations. □

Finally, we use Corollary 4.3.10 and equations (4.15), (4.16) and (4.17) to obtain closed expressions for $Var(\bar{I})$, $E[V]$ and $Var(V)$. The results are summarized in the theorem below, which we state without proof. (The proof uses only elementary and standard algebraic transformations.)

Theorem 4.3.11
Let $f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{Z}_2$ be two balanced functions. Then the expectation and the variance (over all round functions g of ciphers with block length n and round key length k) of $\bar{I}(S^1)$ and $V(S^1)$, where $\bar{I}(S^1)$ is the average-key imbalance of S^1 and $V(S^1)$ is defined by (4.5), where $S^1 = f_1(X) \oplus f_2(g(X,Z))$, and where X is a random variable uniformly distributed on $\mathbb{Z}_2^n$, are

$$E[\bar{I}] = \frac{\binom{2m}{m}^2}{2\binom{4m}{2m}},$$

$$Var(\bar{I}) = \frac{1}{p(4m-1)} - \frac{1}{4p}\frac{\binom{2m}{m}^4}{\binom{4m}{2m}^2} = \frac{1}{p(4m-1)} - \frac{1}{p}(E[\bar{I}])^2,$$

$$E[V] = \frac{p-1}{p}\left(\frac{1}{4m-1} - \frac{\binom{2m}{m}^4}{4\binom{4m}{2m}^2}\right) = (p-1)\,Var(\bar{I}) \text{ and}$$

$$Var(V) = \frac{(p-1)(3-p)}{p^3(4m-1)^2} + \frac{(p-1)^2(3m-2)}{p^3 m(4m-1)(4m-3)}$$

$$- \frac{\binom{2m}{m}^4}{\binom{4m}{2m}^2}\left[\frac{(p-1)^2}{p^3(2m-1)} + \frac{(p-1)(3-2p)}{p^3(4m-1)}\right] + \frac{\binom{2m}{m}^8}{\binom{4m}{2m}^4}\frac{(p-1)(3-2p)}{8p^3},$$

where $m = 2^{n-2}$ and $p = 2^k$. □

The formulas for the four moments are finally quite simple and we wonder whether they could be found in a less fastidious way. Table 4.6 shows the value of the four moments for some block lengths n with key length $k = n$.

n, k	$E[\bar{I}]$	$Var(\bar{I})$	$E[V]$	$Var(V)$
2	$1/3$	$1/18 = 5.\bar{5} * 10^{-2}$	$1/6$	$8.10 * 10^{-3}$
3	$9/35 = 0.257$	$9.59 * 10^{-3}$	$6.71 * 10^{-2}$	$7.98 * 10^{-4}$
4	$1225/6435 = 0.190$	$1.90 * 10^{-3}$	$2.85 * 10^{-2}$	$1.02 * 10^{-4}$
5	$1.38 * 10^{-1}$	$4.15 * 10^{-4}$	$1.29 * 10^{-2}$	$1.25 * 10^{-5}$
6	$9.86 * 10^{-2}$	$9.62 * 10^{-5}$	$6.06 * 10^{-3}$	$1.51 * 10^{-6}$
7	$7.01 * 10^{-2}$	$2.31 * 10^{-5}$	$2.94 * 10^{-3}$	$1.85 * 10^{-7}$
8	$4.97 * 10^{-2}$	$5.66 * 10^{-6}$	$1.44 * 10^{-3}$	$2.29 * 10^{-8}$
10	$2.49 * 10^{-2}$	$3.48 * 10^{-7}$	$3.56 * 10^{-4}$	$3.54 * 10^{-10}$
12	$1.25 * 10^{-2}$	$2.17 * 10^{-8}$	$8.88 * 10^{-5}$	$5.52 * 10^{-12}$
14	$6.23 * 10^{-3}$	$1.35 * 10^{-9}$	$2.22 * 10^{-5}$	$8.62 * 10^{-14}$
16	$3.12 * 10^{-3}$	$8.46 * 10^{-11}$	$5.55 * 10^{-6}$	$1.35 * 10^{-15}$

Table 4.6: Value of the four moments for different block sizes.

We shall soon give a meaning to the results. But before doing so, we want to study the asymptotic behaviour of the four moments as n and k go to infinity.

Theorem 4.3.12
The asymptotic behaviours of $E[\bar{I}]$, $Var(\bar{I})$, $E[V]$ *and* $Var(V)$ *as* $n \to \infty$ *and* $k \to \infty$ *are the following (where* $a(n,k) \overset{n,k}{\sim} b(n,k)$ *means that* $\lim_{n\to\infty} \lim_{k\to\infty} a(n,k)/b(n,k) = 1$, *and so on).*

$$E[\bar{I}] \overset{n}{\sim} \sqrt{\frac{2}{\pi}}\, 2^{-n/2}, \quad \textit{independently of } k;$$

$$Var(\bar{I}) \overset{n}{\sim} \frac{\pi - 2}{\pi}\, 2^{-(n+k)} \quad \textit{for all } k;$$

$$E[V] \overset{n,k}{\sim} \frac{\pi - 2}{\pi}\, 2^{-n};$$

$$Var(V) \overset{n,k}{\sim} 2\,\frac{\pi^2 - 8}{\pi^2}\, 2^{-(2n+k)}.$$

Proof:

By Stirling's approximation of the factorial [7], $\binom{2a}{a} \overset{a}{\sim} 2^{2a}/\sqrt{\pi a}$. Then $\binom{2m}{m}^2 / \binom{4m}{2m} \overset{m}{\sim} \sqrt{2/\pi m}$. Hence, since $p = 2^k$ and $m = 2^{n-2}$ go to infinity

if and only if k and n do, we have, by Theorem 4.3.11,

$$\begin{aligned} E[\overline{I}] &\overset{m}{\sim} \frac{1}{2}\sqrt{\frac{2}{\pi m}} = \sqrt{\frac{2}{\pi}}\,2^{-n/2}, \quad \text{independently of } k; \\ Var(\overline{I}) &\overset{m}{\sim} \frac{1}{p}\left(\frac{1}{4m-1} - \frac{1}{2\pi m}\right) \\ &= \frac{2\pi m - 4m + 1}{p \cdot 2\pi m(4m-1)} \overset{m}{\sim} \frac{\pi-2}{4\pi}\frac{1}{pm} = \frac{\pi-2}{\pi}2^{-(n+k)}; \\ E[V] &= (p-1)Var(\overline{I}) \overset{m}{\sim} \frac{\pi-2}{4\pi}\frac{p-1}{pm} \overset{p}{\sim} \frac{\pi-2}{4\pi}\frac{1}{m} = \frac{\pi-2}{\pi}2^{-n}. \end{aligned}$$

Finally, because $\frac{(p-1)^2}{p^3(2m-1)} + \frac{(p-1)(3-2p)}{p^3(4m-1)} = \frac{(p-1)(p+2m-2)}{p^3(2m-1)(4m-1)} \overset{p,m}{\sim} \frac{p+2m-2}{8p^2m^2}$, we have

$$\begin{aligned} Var(V) &\overset{p,m}{\sim} -\frac{1}{p(4m)^2} + \frac{3}{p(4m)^2} - \frac{2}{\pi m}\frac{p+2m-2}{8p^2m^2} - \frac{4}{\pi^2m^2}\frac{1}{4p} \\ &= \frac{(\pi^2-8)pm - 2\pi p - 4\pi m}{8\pi^2p^2m^3} \overset{p,m}{\sim} \frac{\pi^2-8}{8\pi^2}\frac{1}{pm^2} = 2\frac{\pi^2-8}{\pi^2}2^{-(2n+k)}. \square \end{aligned}$$

4.3.4 Interpretation of The Results

Suppose that the block length n and the key length k are large enough. Then, whatever the choice of the balanced functions f_1, f_2, on the average $\overline{I}(S^1) \approx \sqrt{\frac{2}{\pi}}\,2^{-n/2}$ and $V(S^1) \approx \frac{\pi-2}{\pi}\,2^{-n}$. Moreover, the relative variances $Var(\overline{I}/E[\overline{I}])$ and $Var(V/E[V])$ are of the order of 2^{-k}. Hence, we can say that $\overline{I}(S^1)$ (resp. $V(S^1)$) is very close to $\sqrt{\frac{2}{\pi}}\,2^{-n/2}$ (resp. $\frac{\pi-2}{\pi}\,2^{-n}$) for almost all I/O sums involving f_1 and f_2. But this fact holds for any f_1, f_2. Thus, $\overline{I}(S^1) \approx \sqrt{\frac{2}{\pi}}\,2^{-n/2}$ and $V(S^1) \approx \frac{\pi-2}{\pi}\,2^{-n}$ for almost all I/O sums, and the longer the round keys, the smaller the proportion of I/O sums for which $\overline{I}(S^1)$ and $V(S^1)$ are far from these values.

Remember that $\overline{I}(S^1)$ is the average and $V(S^1)$ is the variance of the key-dependent imbalances. Hence, for almost all I/O sums, the relative variance of the key-dependent imbalances is $V(S^1)/\overline{I}(S^1)^2 \approx \frac{\pi-2}{2} \approx 0.57$. It is now for everyone to decide for himself whether he considers the key-dependent imbalances to be "approximately equal".

4.4 Validity of The Hypothesis of Fixed-Key Equivalence for More Than One Round

In the case of more than one round, we were not able to compute the average and the variance of $\overline{I}$ and of V as we did for one round. The main reason is the following: in the 1-round case, we made intensive use of the fact that defining a round function is equivalent to defining as many invertible functions as there are possible first-round keys. Then we based our proofs on the number of functions of the form $f_1 \oplus f_2 \circ g$ that have a certain imbalance, where f_1 and f_2 are balanced and fixed and g runs over all invertible functions on the binary n-tuples. With that result, we calculated the number of round functions for which the 1-round I/O sum has a certain distribution of the key-dependent imbalances. Counting was feasible because, for any first-round key z_1, the distribution of the functions $g_{z_1} = g(\cdot, z_1)$, as g runs over all possible round functions, is uniform on the set of invertible functions on the binary n-tuples.

In the multiple round case, i-round I/O sums are constructed by means of functions of the form $f_1 \oplus f_2 \circ g_{z_i} \circ \cdots \circ g_{z_1}$, where $g_{z_1}, \ldots, g_{z_i}$ are invertible and thus $g_{z_i} \circ \cdots \circ g_{z_1}$ is also invertible, and f_1 and f_2 are again balanced. However, the distribution of the functions $g_{z_i} \circ \cdots \circ g_{z_1}$ on the set of invertible functions on the binary n-tuples, as g runs over all possible round functions, is no longer uniform for all values $z_1, \ldots, z_i$ of the round keys. Because of this, we were not able to count the number of round functions for which the i-round I/O sum has a certain distribution of the key-dependent imbalances.

We tried to see whether it was possible somehow to calculate the expected value and the variance of $\overline{I}$ and V recursively in the number of rounds. The only moment with which we succeeded was the average of $\overline{I}(f_1(X) \oplus f_2(Y(i))$ over all round functions for fixed balanced functions f_1 and f_2. According to (2.5), we can write the key-dependent imbalance of $S^{1\ldots i}$ as $I(S^{1\ldots i}|z_1, \ldots, z_i) = I(f_1 \oplus f_2 \circ g_{z_1,\ldots,z_i})$. Let ω be the number of possible round functions and let $E_i[\overline{I}]$ be the average of the average-key imbalance of $S^{1\ldots i}$ over all round functions g, with fixed balanced functions f_1 and f_2. Then

$$E_1[\overline{I}] = \frac{1}{\omega}\sum_g \overline{I}(S^1) = \frac{1}{\omega}\sum_g \frac{1}{2^k}\sum_{z_1} I(f_1 \oplus f_2 \circ g_{z_1})$$

$$= \frac{1}{2^k}\sum_{z_1} \frac{1}{\omega}\sum_g I(f_1 \oplus f_2 \circ g_{z_1}) \qquad \text{and}$$

$$\begin{aligned} E_i[\bar{I}] &= \frac{1}{\omega}\sum_g \bar{I}(S^{1\ldots i}) = \frac{1}{\omega}\sum_g \left(\frac{1}{2^k}\right)^i \sum_{z_1,\ldots,z_i} I(f_1 \oplus f_2 \circ g_{z_1,\ldots,z_i}) \\ &= \left(\frac{1}{2^k}\right)^{i-1} \sum_{z_2,\ldots,z_i} \left[\frac{1}{2^k}\sum_{z_1}\frac{1}{\omega}\sum_g I(f_1 \oplus (f_2 \circ g_{z_2,\ldots,z_i}) \circ g_{z_1})\right] \\ &= \left(\frac{1}{2^k}\right)^{i-1} \sum_{z_2,\ldots,z_i} E_1[\bar{I}] = E_1[\bar{I}], \end{aligned}$$

where in the penultimate equality we used the fact that $f_2 \circ g_{z_2,\ldots,z_i}$ is a balanced function that does not depend on z_1. This somewhat surprising equality is summarized in Theorem 4.4.1. But before stating this theorem, we show that one cannot do the same for the other moments. Consider, for instance, the average of $\bar{I}^2(S^{1\ldots i})$ over all ciphers. For $i = 1$ this becomes

$$\begin{aligned} E_1[\bar{I}^2] &= \frac{1}{\omega}\sum_g \bar{I}^2(S^1) = \frac{1}{\omega}\sum_g \left(\frac{1}{2^k}\sum_{z_1} I(f_1 \oplus f_2 \circ g_{z_1})\right)^2 \\ &= \frac{1}{2^{2k}}\sum_{z_1,z_1'}\frac{1}{\omega}\sum_g I(f_1 \oplus f_2 \circ g_{z_1}) \cdot I(f_1 \oplus f_2 \circ g_{z_1'}). \end{aligned}$$

If we try to use the same factorisation as before, we get, for instance for $i = 2$,

$$\begin{aligned} E_2[\bar{I}] &= \frac{1}{\omega}\sum_g \bar{I}^2(S^{12}) = \frac{1}{\omega}\sum_g \left(\left(\frac{1}{2^k}\right)^2 \sum_{z_1,z_2} I(f_1 \oplus f_2 \circ g_{z_1,z_2})\right)^2 \\ &= \frac{1}{2^{2k}}\sum_{z_2,z_2'}\left[\frac{1}{2^{2k}}\sum_{z_1,z_1'}\frac{1}{\omega}\sum_g I(f_1 \oplus (f_2 \circ g_{z_2}) \circ g_{z_1}) \cdot I(f_1 \oplus (f_2 \circ g_{z_2'}) \circ g_{z_1'})\right]. \end{aligned}$$

But now the expression in square brackets is *not* equal to $E_1[\bar{I}^2]$, because in general $f_2 \circ g_{z_2}$ is different from $f_2 \circ g_{z_2'}$. Even if we try to average further over all balanced functions f_2 and/or all balanced functions f_1, we stumble over this obstacle. Nevertheless, as said above, we still have:

Theorem 4.4.1
Let $f_1, f_2 : \mathbb{Z}_2^n \to \mathbb{Z}_2$ be two balanced functions. Then the expectation (over all round functions g of ciphers with block length n and round key length k) of the average-key imbalance for i rounds $\overline{I}(S^{1\ldots i})$, where $S^{1\ldots i} = f_1(X) \oplus f_2(Y(i))$ is an i-round I/O sum, where $Y(i)$ is the output of the i^{th} round of the cipher defined by the round function g and where X is a random variable uniformly distributed on $\mathbb{Z}_2^n$, is independent of i and of the length of the round keys and is equal to

$$E[\overline{I}] = \frac{\binom{2^{n-1}}{2^{n-2}}^2}{2\binom{2^n}{2^{n-1}}}. \tag{4.18}$$

□

This gives one a point of reference when using linear cryptanalysis to attack a cipher. Note that $\binom{2^{n-1}}{2^{n-2}}^2 / 2\binom{2^n}{2^{n-1}}$ is also the joint average over all ciphers and all pairs of balanced functions f_1, f_2. On the other hand, for a fixed cipher, the average of $\overline{I}(f_1(X) \oplus f_2(Y(i))$ over all pairs of balanced functions f_1, f_2 may be different from $\binom{2^{n-1}}{2^{n-2}}^2 / 2\binom{2^n}{2^{n-1}}$.

Suppose now that one has an $(r-1)$-round I/O sum with corresponding balanced functions f_0 and f_{r-1}. If its average-key imbalance is substantially larger than $\binom{2^{n-1}}{2^{n-2}}^2 / 2\binom{2^n}{2^{n-1}}$, then the cipher in question "fits well" to the I/O sum. In general, this does not imply the converse statement that the I/O sum fits well to the cipher, because the average over all f_0 and f_{r-1} of the average-key imbalance may be larger than $\binom{2^{n-1}}{2^{n-2}}^2 / 2\binom{2^n}{2^{n-1}}$ in which case the I/O sum found might have no exceptional property. However, one would expect that, for a cipher that is more resistent than the average cipher against linear cryptanalysis, the average over all pairs of balanced functions f_0, f_{r-1} will be smaller than $\binom{2^{n-1}}{2^{n-2}}^2 / 2\binom{2^n}{2^{n-1}}$ so that the $(r-1)$-round I/O sum found with average-key imbalance substantially larger than $\binom{2^{n-1}}{2^{n-2}}^2 / 2\binom{2^n}{2^{n-1}}$ would be "very good" to use in a linear cryptanalysis attack.

These considerations and Theorem 4.3.12 lead us to formulate the following more precise definition of an effective I/O sum:

Definition 4.4.2
An I/O sum is *effective* if its average key imbalance is substantially larger than $\sqrt{\frac{2}{\pi}}\, 2^{-n/2}$.

Example 4.4.3
In his attack on DES [39], Matsui found a sum of linked threefold sums with imbalance $1.19*2^{-20}$, which lower-bounds the average-key imbalance of a 14-round I/O sum. This is substantially larger than $\sqrt{\frac{2}{\pi}}\,2^{-32}$. To calculate the imbalance, he used the Piling-up Lemma without caring whether the threefold sums were independent. Thus, the true imbalance might be different. However, as explained in Subsection 3.5.4, the true imbalance is probably close to $1.19 * 2^{-20}$. According to our definition above, his 14-round I/O sum is effective.

4.A Proofs

4.A.1 Two Lemmata

Lemma 4.A.1
For any integer $m \geq 0$,

1. $\sum_{k=0}^{m} \binom{m}{k}^2 = \binom{2m}{m}$;
2. $\sum_{k=0}^{m} k\binom{m}{k}^2 = \frac{m}{2}\binom{2m}{m}$;
3. $\sum_{k=0}^{m} k^2\binom{m}{k}^2 = m^2\binom{2(m-1)}{m-1}$;
4. $\sum_{k=0}^{m} k^3\binom{m}{k}^2 = \frac{m^2}{2}(m+1)\binom{2(m-1)}{m-1}$;
5. $\sum_{k=0}^{m} k^4\binom{m}{k}^2 = m^3\binom{2(m-1)}{m-1} + m^2(m-1)^2\binom{2(m-2)}{m-2}$.

Proof:
The first two identities can be found in [45, p. 622].

3. We use $\binom{m}{k} = \frac{m}{k}\binom{m-1}{k-1}$; to be general, we define $g(j,m) := \sum_{k=0}^{m} k^j\binom{m}{k}^2$. Then

$$\sum_{k=0}^{m} k^j\binom{m}{k}^2 = \sum_{k=1}^{m} k^j\frac{m^2}{k^2}\binom{m-1}{k-1}^2 = m^2\sum_{k=1}^{m} k^{j-2}\binom{m-1}{k-1}^2$$
$$= m^2\sum_{k=0}^{m-1}(k+1)^{j-2}\binom{m-1}{k}^2 = m^2\sum_{k=0}^{m-1}\sum_{i=0}^{j-2}\binom{j-2}{i}k^i\binom{m-1}{k}^2$$
$$= m^2\sum_{i=0}^{j-2}\binom{j-2}{i}\sum_{k=0}^{m-1}k^i\binom{m-1}{k}^2$$

so that $g(j,m) = m^2\sum_{i=0}^{j-2}\binom{j-2}{i}g(i,m-1)$. Applied to $j=2$, this gives $g(2,m) = m^2 g(0,m-1)$.

4. $g(3,m) = m^2(g(0,m-1)+g(1,m-1)) = \frac{m^2}{2}(m+1)\binom{2(m-1)}{m-1}$.

5. $g(4,m) = m^2(g(0,m-1)+2g(1,m-1)+g(2,m-1))$
$= m^3\binom{2(m-1)}{m-1} + m^2(m-1)^2\binom{2(m-2)}{m-2}$. □

Lemma 4.A.2
For any integer $m \geq 1$,

1. $\sum_{k=0}^{m-1}\binom{2m}{k}^2(m-k) = \frac{m}{4}\binom{2m}{m}^2$.

2. $\sum_{k=0}^{m-1}\binom{2m}{k}^2(m-k)^2 = \frac{m^2}{2(4m-1)}\binom{4m}{2m}$.

3. $\sum_{k=0}^{m-1}\binom{2m}{k}^2(m-k)^3 = \frac{m^3}{4(2m-1)}\binom{2m}{m}^2$.

4. $\sum_{k=0}^{m-1}\binom{2m}{k}^2(m-k)^4 =$
$\frac{1}{2}\left(-3m^4\binom{4m}{2m}+8m^4\binom{4m-2}{2m-1}+4m^2(2m-1)^2\binom{4m-4}{2m-2}\right)$.

Proof:
We begin with the proofs of 2. and 4. We use Lemma 4.A.1.

$$\begin{aligned}
\sum_{k=0}^{m-1}\binom{2m}{k}^2(m-k)^2 &= \frac{1}{2}\sum_{k=0}^{2m}\binom{2m}{k}^2(m-k)^2 \\
&= \frac{1}{2}\left[m^2\sum_{k=0}^{2m}\binom{2m}{k}^2 - 2m\sum_{k=0}^{2m}k\binom{2m}{k}^2 + \sum_{k=0}^{2m}k^2\binom{2m}{k}^2\right] \\
&= \frac{1}{2}\left[m^2\binom{4m}{2m} - 2m^2\binom{4m}{2m} + 4m^2\binom{4m-2}{2m-1}\right] \\
&= \frac{1}{2}\left[4m^2\binom{4m-2}{2m-1} - m^2\binom{4m}{2m}\right] \\
&= \frac{1}{2}\left[4m^2\binom{4m}{2m}\frac{4m^2}{4m(4m-1)} - m^2\binom{4m}{2m}\right] \\
&= \frac{1}{2}\binom{4m}{2m}\left[\frac{4m^3}{4m-1} - m^2\right] = \frac{m^2}{2(4m-1)}\binom{4m}{2m}.
\end{aligned}$$

Statement 4. is proved in the same way by again using Lemma 4.A.1:

$$\sum_{k=0}^{m-1}\binom{2m}{k}^2(m-k)^4 = \frac{1}{2}\sum_{k=0}^{2m}\binom{2m}{k}^2(m-k)^4$$

$$
\begin{aligned}
&= \frac{1}{2}\left[\sum_{k=0}^{2m}\binom{2m}{k}^2(m^4-4m^3k+6m^2k^2-4mk^3+k^4)\right] \\
&= \frac{1}{2}\left[m^4\binom{4m}{2m}-4m^3m\binom{4m}{2m}+6m^2 4m^2\binom{4m-2}{2m-1}\right. \\
&\qquad -4m\frac{4m^2}{2}(2m+1)\binom{4m-2}{2m-1}+8m^3\binom{4m-2}{2m-1} \\
&\qquad \left.+4m^2(2m-1)^2\binom{4m-4}{2m-2}\right] \\
&= \frac{1}{2}\left[-3m^4\binom{4m}{2m}+8m^4\binom{4m-2}{2m-1}+4m^2(2m-1)^2\binom{4m-4}{2m-2}\right].
\end{aligned}
$$

Statement 1. cannot be proved directly (i.e., by simple sum manipulations). By using Gosper's method [14], one finds T_k such that $T_{k+1}-T_k = \binom{2m}{k}^2(m-k)$, namely $T_k = \frac{k^2}{4m}\binom{2m}{k}^2$. Then $\sum_{k=0}^{m-1}\binom{2m}{k}^2(m-k) = T_m - T_0 = \frac{m}{4}\binom{2m}{m}^2$.

Using the same method, one finds S_k with $S_{k+1}-S_k = \binom{2m}{k}^2(m-k)^3$ to be

$$S_k = \frac{2k^4-2k^3(2m+1)+k^2m(2m+3)}{4(2m-1)}\binom{2m}{k}^2.$$

Then $\sum_{k=0}^{m-1}\binom{2m}{k}^2(m-k)^3 = S_m - S_0 = \frac{m^3}{4(2m-1)}\binom{2m}{m}^2$. □

In the remainder of the appendix, we prove Equation (4.9) and Proposition 4.3.9.

4.A.2 Preliminary Identities

Lemma 4.A.3
For any non-negative integers n, i, and j and any real number x,

$$\sum_{w=0}^{n}\binom{n}{w}\binom{n-w}{j}\binom{w}{i}x^w = \binom{i+j}{j}\binom{n}{i+j}x^i(1+x)^{n-i-j} \qquad (4.19)$$

with the convention $0^0 = 1$.
Proof:

If $x = 0$, then the equation becomes $\binom{n}{0}\binom{n}{j}\binom{0}{i} = \binom{i+j}{j}\binom{n}{i+j}0^i$; this holds because $\binom{0}{i}$ is 0 if $i \neq 0$ and 1 if $i = 0$. Suppose now $x \neq 0$. The summand on the left side is zero if $w < i$ or if $w > n-j$ so we can let the

sum run only from i to $n-j$. Furthermore, multiplying out the binomial coefficients transforms the equation into

$$\sum_{w=i}^{n-j} \frac{n!}{i!j!(w-i)!(n-w-j)!}\, x^w = \frac{n!}{i!j!(n-i-j)!}\, x^i\,(1+x)^{n-i-j}.$$

Multiplying both sides by $i!j!x^{-i}/n!$ yields

$$\sum_{w=i}^{n-j} \frac{(n-i-j)!}{(w-i)!(n-w-j)!} x^{w-i} = (1+x)^{n-i-j}.$$

But the left side of this equation is equal to $\sum_{w=0}^{n-i-j} \binom{n-i-j}{w} x^w$ and hence the equation holds by the binomial formula [7]. □

Lemma 4.A.4
For any integers $n \geq 0$ and k and any complex numbers b, c and x, the following identities hold:

1a) $\sum_{w=0}^{n} \binom{n}{w} x^w = (1+x)^n$;

1b) $\sum_{w=0}^{n} \binom{n}{w} x^w \binom{n-w}{1} = \binom{n}{1}(1+x)^{n-1}$;

1c) $\sum_{w=0}^{n} \binom{n}{w} x^w \binom{n-w}{2} = \binom{n}{2}(1+x)^{n-2}$;

1d) $\sum_{w=0}^{n} \binom{n}{w} x^w \binom{n-w}{3} = \binom{n}{3}(1+x)^{n-3}$;

1e) $\sum_{w=0}^{n} \binom{n}{w} x^w \binom{n-w}{4} = \binom{n}{4}(1+x)^{n-4}$;

2a) $\sum_{w=0}^{n} \binom{n}{w} x^w (b+kw) = (1+x)^n b + \binom{n}{1} x(1+x)^{n-1} k$;

2b) $\sum_{w=0}^{n} \binom{n}{w} x^w (b+kw) \binom{n-w}{1} = \binom{n}{1}(1+x)^{n-1} b + 2\binom{n}{2} x(1+x)^{n-2} k$;

2c) $\sum_{w=0}^{n} \binom{n}{w} x^w (b+kw) \binom{n-w}{2} = \binom{n}{2}(1+x)^{n-2} b + 3\binom{n}{3} x(1+x)^{n-3} k$;

2d) $\sum_{w=0}^{n} \binom{n}{w} x^w (b+kw) \binom{n-w}{3} = \binom{n}{3}(1+x)^{n-3} b + 4\binom{n}{4} x(1+x)^{n-4} k$;

3a) $\sum_{w=0}^{n} \binom{n}{w} x^w (b+kw)^2 = (1+x)^n b^2 + \binom{n}{1} x(1+x)^{n-1}(2bk+k^2)$
$+2\binom{n}{2} x^2 (1+x)^{n-2} k^2$;

3b) $\sum_{w=0}^{n} \binom{n}{w} x^w (b+kw)^2 \binom{n-w}{1} = \binom{n}{1}(1+x)^{n-1} b^2$
$+2\binom{n}{2} x(1+x)^{n-2}(2bk+k^2) + 6\binom{n}{3} x^2 (1+x)^{n-3} k^2$;

3c) $\sum_{w=0}^{n} \binom{n}{w} x^w (b+kw)^2 \binom{n-w}{2} = \binom{n}{2}(1+x)^{n-2} b^2$
$+3\binom{n}{3} x(1+x)^{n-3}(2bk+k^2) + 12\binom{n}{4} x^2 (1+x)^{n-4} k^2$;

4a) $\sum_{w=0}^{n}\binom{n}{w}x^w(b+kw)^3 = (1+x)^n b^3 + \binom{n}{1}x(1+x)^{n-1}(3b^2k+3bk^2+k^3)$
$+2\binom{n}{2}x^2(1+x)^{n-2}(3bk^2+3k^3)+6\binom{n}{3}x^3(1+x)^{n-3}k^3;$

4b) $\sum_{w=0}^{n}\binom{n}{w}x^w(b+kw)^3\binom{n-w}{1} = \binom{n}{1}(1+x)^{n-1}b^3$
$+2\binom{n}{2}x(1+x)^{n-2}(3b^2k+3bk^2+k^3)$
$+6\binom{n}{3}x^2(1+x)^{n-3}(3bk^2+3k^3)+24\binom{n}{4}x^3(1+x)^{n-4}k^3;$

5a) $\sum_{w=0}^{n}\binom{n}{w}x^w(b+kw)^4 = (1+x)^n b^4$
$+\binom{n}{1}x(1+x)^{n-1}(4b^3k+6b^2k^2+4bk^3+k^4)$
$+2\binom{n}{2}x^2(1+x)^{n-2}(6b^2k^2+12bk^3+7k^4)$
$+6\binom{n}{3}x^3(1+x)^{n-3}(4bk^3+6k^4)+24\binom{n}{4}x^4(1+x)^{n-4}k^4;$

6a) $\sum_{w=0}^{n}\binom{n}{w}x^w(b+k^2w)(c+kw) = (1+x)^n bc$
$+\binom{n}{1}x(1+x)^{n-1}(ck^2+bk+k^3)+2\binom{n}{2}x^2(1+x)^{n-2}k^3;$

6b) $\sum_{w=0}^{n}\binom{n}{w}x^w(b+k^2w)(c+kw)\binom{n-w}{1} = \binom{n}{1}(1+x)^{n-1}bc$
$+2\binom{n}{2}x(1+x)^{n-2}(ck^2+bk+k^3)+6\binom{n}{3}x^2(1+x)^{n-3}k^3;$

6c) $\sum_{w=0}^{n}\binom{n}{w}x^w(b+k^2w)(c+kw)^2 = (1+x)^n bc^2$
$+\binom{n}{1}x(1+x)^{n-1}(c^2k^2+2bck+2ck^3+bk^2+k^4)$
$+2\binom{n}{2}x^2(1+x)^{n-2}(2ck^3+bk^2+3k^4)+6\binom{n}{3}x^3(1+x)^{n-3}k^4.$

Proof:

Some of the identities are trivial. The proofs of the others make use of Lemma 4.A.3 and of the following identities, valid for any real number w and any integer n:

- $w^2 = 2\binom{w}{2}+\binom{w}{1}$;
- $w^3 = 6\binom{w}{3}+6\binom{w}{2}+\binom{w}{1}$;
- $w^4 = 24\binom{w}{4}+36\binom{w}{3}+14\binom{w}{2}+\binom{w}{1}$.
- $\binom{n}{1}\binom{n-1}{1} = 2\binom{n}{2}$; $\binom{n}{1}\binom{n-1}{2} = 3\binom{n}{3}$; $\binom{n}{1}\binom{n-1}{3} = 4\binom{n}{4}$;
- $\binom{n}{2}\binom{n-2}{1} = 3\binom{n}{3}$; $\binom{n}{2}\binom{n-2}{2} = 6\binom{n}{4}$;
- $\binom{n}{3}\binom{n-3}{1} = 4\binom{n}{4}$. □

We will use identities 2a), 2b), 2c) and 3a) also with k^2 instead of k.

4.A.3 Proof of (4.9) and of Proposition 4.3.9

All proofs make repeated use of Lemma 4.A.4.

Proposition 4.A.5
For any integers $n \geq 0$, $k \geq 1$ and $a \geq 0$ and any complex number b, let $H_1(n,k,a,b)$ be defined by the recursion

$$H_1(n,k,a,b) := \sum_{\gamma=0}^{n} \binom{n}{\gamma} f(k,k+a)^{\gamma} H_1(n-\gamma, k-1, a+1, b+k\gamma) \tag{4.20}$$

and the initial condition $H_1(n,0,a,b) := b$. Then, for any integer $k \geq 0$,

$$H_1(n,k,a,b) = \tag{4.21}$$
$$\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-1} \left\{ \binom{n}{1} \sum_{j=1}^{k} j f(j,k+a) + \left(1+\sum_{j=1}^{k} f(j,k+a)\right) b \right\}.$$

Proof:

It is easy to see that (4.21) implies $H_1(n,0,a,b) = b$. (Empty sums are treated as zero.) We prove the general case by induction. Let $k \geq 1$ and let the proposition hold up to $k-1$. Then (ignoring the stars $*$; they will be used in the proof of the identity satisfied by the function H_3)

$$\begin{aligned}
H_1(n,k,a,b) &= \sum_{w=0}^{n} \binom{n}{w} f(k,k+a)^{w} H_1(n-w, k-1, a+1, b+k^*w) \\
&= \sum_{w=0}^{n} \binom{n}{w} f(k,k+a)^{w} \left(1+\sum_{j=1}^{k-1} f(j,k+a)\right)^{n-w-1} \times \\
&\quad \left\{ \binom{n-w}{1} \sum_{j=1}^{k-1} j^* f(j,k+a) + \left(1+\sum_{j=1}^{k-1} f(j,k+a)\right)(b+k^*w) \right\} \\
&= \left(1+\sum_{j=1}^{k-1} f(j,k+a)\right)^{n-1} \sum_{w=0}^{n} \binom{n}{w} \left(\frac{f(k,k+a)}{1+\sum_{j=1}^{k-1} f(j,k+a)}\right)^{w} \times \\
&\quad \left\{ \binom{n-w}{1} \sum_{j=1}^{k-1} j^* f(j,k+a) + \left(1+\sum_{j=1}^{k-1} f(j,k+a)\right)(b+k^*w) \right\}
\end{aligned}$$

$$
\begin{aligned}
= \quad & \left(1+\sum_{j=1}^{k-1} f(j,k+a)\right)^{n-1} \Bigg\{ \\
& \sum_{j=1}^{k-1} j^* f(j,k+a) \binom{n}{1} \left(1+\frac{f(k,k+a)}{1+\sum_{j=1}^{k-1} f(j,k+a)}\right)^{n-1} \\
& \quad + \left(1+\sum_{j=1}^{k-1} f(j,k+a)\right) \Bigg[\left(1+\frac{f(k,k+a)}{1+\sum_{j=1}^{k-1} f(j,k+a)}\right)^{n} b \\
& \qquad + \binom{n}{1} \frac{f(k,k+a)}{1+\sum_{j=1}^{k-1} f(j,k+a)} \left(1+\frac{f(k,k+a)}{1+\sum_{j=1}^{k-1} f(j,k+a)}\right)^{n-1} k^* \Bigg] \Bigg\} \\
= \quad & \left(1+\sum_{j=1}^{k-1} f(j,k+a)\right)^{n-1} \Bigg\{ \\
& \sum_{j=1}^{k-1} j^* f(j,k+a) \binom{n}{1} \left(\frac{1+\sum_{j=1}^{k} f(j,k+a)}{1+\sum_{j=1}^{k-1} f(j,k+a)}\right)^{n-1} \\
& \quad + \left(1+\sum_{j=1}^{k-1} f(j,k+a)\right) \Bigg[\left(\frac{1+\sum_{j=1}^{k} f(j,k+a)}{1+\sum_{j=1}^{k-1} f(j,k+a)}\right)^{n} b \\
& \qquad + \binom{n}{1} \frac{f(k,k+a)}{1+\sum_{j=1}^{k-1} f(j,k+a)} \left(\frac{1+\sum_{j=1}^{k} f(j,k+a)}{1+\sum_{j=1}^{k-1} f(j,k+a)}\right)^{n-1} k^* \Bigg] \Bigg\}
\end{aligned}
$$

(we cancel all terms $1+\sum_{j=1}^{k-1} f(j,k+a)$ and factor out some other terms)

$$
\begin{aligned}
= \quad & \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n} b \\
& + \binom{n}{1} \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-1} \left[\sum_{j=1}^{k-1} j^* f(j,k+a) + k^* f(k.k+a)\right] \\
= \quad & \left(1+\sum_{j=1}^{k-1} f(j,k+a)\right)^{n-1} \times \\
& \left\{ \left(1+\sum_{j=1}^{k} f(j,k+a)\right) b + \binom{n}{1} \sum_{j=1}^{k} j^* f(j,k+a) \right\}.
\end{aligned}
$$

□

The proofs of the other identities use even longer formulae. To keep them a little shorter, we introduce some abbreviations.

Definition 4.A.6

- $f_k := f(k, k+a)$;
- $F_{i,l} := \sum_{j=1}^{l} j^i f(j, k+a)$.

The following identities, which are easy to prove, will be very useful.

Lemma 4.A.7

- $1 + \frac{f_k}{F_{0,k-1}} = \frac{1+F_{0,k}}{1+F_{0,k-1}}$;
- $F_{1,k} = kf_k + F_{1,k-1}$;
- $F_{2,k} = k^2 f_k + F_{2,k-1}$;
- $F_{3,k} = k^3 f_k + F_{3,k-1}$;
- $F_{4,k} = k^4 f_k + F_{4,k-1}$;
- $F_{1,k}^2 = k^2 f_k^2 + 2kf_k F_{1,k-1} + F_{1,k-1}^2$;
- $F_{2,k}^2 = k^4 f_k^2 + 2k^2 f_k F_{2,k-1} + F_{2,k-1}^2$;
- $F_{1,k}^3 = k^3 f_k^3 + 3k^2 f_k^2 F_{1,k-1} + 3kf_k F_{1,k-1}^2 + F_{1,k-1}^3$;
- $F_{1,k}^4 = k^4 f_k^4 + 4k^3 f_k^3 F_{1,k-1} + 6k^2 f_k^2 F_{1,k-1}^2 + 4kf_k F_{1,k-1}^3 + F_{1,k-1}^4$;
- $F_{1,k}F_{2,k} = k^3 f_k^2 + k^2 f_k F_{1,k-1} + kf_k F_{2,k-1} + F_{1,k-1}F_{2,k-1}$;
- $F_{1,k}F_{3,k} = k^4 f_k^2 + k^3 f_k F_{1,k-1} + kf_k F_{3,k-1} + F_{1,k-1}F_{3,k-1}$;
- $F_{1,k}^2 F_{2,k} = k^4 f_k^3 + 2k^3 f_k^2 F_{1,k-1} + k^2 f_k F_{1,k-1}^2 + k^2 f_k^2 F_{2,k-1}$
  $+2kf_k F_{1,k-1}F_{2,k-1} + F_{1,k-1}^2 F_{2,k-1}$.

Now we are ready to face and defeat the other monsters.

Proposition 4.A.8
For any integers $n \geq 0$, $k \geq 1$ and $a \geq 0$ and any complex number b such that $-\pi/2 < \arg(b) \leq \pi/2$ or $b = 0$, let $H_2(n, k, a, b^2)$ be defined by the recursion

$$H_2(n,k,a,b^2) := \sum_{\gamma=0}^{n} \binom{n}{\gamma} f(k,k+a)^{\gamma} H_2(n-\gamma, k-1, a+1, (b+k\gamma)^2) \tag{4.22}$$

and the initial condition $H_2(n, 0, a, b^2) := b^2$. Then, for any integer $k \geq 0$,

$$\begin{aligned} H_2(n,k,a,b^2) = {} & \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-2} \Bigg\{ \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{2} \left[b^2\right] \\ & + \binom{n}{1} \left(1+\sum_{j=1}^{k} f(j,k+a)\right) \left[\sum_{j=1}^{k} j^2 f(j,k+a) + 2b \sum_{j=1}^{k} j f(j,k+a)\right] \\ & + 2\binom{n}{2} \left[\left(\sum_{j=1}^{k} j f(j,k+a)\right)^{2}\right] \Bigg\}. \end{aligned} \tag{4.23}$$

Proof:

It is easy to see that (4.23) implies $H_2(n, 0, a, b^2) = b^2$. We prove the general case by induction. Let $k \geq 1$ and assume that the proposition holds up to $k - 1$. First we rewrite (4.23) with the above introduced abbreviations:

$$\begin{aligned} H_2(n,k,a,b^2) = {} & \left(1+F_{0,k}\right)^{n-2} \Big\{ \left(1+F_{0,k}\right)^{2} \left[b^2\right] \\ & + \binom{n}{1} \left(1+F_{0,k}\right) \left[F_{2,k} + 2bF_{1,k}\right] + 2\binom{n}{2} \left[F_{1,k}^2\right] \Big\}. \end{aligned} \tag{4.24}$$

Then

$$\begin{aligned} H_2(n,k,a,b^2) &= \sum_{w=0}^{n} \binom{n}{w} f_k^w H_2(n-w, k-1, a+1, (b+kw)^2) \\ &= \sum_{w=0}^{n} \binom{n}{w} f_k^w \left(1+F_{0,k-1}\right)^{n-w-2} \Bigg\{ \left(1+F_{0,k-1}\right)^{2} \left[(b+kw)^2\right] \\ &\quad + 2\binom{n-w}{2} \left[F_{1,k-1}^2\right] \\ &\quad + \binom{n-w}{1} \left(1+F_{0,k-1}\right) \left[F_{2,k-1} + 2(b+kw)F_{1,k-1}\right] \Bigg\} \end{aligned}$$

$$
\begin{aligned}
&= \left(1+F_{0,k-1}\right)^{n-2}\Bigg\{\sum_{w=0}^{n}\binom{n}{w}\left(\frac{f_k}{1+F_{0,k-1}}\right)^{w}\Bigg[(b+kw)^2\left(1+F_{0,k-1}\right)^{2}\\
&\quad +2\binom{n-w}{2}F_{1,k-1}^2+\left(1+F_{0,k-1}\right)F_{2,k-1}\binom{n-w}{1}\\
&\qquad +2\left(1+F_{0,k-1}\right)F_{1,k-1}(b+kw)\binom{n-w}{1}\Bigg]\Bigg\}\\
&= \left(1+F_{0,k-1}\right)^{n-2}\Bigg\{\left(1+F_{0,k-1}\right)^{2}\Bigg[b^2\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n}\\
&\quad +\binom{n}{1}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}(2bk+k^2)\\
&\quad +2\binom{n}{2}\left(\frac{f_k}{1+F_{0,k-1}}\right)^{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}k^2\Bigg]\\
&\quad +2F_{1,k-1}^2\binom{n}{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\\
&\quad +\left(1+F_{0,k-1}\right)F_{2,k-1}\binom{n}{1}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}\\
&\quad +2\left(1+F_{0,k-1}\right)F_{1,k-1}\Bigg[\binom{n}{1}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}\\
&\qquad +2\binom{n}{2}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Bigg]\Bigg\}\\
&= \left(1+F_{0,k}\right)^{n-2}\Bigg\{\left(1+F_{0,k}\right)^{2}\Big[b^2\Big]\\
&\quad +\binom{n}{1}\left(1+F_{0,k}\right)\Big[2bkf_k+k^2f_k+F_{2,k-1}+2bF_{1,k-1}\Big]\\
&\quad +2\binom{n}{2}\Big[k^2f_k^2+F_{1,k-1}^2+2kf_kF_{1,k-1}\Big]\Bigg\}\\
&= \left(1+F_{0,k}\right)^{n-2}\Bigg\{\left(1+F_{0,k}\right)^{2}\Big[b^2\Big]\\
&\quad +\binom{n}{1}\left(1+F_{0,k}\right)\Big[F_{2,k}+2bF_{1,k}\Big]+2\binom{n}{2}\Big[F_{1,k}^2\Big]\Bigg\}.
\end{aligned}
$$

□

Proposition 4.A.9
For any integers $n \geq 0$, $k \geq 1$ and $a \geq 0$ and any complex number b, let $H_3(n,k,a,b)$ be defined by the recursion

$$H_3(n,k,a,b) := \sum_{\gamma=0}^{n} \binom{n}{\gamma} f(k,k+a)^{\gamma} H_3(n-\gamma, k-1, a+1, b+k^2\gamma) \tag{4.25}$$

and the initial condition $H_3(n,0,a,b) := b$. Then, for any integer $k \geq 0$,

$$H_3(n,k,a,b) = \tag{4.26}$$
$$\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-1} \left\{ b\left(1+\sum_{j=1}^{k} f(j,k+a)\right) + \binom{n}{1} \sum_{j=1}^{k} j^2 f(j,k+a) \right\}.$$

Proof:

The proof is essentially the same as that of Proposition 4.A.5: square there every symbol that has a star, or, if you prefer, set $* = 2$ everywhere.□

Proposition 4.A.10
For any integers $n \geq 0$, $k \geq 1$ and $a \geq 0$ and any complex number b such that $-\pi/2 < \arg(b) \leq \pi/2$ or $b = 0$, let $H_4(n,k,a,b^2)$ be defined by the recursion

$$H_4(n,k,a,b^2) := \sum_{w=0}^{n} \binom{n}{w} f(k,k+a)^{w} H_4(n-w, k-1, a+1, (b+k^2 w)^2) \tag{4.27}$$

and the initial condition $H_4(n,0,a,b^2) := b^2$. Then, for any integer $k \geq 0$,

$$\begin{aligned} H_4(n,k,a,b^2) = \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-2} \Bigg\{ & \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^2 \left[b^2\right] \\ + & \binom{n}{1}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)\left[\sum_{j=1}^{k} j^4 f(j,k+a) + 2b\sum_{j=1}^{k} j^2 f(j,k+a)\right] \\ + & 2\binom{n}{2}\left[\left(\sum_{j=1}^{k} j^2 f(j,k+a)\right)^2\right]\Bigg\}. \end{aligned} \tag{4.28}$$

Proof:

The proof is essentially the same as the one of Proposition 4.A.8: replace in that proof H_2 by H_4, $F_{1,k-1}$ (resp. $F_{1,k}$, $F_{2,k-1}$, $F_{2,k}$) by $F_{2,k-1}$ (resp. $F_{2,k}$, $F_{4,k-1}$, $F_{4,k}$); also replace k (resp. k^2) by k^2 (resp. k^4) at all places except in indices, in limits of a sum or in the second argument of H_2. □

Proposition 4.A.11

For any integers $n \geq 0$, $k \geq 1$ and $a \geq 0$ and any complex number b such that $-\pi/4 < \arg(b) \leq \pi/4$ or $b = 0$, let $H_5(n,k,a,b^4)$ be defined by the recursion

$$H_5(n,k,a,b^4) := \sum_{w=0}^{n} \binom{n}{w} f(k,k+a)^w H_5(n-w,k-1,a+1,(b+k\gamma)^4) \tag{4.29}$$

and the initial condition $H_5(n,0,a,b^4) := b^4$. Then, for any integer $k \geq 0$,

$$\begin{aligned}
H_5(n,k,a,b^4) = {} & \Big(1+\sum_{j=1}^{k} f(j,k+a)\Big)^{n-4} \Bigg\{ \Big(1+\sum_{j=1}^{k} f(j,k+a)\Big)^{4} \Big[b^4\Big] \\
& + \binom{n}{1} \Big(1+\sum_{j=1}^{k} f(j,k+a)\Big)^{3} \Big[4b^3 \sum_{j=1}^{k} j f(j,k+a) \\
& \quad + 6b^2 \sum_{j=1}^{k} j^2 f(j,k+a) + 4b \sum_{j=1}^{k} j^3 f(j,k+a) + \sum_{j=1}^{k} j^4 f(j,k+a)\Big] \\
& + 2\binom{n}{2} \Big(1+\sum_{j=1}^{k} f(j,k+a)\Big)^{2} \Big[\\
& \quad 6b^2 \Big(\sum_{j=1}^{k} j f(j,k+a)\Big)^{2} + 12b \sum_{j=1}^{k} j f(j,k+a) \sum_{j=1}^{k} j^2 f(j,k+a) \\
& \quad + 3\Big(\sum_{j=1}^{k} j^2 f(j,k+a)\Big)^{2} + 4 \sum_{j=1}^{k} j f(j,k+a) \sum_{j=1}^{k} j^3 f(j,k+a)\Big] \\
& + 6\binom{n}{3} \Big(1+\sum_{j=1}^{k} f(j,k+a)\Big) \Big[4b\Big(\sum_{j=1}^{k} j f(j,k+a)\Big)^{3} \\
& \quad + 6\Big(\sum_{j=1}^{k} j f(j,k+a)\Big)^{2} \sum_{j=1}^{k} j^2 f(j,k+a)\Big] \\
& + 24\binom{n}{4} \Big(\sum_{j=1}^{k} j f(j,k+a)\Big)^{4} \Bigg\}.
\end{aligned} \tag{4.30}$$

Proof:

To prove that (4.30) implies $H_5(n,0,a,b^4) = b^4$ is easy. We prove the general case by induction. Let $k \geq 1$ and assume that the proposition holds up to $k-1$. First we rewrite (4.30) with the above introduced abbreviations:

$$\begin{aligned} H_5(n,k,a,b^4) &= \left(1+F_{0,k}\right)^{n-4}\Big\{\left(1+F_{0,k}\right)^4\left[b^4\right] \\ &+\binom{n}{1}\left(1+F_{0,k}\right)^3\left[4b^3F_{1,k}+6b^2F_{2,k}+4bF_{3,k}+F_{4,k}\right] \\ &+2\binom{n}{2}\left(1+F_{0,k}\right)^2\left[6b^2F_{1,k}^2+12bF_{1,k}F_{2,k}+3F_{2,k}^2+4F_{1,k}F_{3,k}\right] \\ &+6\binom{n}{3}\left[4bF_{1,k}^3+6F_{1,k}^2F_{2,k}\right]+24\binom{n}{4}\left[F_{1,k}^4\right]\Big\}. \end{aligned} \tag{4.31}$$

Then

$$H_5(n,k,a,b^4) = \sum_{w=0}^{n}\binom{n}{w}f(k,k+a)^w H_5(n-w,k-1,a+1,(b+k\gamma)^4)$$

$$\begin{aligned} &= \sum_{w=0}^{n}\binom{n}{w}f_k^w\left(1+F_{0,k-1}\right)^{n-w-4}\Big\{\left(1+F_{0,k-1}\right)^4\left[(b+kw)^4\right] \\ &+\binom{n-w}{1}\left(1+F_{0,k-1}\right)^3\Big[4(b+kw)^3F_{1,k-1} \\ &\qquad +6(b+kw)^2F_{2,k-1}+4(b+kw)F_{3,k-1}+F_{4,k-1}\Big] \\ &+2\binom{n-w}{2}\left(1+F_{0,k-1}\right)^2\Big[6(b+kw)^2F_{1,k-1}^2 \\ &\qquad +12(b+kw)F_{1,k-1}F_{2,k-1}+3F_{2,k-1}^2+4F_{1,k-1}F_{3,k-1}\Big] \\ &+6\binom{n-w}{3}\left(1+F_{0,k-1}\right)\left[4(b+kw)F_{1,k-1}^3+6F_{1,k-1}^2F_{2,k-1}\right] \\ &+24\binom{n-w}{4}\left[F_{1,k-1}^4\right]\Big\} \end{aligned}$$

$$
\begin{aligned}
= \quad & \Big(1+F_{0,k-1}\Big)^{n-4} \sum_{w=0}^{n} \binom{n}{w} \Big(\frac{f_k}{1+F_{0,k-1}}\Big)^{w} \Big\{ \\
& \Big(1+F_{0,k-1}\Big)^{4} (b+kw)^4 + 4\Big(1+F_{0,k-1}\Big)^{3} F_{1,k-1} \binom{n-w}{1} (b+kw)^3 \\
& + 6\Big(1+F_{0,k-1}\Big)^{3} F_{2,k-1} \binom{n-w}{1} (b+kw)^2 \\
& + 4\Big(1+F_{0,k-1}\Big)^{3} F_{3,k-1} \binom{n-w}{1} (b+kw) \\
& + \Big(1+F_{0,k-1}\Big)^{3} F_{4,k-1} \binom{n-w}{1} \\
& + 12\Big(1+F_{0,k-1}\Big)^{2} F_{1,k-1}^2 \binom{n-w}{2} (b+kw)^2 \\
& + 24\Big(1+F_{0,k-1}\Big)^{2} F_{1,k-1} F_{2,k-1} \binom{n-w}{2} (b+kw) \\
& + 6\Big(1+F_{0,k-1}\Big)^{2} F_{2,k-1}^2 \binom{n-w}{2} \\
& + 8\Big(1+F_{0,k-1}\Big)^{2} F_{1,k-1} F_{3,k-1} \binom{n-w}{2} \\
& + 24\Big(1+F_{0,k-1}\Big) F_{1,k-1}^3 \binom{n-w}{3} (b+kw) \\
& + 36\Big(1+F_{0,k-1}\Big) F_{1,k-1}^2 F_{2,k-1} \binom{n-w}{3} + 24 F_{1,k-1}^4 \binom{n-w}{4} \Big\} \\
= \quad & \Big(1+F_{0,k-1}\Big)^{n-4} \Big\{ \Big(1+F_{0,k-1}\Big)^{4} \Big[\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n} \Big[b^4\Big] \\
& + \binom{n}{1} \frac{f_k}{1+F_{0,k-1}} \Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-1} \Big[4b^3k + 6b^2k^2 + 4bk^3 + k^4\Big] \\
& + 2\binom{n}{2} \Big(\frac{f_k}{1+F_{0,k-1}}\Big)^{2} \Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-2} \Big[6b^2k^2 + 12bk^3 + 7k^4\Big] \\
& + 6\binom{n}{3} \Big(\frac{f_k}{1+F_{0,k-1}}\Big)^{3} \Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-3} \Big[4bk^3 + 6k^4\Big] \\
& + 24\binom{n}{4} \Big(\frac{f_k}{1+F_{0,k-1}}\Big)^{4} \Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-4} \Big[k^4\Big] \Big]
\end{aligned}
$$

...

(continued on next page)

(continued from previous page)

$$
\begin{aligned}
&+4\Big(1+F_{0,k-1}\Big)^3 F_{1,k-1}\left[\binom{n}{1}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}\Big[b^3\Big]\right.\\
&\qquad+2\binom{n}{2}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Big[3b^2k+3bk^2+k^3\Big]\\
&\qquad+6\binom{n}{3}\left(\frac{f_k}{1+F_{0,k-1}}\right)^2\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\Big[3bk^2+3k^3\Big]\\
&\qquad\left.+24\binom{n}{4}\left(\frac{f_k}{1+F_{0,k-1}}\right)^3\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-4}\Big[k^3\Big]\right]\\
&+6\Big(1+F_{0,k-1}\Big)^3 F_{2,k-1}\left[\binom{n}{1}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}\Big[b^2\Big]\right.\\
&\qquad+2\binom{n}{2}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Big[2bk+k^2\Big]\\
&\qquad\left.+6\binom{n}{3}\left(\frac{f_k}{1+F_{0,k-1}}\right)^2\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\Big[k^2\Big]\right]\\
&+4\Big(1+F_{0,k-1}\Big)^3 F_{3,k-1}\left[\binom{n}{1}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}\Big[b\Big]\right.\\
&\qquad\left.+2\binom{n}{2}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Big[k\Big]\right]\\
&+\Big(1+F_{0,k-1}\Big)^3 F_{4,k-1}\binom{n}{1}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}\\
&+12\Big(1+F_{0,k-1}\Big)^2 F_{1,k-1}^2\left[\binom{n}{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Big[b^2\Big]\right.\\
&\qquad+3\binom{n}{3}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\Big[2bk+k^2\Big]\\
&\qquad\left.+12\binom{n}{4}\left(\frac{f_k}{1+F_{0,k-1}}\right)^2\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-4}\Big[k^2\Big]\right]\\
&+24\Big(1+F_{0,k-1}\Big)^2 F_{1,k-1}F_{2,k-1}\left[\binom{n}{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Big[b\Big]\right.\\
&\qquad\left.+3\binom{n}{3}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\Big[k\Big]\right]
\end{aligned}
$$

(continued on next page)

(continued from previous page) ...

$$
\begin{aligned}
&+6\Big(1+F_{0,k-1}\Big)^2 F_{2,k-1}^2 \binom{n}{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\\
&+8\Big(1+F_{0,k-1}\Big)^2 F_{1,k-1}F_{3,k-1}\binom{n}{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\\
&+24\Big(1+F_{0,k_1}\Big)F_{1.k-1}^3\left[\binom{n}{3}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\Big[b\Big]\right.\\
&\qquad\left.+4\binom{n}{4}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-4}\Big[k\Big]\right]\\
&+36\Big(1+F_{0,k-1}\Big)F_{1,k-1}^2F_{2,k-1}\binom{n}{3}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\\
&+24F_{1,k-1}^4\binom{n}{4}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-4}\Bigg\}
\end{aligned}
$$

(all $\Big(1+F_{0,k-1}\Big)$-expressions inside the curly brackets cancel, and we factor out the binomial coefficients)

$$
\begin{aligned}
=\ &\Big(1+F_{0,k-1}\Big)^{n-4}\Bigg\{\Big(1+F_{0,k}\Big)^4\Big[b^4\Big]\\
&+\binom{n}{1}\Big(1+F_{0,k}\Big)^3\Big[f_k(4b^3k+6b^2k^2+4bk^3+k^4)+4F_{1,k-1}b^3\\
&\quad+6F_{2,k-1}b^2+4F_{3,k-1}b+F_{4,k-1}\Big]\\
&+2\binom{n}{2}\Big(1+F_{0,k}\Big)^2\Big[f_k^2(6b^2k^2+12bk^3+7k^4)\\
&\quad+4F_{1,k-1}f_k(3b^2k+3bk^2+k^3)+6F_{2,k-1}f_k(2bk+k^2)+4F_{3,k-1}f_kk\\
&\quad+6F_{1,k-1}^2b^2+12F_{1,k-1}F_{2,k-1}b+3F_{2,k-1}^2+4F_{1,k-1}F_{3,k-1}\Big]\\
&+6\binom{n}{3}\Big(1+F_{0,k}\Big)\Big[f_k^3(4bk^3+6k^4)+4F_{1,k-1}f_k^2(3bk^2+3k^3)\\
&\quad+6F_{2,k-1}f_k^2(k^2)+6F_{1,k-1}^2f_k(2bk+k^2)\\
&\quad+12F_{1,k-1}F_{2,k-1}f_k(k)+4F_{1,k-1}^3b+6F_{1,k-1}^2F_{2,k-1}\Big]\\
&+24\binom{n}{4}\Big[f_k^4k^4+4F_{1,k-1}f_k^3k^3+6F_{1,k-1}^2f_k^2k^2+4F_{1,k-1}^3f_kk+F_{1,k-1}^4\Big]\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \left(1+F_{0,k-1}\right)^{n-4}\Bigg\{\left(1+F_{0,k}\right)^{4}\left[b^4\right] \\
& +\binom{n}{1}\left(1+F_{0,k}\right)^{3}\Big[4b^3(kf_k+F_{1,k-1})+6b^2(k^2f_k+F_{2,k-1}) \\
& \quad +4b(k^3f_k+F_{3,k-1})+(k^4f_k+F_{4,k-1})\Big] \\
& +2\binom{n}{2}\left(1+F_{0,k}\right)^{2}\Big[6b^2(k^2f_k^2+2kf_kF_{1,k-1}+F_{1,k-1}^2) \\
& \quad +12b(k^3f_k^2+k^2f_kF_{1,k-1}+kf_kF_{2,k-1}+F_{1,k-1}F_{2,k-1}) \\
& \quad +3(k^4f_k^2+2k^2f_kF_{2,k-1}) \\
& \quad +4(k^4f_k^2+k^3f_kF_{1,k-1}+kf_kF_{3,k-1}+F_{1,k-1}F_{3,k-1})\Big] \\
& +6\binom{n}{3}\left(1+F_{0,k}\right)\Big[4b(k^3f_k^3+3k^2f_k^2F_{1,k-1}+3kf_kF_{1,k-1}^2+F_{1,k-1}^3) \\
& \quad +6(k^4f_k^3+2k^3f_k^2F_{1,k-1}+k^2f_k^2F_{2,k-1}+k^2f_kF_{1,k-1}^2 \\
& \qquad +2kf_kF_{1,k-1}F_{2,k-1}+F_{1,k-1}^2F_{2,k-1})\Big] \\
& +24\binom{n}{4}\Big[f_k^4k^4+4F_{1,k-1}f_k^3k^3+6F_{1,k-1}^2f_k^2k^2+4F_{1,k-1}^3f_kk+F_{1,k-1}^4\Big]\Bigg\} \\
= \quad & \left(1+F_{0,k}\right)^{n-4}\Bigg\{\left(1+F_{0,k}\right)^{4}\left[b^4\right] \\
& +\binom{n}{1}\left(1+F_{0,k}\right)^{3}\Big[4b^3F_{1,k}+6b^2F_{2,k}+4bF_{3,k}+F_{4,k}\Big] \\
& +2\binom{n}{2}\left(1+F_{0,k}\right)^{2}\Big[6b^2F_{1,k}^2+12bF_{1,k}F_{2,k}+3F_{2,k}^2+4F_{1,k}F_{3,k}\Big] \\
& +6\binom{n}{3}\left(1+F_{0,k}\right)\Big[4bF_{1,k}^3+6F_{1,k}^2F_{2,k}\Big]+24\binom{n}{4}\Big[F_{1,k}^4\Big]\Bigg\}.
\end{aligned}
$$

□

Proposition 4.A.12
For any integers $n \geq 0$, $k \geq 1$ and $a \geq 0$ and any complex numbers b and c such that $-\pi/2 < \arg(c) \leq \pi/2$ or $c = 0$, let $H_6(n, k, a, b, c^2)$ be defined by the recursion

$$H_6(n,k,a,b,c^2) := \sum_{w=0}^{n} \binom{n}{w} f(k,k+a)^w H_6(n-w,k-1,a+1,b+k^2w,(c+kw)^2) \tag{4.32}$$

and the initial condition $H_6(n,0,a,b,c^2) := bc^2$. Then, for any integer $k \geq 0$,

$$\begin{aligned}
H_6(n,k,a,b,c^2) = {} & \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^{n-3} \Bigg\{ \left(1+\sum_{j=1}^{k} f(j,k+a)\right)^3 \left[bc^2\right] \\
& + \binom{n}{1}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)^2 \Bigg[b\sum_{j=1}^{k} j^2 f(j,k+a) + 2bc\sum_{j=1}^{k} j f(j,k+a) \\
& \quad + 2c\sum_{j=1}^{k} j^3 f(j,k+a) + c^2\sum_{j=1}^{k} j^2 f(j,k+a) + \sum_{j=1}^{k} j^4 f(j,k+a)\Bigg] \\
& + 2\binom{n}{2}\left(1+\sum_{j=1}^{k} f(j,k+a)\right)\Bigg[b\left(\sum_{j=1}^{k} j f(j,k+a)\right)^2 \\
& \quad + 2c\sum_{j=1}^{k} j f(j,k+a)\sum_{j=1}^{k} j^2 f(j,k+a) + \left(\sum_{j=1}^{k} j^2 f(j,k+a)\right)^2 \\
& \quad + 2\sum_{j=1}^{k} j f(j,k+a)\sum_{j=1}^{k} j^3 f(j,k+a)\Bigg] \\
& + 6\binom{n}{3}\left(\sum_{j=1}^{k} j f(j,k+a)\right)^2 \sum_{j=1}^{k} j^2 f(j,k+a)\Bigg\}.
\end{aligned} \tag{4.33}$$

Proof:

To prove that (4.33) implies $H_6(n,0,a,b,c^2) = bc^2$ is easy. We prove the general case by induction. Let $k \geq 1$ and assume that the proposition holds up to $k-1$. First we rewrite (4.33) with the usual abbreviations:

$$
\begin{aligned}
H_6(n,k,a,b,c^2) &= \left(1+F_{0,k}\right)^{n-3}\Bigg\{\left(1+F_{0,k}\right)^3\left[bc^2\right] \\
&\quad+\binom{n}{1}\left(1+F_{0,k}\right)^2\left[bF_{2,k}+2bcF_{1,k}+2cF_{3,k}+c^2F_{2,k}+F_{4,k}\right] \\
&\quad+2\binom{n}{2}\left(1+F_{0,k}\right)\left[bF_{1,k}^2+2cF_{1,k}F_{2,k}+F_{2,k}^2+2F_{1.k}F_{3,k}\right] \\
&\quad+6\binom{n}{3}\left[F_{1,k}^2F_{2,k}\right]\Bigg\}. \qquad (4.34)
\end{aligned}
$$

Then

$$
\begin{aligned}
H_6(n,k,a,b,c^2) &= \sum_{w=0}^{n}\binom{n}{w}f_k^w H_6(n-w,k-1,a+1,b+k^2w,(c+kw)^2) \\
&= \sum_{w=0}^{n}\binom{n}{w}f_k^w\left(1+F_{0,k-1}\right)^{n-w-3}\Bigg\{\left(1+F_{0,k-1}\right)^3\left[(b+kw)(c+kw)^2\right] \\
&\quad+\binom{n-w}{1}\left(1+F_{0,k-1}\right)^2\Big[(b+k^2w)F_{2,k-1}+2(b+kw)(c+kw)F_{1,k-1} \\
&\qquad+2(c+kw)F_{3,k-1}+(c+kw)^2F_{2,k-1}+F_{4,k-1}\Big] \\
&\quad+2\binom{n-w}{2}\left(1+F_{0,k-1}\right)\Big[(b+k^2w)F_{1,k-1}^2+2(c+kw)F_{1,k-1}F_{2,k-1} \\
&\qquad+F_{2,k-1}^2+2F_{1,k-1}F_{3,k-1}\Big] \\
&\quad+6\binom{n-w}{3}\left[F_{1,k-1}^2F_{3,k-1}\right]\Bigg\} \\
&= \left(1+F_{0,k-1}\right)^{n-3}\sum_{w=0}^{n}\binom{n}{w}\left(\frac{f_k}{1+F_{0,k-1}}\right)^w\Bigg\{ \\
&\quad\left(1+F_{0,k-1}\right)^3(b+kw)(c+kw)^2 \\
&\quad+\left(1+F_{0,k-1}\right)^2F_{2,k-1}\binom{n-w}{1}(b+k^2w) \\
&\quad+2\left(1+F_{0,k-1}\right)^2F_{1,k-1}\binom{n-w}{1}(b+k^2w)(c+kw) \\
&\quad\cdots
\end{aligned}
$$

(continued on next page)

(continued from previous page)

$$
\begin{aligned}
&\quad \cdots \\
&+2\Big(1+F_{0,k-1}\Big)^2 F_{3,k-1}\binom{n-w}{1}(c+kw) \\
&+\Big(1+F_{0,k-1}\Big)^2 F_{2,k-1}\binom{n-w}{1}(c+kw)^2 \\
&+\Big(1+F_{0,k-1}\Big)^2 F_{4,k-1}\binom{n-w}{1} \\
&+2\Big(1+F_{0,k-1}\Big) F_{1,k-1}^2\binom{n-w}{2}(b+k^2w) \\
&+4\Big(1+F_{0,k-1}\Big) F_{1,k-1}F_{2,k-1}\binom{n-w}{2}(c+kw) \\
&+2\Big(1+F_{0,k-1}\Big) F_{2,k-1}^2\binom{n-w}{2} \\
&+4\Big(1+F_{0,k-1}\Big) F_{1,k-1}F_{3,k-1}\binom{n-w}{2}+6F_{1,k-1}^2F_{2,k-1}\binom{n-w}{3}\Bigg\} \\
=\ &\Big(1+F_{0,k-1}\Big)^{n-3}\Bigg\{\Big(1+F_{0,k-1}\Big)^3\Bigg[\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^n\Big[bc^2\Big] \\
&\quad+\binom{n}{1}\frac{f_k}{1+F_{0,k-1}}\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-1}\Big[c^2k^2+2bck+2ck^3+bk^2+k^4\Big] \\
&\quad+2\binom{n}{2}\Big(\frac{f_k}{1+F_{0,k-1}}\Big)^2\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-2}\Big[2ck^3+bk^2+3k^4\Big] \\
&\quad+6\binom{n}{3}\Big(\frac{f_k}{1+F_{0,k-1}}\Big)^3\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-3}\Big[k^4\Big]\Bigg] \\
&+\Big(1+F_{0,k-1}\Big)^2F_{2,k-1}\Bigg[\binom{n}{1}\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-1}\Big[b\Big] \\
&\quad+2\binom{n}{2}\frac{f_k}{1+F_{0,k-1}}\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-2}\Big[k^2\Big]\Bigg] \\
&+2\Big(1+F_{0,k-1}\Big)^2F_{1,k-1}\Bigg[\binom{n}{1}\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-1}\Big[bc\Big] \\
&\quad+2\binom{n}{2}\frac{f_k}{1+F_{0,k-1}}\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-2}\Big[ck^2+bk+k^3\Big] \\
&\quad+6\binom{n}{3}\Big(\frac{f_k}{1+F_{0,k-1}}\Big)^2\Big(\frac{1+F_{0,k}}{1+F_{0,k-1}}\Big)^{n-3}\Big[k^3\Big]\Bigg] \\
&\quad\cdots
\end{aligned}
$$

(continued on next page)

(continued from previous page) ...

$$
\begin{aligned}
&+2\Big(1+F_{0,k-1}\Big)^2 F_{3,k-1}\left[\binom{n}{1}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}\Big[c\Big]\right.\\
&\qquad\left.+2\binom{n}{2}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Big[k\Big]\right]\\
&+\Big(1+F_{0,k-1}\Big)^2 F_{2,k-1}\left[\binom{n}{1}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}\Big[c^2\Big]\right.\\
&\qquad+2\binom{n}{2}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Big[2ck+k^2\Big]\\
&\qquad\left.+6\binom{n}{3}\left(\frac{f_k}{1+F_{0,k-1}}\right)^2\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\Big[k^2\Big]\right]\\
&+\Big(1+F_{0,k-1}\Big)^2 F_{4,k-1}\binom{n}{1}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-1}\\
&+2\Big(1+F_{0,k-1}\Big)F_{1,k-1}^2\left[\binom{n}{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Big[b\Big]\right.\\
&\qquad\left.+3\binom{n}{3}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\Big[k^2\Big]\right]\\
&+4\Big(1+F_{0,k-1}\Big)F_{1,k-1}F_{2,k-1}\left[\binom{n}{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\Big[c\Big]\right.\\
&\qquad\left.+3\binom{n}{3}\frac{f_k}{1+F_{0,k-1}}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\Big[k\Big]\right]\\
&+2\Big(1+F_{0,k-1}\Big)F_{2,k-1}^2\binom{n}{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\\
&+4\Big(1+F_{0,k-1}\Big)F_{1,k-1}F_{3,k-1}\binom{n}{2}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-2}\\
&+6F_{1,k-1}^2F_{2,k-1}\binom{n}{3}\left(\frac{1+F_{0,k}}{1+F_{0,k-1}}\right)^{n-3}\Bigg\}
\end{aligned}
$$

(all $\Big(1+F_{0,k-1}\Big)$-expressions inside the curly brackets cancel, and we factor out the binomial coefficients)

$$
\begin{aligned}
= \quad & \left(1+F_{0,k-1}\right)^{n-3}\Bigg\{\left(1+F_{0,k}\right)^{3}\left[bc^2\right] \\
& +\binom{n}{1}\left(1+F_{0,k}\right)^{2}\Big[f_k(c^2k^2+2bck+2ck^3+bk^2+k^4)+F_{2,k-1}b \\
& \qquad +2F_{1,k-1}bc+2F_{3,k-1}c+F_{2,k-1}c^2+F_{4,k-1}\Big] \\
& +2\binom{n}{2}\left(1+F_{0,k}\right)\Big[f_k^2(2ck^3+bk^2+3k^4)+F_{2,k-1}f_kk^2 \\
& \qquad +2F_{1,k-1}f_k(ck^2+bk+k^3)+2F_{3,k-1}f_kk+F_{2,k-1}f_k(2ck+k^2) \\
& \qquad +F_{1,k-1}^2b+2F_{1,k-1}F_{2,k-1}c+F_{2,k-1}^2+2F_{1,k-1}F_{3,k-1}\Big] \\
& +6\binom{n}{3}\Big[f_k^3k^4+2F_{1,k-1}f_k^2k^3+F_{2,k-1}f_k^2k^2+F_{1,k-1}^2f_kk^2 \\
& \qquad +2F_{1,k-1}F_{2,k-1}f_kk+F_{1,k-1}^2F_{2,k-1}\Big]\Bigg\} \\
= \quad & \left(1+F_{0,k-1}\right)^{n-3}\Bigg\{\left(1+F_{0,k}\right)^{3}\left[bc^2\right] \\
& +\binom{n}{1}\left(1+F_{0,k}\right)^{2}\Big[b(k^2f_k+F_{2,k-1})+2bc(kf_k+F_{1,k-1}) \\
& \qquad +2c(k^3f_k+F_{3,k-1})+c^2(k^2f_k+F_{2,k-1})+(k^4f_k+F_{4,k-1})\Big] \\
& +2\binom{n}{2}\left(1+F_{0,k}\right)\Big[b(k^2f_k^2+2kf_kF_{1,k-1}+F_{1,k-1}^2) \\
& \qquad +2c(k^3f_k^2+k^2f_kF_{1,k-1}+kf_kF_{2,k-1}+F_{1,k-1}F_{2,k-1}) \\
& \qquad +(k^4f_k^2+k^2f_kF_{2,k-1}+k^2f_kF_{2,k-1}+F_{2,k-1}^2) \\
& \qquad +2(k^2f_k^2+k^3f_kF_{1,k-1}+kf_kF_{3,k-1}+F_{1,k-1}F_{3,k-1})\Big] \\
& +6\binom{n}{3}\Big[k^4f_k^3+2k^3f_k^2F_{1,k-1}+k^2f_k^2F_{2,k-1}+k^2f_kF_{1,k-1}^2 \\
& \qquad +2kf_kF_{1.k-1}F_{2.k-1}+F_{1,k-1}^2F_{2,k-1}\Big]\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \left(1+F_{0,k}\right)^{n-3}\left\{\left(1+F_{0,k}\right)^{3}\left[bc^2\right]\right. \\
& +\binom{n}{1}\left(1+F_{0,k}\right)^{2}\left[bF_{2,k}+2bcF_{1,k}+2cF_{3,k}+c^2F_{2,k}+F_{4,k}\right] \\
& +2\binom{n}{2}\left(1+F_{0,k}\right)\left[bF_{1,k}^2+2cF_{1,k}F_{2,k}+F_{2,k}^2+2F_{1.k}F_{3,k}\right] \\
& \left.+6\binom{n}{3}\left[F_{1,k}^2F_{2,k}\right]\right\}.
\end{aligned}
$$

□

Chapter 5

The Piling-Up Hypothesis

Matsui's Piling-up Lemma is used in the binary generalisation of linear cryptanalysis (and also in Matsui's original linear cryptanalysis) to compute the probability of success of the attack (see Chapter 2). Harpes has shown [16] that in the group generalisation of linear cryptanalysis, which is briefly described below, the imbalance of a product of independent random variables is not in general equal to the product of their imbalances. However, he noticed that in most cases both values were very close to each other. This led him to propose a conjecture that he called the *piling-up hypothesis*. The purpose of this chapter is to study whether this hypothesis holds.

In Section 5.1, we present briefly the attack called the group generalisation of linear cryptanalysis and state the piling-up hypothesis. Section 5.2 is concerned with the validity of this hypothesis for a certain imbalance measure; we show for this measure that the hypothesis holds by proving that on average the imbalance of a product of two random variables is equal to the product of the imbalances of these random variables and by proving that on average the squared distance between the imbalance of the product and the product of the imbalances is small. We relegate a long proof to the appendix.

5.1 Group Generalisation of Linear Cryptanalysis

This attack [16] is a further generalisation of linear cryptanalysis in which one replaces:

- binary-valued balanced functions by m-ary balanced functions ($m \geq 2$), that is, functions that take on each value in $\{0, 1, \ldots, m-1\}$ for the same number of arguments;
- mod 2-addition $\oplus$ by a group operation $*$ on $\mathbb{Z}_m$;
- I/O sums $S = f(X) \oplus g(Y)$ by *I/O differences* $D = f(X) * g(Y)^{-1}$, where $g(Y)^{-1}$ is the inverse of $g(Y)$ under the group operation $*$;
- threefold sums $T = f(X) \oplus g(Y) \oplus h(K)$ by *I/K/O combinations* $C = f(X) * h(K) * g(Y)^{-1}$.

Also, the binary imbalance $I(X) = |2P[X = 0] - 1|$ is replaced by one of the following.

$$\text{(l^1-imbalance)} \qquad I_1(\boldsymbol{p}) := \frac{m}{2(m-1)} \sum_{i=0}^{m-1} \left| p_i - \frac{1}{m} \right| \tag{5.1}$$

$$\text{(Euclidian imbalance)} \qquad I_2(\boldsymbol{p}) := \sqrt{\frac{m}{m-1} \sum_{i=0}^{m-1} p_i^2 - \frac{1}{m-1}} \tag{5.2}$$

$$\text{(l^∞-imbalance)} \qquad I_\infty(\boldsymbol{p}) := \frac{m}{m-1} \max_i \left| p_i - \frac{1}{m} \right| \tag{5.3}$$

$$\text{(peak imbalance)} \qquad I_p(\boldsymbol{p}) := \frac{m}{m-1} \max_i \left(p_i - \frac{1}{m} \right) \tag{5.4}$$

$$\text{(peak-to-peak imbalance)} \qquad I_{pp}(\boldsymbol{p}) := \max_i p_i - \min_i p_i \tag{5.5}$$

$$\text{(redundancy imbalance)} \qquad I_{red}(\boldsymbol{p}) := 1 - \frac{H(\boldsymbol{p})}{\log m} \tag{5.6}$$

where $H(\boldsymbol{p})$ is the entropy of a random variable with probability vector $\boldsymbol{p}$. A *probability vector* $\boldsymbol{p} = (p_0, p_1, \ldots, p_{m-1})$ is a vector whose components are non-negative and sum up to 1. It corresponds to the probability distribution of some m-ary random variable X with $P_X(i) = p_i$, $i = 0, 1, \ldots, m-1$. We shall talk of $\boldsymbol{p}$ as a probability vector when $\boldsymbol{p}$ is any point in $\mathbb{R}^m$ with the above property and more specifically as a probability distribution when $\boldsymbol{p}$ is related to some random variable. Both concepts are equivalent. Note that the value of the imbalances does not change if one permutes the indices of $\boldsymbol{p}$. (We call that operation a *permutation of* $\boldsymbol{p}$.) For the imbalance of a random variable X with probability distribution $\boldsymbol{p}$, we shall write interchangeably $I(X)$ or $I(\boldsymbol{p})$, depending on which is more convenient. Harpes has shown [16] that the following properties hold for all of the above imbalances:

- $0 \leq I(\boldsymbol{p}) \leq 1$ for all probability vectors $\boldsymbol{p}$;

- $I(\boldsymbol{p}) = 0 \Leftrightarrow p_i = \frac{1}{m}$ for all i;
- $I(\boldsymbol{p}) = 1 \Leftrightarrow$ there is an i with $p_i = 1$ and $p_j = 0$ for all $j \neq i$.

Hereafter, we omit the redundancy imbalance from consideration because it does not possess some natural properties that the others do; for instance, as can be easily seen, all the others reduce for $m = 2$ to the binary imbalance defined in Chapter 2. Also, we will write I when we mean that the imbalance considered may be any of the five imbalances I_1, I_2, I_∞, I_p, and I_{pp}.

Remark 5.1.1
The generalisation of linear cryptanalysis applies only for m equal to a power of 2 when, as we consider, both the plaintext and the ciphertext are binary n-tuples and thus have 2^n possible values. In that case, m-ary balanced functions must take on each value for $2^n/m$ arguments; but this number is an integer only if $m = 2^l$ with $0 \leq l \leq n$. Nevertheless, because in this chapter we are concerned only with imbalances, we will not restrict ourselves to powers of two for m. Everything in this chapter, unless stated otherwise, is valid for any integer $m \geq 2$.

Harpes has shown other properties of the imbalances, valid for any group operation $*$ on $\mathbb{Z}_m$ and any independent m-ary random variables X and Y [16]:

a) As a function on m-ary probability distributions, I is convex-$\cup$;

b) $I(X * Y) \leq \min(I(X), I(Y))$;

c) $I_1(X * Y) \leq \frac{2(m-1)}{m} I_1(X) \cdot I_1(Y)$;

d) $I_2(X * Y) \leq \sqrt{m-1}\, I_2(X) \cdot I_2(Y)$.

e) The *one-peak distribution* $\boldsymbol{p} = (\frac{1+(m-1)\varepsilon}{m}, \frac{1-\varepsilon}{m}, \ldots, \frac{1-\varepsilon}{m})$, where $-\frac{1}{m-1} \leq \varepsilon \leq 1$, has imbalance $I(\boldsymbol{p}) = |\varepsilon|$.

f) If X and Y have one-peak distributions, then $X * Y$ also has a one-peak distribution and

$$I(X * Y) = I(X) \cdot I(Y). \tag{5.7}$$

However, (5.7) does not hold in general for independent random variables X and Y, as can be seen from the following example, taken from [16]:

Example 5.1.2
Let $\boldsymbol{p} = (\frac{1}{2}, \frac{1}{2}, 0, 0)$ and $\boldsymbol{q} = (\frac{1}{2}, 0, \frac{1}{2}, 0)$ be probability distributions for independent random variables X and Y. Let the group operation $*$ be $\oplus_4$, addition mod 4. Then $X \oplus_4 Y$ is uniformly distributed and thus $I(X \oplus_4 Y) = 0$. However, $I_1(X) = \frac{2}{3}, I_2(X) = \sqrt{\frac{1}{3}}, I_\infty(X) = \frac{1}{3}, I_p(X) = \frac{1}{3}$ and $I_{pp}(X) = \frac{1}{2}$ and Y has the same imbalance as X. Thus $I(X \oplus_4 Y) = 0 \neq I(X)I(Y)$.

However, it seems that in most cases the imbalance of a product of independent random variables is approximately equal to the product of the imbalances of the individual random variables. This fact led Harpes to suggest the following hypothesis:

Conjecture 5.1.3 (Piling-up Hypothesis)
For any integer $r \geq 2$, for any group operation $*$ on $\mathbb{Z}_m$,

$$I(X_1 * \cdots * X_r) \approx \prod_{j=1}^{r} I(X_j). \tag{5.8}$$

for almost all m-ary independent random variables $X_1, \ldots, X_r$.

(In fact, Harpes restricted his hypothesis to products of I/K/O combinations, but if the hypothesis holds in our broader sense, it also holds in Harpes's sense.) The statement of this hypothesis is again not very precise. Our goal is to give a quantitative meaning to the approximation sign. It is enough to consider the case where $r = 2$, i.e., only two random variables, since if (5.8) holds for $r = 2$, then it holds for any $r \geq 2$ by induction. If we know how well (5.8) holds for $r = 2$, then we also know how well it holds when $r > 2$.

Remark 5.1.4
It is enough to consider the piling-up hypothesis (5.8) for random variables that are not balanced: if one of them is balanced, then their product is balanced and we have equality with 0 on both sides. Thus, in what follows, all random variables will be assumed to have non-zero imbalance. This allows us to make the following definition.

Definition 5.1.5
For any imbalance measure I, define the *piling-up factor* as

$$\Delta(X_1, X_2) := \frac{I(X_1 * X_2)}{I(X_1)I(X_2)}. \tag{5.9}$$

We will write Δ_1 for the piling-up factor based on the imbalance I_1, and so on.

The piling-up factor normalizes the departure from the piling-up hypothesis and allows one to compare in particular cases how far one is from the equality $I(X_1 * X_2) = I(X_1)I(X_2)$. The piling-up hypothesis for two random variables reads:

For any group operation $*$ on $\mathbb{Z}_m$,

$$\Delta(X_1, X_2) \approx 1 \tag{5.10}$$

for almost all m-ary independent random variables X_1, X_2.

Remark 5.1.6
The bounds c) and d) on Page 5.1 become now:

c) $\Delta_1(X, Y) \leq \frac{2(m-1)}{m}$;

d) $\Delta_2(X, Y) \leq \sqrt{m-1}$.

Remark 5.1.7
All imbalance measures that are convex-$\cup$ on the set of all probability distributions, equal to 1 for a constant random variable, and equal to 0 for a uniformly distributed random variable are in a sense equally appropriate for measuring the usefulness of an expression (an I/O sum or an I/O difference or a threefold sum or an I/K/O combination) for use in an attack. Imbalance does not provide an absolute goodness of such an expression but rather a relative goodness that allows one to compare different expressions. Thus, in a particular situation, it is enough to prove that the piling-up hypothesis holds for one imbalance measure and then to use that imbalance measure to estimate the probability of success of the attack. Note that this does not imply that if the hypothesis holds for one imbalance measure, then it also holds for the other imbalance measures.

5.2 Validity of the Hypothesis for the Imbalance I_2^2

In the remainder of the chapter, we prove that the piling-up hypothesis holds for the imbalance measure I_2^2 (that is, the square of the Euclidian

imbalance I_2). The imbalance I_2^2 is convex-$\cup$ because it is the square of a non-negative convex-$\cup$ function. We have

$$I_2^2(\boldsymbol{p}) = \frac{m}{m-1}\sum_{i=0}^{m-1} p_i^2 - \frac{1}{m-1} = \frac{m}{m-1}\sum_{i=0}^{m-1}\left(p_i - \frac{1}{m}\right)^2. \quad (5.11)$$

Sometimes, one of the above forms is more convenient than the other. The corresponding piling-up factor is

$$\Delta_2^2(X_1, X_2) = \frac{I_2^2(X_1 * X_2)}{I_2^2(X_1) I_2^2(X_2)}. \quad (5.12)$$

Matsui's Piling-up Lemma says that $\Delta_2^2(X_1, X_2) = 1$ for all independent random variables X_1, X_2 when $m = 2$. Harpes has shown [16] that the same holds for $m = 3$. For that reason, we could consider only $m \geq 4$; this we will do, but only from that point where by this restriction we can avoid laborious case distinctions.

5.2.1 Averaging Over One Random Variable

In this subsection, we consider a random variable X_1 with probability distribution $\boldsymbol{p}$, a second probability vector $\boldsymbol{q}$, and the average of $I_2^2(X_1 * X_2)$ over all random variables X_2 independent of X_1 that have $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$ as their probability distribution. We show that for any $m \geq 2$ and any group operation $*$ on $\mathbb{Z}_m$, this average is always equal to $I_2^2(\boldsymbol{p}) I_2^2(\boldsymbol{q})$.

Definition 5.2.1
For any probability vector $\boldsymbol{p}$ of length m and any permutation π of $\{0, \ldots, m-1\}$, let $\boldsymbol{p}_\pi$ be the probability vector defined by $(\boldsymbol{p}_\pi)_i := p_{\pi(i)}$. Then $I_2^2(\boldsymbol{p}_\pi) = I_2^2(\boldsymbol{p})$.

Remark 5.2.2
If X_1 and X_2 are independent and X_1 (resp. X_2) has $\boldsymbol{p}$ (resp. $\boldsymbol{q}$) as its probability distribution where $p_i = P_{X_1}(i)$ and $q_i = P_{X_2}(i)$, then the probability that $X_1 * X_2$ be equal to i is $P_{X_1 * X_2}(i) = \sum_{j=0}^{m-1} P_{X_1}(j) P_{X_2}(j^{-1} * i) = \sum_{j=0}^{m-1} p_j q_{j^{-1} * i}$, where j^{-1} is the inverse of j under the group operation $*$. Thus, the imbalance of $X_1 * X_2$ is given by

$$I_2^2(X_1 * X_2) = \frac{m}{m-1}\sum_{i=0}^{m-1}\left(\sum_{j=0}^{m-1} p_j q_{j^{-1} * i} - \frac{1}{m}\right)^2. \quad (5.13)$$

Now the number of random variables that have probability distribution $\boldsymbol{q}$ is the same as the number of random variables that have probability distribution $\boldsymbol{q}_\pi$, and this holds for every permutation $\boldsymbol{q}_\pi$ of $\boldsymbol{q}$. Thus, in order to calculate the average of (5.13) over all random variables X_2 that are independent of X_1 and have $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$ as their probability distribution, it is enough, for each permutation of $\boldsymbol{q}$ (including the identity permutation), to consider a single random variable X_2 having that probability distribution. Thus, we must average the right side of (5.13) over all probability vectors $\boldsymbol{q}_\pi$ with $\pi \in S_m$, where S_m is the symmetric group of order m. Nevertheless, we shall continue talking about "the average over all random variables X_2 that are independent of X_1 and have $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$ as probability distribution".

We seek the value of

$$E[I_2^2(X_1 * X_2)] = \frac{1}{m!} \sum_{\pi \in S_m} \frac{m}{m-1} \sum_{i=0}^{m-1} \left(\sum_{j=0}^{m-1} p_j q_{\pi(j^{-1} * i)} - \frac{1}{m} \right)^2, \quad (5.14)$$

where we will later discuss the validity of using expected value notation here. The product of the imbalance of X_1 and X_2 is

$$I_2^2(X_1) \cdot I_2^2(X_2) = \left(\frac{m}{m-1} \right)^2 \left(\sum_{j=0}^{m-1} p_j^2 - \frac{1}{m} \right) \left(\sum_{i=0}^{m-1} q_i^2 - \frac{1}{m} \right). \quad (5.15)$$

We will prove that (5.14) is identical to (5.15). We need the following Lemmata.

Lemma 5.2.3
For any m-ary probability vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ and any $i \in \mathbb{Z}_m$,

$$\sum_{\pi} \left(\sum_{j=0}^{m-1} p_j q_{\pi(j^{-1} * i)} - \frac{1}{m} \right)^2 = \sum_{\pi} \left(\sum_{j=0}^{m-1} p_j q_{\pi(j)} - \frac{1}{m} \right)^2 \quad (5.16)$$

where $\sum_\pi$ is the sum over all permutations π of $\{0, \ldots, m-1\}$.
Proof:

For any permutation π, define the set of pairs $M_\pi := \{(j, \pi(j)) | \, 0 \leq j \leq m-1\}$, and consider the functions $h : \mathbb{Z}_m \to \mathbb{Z}_m, h(j) := j^{-1} * i$, and $f : S_m \to S_m, f(\pi) := \pi \circ h$. Both h and f are invertible and thus the function $F : M_\pi \mapsto M_{f(\pi)}$ is also invertible. Explicitely, $M_{f(\pi)} = \{(j, \pi(j^{-1} * i)) | \, 0 \leq j \leq m-1\}$. Thus, the set of all $M_{f(\pi)}$, $\pi \in S_m$, is equal to the set of all M_π, $\pi \in S_m$. Now consider the pairs of indices of the products pq in (5.16). On the left, one sums over all $M_{f(\pi)}, \pi \in S_m$,

and over all pairs of indices in $M_{f(\pi)}$, while on the right, one sums over all $M_\pi, \pi \in S_m$, and over all pairs of indices in M_π. Thus, equality must hold. □

Lemma 5.2.4
Let $a_0, a_1, \ldots, a_{m-1}$ be real numbers. Then, for any $0 \le k \ne l \le m-1$,

$$\sum_\pi a_{\pi(k)} = (m-1)! \sum_{i=0}^{m-1} a_i \quad \text{and} \quad \sum_\pi a_{\pi(k)} a_{\pi(l)} = (m-2)! \sum_{0 \le i \ne j \le m-1} a_i a_j,$$

where $\sum_\pi$ is the sum over all permutations in S_m. (There are analogous identities for more than two indices.)

Proof:

For any $0 \le k \ne l \le m-1$ and any $0 \le i \ne j \le m-1$, there are $(m-1)!$ permutations that map k to i and there are $(m-2)!$ permutations that map k to i and l to j. □

We are now ready to prove the following theorem.

Theorem 5.2.5
Let $m \ge 2$ be an integer and I_2^2 be defined by (5.11); let $\boldsymbol{q}$ be a probability vector with m elements and X_1 be a random variable with values on $\langle \mathbb{Z}_m, * \rangle$. Then the average of $I_2^2(X_1 * X_2)$ over all random variables X_2 independent of X_1 that have $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$ as their probability distribution is

$$E[I_2^2(X_1 * X_2)] = I_2^2(X_1) \cdot I_2^2(\boldsymbol{q}). \tag{5.17}$$

Proof:

Let $\boldsymbol{p}$ be the probability distribution of X_1. Then the aforementioned average is

$$E[I_2^2(X_1 * X_2)] = \frac{1}{m!} \sum_\pi \frac{m}{m-1} \sum_{i=0}^{m-1} \left(\sum_{j=0}^{m-1} p_j q_{\pi(j^{-1} * i)} - \frac{1}{m} \right)^2 \text{(by (5.14))}$$

$$= \sum_{i=0}^{m-1} \frac{1}{m!} \frac{m}{m-1} \sum_\pi \left(\sum_{j=0}^{m-1} p_j q_{\pi(j)} - \frac{1}{m} \right)^2 \qquad \text{(by Lemma 5.2.3)}$$

$$= \frac{1}{(m-1)!} \frac{m}{m-1} \sum_\pi \left(\sum_{j=0}^{m-1} p_j q_{\pi(j)} - \frac{1}{m} \right)^2$$

$$= \frac{1}{(m-1)!} \frac{m}{m-1} \sum_\pi \left(\left(\sum_{j=0}^{m-1} p_j q_{\pi(j)} \right)^2 - \frac{2}{m} \sum_{j=0}^{m-1} p_j q_{\pi(j)} + \frac{1}{m^2} \right)$$

$$
\begin{aligned}
&= \frac{1}{(m-1)!}\frac{m}{m-1}\sum_{0\le i,j\le m-1} p_i p_j \sum_{\pi} q_{\pi(i)} q_{\pi(j)} \\
&\quad - \frac{1}{(m-1)!}\frac{2}{m-1}\sum_{j=0}^{m-1} p_j \sum_{\pi} q_{\pi(j)} + \frac{1}{(m-1)!}\frac{m}{m-1}\sum_{\pi}\frac{1}{m^2} \\
&= \frac{1}{(m-1)!}\frac{m}{m-1}\sum_{j=0}^{m-1} p_j^2 \sum_{\pi} q_{\pi(j)}^2 \\
&\quad + \frac{1}{(m-1)!}\frac{m}{m-1}\sum_{0\le i\ne j\le m-1} p_i p_j \sum_{\pi} q_{\pi(i)} q_{\pi(j)} \\
&\quad - \frac{1}{(m-1)!}\frac{2}{m-1}\sum_{j=0}^{m-1} p_j \sum_{\pi} q_{\pi(j)} + \frac{1}{(m-1)!}\frac{m}{m-1}\sum_{\pi}\frac{1}{m^2} \\
&= \frac{1}{(m-1)!}\frac{m}{m-1}\sum_{j=0}^{m-1} p_j^2 \,(m-1)! \sum_{i=0}^{m-1} q_i^2 \qquad \text{(by Lemma 5.2.4)} \\
&\quad + \frac{1}{(m-1)!}\frac{m}{m-1}\sum_{0\le i\ne j\le m-1} p_i p_j \,(m-2)! \sum_{0\le k\ne l\le m-1} q_k q_l \\
&\quad - \frac{1}{(m-1)!}\frac{2}{m-1}\sum_{j=0}^{m-1} p_j \,(m-1)! \sum_{i=0}^{m-1} q_i + \frac{1}{(m-1)!}\frac{m}{m-1}\frac{m!}{m^2} \\
&= \frac{m}{m-1}\sum_{j=0}^{m-1} p_j^2 \sum_{i=0}^{m-1} q_i^2 + \frac{1}{m-1}\frac{m}{m-1}\sum_{0\le i\ne j\le m-1} p_i p_j \sum_{0\le k\ne l\le m-1} q_k q_l \\
&\quad - \frac{2}{m-1} + \frac{1}{m-1} \\
&= \frac{m}{m-1}\sum_{j=0}^{m-1} p_j^2 \sum_{i=0}^{m-1} q_i^2 + \frac{1}{m-1}\frac{m}{m-1}\left(1-\sum_{j=0}^{m-1} p_j^2\right)\left(1-\sum_{i=0}^{m-1} q_i^2\right) - \frac{1}{m-1} \\
&= \frac{m}{m-1}\left(\sum_{j=0}^{m-1} p_j^2 \sum_{i=0}^{m-1} q_i^2 + \frac{1}{m-1} - \frac{1}{m-1}\sum_{j=0}^{m-1} p_j^2 - \frac{1}{m-1}\sum_{i=0}^{m-1} q_i^2 \right.\\
&\qquad \left. + \frac{1}{m-1}\sum_{j=0}^{m-1} p_j^2 \sum_{i=0}^{m-1} q_i^2 - \frac{1}{m}\right) \\
&= \left(\frac{m}{m-1}\right)^2\left(\sum_{j=0}^{m-1} p_j^2 \sum_{i=0}^{m-1} q_i^2 - \frac{1}{m}\sum_{j=0}^{m-1} p_j^2 - \frac{1}{m}\sum_{i=0}^{m-1} q_i^2 + \frac{1}{m^2}\right)
\end{aligned}
$$

$$= \left(\frac{m}{m-1}\right)^2 \left(\sum_{j=0}^{m-1} p_j^2 - \frac{1}{m}\right)\left(\sum_{i=0}^{m-1} q_i^2 - \frac{1}{m}\right)$$
$$= I_2^2(\boldsymbol{p}) \cdot I_2^2(\boldsymbol{q}) = I_2^2(X_1) \cdot I_2^2(\boldsymbol{q}). \qquad \square$$

Remark 5.2.6
We wrote an expected value in (5.17) for the following reason: we can define a random variable Y whose values are m-ary random variables and which is uniformly distributed on the set of all random variables that have $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$ as their probability distribution. That is, if $Y = y$ then y is an m-ary random variable that has $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$ as its probability distribution. Then an expression like $I_2^2(X_1 * y)$ has a meaning, and the theorem says that $E_Y[I_2^2(X_1 * Y)] = I_2^2(X_1) I_2^2(\boldsymbol{q})$. (The subscript Y in E_Y means that the expectation is taken on Y.)

Theorem 5.2.5 still does not provide an m-ary analogue to Matsui's Piling-up Lemma since we do not yet know how far the values $I_2^2(X_1 * X_2)$ are in general from $I_2^2(X_1) \cdot I_2^2(X_2)$. In order to find this, we reuse our model of a random variable Y uniformly distributed on the set of all random variables having $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$ as probability distribution, but this time we look at the variance rather than the expected value.

We first consider $E_Y[I_2^4(X_1 * Y)]$. We have, for X_1 and X_2 independent,

$$I_2^4(X_1 * X_2) = \left(\frac{m}{m-1}\right)^2 \left(\sum_{i=0}^{m-1}\left(\sum_{j=0}^{m-1} p_j q_{j^{-1}*i} - \frac{1}{m}\right)^2\right)^2$$
$$= \left(\frac{m}{m-1}\right)^2 \sum_{i=0}^{m-1}\sum_{k=0}^{m-1}\left(\sum_{j=0}^{m-1} p_j q_{j^{-1}*i} - \frac{1}{m}\right)^2 \left(\sum_{l=0}^{m-1} p_l q_{l^{-1}*k} - \frac{1}{m}\right)^2.$$

Because k^{-1} runs over $\mathbb{Z}_m$ as k runs over $\mathbb{Z}_m$, we can replace $q_{l^{-1}*k}$ by $q_{l^{-1}*k^{-1}}$. Then the expected value is

$$E_Y[I_2^4(X_1 * Y)] = \left(\frac{m}{m-1}\right)^2 \frac{1}{m!} \times \tag{5.18}$$
$$\sum_{i=0}^{m-1}\sum_{k=0}^{m-1}\sum_{\pi \in S_m}\left(\sum_{j=0}^{m-1} p_j q_{\pi(j^{-1}*i)} - \frac{1}{m}\right)^2 \left(\sum_{l=0}^{m-1} p_l q_{\pi(l^{-1}*k^{-1})} - \frac{1}{m}\right)^2.$$

Hereafter in this chapter, we omit the group operation $*$ in subscripts where no misunderstanding can occur. The following Lemma simplifies the indices in (5.18).

Lemma 5.2.7
For any m-ary probability distributions $\boldsymbol{p}$ *and* $\boldsymbol{q}$,

$$E_Y[I_2^4(X_1 * Y)] = \tag{5.19}$$
$$\left(\frac{m}{m-1}\right)^2 \frac{1}{m!} \sum_{i=0}^{m-1} \sum_{k=0}^{m-1} \sum_{\pi \in S_m} \left(\sum_{j=0}^{m-1} p_j q_{\pi(j)} - \frac{1}{m} \right)^2 \left(\sum_{l=0}^{m-1} p_l q_{\pi(ikl)} - \frac{1}{m} \right)^2 .$$

Proof:

We show that the expressions inside the sum over k are equal in (5.18) and in (5.19). We use the same technique as in Lemma 5.2.3. Let i and k be fixed. For any permutations π and $\tilde{\pi}$, define the sets of four-tuples $M_{\pi,\tilde{\pi}} := \{(j, \pi(j), l, \tilde{\pi}(l)) \,|\, 0 \leq j, l \leq m-1\}$ and consider the functions $h_1, h_2 : \mathbb{Z}_m \to \mathbb{Z}_m, h_1(j) := j^{-1} * i, h_2(l) := l^{-1} * k^{-1}$, and $f_1, f_2 : S_m \to S_m, f_1(\pi) := \pi \circ h_1, f_2(\pi) := \pi \circ h_2$. The four indices of the p's and q's in (5.18) run over all $\pi \in S_m$ and all elements of $M_{f_1(\pi),f_2(\pi)}$. Since $(f_2 \circ f_1^{-1})(\pi)(l) = (\pi \circ h_1^{-1} \circ h_2)(l) = \pi(i * k * l)$, the four indices of the p's and q's in (5.19) run over all $\pi \in S_m$ and all elements of $M_{\pi,(f_2 \circ f_1^{-1})(\pi)}$. But because h_1, h_2, f_1 and f_2 are invertible, $\{M_{f_1(\pi),f_2(\pi)} | \pi \in S_m\} = \{M_{\pi,(f_2 \circ f_1^{-1})(\pi)} | \pi \in S_m\}$. Thus, both expressions inside the sum over k must be the same. □

The averages we are taking for the moment are not symmetrical in X_1 and X_2, respectively in $\boldsymbol{p}$ and $\boldsymbol{q}$. Thus, we provisionally distinguish between probability vectors we write with $\boldsymbol{p}$ and such we write with $\boldsymbol{q}$. The following definitions will help keeping the formulas shorter.

Definition 5.2.8
For a probability vector $\boldsymbol{q}$, define

$$Q_i := \sum_r q_r^i; \qquad Q_{ij} := \sum_{r \neq s} q_r^i q_s^j;$$
$$Q_{ijk} := \sum_{r,s,t \text{ all different}} q_r^i q_s^j q_t^k; \qquad Q_{ijkl} := \sum_{r,s,t,u \text{ all different}} q_r^i q_s^j q_t^k q_u^l;$$

All Q's with more than one index can be written as functions of the Q_i. We need the following:

Lemma 5.2.9
1. $Q_{11} = 1 - Q_2$;
2. $Q_{21} = Q_2 - Q_3$;
3. $Q_{31} = Q_3 - Q_4$;

4. $Q_{22} = Q_2^2 - Q_4$;

5. $Q_{111} = 1 - 3Q_2 + 2Q_3$;

6. $Q_{211} = Q_2 - 2Q_3 + 2Q_4 - Q_2^2$;

7. $Q_{1111} = 1 - 6Q_2 + 8Q_3 + 3Q_2^2 - 6Q_4$. □

We will also use the following abbreviations:

Definition 5.2.10
For an m-ary probability vector $\boldsymbol{p}$ and for any $j \in \mathbb{Z}_m$, define

- $P_r := \sum_n p_n^r$;
- $P_{rs} := P_{rs}(j) := \sum_n p_n^r p_{jn}^s$;
- $P_{rst} := P_{rst}(j) := \sum_n p_n^r p_{jn}^s p_{j^2n}^t$;

Notice that the P's are not defined the same way as the Q's. Moreover, they depend on j. They have the following properties, that we state without their simple proofs.

Lemma 5.2.11
Let e_* be the neutral element, j be an element of order two, and i be any element, of $\langle \mathbb{Z}_m, * \rangle$. Then

1. $P_{rs}(e_*) = P_{r+s}$ *and* $P_{rst}(e_*) = P_{r+s+t}$;

2. $P_{rs}(i^{-1}) = P_{sr}(i)$;

3. $P_{rs}(j) = P_{rs}(j^{-1}) = P_{sr}(j)$ *and* $P_{rst}(j) = P_{r+t,s}(j)$. □

If we fix i and k, the inner summation including only π, j and l in (5.19) can be decomposed into nine parts as follows:

$$\sum_{\pi \in S_m} \left(\sum_{j=0}^{m-1} p_j q_{\pi(j)} - \frac{1}{m} \right)^2 \left(\sum_{l=0}^{m-1} p_l q_{\pi(ikl)} - \frac{1}{m} \right)^2 =$$

$$\textcircled{1}(ik) - \frac{2}{m}\textcircled{2}(ik) + \frac{1}{m^2}\textcircled{3}(ik) - \frac{2}{m}\textcircled{4}(ik) + \frac{4}{m^2}\textcircled{5}(ik) - \frac{2}{m^3}\textcircled{6}(ik)$$
$$+\frac{1}{m^2}\textcircled{7}(ik) - \frac{2}{m^3}\textcircled{8}(ik) + \frac{1}{m^4}\textcircled{9}(ik) \qquad (5.20)$$

where

$$
\begin{aligned}
①(ik) &= \sum_{\pi}\Big(\sum_{j} p_j q_{\pi(j)}\Big)^2\Big(\sum_{l} p_l q_{\pi(ikl)}\Big)^2,\\
②(ik) &= \sum_{\pi}\sum_{j} p_j q_{\pi(j)}\Big(\sum_{l} p_l q_{\pi(ikl)}\Big)^2,\\
③(ik) &= \sum_{\pi}\Big(\sum_{l} p_l q_{\pi(ikl)}\Big)^2,\\
④(ik) &= \sum_{\pi}\Big(\sum_{j} p_j q_{\pi(j)}\Big)^2\sum_{l} p_l q_{\pi(ikl)},\\
⑤(ik) &= \sum_{\pi}\sum_{j} p_j q_{\pi(j)}\sum_{l} p_l q_{\pi(ikl)},\\
⑥(ik) &= \sum_{\pi}\sum_{l} p_l q_{\pi(ikl)},\\
⑦(ik) &= \sum_{\pi}\Big(\sum_{j} p_j q_{\pi(j)}\Big)^2,\\
⑧(ik) &= \sum_{\pi}\sum_{j} p_j q_{\pi(j)},\\
⑨(ik) &= \sum_{\pi} 1.
\end{aligned}
$$

We show in the appendix to this chapter (Proposition 5.A.1) that

$$
\begin{aligned}
①(ik) &= (m-1)!Q_4P_{22}(ik)\\
&+ (m-2)!Q_{31}(2P_{12}(ik)+2P_{21}(ik)-4P_{22}(ik))\\
&+ (m-2)!Q_{22}(P_2^2+2P_{11}^2(ik)-3P_{22}(ik))\\
&+ (m-3)!Q_{211}(2P_2+4P_{11}(ik)-6P_{12}(ik)-6P_{21}(ik)\\
&\qquad +12P_{22}(ik)-2P_2^2-4P_{11}^2(ik))\\
&+ (m-4)!Q_{1111}(1-2P_2-4P_{11}(ik)+4P_{12}(ik)+4P_{21}(ik)\\
&\qquad +P_2^2+2P_{11}^2(ik)-6P_{22}(ik))\\
②(ik) &= (m-1)!Q_3P_{21}(ik)+(m-2)!Q_{21}(P_2+2P_{11}(ik)-3P_{21}(ik))\\
&+ (m-3)!Q_{111}(1-P_2-2P_{11}(ik)+2P_{21}(ik))\\
③(ik) &= (m-1)!Q_2P_2+(m-2)!Q_{11}(1-P_2)\\
④(ik) &= (m-1)!Q_3P_{12}(ik)+(m-2)!Q_{21}(P_2+2P_{11}(ik)-3P_{12}(ik))\\
&+ (m-3)!Q_{111}(1-P_2-2P_{11}(ik)+2P_{12}(ik))\\
⑤(ik) &= (m-1)!Q_2P_{11}(ik)+(m-2)!Q_{11}(1-P_{11}(ik))
\end{aligned}
$$

$$\begin{aligned} ⑥(ik) &= (m-1)! \\ ⑦(ik) &= ③(ik) \\ ⑧(ik) &= ⑥(ik) \\ ⑨(ik) &= m! \end{aligned}$$

for all i, k. At this point we restrict our study to the interesting cases $m \geq 4$. (Indeed, what happens to $①(ik)$ if $m < 4$? Factorials of negative integers are not defined. In fact, if one does the calculation for $m < 4$, then one finds that the term beginning with $(m-4)!$ does not occur if $m < 4$ and that the one beginning with $(m-3)!$ does not occur if $m < 3$. The same holds for $②(ik)$ and $④(ik)$.) In the next step we sum the above result over all i and k in $\mathbb{Z}_m$. We use the fact that

$$\begin{aligned} \sum_{i,k} P_{rs}(ik) &= \sum_{i,k} \sum_{n} p_n^r p_{ikn}^s = \sum_n p_n^r \sum_i \sum_k p_{ikn}^s = \sum_n p_n^r \sum_i P_s \\ &= \sum_n p_n^r m\, P_s = m\, P_r P_s. \end{aligned} \tag{5.21}$$

In particular, $\sum_{i,k} P_{1s}(ik) = mP_s$. Hence, if we sum each of the nine parts over i and k and make use of Proposition 5.2.9, we get

$$\begin{aligned} \sum_{i,k} ①(ik) &= P_2^2 (m-4)! \Big[2m^3 Q_4 - 8m^2 Q_3 - (2m^3 - 12m^2) Q_2 \\ &\quad + (m^4 - 6m^3 + 6m^2) Q_2^2 + (m^2 - 6m) \Big] \\ &+ P_2 (m-4)! \Big[-4m^2 Q_4 + 16m Q_3 + (2m^3 - 6m^2 - 12m) Q_2 \\ &\quad - (2m^3 - 12m^2 + 12m) Q_2^2 - (2m^2 - 8m) \Big] \\ &+ \sum_{i,k} P_{11}^2(ik)\, (m-4)! \Big[-(2m^2 - 2m) Q_4 + (8m - 8) Q_3 \\ &\quad - 4m Q_2 + (2m^2 - 6m + 6) Q_2^2 + 2 \Big] \\ &+ (m-4)! \Big[2m^2 Q_4 - 8m Q_3 - (2m^2 - 12m) Q_2 \\ &\quad - m^2 Q_2^2 + (m^2 - 4m) \Big] \end{aligned}$$

$$\begin{aligned}
\sum_{i,k} ②(ik) &= m(m-2)!\,(1-Q_2-P_2+m\,Q_2P_2)\\
\sum_{i,k} ③(ik) &= m^2(m-2)!\,(1-Q_2-P_2+m\,Q_2P_2)\\
\sum_{i,k} ④(ik) &= \sum_{i,k} ②(ik)\\
\sum_{i,k} ⑤(ik) &= m!\\
\sum_{i,k} ⑥(ik) &= m^2(m-1)!\\
\sum_{i,k} ⑦(ik) &= \sum_{i,k} ③(ik)\\
\sum_{i,k} ⑧(ik) &= \sum_{i,k} ⑥(ik)\\
\sum_{i,k} ⑨(ik) &= m^2m!
\end{aligned}$$

Then

$$\begin{aligned}
&\sum_{i,k}\Big(①(ik)-\frac{2}{m}②(ik)+\frac{1}{m^2}③(ik)-\frac{2}{m}④(ik)+\frac{4}{m^2}⑤(ik)-\frac{2}{m^3}⑥(ik)\\
&\qquad+\frac{1}{m^2}⑦(ik)-\frac{2}{m^3}⑧(ik)+\frac{1}{m^4}⑨(ik)\Big)\\
&= P_2^2(m-4)!\Big[2m^3Q_4-8m^2Q_3-(2m^3-12m^2)Q_2\\
&\qquad+(m^4-6m^3+6m^2)Q_2^2+(m^2-6m)\Big]\\
&+ P_2(m-4)!\Big[-4m^2Q_4+16mQ_3+(4m^2-24m)Q_2\\
&\qquad-(2m^3-12m^2+12m)Q_2^2-(2m-12)\Big]\\
&+ \sum_{i,k} P_{11}^2(ik)\,(m-4)!\Big[-(2m^2-2m)Q_4+(8m-8)Q_3\\
&\qquad-4mQ_2+(2m^2-6m+6)Q_2^2+2\Big]\\
&+ (m-4)!\Big[2m^2Q_4-8mQ_3-(2m+12)Q_2-m^2Q_2^2-\frac{m+6}{m}\Big].
\end{aligned}$$

Finally, multiplying by $(\frac{m}{m-1})^2\,\frac{1}{m!}$ gives the desired expected value so we have:

Proposition 5.2.12
Let $m \geq 4$ be an integer and I_2^2 be defined by (5.11); let $\boldsymbol{p}$ and $\boldsymbol{q}$ be probability vectors with m components and X_1 be a random variable with values on $\langle \mathbb{Z}_m, * \rangle$ and probability distribution $\boldsymbol{p}$. Then the average of $I_2^4(X_1 * X_2)$ over all random variables X_2 independent of X_1 and having $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$ as their probability distribution is

$$\begin{aligned} E[I_2^4(X_1 * X_2)] &= \frac{1}{(m-1)^2(m-2)^2(m-3)^2}\Big\{ \\ &\quad P_2^2\Big[2m^3Q_4 - 8m^2Q_3 - (2m^3 - 12m^2)Q_2 \\ &\qquad + (m^4 - 6m^3 + 6m^2)Q_2^2 + (m^2 - 6m)\Big] \\ &\quad + P_2\Big[-4m^2Q_4 + 16mQ_3 + (4m^2 - 24m)Q_2 \\ &\qquad - (2m^3 - 12m^2 + 12m)Q_2^2 - (2m - 12)\Big] \\ &\quad + \sum_{i,k} P_{11}^2(ik)\Big[-(2m^2 - 2m)Q_4 + (8m - 8)Q_3 \\ &\qquad - 4mQ_2 + (2m^2 - 6m + 6)Q_2^2 + 2\Big] \\ &\quad + \Big[2m^2Q_4 - 8mQ_3 - (2m + 12)Q_2 - m^2Q_2^2 - \frac{m+6}{m}\Big]\Big\}. \end{aligned} \tag{5.22}$$

□

However, this is not yet fully satisfying because $\sum_{i,k} P_{11}^2(ik)$ still depends on the probability vector $\boldsymbol{p}$. Thus, if we take a permutation of $\boldsymbol{p}$, then $E[I_2^4(X_1 * X_2)]$ changes. Averaging over all permutations of $\boldsymbol{p}$ will allow us to get rid of this inconvenience and to obtain an expression symmetrical in P and Q.

5.2.2 Averaging Over Two Random Variables

In the preceding subsection, the random variable X_1 was fixed. We now want to average further the expressions found above, namely over all X_1 having $\boldsymbol{p}$ or a permutation of $\boldsymbol{p}$ as probability distribution. We do this by considering a random variable X which is uniformly distributed over the set of all random variables having $\boldsymbol{p}$ or a permutation of $\boldsymbol{p}$ as probability distribution. This means that in the end we obtain averages over all pairs of independent random variables (X_1, X_2) such that X_1 has $\boldsymbol{p}$ or a permutation of $\boldsymbol{p}$ as its probability distribution and X_2 has $\boldsymbol{q}$ or a

permutation of $\boldsymbol{q}$ as its probability distribution. We seek expressions for such an average of $I_2^2(X_1 * X_2)$ and of $I_2^4(X_1 * X_2)$.

The first of these averages is easy to obtain.

Theorem 5.2.13
Let $m \geq 4$ be an integer and I_2^2 be defined by (5.11); let $\boldsymbol{p}$ and $\boldsymbol{q}$ be probability vectors with m elements. Then the average of $I_2^2(X_1 * X_2)$ over all pairs of independent random variables (X_1, X_2) such that X_1 has $\boldsymbol{p}$ or a permutation of $\boldsymbol{p}$ as probability distribution and X_2 has $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$ as probability distribution, is

$$E[I_2^2(X_1 * X_2)] = I_2^2(\boldsymbol{p}) \cdot I_2^2(\boldsymbol{q}). \tag{5.23}$$

Proof:

This follows immediately from Theorem 5.2.5 since what one must do is to average the right side of (5.17) over all such random variables X_1. But $I_2^2(\boldsymbol{q})$ is a constant and $I_2^2(X_1) = I_2^2(\boldsymbol{p})$ for all the random variables X_1 considered. □

To obtain the average of $I_2^4(X_1 * X_2)$, we must average the right side of (5.22) over all random variables X_1 having $\boldsymbol{p}$ or a permutation of $\boldsymbol{p}$ as probability distribution. Again, by the same argument as for the average over X_2, it is enough to consider only the average over all permutations of $\boldsymbol{p}$. Since P_i remains constant when one permutes the components of $\boldsymbol{p}$, the only part of the expression that will be affected is $\sum_{i,k} P_{11}^2(ik)$. Averaging $\sum_{i,k} P_{11}^2(ik)$ over all probability vectors obtained from a permutation of $\boldsymbol{p}$ (including the identity permutation) gives

$$\begin{aligned} \frac{1}{m!} \sum_{\pi} \sum_{i,k} P_{11}^2(ik) &= \frac{1}{m!} \sum_{\pi} \sum_{i,k} \left(\sum_{n} p_{\pi(n)} p_{\pi(ikn)} \right)^2 \\ &= \frac{1}{m!} \sum_{i,k} \sum_{l,n} \sum_{\pi} p_{\pi(l)} p_{\pi(n)} p_{\pi(ikl)} p_{\pi(ikn)}. \end{aligned} \tag{5.24}$$

Here we drop the distinction made between the probability distributions written as $\boldsymbol{p}$ as those written as $\boldsymbol{q}$. Further considerations are based on the following Lemma.

Lemma 5.2.14
Let e_* be the neutral element and j be any element of $\langle \mathbb{Z}_m, * \rangle$, and let $\boldsymbol{p}$ be an m-ary probability vector. Then

$$\sum_{l=0}^{m-1} \sum_{n=0}^{m-1} \sum_{\pi \in S_m} p_{\pi(l)} p_{\pi(n)} p_{\pi(jl)} p_{\pi(jn)} = \tag{5.25}$$

$$\begin{cases} m!\, P_2^2, & j = e_* \\ \frac{m!}{(m-1)(m-3)} \left(1 - 6P_2 + 8P_3 + (2m-3)P_2^2 - 2mP_4 \right), & \mathrm{ord}(j) = 2 \\ \frac{m!}{(m-1)(m-2)} \left(1 - 4P_2 + 4P_3 + (m-1)P_2^2 - mP_4 \right), & \mathrm{ord}(j) > 2. \end{cases}$$

Proof:
If $j = e_*$, then

$$\sum_{l=0}^{m-1} \sum_{n=0}^{m-1} \sum_{\pi \in S_m} p_{\pi(l)} p_{\pi(n)} p_{\pi(jl)} p_{\pi(jn)} = \sum_{\pi \in S_m} \sum_{l=0}^{m-1} \sum_{n=0}^{m-1} p_{\pi(l)}^2 p_{\pi(n)}^2$$

$$= \sum_{\pi \in S_m} \left(\sum_{l=0}^{m-1} p_{\pi(l)}^2 \right)^2 = m!\, P_2^2.$$

If $\mathrm{ord}(j) = 2$, then we split the sum into partial sums where $l = n$; $l = jn$; and where l, n, jl, and jn are all different. If $l = n$ (which happens for m pairs (l, n)), then, by Lemmata 5.2.4 and 5.2.9,

$$\sum_{\pi \in S_m} p_{\pi(l)} p_{\pi(n)} p_{\pi(jl)} p_{\pi(jn)} = \sum_{\pi \in S_m} p_{\pi(l)}^2 p_{\pi(jl)}^2$$

$$= (m-2)! \sum_{0 \le i \ne j \le m-1} p_i^2 p_j^2 = (m-2)!(P_2^2 - P_4).$$

If $l = jn$, which also occurs for m pairs (l, n), then $n = jl$ and we have

$$\sum_{\pi \in S_m} p_{\pi(l)} p_{\pi(n)} p_{\pi(jl)} p_{\pi(jn)} = \sum_{\pi \in S_m} p_{\pi(l)}^2 p_{\pi(n)}^2 = (m-2)!(P_2^2 - P_4).$$

And if l, n, jl and jn are all different, which occurs for the remaining $m(m-2)$ pairs (l,n), then, by Lemmata 5.2.4 and 5.2.9,

$$\begin{aligned}\sum_{\pi\in S_m} p_{\pi(l)}p_{\pi(n)}p_{\pi(jl)}p_{\pi(jn)} &= (m-4)! \sum_{r,s,t,u \text{ all different}} p_r p_s p_s p_u \\ &= (m-4)!(1-6P_2+8P_3+3P_2^2-6P_4).\end{aligned}$$

Now summing over all pairs (l,n) gives

$$\begin{aligned}&\sum_{l,n}\sum_{\pi\in S_m} p_{\pi(l)}p_{\pi(n)}p_{\pi(jl)}p_{\pi(jn)} \\ &= 2m(m-2)!(P_2^2-P_4) \\ &\quad +m(m-2)(m-4)!(1-6P_2+8P_3+3P_2^2-6P_4) \\ &= \frac{m!}{(m-1)(m-3)}\Big(1-6P_2+8P_3+(2m-3)P_2^2-2mP_4\Big).\end{aligned}$$

The proof for the case $\mathrm{ord}(j) > 2$ is similar. The sum over all pairs (l,n) is split into partial sums over pairs where: $l=n$; $n=jl$; $l=jn$ (each time m pairs); l, n, jl, jn all different ($m(m-3)$ pairs). The first partial sum gives $(m-2)!(P_2^2-P_4)$, the second and the third one $(m-3)!(P_2-2P_3+2P_4-P_2^2)$, and the last one $(m-4)!(1-6P_2+8P_3+3P_2^2-6P_4)$.□

Now the multiset $\{ik \,|\, i,k \in \mathbb{Z}_m\}$ contains each element of $\langle \mathbb{Z}_m, * \rangle$ m times. Let n be the number of elements of order two in $\langle \mathbb{Z}_m, * \rangle$. Then, by Lemma 5.2.14 and with (5.24),

$$\begin{aligned}&\frac{1}{m!}\sum_{\pi}\sum_{i,k} P_{11}^2(ik) \\ &= mP_2^2 + \frac{mn}{(m-1)(m-3)}\Big(1-6P_2+8P_3+(2m-3)P_2^2-2mP_4\Big) \\ &+ \frac{m(m-n-1)}{(m-1)(m-2)}\Big(1-4P_2+4P_3+(m-1)P_2^2-mP_4\Big). \qquad (5.26)\end{aligned}$$

Armed with this result and Proposition 5.2.12, we can state:

Proposition 5.2.15
Let $m \geq 4$ be an integer, let $*$ be a group operation on $\mathbb{Z}_m$ such that the group $\langle \mathbb{Z}_m, * \rangle$ has n elements of order two and let I_2^2 be defined by (5.11); let $\boldsymbol{p}$ and $\boldsymbol{q}$ be probability vectors with m components and let (X, Y) be a uniformly distributed random variable on the set of all pairs of independent random variables whose first component has $\boldsymbol{p}$ or a permutation of $\boldsymbol{p}$, and whose second component has $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$, as probability distribution. Then

$$E_{XY}[I_2^4(X * Y)] = A_m(\boldsymbol{p}, \boldsymbol{q}) + nB_m(\boldsymbol{p}, \boldsymbol{q})$$

where

$$\begin{aligned}
A_m(\boldsymbol{p}, \boldsymbol{q}) \;=\; & \frac{m}{(m-1)^3(m-2)^2(m-3)} \times \\
& \Big\{ \quad Q_2^2 P_2^2 \Big(m^5 - 4m^4 + 18m^2 - 18m\Big) \\
& + m\Big(m^2 - 4m + 6\Big)\Big[Q_2^2\Big(1 - 2mP_2\Big) + P_2^2\Big(1 - 2mQ_2\Big)\Big] \\
& + 8m\Big(m^2 - 3m + 3\Big)\Big[Q_2^2\Big(P_3 - \frac{m}{4}P_4\Big) + P_2^2\Big(Q_3 - \frac{m}{4}Q_4\Big)\Big] \\
& + 8m\Big[\Big(Q_3 - \frac{m}{4}Q_4\Big)\Big(1 - 2mP_2\Big) + \Big(P_3 - \frac{m}{4}P_4\Big)\Big(1 - 2mQ_2\Big)\Big] \\
& + 32m\Big(m - 1\Big)\Big(Q_3 - \frac{m}{4}Q_4\Big)\Big(P_3 - \frac{m}{4}P_4\Big) \\
& + \frac{1}{m}\Big(m^2 - 4m + 12\Big)\Big(1 - 2mQ_2\Big)\Big(1 - 2mP_2\Big)\Big\}
\end{aligned}$$

and

$$\begin{aligned}
B_m(\boldsymbol{p}, \boldsymbol{q}) \;=\; & \frac{2m^2}{(m-1)^4(m-2)^2(m-3)^2} \\
& \times \Big[Q_2^2\Big(m^2 - 3m + 3\Big) + \Big(1 - 2mQ_2\Big) + 4(m-1)\Big(Q_3 - \frac{m}{4}Q_4\Big)\Big] \\
& \times \Big[P_2^2\Big(m^2 - 3m + 3\Big) + \Big(1 - 2mP_2\Big) + 4(m-1)\Big(P_3 - \frac{m}{4}P_4\Big)\Big].
\end{aligned}$$

Proof:
One averages the right side of (5.22) over all permutations of $\boldsymbol{p}$. With (5.26), the proof consists only of expansions and factorisations. □

Notice that $E_{XY}[I_2^4(X * Y)]$ depends only on the number of elements of order two in $\langle \mathbb{Z}_m, * \rangle$. We are still not finished: this expression is too complicated and what we are really interested in is the variance of $I_2^2(X_1 * X_2)$, not the average of $I_2^4(X_1 * X_2)$.

Lemma 5.2.16
For any m and any probability vectors $\boldsymbol{p}$ and $\boldsymbol{q}$,

$$A_m(\boldsymbol{p}, \boldsymbol{q}) - (m-3)B_m(\boldsymbol{p}, \boldsymbol{q}) = I_2^4(\boldsymbol{p})I_2^4(\boldsymbol{q}).$$

Proof:

Equation (5.11) can also be written as $I_2^2(\boldsymbol{p}) = (mP_2 - 1)/(m-1)$ and similarly for $I_2^2(\boldsymbol{q})$. A comparison between A_m and B_m shows that, for the components where $Q_3 - \frac{m}{4}Q_4$ or $P_3 - \frac{m}{4}P_4$ occur, the coefficients in A_m are always $m-3$ times as large as those in B_m. Thus, $A_m - (m-3)B_m$ is composed of coefficients containing only Q_2, Q_2^2, P_2, and P_2^2. One shows that in $A_m - (m-3)B_m$ the coefficient of $Q_2^2P_2^2$ is $\frac{m^4}{(m-1)^4}$, the coefficients of $Q_2^2(1-2mP_2)$ and of $P_2^2(1-2mQ_2)$ are $\frac{m^2}{(m-1)^4}$, and the coefficient of $(1-2mQ_2)(1-2mP_2)$ is $\frac{1}{(m-1)^4}$. Therefore,

$$\begin{aligned} & A_m(\boldsymbol{p}, \boldsymbol{q}) - (m-3)B_m(\boldsymbol{p}, \boldsymbol{q}) \\ &= \frac{1}{(m-1)^4}\left[m^4Q_2^2P_2^2 + m^2Q_2^2(1-2mP_2) + m^2P_2^2(1-2mQ_2)\right. \\ &\qquad \left. + (1-2mQ_2)(1-2mP_2)\right] \\ &= \frac{1}{(m-1)^4}(mQ_2-1)^2(mP_2-1)^2 \;=\; I_2^4(\boldsymbol{p})I_2^4(\boldsymbol{q}). \end{aligned}$$

□

After the next definition, we will be able to write the variance of $I_2^2(X_1 * X_2)$ on a more compact way.

Definition 5.2.17
For any non-uniform probability vector $\boldsymbol{p}$, define

$$F(\boldsymbol{p}) := \frac{4P_3 - mP_4 - 3P_2^2}{I_2^4(\boldsymbol{p})} = \frac{4P_3 - mP_4 - 3P_2^2}{(mP_2-1)^2}(m-1)^2. \qquad (5.27)$$

At last, we have:

Theorem 5.2.18
Let $m \geq 4$ be an integer, let $*$ be a group operation on $\mathbb{Z}_m$ such that the group $\langle \mathbb{Z}_m, * \rangle$ has n elements of order two and let I_2^2 be defined by (5.11); let $\boldsymbol{p}$ and $\boldsymbol{q}$ be probability vectors with m elements and (X, Y) be a uniformly distributed random variable on the set of all pairs of independent random variables whose first component has $\boldsymbol{p}$ or a permutation of $\boldsymbol{p}$, and whose second component $\boldsymbol{q}$ or a permutation of $\boldsymbol{q}$, as probability distribution. Then

$$\begin{aligned} E_{XY}[\Delta_2^2(X * Y)] &= 1 \quad \text{and} \\ Var_{XY}(\Delta_2^2(X * Y)) &= \frac{2m^2(n+m-3)}{(m-1)^2(m-2)^2(m-3)^2} \times \\ &\quad \Big((m-1) + F(\boldsymbol{p})\Big)\Big((m-1) + F(\boldsymbol{q})\Big). \end{aligned}$$

Proof:

The first equation is a restatement of Theorem 5.2.13. As to the second equation, we first have

$$\begin{aligned} Var(\Delta_2^2) &= E[\Delta_2^4] - E[\Delta_2^2]^2 = \frac{E_{XY}[I_2^4(X * Y)]}{I_2^4(\boldsymbol{p}) I_2^4(\boldsymbol{q})} - 1 \\ &= \frac{A_m(\boldsymbol{p}, \boldsymbol{q}) + nB_m(\boldsymbol{p}, \boldsymbol{q})}{A_m(\boldsymbol{p}, \boldsymbol{q}) - (m-3)B_m(\boldsymbol{p}, \boldsymbol{q})} - 1 \\ &= \frac{(n+m-3)B_m(\boldsymbol{p}, \boldsymbol{q})}{A_m(\boldsymbol{p}, \boldsymbol{q}) - (m-3)B_m(\boldsymbol{p}, \boldsymbol{q})} = \frac{(n+m-3)B_m(\boldsymbol{p}, \boldsymbol{q})}{I_2^4(\boldsymbol{p}) I_2^4(\boldsymbol{q})} \\ &= (n+m-3)B_m(\boldsymbol{p}, \boldsymbol{q}) \frac{(m-1)^4}{(mQ_2 - 1)^2 (mP_2 - 1)^2}. \end{aligned}$$

Moreover, we can rewrite $B_m(\boldsymbol{p}, \boldsymbol{q})$ differently; for $\boldsymbol{q}$, we have

$$\begin{aligned} &Q_2^2\Big(m^2 - 3m + 3\Big) + \Big(1 - 2mQ_2\Big) + 4(m-1)\Big(Q_3 - \frac{m}{4}Q_4\Big) \\ &\quad = (mQ_2 - 1)^2 - 3(m-1)Q_2^2 + 4(m-1)Q_3 - m(m-1)Q_4 \\ &\quad = \frac{(mQ_2-1)^2}{m-1}\left((m-1) + \frac{4(m-1)^2Q_3 - m(m-1)^2Q_4 - 3(m-1)Q_2^2}{(mQ_2-1)^2}\right) \\ &\quad = \frac{(mQ_2-1)^2}{m-1}\Big((m-1) + F(\boldsymbol{q})\Big) \end{aligned}$$

and accordingly for $\boldsymbol{p}$. Thus,

$$
\begin{aligned}
B_m(\boldsymbol{p},\boldsymbol{q}) &= \frac{2m^2}{(m-1)^4(m-2)^2(m-3)^2} \\
&\times\Big[Q_2^2\Big(m^2-3m+3\Big)+\Big(1-2mQ_2\Big)+4(m-1)\Big(Q_3-\frac{m}{4}Q_4\Big)\Big] \\
&\times\Big[P_2^2\Big(m^2-3m+3\Big)+\Big(1-2mP_2\Big)+4(m-1)\Big(P_3-\frac{m}{4}P_4\Big)\Big] \\
&= \frac{2m^2}{(m-1)^4(m-2)^2(m-3)^2}\,\frac{(mQ_2-1)^2(mP_2-1)^2}{(m-1)^2} \\
&\times\Big((m-1)+F(\boldsymbol{p})\Big)\Big((m-1)+F(\boldsymbol{q})\Big).
\end{aligned}
$$

Then

$$
\begin{aligned}
Var(\Delta_2^2) &= (n+m-3)B_m(\boldsymbol{p},\boldsymbol{q})\,\frac{(m-1)^4}{(mQ_2-1)^2(mP_2-1)^2} \\
&= \frac{2m^2(n+m-3)}{(m-1)^2(m-2)^2(m-3)^2}\Big((m-1)+F(\boldsymbol{p})\Big)\Big((m-1)+F(\boldsymbol{q})\Big). \quad \square
\end{aligned}
$$

Remark 5.2.19
If $m \geq 4$ is a power of 2, then if the group operation is $\oplus_m$, addition mod m, we have $n = 1$, and if it is $\oplus$, bitwise addition mod 2, we have $n = m - 1$. Thus, $Var(\Delta_2^2)$ is always twice as large for $\oplus$ as for $\oplus_m$, and for all other groups it lies between these values. (All groups whose order is a power of 2 must have at least one element of order 2: all elements have as order a power of 2; let $\mathrm{ord}(a) = 2^k$; then $\mathrm{ord}(a^{2^{k-1}}) = 2$.)

Finally, we prove that $Var_{XY}(\Delta_2^2(X * Y))$ is upper-bounded by an expression that behaves like $1/m$ for large m. We need the following inequality.

Lemma 5.2.20
For any positive integer m and any vector $\boldsymbol{r} = (r_0, \ldots, r_{m-1})$ in $\mathbb{R}^m$, we have

$$
m^3R_4 - 4m^2R_3R_1 - m^2R_2^2 + 8mR_2R_1^2 - 4R_1^4 \geq 0\,, \tag{5.28}
$$

where $R_j = \sum_{i=0}^{m-1} r_i^j$, with equality if and only if either all r_i are equal, or m is even and half of the r_i are equal to some real number s_1 while the other r_i's are equal to some real number s_2.
Proof:

Let $G(\boldsymbol{r}) := m^3R_4 - 4m^2R_3R_1 - m^2R_2^2 + 8mR_2R_1^2 - 4R_1^4$. It is easily seen that $G(\alpha\boldsymbol{r}) = \alpha^4 G(\boldsymbol{r})$ for any real number α. Let $\mathbf{1}$ denote the all-one

vector in $\mathbb{R}^m$. We show next that $G(\boldsymbol{r}+\mathbf{1}) = G(\boldsymbol{r})$ for all $\boldsymbol{r}$ in $\mathbb{R}^m$. Let $t_i = r_i + 1$, $i = 0, 1, \ldots, m-1$, $\boldsymbol{t} = (t_0, \ldots, t_{m-1})$, and $T_j = \sum_{i=10}^{m-1} t_i^j$. Then

$$\begin{aligned} T_1 &= R_1 + m, & T_2 &= R_2 + 2R_1 + m, \\ T_3 &= R_3 + 3R_2 + 3R_1 + m, & T_4 &= R_4 + 4R_3 + 6R_2 + 4R_1 + m \end{aligned}$$

and

$$\begin{aligned} G(\boldsymbol{t}) &= m^3T_4 - 4m^2T_3T_1 - m^2T_2^2 + 8mT_2T_1^2 - 4T_1^4 \\ &= m^3(R_4 + 4R_3 + 6R_2 + 4R_1 + m) \\ &\quad - 4m^2(R_3 + 3R_2 + 3R_1 + m)(R_1 + m) - m^2(R_2 + 2R_1 + m)^2 \\ &\quad + 8m(R_2 + 2R_1 + m)(R_1 + m)^2 - 4(R_1 + m)^4 \\ &= R_4(m^3) + R_3(4m^3 - 4m^3) + R_2(6m^3 - 12m^3 - 2m^3 + 8m^3) \\ &\quad + R_1(4m^3 - 4m^3 - 12m^3 - 4m^3 + 16m^3 + 16m^3 - 16m^3) \\ &\quad + 1(m^4 - 4m^4 - m^4 + 8m^4 - 4m^4) \\ &\quad + R_1^2(-12m^2 - 4m^2 + 32m^2 + 8m^2 - 24m^2) \\ &\quad + R_1R_2(-12m^2 - 4m^2 + 16m^2) + R_1R_3(-4m^2) \\ &\quad + R_2^2(-m^2) + R_1^3(16m - 16m) + R_1^2R_2(8m) + R_1^4(-4) \\ &= m^3R_4 - 4m^2R_1R_3 - m^2R_2^2 + 8mR_1^2R_2 - 4R_1^4 \\ &= G(\boldsymbol{r}). \end{aligned}$$

Now $G(\boldsymbol{r}+\alpha\mathbf{1}) = G(\boldsymbol{r})$ for all $\boldsymbol{r} \in \mathbb{R}^m$ and all $\alpha \in \mathbb{R}$, because $G(\boldsymbol{r}+\alpha\mathbf{1}) = \alpha^4 G(\frac{1}{\alpha}\boldsymbol{r} + \mathbf{1}) = \alpha^4 G(\boldsymbol{r}/\alpha) = G(\boldsymbol{r})$. Thus, it is enough to prove (5.28) on the hyperplane

$$\Big\{\alpha \cdot \mathbf{1} \,\Big|\, \alpha \in \mathbb{R}\Big\}^{\perp} = \Big\{\boldsymbol{r} \,\Big|\, \sum_{i=0}^{m-1} r_i = 0\Big\} = \Big\{\boldsymbol{r} \,\Big|\, R_1 = 0\Big\}.$$

But if $R_1 = 0$, then $G(\boldsymbol{r}) = m^3R_4 - m^2R_2^2$. Thus, we have to show that $m^3R_4 - m^2R_2^2 \geq 0$ for all $\boldsymbol{r}$ with $R_1 = 0$. But in fact, $m^3R_4 - m^2R_2^2 \geq 0$ for all $\boldsymbol{r}$ in $\mathbb{R}^m$: consider a random variable X such that $P_X(r_i^2) = 1/m$, $i = 0, \ldots, m-1$. Then $E[X] = \frac{1}{m}\sum_{i=0}^{m-1} r_i^2 = \frac{1}{m}R_2$ and $E[X^2] = \frac{1}{m}\sum_{i=0}^{m-1} r_i^4 = \frac{1}{m}R_4$ from which it follows that $m^3R_4 - m^2R_2^2 = m^4Var(X) \geq 0$.

As to equality in (5.28): it is easy to verify that the above condition is sufficient. Now let $G(\boldsymbol{r}) = 0$ but do not let all r_i be equal. Let $\boldsymbol{t} = \boldsymbol{r} - \frac{R_1}{m}\mathbf{1}$. Then $G(\boldsymbol{t}) = 0$ and $T_1 = 0$. Again, consider a random variable X such that $P_X(t_i^2) = 1/m$, $i = 0, \ldots, m-1$; then $0 = G(\boldsymbol{t}) = m^4Var(X)$ so X

is constant. Thus, all t_i^2 are equal. Because $\sum_{i=0}^{m-1} t_i = 0$, it follows that half of the t_i are equal to some real number t and the other t_i are equal to $-t$, and that m is even. Then half of the r_i are equal to $s_1 := t + R_1/m$ and the others are equal to $s_2 := -t + R_1/m$. □

Corollary 5.2.21
For any integer $m \geq 2$ and any non-uniform m-ary probability distribution $\boldsymbol{p}$,

$$F(\boldsymbol{p}) \leq -4\left(\frac{m-1}{m}\right)^2. \tag{5.29}$$

Proof:

$$F(\boldsymbol{p}) = -\left(\frac{G(\boldsymbol{p})}{(mP_2 - 1)^2} + 4\right)\frac{(m-1)^2}{m^2} \leq -4\left(\frac{m-1}{m}\right)^2. \qquad \square$$

It follows now that $(m-1) + F(\boldsymbol{p}) \leq \frac{(m-1)(m-2)^2}{m^2}$ and hence that

$$Var_{XY}(\Delta_2^2) \leq \frac{2(n+m-3)(m-2)^2}{m^2(m-3)^2} \underset{\sim}{m} \frac{2(n+m-3)}{m^2}. \tag{5.30}$$

When m is a power of two, then $1 \leq n \leq m-1$; it follows from (5.30) and Theorem 5.2.18 that if m is a large enough power of two, then $\Delta_2^2(X_1 * X_2)$ is almost always equal to one, that is, that the piling-up hypothesis holds for two random variables. Actually, depending on what one means by "$\approx$", values of m like 8 or 16 might be enough; for instance, if it is enough for a random variable with expected value 1 to have a variance of at most 0.3 in order that it be declared as "more or less constant", then the piling-up hypothesis must be considered as valid in the following two examples: if $m = 8$ and $*$ is mod 8-addition, then $n = 1$ and the upper bound in (5.30) is 432/1600, which is small enough; or if $m = 16$ and $*$ is the bitwise addition mod 2, then $n = 15$ and the upper bound is 10976/43264. In both cases, such a "small" m would be enough. If m is not a power of two, we also have $n \leq m-1$ (although the upper bound cannot always be reached) so we can draw the same conclusions.

We summarize the results of the chapter as:

Theorem 5.2.22
Let m be a large integer and $*$ be a group operation on $\mathbb{Z}_m$. Then

$$I_2^2(X_1 * X_2) \approx I_2^2(X_1) \cdot I_2^2(X_2)$$

for almost all independent m-ary random variables X_1 and X_2. □

Remark 5.2.23
Because the piling-up hypothesis holds for I_2^2, we believe that this imbalance measure is the "right one". It is also easy to compute for any probability distribution. Moreover, in the binary case, as mentioned in Chapter 2, the probability of success of the attack, for some fixed value of the key, increases with the square of the key-dependent imbalance; but the square of the imbalance is precisely I_2^2, since I_2 reduces to the binary imbalance when $m = 2$. This strengthens our conviction that I_2^2 is the right measure to use both in the binary generalisation of linear cryptanalysis and in the group generalisation of linear cryptanalysis. This should also allow a better comparison of both attacks.

5.A Proof of The Nine Parts

Proposition 5.A.1
Let $m \geq 4$ be an integer, $\mathbf{p}$ and $\mathbf{q}$ probability vectors of length m and i, k two elements of the group $\langle \mathbb{Z}_m, * \rangle$. Define

$$\begin{aligned}
①(ik) &= \sum_{\pi} \left(\sum_{j} p_j q_{\pi(j)} \right)^2 \left(\sum_{l} p_l q_{\pi(ikl)} \right)^2, \\
②(ik) &= \sum_{\pi} \sum_{j} p_j q_{\pi(j)} \left(\sum_{l} p_l q_{\pi(ikl)} \right)^2, \\
③(ik) &= \sum_{\pi} \left(\sum_{l} p_l q_{\pi(ikl)} \right)^2, \\
④(ik) &= \sum_{\pi} \left(\sum_{j} p_j q_{\pi(j)} \right)^2 \sum_{l} p_l q_{\pi(ikl)}, \\
⑤(ik) &= \sum_{\pi} \sum_{j} p_j q_{\pi(j)} \sum_{l} p_l q_{\pi(ikl)}, \\
⑥(ik) &= \sum_{\pi} \sum_{l} p_l q_{\pi(ikl)}, \\
⑦(ik) &= \sum_{\pi} \left(\sum_{j} p_j q_{\pi(j)} \right)^2, \\
⑧(ik) &= \sum_{\pi} \sum_{j} p_j q_{\pi(j)}, \\
⑨(ik) &= \sum_{\pi} 1,
\end{aligned}$$

(continued on next page)

(continued from previous page)

where $\sum_\pi$ *is the sum over all permutations of* $\{0, 1, \ldots, m-1\}$. *Then*

$$
\begin{aligned}
①(ik) &= (m-1)!Q_4P_{22}(ik) \\
&+ (m-2)!Q_{31}(2P_{12}(ik) + 2P_{21}(ik) - 4P_{22}(ik)) \\
&+ (m-2)!Q_{22}(P_2^2 + 2P_{11}^2(ik) - 3P_{22}(ik)) \\
&+ (m-3)!Q_{211}(2P_2 + 4P_{11}(ik) - 6P_{12}(ik) - 6P_{21}(ik) \\
&\qquad + 12P_{22}(ik) - 2P_2^2 - 4P_{11}^2(ik)) \\
&+ (m-4)!Q_{1111}(1 - 2P_2 - 4P_{11}(ik) + 4P_{12}(ik) + 4P_{21}(ik) \\
&\qquad + P_2^2 + 2P_{11}^2(ik) - 6P_{22}(ik)) \\
②(ik) &= (m-1)!Q_3P_{21}(ik) + m - 2)!Q_{21}(P_2 + 2P_{11}(ik) - 3P_{21}(ik)) \\
&+ (m-3)!Q_{111}(1 - P_2 - 2P_{11}(ik) + 2P_{21}(ik)) \\
③(ik) &= (m-1)!Q_2P_2 + (m-2)!Q_{11}(1 - P_2) \\
④(ik) &= (m-1)!Q_3P_{12}(ik) + (m-2)!Q_{21}(P_2 + 2P_{11}(ik) - 3P_{12}(ik)) \\
&+ (m-3)!Q_{111}(1 - P_2 - 2P_{11}(ik) + 2P_{12}(ik)) \\
⑤(ik) &= (m-1)!Q_2P_{11}(ik) + (m-2)!Q_{11}(1 - P_{11}(ik)) \\
⑥(ik) &= (m-1)! \\
⑦(ik) &= ③(ik) \\
⑧(ik) &= ⑥(ik) \\
⑨(ik) &= m!,
\end{aligned}
$$

where the Q's and the P's are defined by Definitions 5.2.8 and 5.2.10, respectively.

Proof:

We make repeated use of Lemma 5.2.4 and of identities similar to the ones proved there. In what follows, e_* will be the neutral element of $\langle \mathbb{Z}_m, * \rangle$. We begin by proving the simpler expressions. The expression for ⑨(ik) is obvious. The identities for ⑧(ik) and ⑦(ik) are proved as follows:

$$
\begin{aligned}
⑧(ik) &= \sum_j p_j \sum_\pi q_{\pi(j)} = \sum_j p_j (m-1)! \sum_i q_i = (m-1)! \\
⑦(ik) &= \sum_\pi \left(\sum_j p_j q_{\pi(j)} \right)^2 = \sum_\pi \sum_{i,j} p_i p_j q_{\pi(i)} q_{\pi(j)} \\
&= \sum_{i,j} p_i p_j \sum_\pi q_{\pi(i)} q_{\pi(j)} = \sum_j p_j^2 \sum_\pi q_{\pi(j)}^2 + \sum_{i \neq j} p_i p_j \sum_\pi q_{\pi(i)} q_{\pi(j)}
\end{aligned}
$$

$$\begin{aligned} &= \sum_j p_j^2 (m-1)! \sum_i q_i^2 + \sum_{i \neq j} p_i p_j \sum_{k \neq l} q_k q_l \\ &= (m-1)!\, Q_2 P_2 + (m-2)!\, Q_{11}(1 - P_2) \end{aligned}$$

since $\sum_{i \neq j} p_i p_j = \sum_{i,j} p_i p_j - \sum_j p_j^2 = 1 - P_2$. The identities for ⑥$(ik)$ and ③(ik) are proved in a similar way. Next, we are concerned with ⑤(ik). We have

$$⑤(ik) = \sum_\pi \sum_j p_j q_{\pi(j)} \sum_l p_l q_{\pi(ikl)} = \sum_j \sum_l p_j p_l \sum_\pi q_{\pi(j)} q_{\pi(ikl)}.$$

Here we must begin to distinguish between different values of ik.

1. $ik = e_*$. Then

$$\begin{aligned} ⑤(ik) &= ⑤(e_*) = \sum_{j,l} p_j p_l \sum_\pi q_{\pi(j)} q_{\pi(l)} \\ &= (m-1)!\, Q_2 P_2 + (m-2)!\, Q_{11}(1 - P_2). \end{aligned} \tag{5.31}$$

   (Just repeat the proof of the identity for ⑦(ik).)

2. $ik \neq e_*$. We break up the sum over all j, l into sums over: $j = l \neq ikl$; $j = ikl \neq l$; and j, l, ikl all different. This gives

$$\begin{aligned} ⑤(ik) &= \sum_j p_j^2 \sum_\pi q_{\pi(j)} q_{\pi(ikj)} + \sum_{\substack{j \neq l \\ j = ikl}} p_j p_l \sum_\pi q_{\pi(j)}^2 \\ &\quad + \sum_{\substack{j \neq l \\ j \neq ikl}} p_j p_l \sum_\pi q_{\pi(j)} q_{\pi(ikl)} \\ &= \sum_j p_j^2 (m-2)! \sum_{r \neq s} q_r q_s + \sum_l p_{ikl} p_l (m-1)! \sum_i q_i^2 \\ &\quad + \sum_{\substack{j \neq l \\ j \neq ikl}} p_j p_l (m-2)! \sum_{r \neq s} q_r q_s \\ &= (m-2)! Q_{11} P_2 + (m-1)! Q_2 P_{11}(ik) \\ &\quad + (m-2)! Q_{11}(1 - P_2 - P_{11}(ik)) \\ &= (m-1)!\, Q_2 P_{11}(ik) + (m-2)!\, Q_{11}(1 - P_{11}(ik)) \end{aligned} \tag{5.32}$$

   where in the penultimate equality we used the fact that

$$\sum_{\substack{j \neq l \\ j \neq ikl}} p_j p_l = \sum_{j,l} p_j p_l - \sum_l p_l^2 - \sum_l p_{ikl} p_l = 1 - P_2 - P_{11}(ik).$$

But (5.32) reduces to (5.31) if we put $ik = e_*$ since $P_{11}(e_*) = P_2$ (Lemma 5.2.11); thus, (5.32) is valid for both $ik = e_*$ and $ik \neq e_*$.

We now consider ④(ik). We have

$$④(ik) = \sum_{\pi}\Big(\sum_{j} p_j q_{\pi(j)}\Big)^2 \sum_{l} p_l q_{\pi(ikl)} = \sum_{j,l,n} p_j p_l p_n \sum_{\pi} q_{\pi(j)} q_{\pi(ikl)} q_{\pi(n)}. \tag{5.33}$$

We again distinguish between different values of ik. Let first $ik = e_*$. Then

$$④(e_*) = \sum_{j,l,n} p_j p_l p_n \sum_{\pi} q_{\pi(j)} q_{\pi(l)} q_{\pi(n)}.$$

We break up the sum into sums where: $j = l = n$; $j = l \neq n$; $j = n \neq l$; $l = n \neq j$; j, l, n all different. The first partial sum is

$$\sum_j p_j^3 \sum_{\pi} q_{\pi(j)}^3 = \sum_j p_j^3 (m-1)! \sum_i q_i^3 = (m-1)! Q_3 P_3;$$

The second is

$$\begin{aligned}\sum_{j\neq n} p_j^2 p_n \sum_{\pi} q_{\pi(j)}^2 q_{\pi(n)} &= \sum_{j\neq n} p_j^2 p_n (m-2)! \sum_{r\neq s} q_r^2 q_s \\ &= (m-2)! \sum_j p_j^2 (1-p_j) Q_{21} = (m-2)! Q_{21}(P_2 - P_3).\end{aligned}$$

The third and fourth give the same as the second, because $p_j p_l p_n \sum_{\pi} q_{\pi(j)} q_{\pi(l)} q_{\pi(n)}$ is symmetrical in j, l and n. The last partial sum is

$$\begin{aligned}&\sum_{j,l,n \text{ all different}} p_j p_l p_n \sum_{\pi} q_{\pi(j)} q_{\pi(l)} q_{\pi(n)} \\ &= \sum_{j,l,n \text{ all different}} p_j p_l p_n (m-3)! \sum_{r,s,t \text{ all different}} q_r q_s q_t \\ &= (m-3)! \sum_{j\neq l} p_j p_l (1 - p_j - p_l) Q_{111} \\ &= (m-3)! Q_{111}\Big(\sum_{j\neq l} p_j p_l - \sum_{j\neq l} p_j^2 p_l - \sum_{j\neq l} p_j p_l^2\Big) \\ &= (m-3)! Q_{111}\Big(1 - P_2 - 2(P_2 - P_3)\Big) \\ &= (m-3)! Q_{111}\Big(1 - 3P_2 + 2P_3\Big).\end{aligned}$$

All in all, we have

$$\begin{aligned} ④(e_*) &= (m-1)!Q_3P_3 + 3(m-2)!Q_{21}(P_2 - P_3) \\ &+ (m-3)!Q_{111}(1 - 3P_2 + 2P_3). \end{aligned} \tag{5.34}$$

Before we go on, we introduce a new notation for subsets of the set of all n-tuples: $[j]$ will stand for the set of all j; $[j,l]$ for the set of all pairs (j,l) with $j \neq l$; $[j,l=n]$ for the set of all triples (j,l,n) such that $l = n$ but j is different from l and n; $[j,l=n,m]$ for the set of all 4-tuples (j,l,n,m) all components of whose are different except for $l = n$; and so on. (Two indices separated by at least one comma are always different.)

Let now $ik \neq e_*$. We break up the sum as follows (these are all the possibilities for the indices j, l, n and ikl, since $l \neq ikl$):

Ⓐ $j = l = n$	Ⓔ $[j = n, l, ikl]$	Ⓗ $[j = ikl, l, n]$
Ⓑ $[j = l, n = ikl]$	Ⓕ $[j = ikl, l = n]$	Ⓘ $[j, l, n = ikl]$
Ⓒ $[j = l, n, ikl]$	Ⓖ $[j, l = n, ikl]$	Ⓙ $[j, l, n, ikl]$
Ⓓ $j = n = ikl$		

and we denote the corresponding partial sums with ④Ⓐ to ④Ⓙ. In the same way as above, one proves now that

$$\begin{aligned}
④Ⓐ &= (m-2)!P_3Q_{21} \\
④Ⓑ &= (m-2)!P_{21}(ik)Q_{21} \\
④Ⓒ &= (m-3)!(P_2 - P_3 - P_{21}(ik))Q_{111} \\
④Ⓓ &= (m-1)!P_{12}(ik)Q_3 \\
④Ⓔ &= (m-2)!(P_2 - P_3 - P_{12}(ik))Q_{21} \\
④Ⓕ &= (m-2)!P_{21}(ik)Q_{21} \\
④Ⓖ &= (m-3)!(P_2 - P_3 - P_{21}(ik))Q_{111} \\
④Ⓗ &= (m-2)!(P_{11}(ik) - P_{21}(ik) - P_{12}(ik))Q_{21} \\
④Ⓘ &= (m-2)!(P_{11}(ik) - P_{21}(ik) - P_{12}(ik))Q_{21} \\
④Ⓙ &= (m-3)!((1-3P_2+2P_3-2P_{11}(ik)+2P_{12}(ik)+2P_{21}(ik))Q_{111}.
\end{aligned}$$

Note that we do not have to calculate all of these expressions because the quantity $p_jp_lp_n \sum_\pi q_{\pi(j)}q_{\pi(ikl)}q_{\pi(n)}$ inside of the sum over all j,l,n in (5.33) is symmetrical in j and n so we have ④Ⓕ = ④Ⓑ, ④Ⓖ = ④Ⓒ

and ④Ⓘ = ④Ⓗ. Adding these expressions gives

$$\begin{aligned} ④(ik) &= (m-1)!Q_3P_{12}(ik) + (m-2)!Q_{21}(P_2 + 2P_{11}(ik) - 3P_{12}(ik)) \\ &+ (m-3)!Q_{111}(1 - P_2 - 2P_{11}(ik) + 2P_{12}(ik)). \end{aligned} \tag{5.35}$$

This expression can also be used when $ik = e_*$ because, if we replace $P_{ij}(ik) = P_{ij}(e_*)$ by P_{i+j}, then (5.35) reduces to (5.34).

We next consider ②(ik). Let $\tilde{p}_j := p_{(ik)^{-1}j}$, $\tilde{P}_r := \sum_n \tilde{p}_n^r$ and $\tilde{P}_{rs}(j) := \sum_n \tilde{p}_n^r \tilde{p}_{jn}^s$ and note that $\tilde{P}_j = P_j$ and $\tilde{P}_{rs}(j) = P_{rs}((ik)^{-1}j(ik))$. Then

$$\begin{aligned} ②(ik) &= \sum_\pi \sum_j p_j q_{\pi(j)} \left(\sum_l p_l q_{\pi(ikl)} \right)^2 \\ &= \sum_\pi \sum_j \tilde{p}_j q_{\pi((ik)^{-1}j)} \left(\sum_l \tilde{p}_l q_{\pi(l)} \right)^2, \end{aligned}$$

i.e., this is ④$((ik)^{-1})$ with $\tilde{p}$'s instead of p's; hence,

$$\begin{aligned} ②(ik) &= (m-1)!\, Q_3 \tilde{P}_{12}((ik)^{-1}) \\ &+ (m-2)!\, Q_{21}(\tilde{P}_2 + 2\tilde{P}_{11}((ik)^{-1}) - 3\tilde{P}_{12}((ik)^{-1})) \\ &+ (m-3)!\, Q_{111}(1 - \tilde{P}_2 - 2\tilde{P}_{11}((ik)^{-1}) + 2\tilde{P}_{12}((ik)^{-1})). \end{aligned}$$

Making the replacements $\tilde{P}_j = P_j$, $\tilde{P}_{11}((ik)^{-1}) = P_{11}((ik)^{-1}) = P_{11}(ik)$ and $\tilde{P}_{12}((ik)^{-1}) = P_{12}((ik)^{-1}) = P_{21}(ik)$ finishes the proof of this identity.

Finally, we look at ①(ik). We have

$$\begin{aligned} ①(ik) &= \sum_\pi \left(\sum_j p_j q_{\pi(j)} \right)^2 \left(\sum_l p_l q_{\pi(ikl)} \right)^2 \\ &= \sum_{j,l,n,z} p_j p_l p_n p_z \sum_\pi q_{\pi(j)} q_{\pi(l)} q_{\pi(ikn)} q_{\pi(ikz)}. \end{aligned}$$

We consider separately the cases $ik = e_*$, $\mathrm{ord}(ik) = 2$, and $\mathrm{ord}(ik) > 2$.

Let $ik = e_*$. In this case, we have

$$①(e_*) = \sum_{j,l,n,z} p_j p_l p_n p_z \sum_\pi q_{\pi(j)} q_{\pi(l)} q_{\pi(n)} q_{\pi(z)}.$$

We split in the usual way into the cases:

$\boxed{\text{A}}$ $[j=l=n=z]$	$\boxed{\text{F}}$ $[j=n=z,l]$	$\boxed{\text{K}}$ $[j,l=n,z]$
$\boxed{\text{B}}$ $[j=l=n,z]$	$\boxed{\text{G}}$ $[j=n,l=z]$	$\boxed{\text{L}}$ $[j=z,l,n]$
$\boxed{\text{C}}$ $[j=l=z,n]$	$\boxed{\text{H}}$ $[j=n,l,z]$	$\boxed{\text{M}}$ $[j,l=z,n]$
$\boxed{\text{D}}$ $[j=l,n=z]$	$\boxed{\text{I}}$ $[j=z,l=n]$	$\boxed{\text{N}}$ $[j,l,n=z]$
$\boxed{\text{E}}$ $[j=l,n,z]$	$\boxed{\text{J}}$ $[j,l=n=z]$	$\boxed{\text{O}}$ $[j,l,n,z]$

and call the corresponding expressions ①$\boxed{\text{A}}$ to ①$\boxed{\text{O}}$. But note that $p_j p_l p_n p_z \sum_\pi q_{\pi(j)} q_{\pi(l)} q_{\pi(ikn)} q_{\pi(ikz)}$ is symmetrical in j and l, as well as in n and z so that we have ①$\boxed{\text{B}}$ = ①$\boxed{\text{C}}$ = ①$\boxed{\text{F}}$ = ①$\boxed{\text{J}}$, ①$\boxed{\text{D}}$ = ①$\boxed{\text{G}}$ = ①$\boxed{\text{I}}$ and ①$\boxed{\text{E}}$ = ①$\boxed{\text{H}}$ = ①$\boxed{\text{K}}$ = ①$\boxed{\text{L}}$ = ①$\boxed{\text{M}}$ = ①$\boxed{\text{N}}$. Now

$$
\begin{aligned}
①\boxed{\text{A}} &= \sum_j p_j^4 \sum_\pi q_{\pi(j)}^4 = \sum_j p_j^4 (m-1)! \sum_i q_i^4 = (m-1)! P_4 Q_4; \\
①\boxed{\text{B}} &= \sum_{z \neq j} p_j^3 p_z \sum_\pi q_{\pi(j)}^3 q_{\pi(z)} = \sum_j p_j^3 (1-p_j)(m-2)! \sum_{r \neq s} q_r^3 q_s \\
&= (m-2)!(P_3 - P_4) Q_{31}; \\
①\boxed{\text{D}} &= \sum_{z \neq j} p_j^2 p_z^2 \sum_\pi q_{\pi(j)}^2 q_{\pi(z)}^2 = \sum_j p_j^2 \Big(\sum_z p_z^2 - p_j^2 \Big) (m-2)! \sum_{r \neq s} q_r^2 q_s^2 \\
&= (m-2)!(P_2^2 - P_4) Q_{22}; \\
①\boxed{\text{E}} &= \sum_{[j,n,z]} p_j^2 p_n p_z \sum_\pi q_{\pi(j)}^2 q_{\pi(n)} q_{\pi(z)} \\
&= \sum_{[j,n,z]} p_j^2 p_n p_z (m-3)! \sum_{[r,s,t]} q_r^2 q_s q_t \\
&= (m-3)! Q_{211} \sum_{n \neq j} p_j^2 p_n (1 - p_j - p_n) \\
&= (m-3)! Q_{211} \sum_{n \neq j} \left(p_j^2 p_n - p_j^3 p_n - p_j^2 p_n^2 \right) \\
&= (m-3)!(P_2 - 2P_3 + 2P_4 - P_2^2) Q_{211}; \\
①\boxed{\text{O}} &= \sum_{[j,l,n,z]} p_j p_l p_n p_z \sum_\pi q_{\pi(j)} q_{\pi(l)} q_{\pi(n)} q_{\pi(z)} \\
&= \sum_{[j,l,n,z]} p_j p_l p_n p_z (m-4)! \sum_{[r,s,t,w]} q_r q_s q_t q_w \\
&= (m-4)!(1 - 6P_2 + 8P_3 + 3P_2^2 - 6P_4) Q_{1111}.
\end{aligned}
$$

Adding all expressions, we get

$$\begin{aligned}
①(e_*) &= (m-1)!P_4Q_4 + 4(m-2)!(P_3-P_4)Q_{31} \\
&+ 3(m-2)!(P_2^2-P_4)Q_{22} \\
&+ 6(m-3)!(P_2-2P_3+2P_4-P_2^2)Q_{211} \\
&+ (m-4)!(1-6P_2+8P_3+3P_2^2-6P_4)Q_{1111}. \qquad (5.36)
\end{aligned}$$

Now let $\mathrm{ord}(ik) = 2$. Since the technique we use is always the same, we write only how we split the sum, which of the expressions obtained are equal because of symmetry and their dependence on the P's and Q's. Here

$$①(ik) = \sum_{j,l,n,z} p_j p_l p_n p_z \sum_{\pi} q_{\pi(j)} q_{\pi(l)} q_{\pi(ikn)} q_{\pi(ikz)}$$

and the sum is split as follows:

1 $[j=l=n=z, ikn=ikz]$	17 $[j=n=ikz, l, z=ikn]$
2 $[j=l=n=ikz, z=ikn]$	18 $[j=n, l=ikz, n, ikn]$
3 $[j=l=n, z, ikn, ikz]$	19 $[j=n, l, z, ikn, ikz]$
4 $[j=l=z=ikn, n=ikz]$	20 $[j=z=ikn, l=n=ikz]$
5 $[j=l=z, n, ikn, ikz]$	21 $[j=z, l=n, ikn, ikz]$
6 $[j=l=ikn=ikz, n=z]$	22 $[j=ikn=ikz, l=n=z]$
7 $[j=l, n=z, ikn=ikz]$	23 $[j, l=n=z, ikn=ikz]$
8 $[j=l=ikn, n, z, ikz]$	24 $[j=ikn, l=n, z, ikz]$
9 $[j=l, n=ikz, z=ikn]$	25 $[j, l=n=ikz, z=ikn]$
10 $[j=l=ikz, n, z, ikn]$	26 $[j=ikz, l=n, z, ikn]$
11 $[j=l, n, z, ikn, ikz]$	27 $[j, l=n, z, ikn, ikz]$
12 $[j=n=z, l=ikn=ikz]$	28 $[j=z=ikn, l, n=ikz]$
13 $[j=n=z, l, ikn=ikz]$	29 $[j=z, l=ikn, n, ikz]$
14 $[j=n=ikz, l=z=ikn]$	30 $[j=z, l=ikz, n, ikn]$
15 $[j=n, l=z, ikn, ikz]$	31 $[j=z, l, n, ikn, ikz]$
16 $[j=n, l=ikn, z, ikz]$	32 $[j=ikn, l=z, n, ikz]$

$\boxed{33}$ $[j, l = z = ikn, n = ikz]$

$\boxed{34}$ $[j = ikz, l = z, n, ikn]$

$\boxed{35}$ $[j, l = z, n, ikn, ikz]$

$\boxed{36}$ $[j = ikn = ikz, l, n = z]$

$\boxed{37}$ $[j, l = ikn = ikz, n = z]$

$\boxed{38}$ $[j, l, n = z, ikn = ikz]$

$\boxed{39}$ $[j = ikn, l = ikz, n, z]$

$\boxed{40}$ $[j = ikn, l, n, z, ikz]$

$\boxed{41}$ $[j = ikz, l = ikn, n, z]$

$\boxed{42}$ $[j, l = ikn, n, z, ikz]$

$\boxed{43}$ $[j, l, n = ikz, z = ikn]$

$\boxed{44}$ $[j = ikz, l, n, z, ikn]$

$\boxed{45}$ $[j, l = ikz, n, z, ikn]$

$\boxed{46}$ $[j, l, n, z, ikn, ikz]$

Because $p_j p_l p_n p_z \sum_\pi q_{\pi(j)} q_{\pi(l)} q_{\pi(ikn)} q_{\pi(ikz)}$ is symmetrical in j and l, as well as in n and z, the following expressions are equal:

- ①$\boxed{2}$ and ①$\boxed{4}$;
- ①$\boxed{3}$ and ①$\boxed{5}$;
- ①$\boxed{8}$ and ①$\boxed{10}$;
- ①$\boxed{12}$ and ①$\boxed{22}$;
- ①$\boxed{13}$ and ①$\boxed{23}$;
- ①$\boxed{14}$ and ①$\boxed{20}$;
- ①$\boxed{15}$ and ①$\boxed{21}$;
- ①$\boxed{16}$, ①$\boxed{24}$, ①$\boxed{30}$ and ①$\boxed{34}$;
- ①$\boxed{17}$, ①$\boxed{25}$, ①$\boxed{28}$ and ①$\boxed{33}$;
- ①$\boxed{18}$, ①$\boxed{26}$, ①$\boxed{29}$ and ①$\boxed{32}$;
- ①$\boxed{19}$, ①$\boxed{27}$, ①$\boxed{31}$ and ①$\boxed{35}$;
- ①$\boxed{36}$ and ①$\boxed{37}$;
- ①$\boxed{39}$ and ①$\boxed{41}$;
- ①$\boxed{40}$, ①$\boxed{42}$, ①$\boxed{44}$ and ①$\boxed{45}$.

The expressions are equal to:

$$
\begin{aligned}
①\boxed{1} &= (m-2)!P_4Q_4;\\
①\boxed{2} &= (m-2)!P_{31}(ik)Q_{31};\\
①\boxed{3} &= (m-3)!(P_3-P_4-P_{31})Q_{211};\\
①\boxed{6} &= (m-1)!P_{22}(ik)Q_4;\\
①\boxed{7} &= (m-2)!(P_2^2-P_4-P_{22}(ik))Q_{22};\\
①\boxed{8} &= (m-2)!(P_{21}(ik)-P_{22}(ik)-P_{31}(ik))Q_{31};\\
①\boxed{9} &= (m-3)!(P_2P_{11}(ik)-2P_{31}(ik))Q_{211};\\
①\boxed{11} &= (m-3)!(P_2-2P_3+2P_4-P_2^2-2P_{21}(ik)+4P_{31}(ik)\\
&\qquad +2P_{22}(ik)-P_2P_{11}(ik))Q_{211};\\
①\boxed{12} &= (m-2)!P_{31}(ik)Q_{31};\\
①\boxed{13} &= (m-3)!(P_3-P_4-P_{31}(ik))Q_{211};\\
①\boxed{14} &= (m-2)!P_{22}(ik)Q_{22};\\
①\boxed{15} &= (m-4)!(P_2^2-P_4-P_{22}(ik))Q_{1111};\\
①\boxed{16} &= (m-3)!(P_{21}(ik)-P_{31}(ik)-P_{22}(ik))Q_{211};\\
①\boxed{17} &= (m-3)!(P_{21}(ik)-P_{22}(ik)-P_{31}(ik))Q_{211};\\
①\boxed{18} &= (m-3)!(P_2P_{11}(ik)-2P_{31}(ik))Q_{211};\\
①\boxed{19} &= (m-4)!(P_2-2P_3+2P_4-P_2^2-2P_{21}(ik)+4P_{31}(ik)\\
&\qquad +2P_{22}(ik)-P_2P_{11}(ik))Q_{1111};\\
①\boxed{36} &= (m-2)!(P_{21}(ik)-P_{31}(ik)-P_{22}(ik))Q_{31};\\
①\boxed{38} &= (m-3)!(P_2-2P_3+2P_4-P_2^2-2P_{21}(ik)+2P_{31}(ik)\\
&\qquad +2P_{22}(ik))Q_{211};\\
①\boxed{39} &= (m-2)!(P_{11}^2(ik)-2P_{22}(ik))Q_{22};\\
①\boxed{40} &= (m-3)!(P_{11}(ik)-P_2P_{11}(ik)-P_{11}^2(ik)-4P_{21}(ik)\\
&\qquad +4P_{31}(ik)+4P_{22}(ik))Q_{211};\\
①\boxed{43} &= (m-4)!(P_{11}(ik)-P_2P_{11}(ik)-4P_{21}(ik)+P_{31}(ik)\\
&\qquad +2P_{22}(ik))Q_{1111};\\
①\boxed{46} &= (m-4)!(1-6P_2+8P_3+3P_2^2-6P_4-5P_{11}(ik)+20P_{21}(ik)\\
&\qquad -14P_{22}(ik)-20P_{31}(ik)\\
&\qquad +5P_2P_{11}(ik)+2P_{11}^2(ik))Q_{1111},
\end{aligned}
$$

that is,

$$\begin{aligned}
①(ik) &= (m-1)!Q_4P_{22}(ik)\\
&+ 4(m-2)!Q_{31}(P_{21}(ik)-P_{22}(ik))\\
&+ (m-2)!Q_{22}(P_2^2+2P_{11}^2(ik)-3P_{22}(ik))\\
&+ (m-3)!Q_{211}(2P_2-2P_2^2+4P_{11}(ik)-4P_{11}^2(ik)\\
&\qquad +12P_{22}(ik)-12P_{21}(ik))\\
&+ (m-4)!Q_{1111}(1-2P_2+P_2^2-4P_{11}(ik)+2P_{11}^2(ik)\\
&\qquad +8P_{21}(ik)-6P_{22}(ik)); \qquad (5.37)
\end{aligned}$$

we again have the phenomenon that this reduces to (5.36) if one sets $ik = e_*$. Now let $\mathrm{ord}(ik) > 2$. Then the sum is split as follows:

$\boxed{1}$ $[j = l = n = z, ikn = ikz]$ $\boxed{18}$ $[j = n = z, l, ikn = ikz]$

$\boxed{2}$ $[j = l = n, z = ikn, ikz]$ $\boxed{19}$ $[j = n, l = z = ikn, ikz]$

$\boxed{3}$ $[j = l = n = ikz, z, ikn]$ $\boxed{20}$ $[j = n = ikz, l = z, ikn]$

$\boxed{4}$ $[j = l = n, z, ikn, ikz]$ $\boxed{21}$ $[j = n, l = z, ikn, ikz]$

$\boxed{5}$ $[j = l = z = ikn, n, ikz]$ $\boxed{22}$ $[j = n = ikz, l = ikn, z]$

$\boxed{6}$ $[j = l = z, n = ikz, ikn]$ $\boxed{23}$ $[j = n, l = ikn, z, ikz]$

$\boxed{7}$ $[j = l = z, n, ikn, ikz]$ $\boxed{24}$ $[j = n, l = ikz, z = ikn]$

$\boxed{8}$ $[j = l = ikn = ikz, n = z]$ $\boxed{25}$ $[j = n, l, z = ikn, ikz]$

$\boxed{9}$ $[j = l, n = z, ikn = ikz]$ $\boxed{26}$ $[j = n = ikz, l, z, ikn]$

$\boxed{10}$ $[j = l = ikn, n = ikz, z]$ $\boxed{27}$ $[j = n, l = ikz, z, ikn]$

$\boxed{11}$ $[j = l = ikn, n, z, ikz]$ $\boxed{28}$ $[j = n, l, z, ikn, ikz]$

$\boxed{12}$ $[j = l = ikz, n, z = ikn]$ $\boxed{29}$ $[j = z = ikn, l = n, ikz]$

$\boxed{13}$ $[j = l, n, z = ikn, ikz]$ $\boxed{30}$ $[j = z, l = n = ikz, ikn]$

$\boxed{14}$ $[j = l = ikz, n, z, ikn]$ $\boxed{31}$ $[j = z, l = n, ikn, ikz]$

$\boxed{15}$ $[j = ln = ikz, z, ikn]$ $\boxed{32}$ $[j = ikn = ikz, l = n = z]$

$\boxed{16}$ $[j = l, n, z, ikn, ikz]$ $\boxed{33}$ $[j, l = n = z, ikn = ikz]$

$\boxed{17}$ $[j = n = z, l = ikn = ikz]$ $\boxed{34}$ $[j = ikn, l = n = ikz, z]$

35 $[j = ikn, l = n, z, ikz]$

36 $[j = ikz, l = n, z = ikn]$

37 $[j, l = n, z = ikn, ikz]$

38 $[j = ikz, l = n, z, ikn]$

39 $[j, l = n = ikz, z, ikn]$

40 $[j, l = n, z, ikn, ikz]$

41 $[j = z = ikn, l = ikz, n]$

42 $[j = z = ikn, l, n, ikz]$

43 $[j = z, l = ikn, n = ikz]$

44 $[j = z, l = ikn, n, ikz]$

45 $[j = z, l = ikz, n, ikn]$

46 $[j = z, n = ikz, l, ikn]$

47 $[j = z, l, n, ikn, ikz]$

48 $[j = ikn, l = z, n = ikz]$

49 $[j = ikn, l = z, n, ikz]$

50 $[j = ikz, l = z = ikn, n]$

51 $[j, l = z = ikn, n, ikz]$

52 $[j = ikz, l = z, n, ikn]$

53 $[j, l = z, n = ikz, ikn]$

54 $[j, l = z, n, ikn, ikz]$

55 $[j = ikn = ikz, l, n = z]$

56 $[j, l = ikn = ikz, n = z]$

57 $[j, l, n = z, ikn = ikz]$

58 $[j = ikn, l = ikz, n, z]$

59 $[j = ikn, l, n = ikz, z]$

60 $[j = ikn, l, n, z, ikz]$

61 $[j = ikz, l = ikn, n, z]$

62 $[j, l = ikn, n = ikz, z]$

63 $[j, l = ikn, n, z, ikz]$

64 $[j = ikz, l, n, z = ikn]$

65 $[j, l = ikz, n, z = ikn]$

66 $[j, l, n, z = ikn, ikz]$

67 $[j = ikz, l, n, z, ikn]$

68 $[j, l = ikz, n, z, ikz]$

69 $[j, l, n = ikz, z, ikn]$

70 $[j, l, n, z, ikn, ikz]$

Again, because $p_j p_l p_n p_z \sum_\pi q_{\pi(j)} q_{\pi(l)} q_{\pi(ikn)} q_{\pi(ikz)}$ is symmetrical in j and l, as well as in n and z, the following expressions are equal:

- ①2 and ①6;
- ①3 and ①5;
- ①4 and ①7;
- ①10 and ①12;
- ①11 and ①14;
- ①13 and ①15;
- ①17 and ①32;
- ①18 and ①33;

- ①19, ①20, ①29 and ①30;
- ①21 and ①31;
- ①22, ①34, ①41 and ①50;
- ①23, ①35, ①45 and ①52;
- ①24, ①36, ①43 and ①48;
- ①25, ①37, ①46 and ①53;
- ①26, ①39, ①42 and ①51;
- ①27, ①38, ①44 and ①49;
- ①28, ①40, ①47 and ①54;
- ①55 and ①56;
- ①58 and ①61;
- ①59, ①62, ①64 and ①65;
- ①60, ①63, ①67 and ①68;
- ①66 and ①69.

The expressions are equal to:

$$
\begin{aligned}
①1 &= (m-2)!P_4Q_{22};\\
①2 &= (m-3)!P_{31}(ik)Q_{211};\\
①3 &= (m-2)!P_{13}(ik)Q_{31};\\
①4 &= (m-3)!(P_3 - P_4 - P_{31}(ik) - P_{13}(ik))Q_{211};\\
①8 &= (m-1)!P_{22}(ik)Q_4;\\
①9 &= (m-2)!(P_2^2 - P_4 - P_{22}(ik))Q_{22};\\
①10 &= (m-2)!P_{112}(ik)Q_{31};\\
①11 &= (m-2)!(P_{12}(ik) - P_{22}(ik) - P_{13}(ik) - P_{112}(ik))Q_{31};\\
①13 &= (m-3)!(P_2P_{11}(ik) - P_{31}(ik) - P_{13}(ik) - P_{112}(ik))Q_{211};
\end{aligned}
$$

$$
\begin{aligned}
①\boxed{16} &= (m-3)!(P_2 - 2P_3 + 2P_4 - P_2^2 - 2P_{12}(ik) + 4P_{13}(ik) \\
&\qquad + 2P_{31}(ik) + 2P_{22}(ik) + 2P_{112}(ik) - 2P_2P_{11}(ik))Q_{211}; \\
①\boxed{17} &= (m-2)!P_{31}(ik)Q_{31}; \\
①\boxed{18} &= (m-3)!(P_3 - P_4 - P_{31}(ik))Q_{211}; \\
①\boxed{19} &= (m-3)!P_{22}(ik)Q_{211}; \\
①\boxed{21} &= (m-4)!(P_2^2 - P_4 - 2P_{22}(ik))Q_{1111}; \\
①\boxed{22} &= (m-3)!P_{121}(ik)Q_{22}; \\
①\boxed{23} &= (m-3)!(P_{21}(ik) - P_{22}(ik) - P_{31}(ik) - P_{121}(ik))Q_{211}; \\
①\boxed{24} &= (m-3)!P_{211}(ik)Q_{211}; \\
①\boxed{25} &= (m-4)!(P_{21}(ik) - P_{31}(ik) - P_{22}(ik) - P_{211}(ik))Q_{1111}; \\
①\boxed{26} &= (m-3)!(P_{12}(ik) - P_{22}(ik) - P_{13}(ik) - P_{121}(ik))Q_{211}; \\
①\boxed{27} &= (m-3)!(P_2P_{11}(ik) - P_{13}(ik) - P_{31}(ik) - P_{211}(ik))Q_{211}; \\
①\boxed{28} &= (m-4)!(P_2 - 2P_3 + 2P_4 - P_2^2 - P_{12}(ik) + 2P_{13}(ik) \\
&\qquad - 2P_{21}(ik) + 3P_{22}(ik) + 3P_{31}(ik) + P_{121}(ik) \\
&\qquad + P_{211}(ik) - P_2P_{11}(ik))Q_{1111}; \\
①\boxed{55} &= (m-2)!(P_{21}(ik) - P_{31}(ik) - P_{22}(ik))Q_{31}; \\
①\boxed{57} &= (m-3)!(P_2 - 2P_3 + 2P_4 - P_2^2 - 2P_{21}(ik) + 2P_{22}(ik) \\
&\qquad + 2P_{31}(ik))Q_{211}; \\
①\boxed{58} &= (m-2)!(P_{11}^2(ik) - P_{22}(ik) - 2P_{121}(ik))Q_{22}; \\
①\boxed{59} &= (m-3)!(P_{111}(ik) - P_{211}(ik) - P_{121}(ik) - P_{112}(ik))Q_{211}; \\
①\boxed{60} &= (m-3)!(P_{11}(ik) - 2P_{12}(ik) + 2P_{13}(ik) - 2P_{21}(ik) + 3P_{22}(ik) \\
&\qquad + 2P_{31}(ik) - P_{11}^2(ik) - P_{111}(ik) + P_{112}(ik) + P_{211}(ik) \\
&\qquad + 3P_{121}(ik) - P_2P_{11}(ik))Q_{211}; \\
①\boxed{66} &= (m-4)!(P_{11}(ik) - 2P_{12}(ik) + 2P_{13}(ik) - 2P_{21}(ik) + 2P_{22}(ik) \\
&\qquad + 2P_{31}(ik) - 2P_{111}(ik) + 2P_{112}(ik) \\
&\qquad + 2P_{121}(ik) + 2P_{211}(ik) - P_2P_{11}(ik))Q_{1111}; \\
①\boxed{70} &= (m-4)!(1 - 6P_2 + 8P_3 - 6P_4 + 3P_2^2 - 6P_{11}(ik) \\
&\qquad + 12P_{12}(ik) - 12P_{13}(ik) + 12P_{21}(ik) - 14P_{22}(ik) \\
&\qquad - 12P_{31}(ik) + 6P_2P_{11}(ik) + 2P_{11}^2(ik) + 4P_{111}(ik) \\
&\qquad - 4P_{112}(ik) - 8P_{121}(ik) - 4P_{211}(ik))Q_{1111}.
\end{aligned}
$$

Everything added together gives

$$\begin{aligned}
①(ik) &= (m-1)!Q_4P_{22}(ik) \\
&+ (m-2)!Q_{31}(2P_{12}(ik)+2P_{21}(ik)-4P_{22}(ik)) \\
&+ (m-2)!Q_{22}(P_2^2+2P_{11}^2(ik)-3P_{22}(ik)) \\
&+ (m-3)!Q_{211}(2P_2+4P_{11}(ik)-6P_{12}(ik)-6P_{21}(ik) \\
&\qquad +12P_{22}(ik)-2P_2^2-4P_{11}^2(ik)) \\
&+ (m-4)!Q_{1111}(1-2P_2-4P_{11}(ik)+4P_{12}(ik)+P_{21}(ik) \\
&\qquad +P_2^2+2P_{11}^2(ik)-6P_{22}(ik)). \qquad (5.38)
\end{aligned}$$

Moreover, this formula can be used for all ik since when $\mathrm{ord}(ik)$ is one or two, then the P's are changed as stated in Lemma 5.2.11 and the formula reduces precisely to the ones we had in (5.36) and (5.37). □

Chapter 6

Concluding Remarks

The linear cryptanalysis attack is coming of age. It has been tried out on many ciphers, with greater or less success; it has been formalized and a few attempts to generalise it have been made; ways to design ciphers or at least to make cipher primitives resistant to linear cryptanalysis have been investigated; and all ciphers designed in the last few years take this attack into account. Like many aspects of cryptology, linear cryptanalysis and its generalisations rely on a number of assumptions. In this work, three of these have been studied.

Linked threefold sums can be used to lower-bound the probability of success of an attack using the binary generalisation of linear cryptanalysis. In order to apply the Piling-up Lemma, these threefold sums must be independent. However, their independence is very difficult to determine in practical cases. By an averaging argument, we showed that, in virtually all cases, the Piling-up Lemma can be used validly with dependent threefold sums if one replaces equality by approximation.

The hypothesis of fixed-key equivalence assures the cryptanalyst that the probability of success of the attack is more or less independent of the key used in the encryption and that the average-key imbalance of the I/O sum used in the attack is a good measure for the probability of success. Again by an averaging argument, we showed that, for virtually all one-round I/O sums, the key-dependent imbalances are approximately equal to the average-key imbalance. However, we were not able to specialize this assertion to the effective I/O sums. An important by-product of our investigation, valid for any number of rounds, was a quantitative definition of an effective I/O sum.

In the m-ary variant of linear cryptanalysis called the group generalisation of linear cryptanalysis, there is no analogon to the Piling-up Lemma.

However, once more by an averaging argument, we were able to show that, if m is large enough, one can state a similar identity for virtually all cases by replacing the equality by an approximation.

Several open problems related to our work are worth investigating further. We wish to mention the following ones:

- As to the topic of Chapter 4, one could restrict the averaging to block ciphers whose round functions g have the property that, for all values of the key z, the functions $g(\cdot, z)$ belong to the same subgroup of the group of invertible functions. Accordingly, the definition of an effective I/O sum could be modified for I/O sums derived from a cipher belonging to such a subset of all block ciphers.

- We proved that, for virtually all I/O sums, the key-dependent imbalances were approximately equal to the average-key imbalance for virtually all keys. However, our procedure does not allow us to say that this holds also for virtually all effective I/O sums since most I/O sums are not effective. It is therefore of interest to study the fixed-key equivalence condition for I/O sums whose average-key imbalance is larger than a certain threshold. The threshold should be different for all blocklengths.

- We proved the piling-up hypothesis for the imbalance measure I_2^2 only for two random variables and concluded that it holds for any number of random variables by induction if one changes the definition of the approximation sign. Although this question is not very important and the calculation would no doubt be tedious, we are interested in computing the same kind of averages for more than two random variables as we did for two random variables.

- Finally, the hypothesis of wrong-key randomization presented in Chapter 2, which is the most important of all the hypotheses since it ensures that the attack can find the key, remains untouched by our work.

Bibliography

[1] Eli Biham, "On Matsui's Linear Cryptanalysis", in *Advances in Cryptology – Eurocrypt'94*, Lecture Notes in Computer Science 950, pp. 341-355, Springer 1994. ISBN 3-540-60176-7.

[2] Eli Biham and Alex Biryukov, "An Improvement of Davies' Attack on DES", in *Advances in Cryptology – Eurocrypt'94*, Lecture Notes in Computer Science 950, pp. 461-467, Springer 1994. ISBN 3-540-60176-7.

[3] Eli Biham and Adi Shamir, "Differential Cryptanalysis of DES-like Cryptosystems", in *Advances in Cryptology – CRYPTO'90*, Lecture Notes in Computer Science 537, pp. 2-21, Springer 1991. ISBN 3-540-54508-5.

[4] Eli Biham and Adi Shamir, "Differential Cryptanalysis of Feal and N-Hash", in *Advances in Cryptology – Eurocrypt'91*, Lecture Notes in Computer Science 547, pp. 1-16, Springer 1991. ISBN 3-540-54620-0.

[5] Eli Biham and Adi Shamir, "Differential Cryptanalysis of the Full 16-Round DES", in *Advances in Cryptology – CRYPTO'92*, Lecture Notes in Computer Science 740, pp. 487-496, Springer 1993. ISBN 3-540-57340-2.

[6] Johan Borst, Lars R. Knudsen, and Vincent Rijmen, "Two Attacks on Reduced IDEA", in *Advances in Cryptology – Eurocrypt'97*, Lecture Notes in Computer Science 1233, pp. 1-13, Springer 1997. ISBN 3-540-62975-0.

[7] Ilja N. Bronstein, Konstantin A. Semendjajew, Gerhard Musiol, Heiner Mühlig, *Taschenbuch der Mathematik*, Verlag Harri Deutsch, Frankfurt am Main, 1993. ISBN 3-8171-2001-X.

[8] Lawrence Brown, Matthew Kwan, Josef Pieprzyk, and Jennifer Seberry, "Improving Resistence to Different Cryptanalysis and the Redesign of LOKI", in *Advances in Cryptology - AsiaCrypt '91*, Lecture Notes in Computer Science 739, pp. 36-50, Springer 1993. ISBN 3-540-57332-1.

[9] Lawrence Brown, Josef Pieprzyk, and Jennifer Seberry, "LOKI: A Cryptographic Primitive for Authentication and Secrecy Applications" in *Advances in Cryptology – AUSCRYPT'90*, Lecture Notes in Computer Science 453, pp. 229-236, Springer 1990. ISBN 3-540-53000-2.

[10] Florent Chabaud and Serge Vaudenay, *Links between Differential and Linear Cryptanalysis*, technical report LIENS-94-3 of the Laboratoire d'Informatique de l'Ecole Normale Supérieure, 1994.

[11] "Data Encryption Standard", in Federal Information Processing Standards Publications, No. 46, U. S. Department of Commerce, National Bureau of Standards, January 1977.

[12] Whitfield Diffie and Martin E. Hellman, "New Directions in Cryptography", in *IEEE Transactions in Information Theory*, Vol. 22, pp. 644-654, November 1976.

[13] Robert G. Gallager, *Discrete Stochastic Processes*, Kluwer Academic Publishers, Dordrecht, 1996. ISBN 0-7923-9583-2.

[14] R. William Gosper, Jr., "Decision procedure for indefinite hypergeometric summation", in *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 75, pp. 40-42, 1978.

[15] Ronald L. Graham, Donald E. Knuth and Oren Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, 1989. ISBN 0-201-55802-5.

[16] Carlo Harpes, *Cryptanalysis of Iterated Block Ciphers*, Vol. 7 of ETH Series in Information Processing, Ed. J.L. Massey, Hartung-Gorre Verlag, Konstanz, 1996. ISBN 3-89649-079-6.

[17] Carlo Harpes, Gerhard G. Kramer, and James L. Massey, "A generalization of linear cryptanalysis and the applicability of Matsui's piling-up lemma", in *Advances in Cryptology – Eurocrypt'95*, Lecture Notes in Computer Science 921, pp. 24-38, Springer 1995. ISBN 3-540-59409-4.

[18] Philip Hawkes, "Differential-Linear Weak Key Classes of IDEA", in *Advances in Cryptology – Eurocrypt'98*, Lecture Notes in Computer Science 1403, pp. 112-126, Springer 1998. ISBN 3-540-64518-7.

[19] Thomas Jakobsen, *Correlation Attacks on Block Ciphers*, Master Thesis, Department of Mathematics, Technical University of Denmark, January 1996.

[20] Burton S. Kaliski and Y. L. Lin, "On Differential and Linear Cryptanalysis of the RC5 Encryption Algorithm", in *Advances in Cryptology – CRYPTO'95*, Lecture Notes in Computer Science 963, pp. 171-184, Springer 1995. ISBN 3-540-60221-6.

[21] Burton S. Kaliski Jr. and Matthew J. B. Robshaw, "Linear Cryptanalysis Using Multiple Approximations", in *Advances in Cryptology – CRYPTO'94*, Lecture Notes in Computer Science 839, pp. 26-39, Springer 1994. ISBN 3-540-58333-5.

[22] Burton S. Kaliski Jr. and Matthew J. B. Robshaw, "Linear Cryptanalysis Using Multiple Approximations and FEAL", in *Fast Software Encryption* (Ed. Bart Preneel), Lecture Notes in Computer Science 1008, pp. 249-264, Springer 1995. ISBN 3-540-60590-8.

[23] Lars R. Knudsen, "Cryptanalysis of Loki91", in *Advances in Cryptology – AUSCRYPT'92*, Lecture Notes in Computer Science 718, pp. 196-208, Springer 1993. ISBN 3-540-57220-1.

[24] Lars R. Knudsen, *Block Ciphers – Analysis, Design and Applications*, PhD Thesis, Computer Science Department, Aarhus University, Denmark, November 1994.

[25] Lars R. Knudsen, "Truncated and Higher Order Differentials", in *Fast Software Encryption* (Ed. Bart Preneel), Lecture Notes in Computer Science 1008, pp. 196-211, Springer 1995. ISBN 3-540-60590-8.

[26] Lars R. Knudsen and Willi Meier, "Improved Differential Attacks on RC5", in *Advances in Cryptology – CRYPTO'96*, Lecture Notes in Computer Science 1109, pp. 216-228, Springer 1996. ISBN 3-540-61512-1.

[27] Lars R. Knudsen and Matthew J. B. Robshaw, "Non-Linear Approximations in Linear Cryptanalysis", in *Advances in Cryptology – Eurocrypt'96*, Lecture Notes in Computer Science 1070, pp. 224-236, Springer 1996. ISBN 3-540-61186-X.

[28] Erwin Kreyszig, *Introductory Functional Analysis with Applications*, Wiley & Sons, New York, 1978.

[29] Xuejia Lai, *On the Design and Security of Block Ciphers*, Vol. 1 of ETH Series in Information Processing, Ed. J.L. Massey, Hartung-Gorre Verlag, Konstanz, 1992. ISBN 3-89191-573-X.

[30] Susan K. Langford and Martin E. Hellman, "Differential-Linear Cryptanalysis", in *Advances in Cryptology – CRYPTO'94*, Lecture Notes in Computer Science 839, pp. 17-25, Springer 1994. ISBN 3-540-58333-5.

[31] James L. Massey, *Threshold Decoding*, Cambridge, MA, MIT Press, 1963.

[32] James L. Massey, *Cryptography: Fundamentals and Applications*, copies of transparencies, Advanced Technology Seminars, 1997.

[33] James L. Massey, "SAFER K-64: A Byte-Oriented Block-Ciphering Algorithm", in *Fast Software Encryption* (Ed. Ross Anderson), Lecture Notes in Computer Science 809, pp. 1-17, Springer 1994. ISBN 3-540-58108-1.

[34] James L. Massey, "SAFER K-64: One Year Later", in *Fast Software Encryption* (Ed. Bart Preneel), Lecture Notes in Computer Science 1008, pp. 212-241, Springer 1995. ISBN 3-540-60590-8.

[35] James L. Massey, *SAFER+, Cylink Corporation's Submission for the Advanced Encryption Standard*, presentation at the First Advanced Encryption Standard Candidate Conference, Ventura, CA, August 20-22, 1998; copies of transparencies downloaded from `www.cylink.com/Internet/objects.nsf/reference/SAFERPPT/$file/SAFERPPT.pdf`.

[36] Xuejia Lai, James L. Massey, and Sean Murphy, "Markov Ciphers and Differential Cryptanalysis", in *Advances in Cryptology - Eurocrypt '91*, Lecture Notes in Computer Science 547, pp. 17-38, Springer 1991. ISBN 3-540-54620-0.

[37] Mitsuru Matsui, "On Correlation Between the Order of S-boxes and the Strength of DES", in *Advances in Cryptology – Eurocrypt'94*, Lecture Notes in Computer Science 950, pp. 366-375, Springer 1994. ISBN 3-540-60176-7.

[38] Mitsuru Matsui, "Linear cryptanalysis method for DES cipher", in *Advances in Cryptology – Eurocrypt'93*, Lecture Notes in Computer Science 765, pp. 386-397, Springer 1993. ISBN 3-540-57600-2.

[39] Mitsuru Matsui, "The first experimental cryptanalysis of the data encryption standard", in *Advances in Cryptology – CRYPTO'94*, Lecture Notes in Computer Science 839, pp. 1-11, Springer 1994. ISBN 3-540-58333-5.

[40] Mitsuru Matsui and Atsuhiro Yamagishi, "A New Method for Known Plaintext Attack of FEAL Cipher", in *Advances in Cryptology – Eurocrypt'92*, Lecture Notes in Computer Science 658, pp. 81-91, Springer 1993. ISBN 3-540-56413-6.

[41] Shōji Miyaguchi and Akihiro Shimizu, "Fast Data Encipherment Algorithm FEAL", in *Advances in Cryptology – Eurocrypt'87*, Lecture Notes in Computer Science 304, pp. 267-278, Springer 1988. ISBN 3-540-19102-X.

[42] Shōji Miyaguchi, "The FEAL-8 Cryptosystem and a Call for Attack", in *Advances in Cryptology – CRYPTO'89*, Lecture Notes in Computer Science 435, pp. 624-627, Springer 1990. ISBN 3-540-97317-6.

[43] Shōji Miyaguchi, "The FEAL Cipher Family", in *Advances in Cryptology – CRYPTO'90*, Lecture Notes in Computer Science 537, pp. 627-638, Springer 1991. ISBN 3-540-54508-5.

[44] Sean Murphy, "The Cryptanalysis of FEAL-4 with 20 Chosen Plaintexts", in *Journal of Cryptology*, Vol. 2, No. 3, pp. 145-154, 1990.

[45] Anatolij P. Prudnikov, Yurij A. Brychkov, and Oleg I. Marichev, *Integrals and Series*, Vol. 1, Gordon and Breach Science Publishers, New York, 1986.

[46] Ron Rivest, "The RC5 encryption Algorithm", in *Fast Software Encryption* (Ed. Bart Preneel), Lecture Notes in Computer Science 1008, pp. 86-96, Springer 1995. ISBN 3-540-60590-8.

[47] Press Release of the RSA Data Security Conference, San Jose, CA, January 19th, 1999.

[48] Bruce Schneier, *Applied Cryptography*, Wiley, New York 1996. ISBN 0-471-11709-9.

[49] Claude E. Shannon, "Communication Theory of Secrecy Systems", in *Bell System Technical Journal*, Vol. 28, No. 4, pp. 656-715, 1949.

[50] Gustavus Simmons, *Contemporary Cryptology: The Science of Information Integrity.* IEEE Press, Piscataway, NJ, 1992. ISBN 0-87942-277-7.

[51] Serge Vaudenay, *An Experiment on DES Statistical Cryptanalysis*, technical report LIENS-95-29 of the Laboratoire d'Informatique de l'Ecole Normale Supérieure, 1995.

[52] Doron Zeilberger, "The Method of Creative Telescoping", in *Journal of Symbolic Computation*, Vol. 11, pp. 195-204, 1991.

Lebenslauf

Ich wurde am 23. April 1970 in Chêne-Bougeries im Kanton Genf geboren. Bis 1988 wohnte ich in diesem Kanton, wo ich 1976-1981 die Primarschule und danach die Sekundar- und Mittelschule besuchte, die ich 1988 mit der Maturität Typus A (alte Sprachen) abschloss. Darauf begann ich ein Mathematikstudium an der ETH Zürich, das ich 1993 mit dem Diplom abschloss.

Nach einem kurzen Abstecher ins militärische Leben beschloss ich, meine Ausbildung mit etwas konkreterem als reiner Mathematik zu ergänzen und schrieb mich im Herbst 1995 ins Nachdiplomstudium in Informationstechnik ein. Ich kam dabei zum ersten Mal in Berührung mit der Informationstheorie. Getroffen von ihrer Schönheit fragte ich Prof. J. L. Massey, ob ich die Nachdiplomarbeit unter seiner Leitung schreiben dürfe. Nach einer gewissen Zeit und einigen erfolgreich abgelegten Prüfungen wagte ich, ihm die gleiche Frage bezüglich einer Dissertation zu stellen, und er sagte zu.

Damit wurde ich im April 1997 vom Institut für Signal- und Informationsverarbeitung angestellt und betreute zweimal die ADIT-Vorlesung von Prof. Massey. Anfang 1998 erfuhr ich, dass man die Doktorprüfung spätestens ein Jahr nach der Pensionierung des Referenten ablegen muss, also bis Ende März 1999. Statt den Referenten zu wechseln, habe ich die Herausforderung angenommen, mit dem vorliegenden Resultat.