No-Signalling Attacks and Implications for (Quantum) Nonlocality Distillation

Doctoral Dissertation submitted to the
Faculty of Informatics of the Università della Svizzera italiana
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

presented by
M.Sc. Leonhard Benno Salwey

under the supervision of
Prof. Stefan Wolf and Prof. Gilles Brassard

May 2015

Dissertation Committee

Prof. Nicolas Brunner	Université de Genève, Genève, Switzerland
Prof. Omar Fawzi	École Normale Supérieure de Lyon, Lyon, France
Prof. Marc Langheinrich	Università della Svizzera italiana, Lugano, Switzerland
Prof. Igor Pivkin	Università della Svizzera italiana, Lugano, Switzerland

15th of May 2015

Research Advisor	Co-Advisor
Prof. Stefan Wolf	**Prof. Gilles Brassard**

PhD Program Director

Prof. Igor Pivkin

I certify that except where due acknowledgement has been given, the work presented in this thesis is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; and the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program.

M.Sc. Leonhard Benno Salwey
Lugano, Yesterday May 2015

To my mother, for her unconditional love. To my father, who would have been much more proud of me than I am myself.

You can’t always get what you want...

The Rolling Stones

Abstract

The phenomenon of *nonlocality*, which can arise when *entangled* quantum systems are suitably measured, is perhaps one of the most puzzling features of quantum theory to the philosophical mind. It implies that these measurement statistics cannot be explained by *hidden variables*, as requested by Einstein, and it thus suggests that our universe may *not* be, in principle, a well-determined entity where the uncertainty we perceive in physical observations stems only from our lack of knowledge of the whole.

Besides its philosophical impact, nonlocality is also a resource for information-theoretic tasks since it implies *secrecy*: If nonlocality limits the predictive power that *any* hidden variable (in the universe) can have about some observations, then it limits in particular the predictive power of a hidden variable held by an adversary in a cryptographic scenario. We investigate whether nonlocality alone can empower two parties to perform *unconditionally secure communication* in a feasible manner when only a provably minimal set of assumptions are made for such a task to be possible — independently of the validity of any physical theory (such as quantum theory).

Nonlocality has also been of interest in the study of foundations of quantum theory and the principles that stand beyond its mathematical formalism. In an attempt to single out quantum theory within a broader set of theories, the study of nonlocality may help to point out intuitive principles that distinguish it from the rest. In theories where the limits by which quantum theory constrains the strength of nonlocality are surpassed, many "principles" on which an information theorist would rely on are shattered — one example is the hierarchy of communication complexity as the latter becomes completely trivial once a certain degree of nonlocality is overstepped.

In order to study the structure of such *super-quantum theories* — beyond their aforementioned secrecy aspects — we investigate the phenomenon of *distillation of nonlocality*, the ability to distill stronger forms of nonlocality from weaker ones. By exploiting the inherent connection between nonlocality and secrecy, we provide a novel way of deriving bounds on nonlocality-distillation protocols through an ad-

versarial view to the problem.

Acknowledgements

I would like to thank my supervisors Gilles Brassard and Stefan Wolf for their enduring patience. I would also like to thank the members of the LITQ group in Montréal and the group of Stefan Wolf in Lugano for the good times we had and, in particular, Ämin Baumeler, Arne Hansen, and Dave Touchette for their emotional, scientific, and technical support. Further thanks go to Rotem Arnon-Friedman, Omar Fawzi, and Gregor Schaumann for stimulating discussions and helping me to formalise my rather volatile thoughts.

Contents

Figures

Chapter 1

Introduction

"A modern mathematical proof is not very different from a modern machine, or a modern test setup: The simple fundamental principles are hidden and almost invisible under a mass of technical details." — Hermann Weyl

1.1 Nonlocality and foundations of quantum theory

Why is quantum theory as it is? Although very useful, is not at ease to the common scholar to swallow the fact that some classical information in real world experiments, *i.e.*, the state of a system and the choice of the tested physical property, can be assigned to a normalised positive operator acting on an Hilbert space of some dimension in the first case and a positive operator valued measure (POVM) in the second. But through the *Born rule*, they correctly predict the conditional probabilities $\mathsf{P}(a\,|\,x)$ inferred from real world experiments for the outcomes of an experiment finding the value a of the chosen state-observable combination x. In Weyl's spirit we need to ask: What are the fundamental principles behind the mathematical formalism of quantum theory, and what are the distinguishing properties of the conditional probabilities $\mathsf{P}(a\,|\,x)$ that quantum theory predicts and that we observe in nature?

One of the most puzzling features of quantum theory is *nonlocality*. Bell [Bel64] showed that the distributions $\mathsf{P}(ab\,|\,xy)$ (we call them also boxes from now on) quantum mechanics produces when several observables are measured on a bipartite system — $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ denote the choices of observables (for a fixed state) and $a \in A$ and $b \in B$ the respective measurement outcomes — cannot be described by local hidden variables, as conjectured by Einstein, Podolsky, and Rosen [EPR35]. Let us illustrate Bell's reasoning by an example. Assume that two spatially sep-

arated players, Alice and Bob, have access to a (black) box with unknown inner workings and input and output panels on both sides. Upon insertion of input x, Alice immediately obtains the output a, and similarly Bob obtains b upon insertion of y with the conditional input-output distribution $\mathsf{P}(ab\,|\,xy)$, where we assume that the inputs and outputs are binary. We assume that both players have a free choice for their inputs, which implies that the output of one party has to be independent of the input of the other party — otherwise, *e.g.*, Alice could use her part of the box first to obtain a and thus predict Bob's input y, which is in contradiction to him choosing y freely. Formally, this means that the marginal distributions of the outputs satisfy $\mathsf{P}(a\,|\,xy) = \mathsf{P}(a\,|\,x)$ and $\mathsf{P}(b\,|\,xy) = \mathsf{P}(b\,|\,y)$ or, equivalently, the equations

$$\begin{aligned} \sum_b \mathsf{P}(ab\,|\,xy) &= \sum_b \mathsf{P}(ab\,|\,xy') \quad \forall\ a, x, y, y' \\ \sum_a \mathsf{P}(ab\,|\,xy) &= \sum_a \mathsf{P}(ab\,|\,x'y) \quad \forall\ b, x, x', y\ . \end{aligned} \tag{1.1}$$

The conditions 1.1 are called *no-signalling conditions* — would they not be satisfied, the players could use the box for *instantaneous communication* and be in contradiction with Einstein's theory of relativity. Let us define the quantity

$$\mathrm{CHSH}(\mathsf{P}) := \frac{1}{4} \sum_{xy} \mathsf{P}(a \oplus b = x \cdot y\,|\,xy)\ , \tag{1.2}$$

which describes the average probability that the parity of the outputs is equal to the product of the inputs of the box. We assume now that the internal workings of the box are chosen to maximise $\mathrm{CHSH}(\mathsf{P})$. In classical probability theory, any strategy to produce an outcome a (or respectively b) can be assumed a probabilistic mixture of *deterministic* strategies. If a deterministic strategy to produce the bit a has to be no-signalling, it can be represented by a function $a(x)$ since it must be independent from y. As $a(x)$ is locally computed we refer to it also as a *local deterministic strategy*. Let us write out the winning condition that the local deterministic strategies $a(x)$ and $b(y)$ have to satisfy for the four inputs (x, y)

$$a(0) \oplus b(0) = 0 \tag{1.3}$$
$$a(1) \oplus b(0) = 0 \tag{1.4}$$
$$a(0) \oplus b(1) = 0 \tag{1.5}$$
$$a(1) \oplus b(1) = 1\ . \tag{1.6}$$

If we sum over the four equations (1.3) to (1.6) and take the parity on both sides, we obtain 0 on the left-hand side and 1 on the right-hand side: For any assignment

of $a(x)$ and $b(y)$ we arrive at a contradiction. Consequently, any local deterministic strategy $(a(x), b(y))$ can satisfy at best three of the four equations (1.3) to (1.6), which is achieved by $a(x) = b(y) = 0$, and reach at best a value of $3/4$ for the quantity defined in (1.2). This holds then also for convex combinations of such strategies and implies that any $\mathsf{P}(ab\,|\,xy)$ that is no-signalling (or local) and arises from classical probability theory must satisfy the so-called CHSH inequality [CHSH69]

$$\mathrm{CHSH}(\mathsf{P}) \leq \frac{3}{4} \,. \tag{1.7}$$

Let us assume that internally the black box can perform measurements on a joint quantum system in the state ρ_{AB}. The two players are spatially separated, hence, we may assume that on each side the box interacts locally with the respective subsystem, *i.e.*, *local* observables are measured. As mentioned above, the conditional probabilities $\mathsf{P}(ab\,|\,xy)$ that arise from quantum mechanics are distributed according to the Born rule

$$\mathsf{P}(ab\,|\,xy) = \mathrm{Tr}\left(\Pi_a^x \otimes \Pi_b^y \rho_{AB}\right) \,, \tag{1.8}$$

where Π_a^x is a projector on the subspace of eigenvalue a that corresponds to the outcome a for the chosen local observable indexed with the value x. Real-world experiments [ADR82] have convincingly shown a violation of (1.7) up to almost a value of ≈ 0.85 by making measurements of different angles of polarisation of entangled photons. This is also the maximum value quantum mechanics allows for, *i.e.*, if we require the inner workings of box $\mathsf{P}(ab\,|\,xy)$ to obey the laws of quantum mechanics, it must satisfy the so-called *Cirelson's bound* [Cir80]

$$\mathrm{CHSH}(\mathsf{P}) \leq \frac{2+\sqrt{2}}{4} \approx 0.85 \,. \tag{1.9}$$

Quantum systems that exceed the bound for local (classical) strategies (1.7) are thus said to behave *non-locally*. We are interested in the features of such *nonlocality*, which does not exist in classical probability theory, and if it exists, why it is still limited in quantum theory by the bound (1.9)?

Let us come back to our starting point and the question 'why is quantum mechanics as it is?', which we like to inquire from a purely *information-theoretic* point of view. One desires intuitive reasons why the boxes $\mathsf{P}(ab\,|\,xy)$ not allowed by the Born rule (1.8) also do not exist in nature. Reasons that stand candidates for the fundamental principles on which one can base a more intuitive axiomatic approach to quantum theory. In standard textbooks the axioms on which quantum theory is based are more or less the mathematical description of its abstract formalism. It is

largely desired to find a more intuitive and less formal axiomatic system on which to base quantum theory — as is the case for Einstein's general theory of relativity. Especially as such intuition for the underlying principles may also aid to tackle one of the great challenges in theoretical physics, to find a unified theory of quantum mechanics and Einstein's theory of relativity. One starting point, also philosophically appealing, is the *axiom of free choice* that, as mentioned above, implies that multi-partite distributions $\mathsf{P}(ab\,|\,xy)$ have to be no-signalling.[1] Of course, the latter is also necessary to avoid contradiction with relativity theory. Popescu and Rohrlich showed that there exist boxes, denoted PR in their honour, which are no-signalling yet exceed the quantum bound: they violate (1.7) up to the algebraic maximum of 1 and are defined as

$$\mathsf{PR}(ab\,|\,xy) := \begin{cases} \frac{1}{2} & \text{if } a \oplus b = x \cdot y \\ 0 & \text{otherwise .} \end{cases} \tag{1.10}$$

What are the information-theoretic consequences of the existence of such super-quantum distributions? A first answer was given in [vD99] showing that distributions as (1.10) render communication complexity trivial: using PR boxes as a resource, Alice and Bob can compute any distributed Boolean function with just one bit of classical communication. In [BBL$^+$06], this result was extended to the probabilistic setting and it was shown that noisy versions of (1.10), which we denote PR_ε,

$$\mathsf{PR}_\varepsilon(ab\,|\,xy) := \begin{cases} \frac{1-\varepsilon}{2} & \text{if } a \oplus b = x \cdot y \\ \frac{\varepsilon}{2} & \text{otherwise .} \end{cases} \tag{1.11}$$

allow Alice and Bob to compute any distributed Boolean function with arbitrarily low error probability and a constant amount of communication if $\varepsilon \lesssim 0.09$. Note that the nonlocality of PR_ε is $\mathrm{CHSH}(\mathsf{PR}_\varepsilon) = 1 - \varepsilon$, and through a *depolarisation protocol* [MAG06], by use of three bits of classical shared randomness *any* box $\mathsf{P}(ab\,|\,xy)$ with $\mathrm{CHSH}(\mathsf{P}) = 1 - \varepsilon$ can be converted into a PR_ε without communication. Thus, an axiom of *non-trivial communication complexity* would rule out any box $\mathsf{P}(ab\,|\,xy)$ with such strong nonlocality $\mathrm{CHSH}(\mathsf{P}) \gtrsim 0.91$ to exist in nature. This stimulated a

[1] Actually, the argument relies rather on the precursor of free choice, the existence of *free randomness*, where we refer to the notion of Colbeck and Renner [CR11]. Their definition of free randomness boils down to a random variable x being independent from any other random variable y upon its moment of instantiation. Of course, later other (non-free) random variables can be created which are correlated to x as, otherwise, x would not effect any process in the universe and thus not be of any interest at all. Note that such a definition requires itself already an underlying causal (pre-) ordering of events to define when a random variable lies in the future of another and thus the authors speak of "space-time variables".

series of further works with reasonably compelling principles [LPSW07], [PPK+09], [NW10], [DLR12], [FSA+13], which hold in quantum theory but are violated *exactly if* there exist boxes $\mathsf{P}(ab\,|\,xy)$ that have a stronger degree of nonlocality than allowed by quantum theory and do violate Cirelson's bound (1.9). However, none of these could recover exactly the set of quantum boxes generated by (1.8) (see Figure 1.1 for a pictographic view).

Further works showed that nonlocality can be *distilled*. Forster, Winkler, and Wolf presented a protocol, where, by using many identical no-signalling boxes $\mathsf{P}(ab\,|\,xy)$ with a certain degree of nonlocality CHSH(P), Alice and Bob can simulate a box $\hat{\mathsf{P}}(ab\,|\,xy)$ with a higher degree of nonlocality CHSH($\hat{\mathsf{P}}$) > CHSH(P) without communication [FWW09]. Boxes $\mathsf{P}(ab\,|\,xy)$ from which such distillation of nonlocality is possible have been coined *distillable*. Then in [BS09] a protocol was presented that enables Alice and Bob to use *correlated boxes* P, with CHSH(P) = $3/4 + \delta$ for an δ arbitrarily small, to simulate PR boxes with arbitrary precision. Correlated boxes are mixtures between a PR box and a box that always outputs perfectly random but correlated bits A and B for any input. Therefore, boxes arbitrarily close to the set of quantum boxes render communication complexity trivial as well. The question remained open whether potentially all post-quantum distributions $\mathsf{P}(ab\,|\,xy)$ would violate the principle of communication complexity being non-trivial. This stimulated further research on the distillation of nonlocality. Other distillation protocols were found [ABL+09], [Ras12] and this allowed to exclude more super-quantum boxes $\mathsf{P}(ab\,|\,xy)$. But indications in [ABL+09] and common intuition conjectured that PR_ε boxes are not distillable, *i.e.*, it is not possible to simulate boxes $\hat{\mathsf{P}}$ with CHSH($\hat{\mathsf{P}}$) > CHSH(PR_ε) = $1 - \varepsilon$.

Chapter 4 focuses on the limits of nonlocality distillation protocols using PR_ε. In [Sho09], it has been shown that distillation is impossible by protocols using only two PR_ε. By numerical analysis, this no-go result could be extended to protocols using up to nine PR_ε as resources [For11]. Using an unbounded number of resources, impossibility of distillation was shown in [HR10] for a slightly restricted subset of *non-adaptive* distillation protocols, where the inputs to the resources are chosen independently of other outputs. Dukaric and Wolf presented an intricate argument, which relates distillation of nonlocality to (non-interactive) *distillation of entanglement* and is thus based on the mathematical formalism of quantum theory, that showed that as long as the resources PR_ε are quantum (see (1.8) and Figure 1.1) distillation is strongly limited [DW08].

Only recently, a breakthrough was achieved by Beigi and Gohari [BG14]. Through an elaborate argument, no less intricate than the one from Ducaric and Wolf, for which they introduce an additional mathematical formalism, the authors prove *complete* impossibility of nonlocality distillation by general protocols when the resources

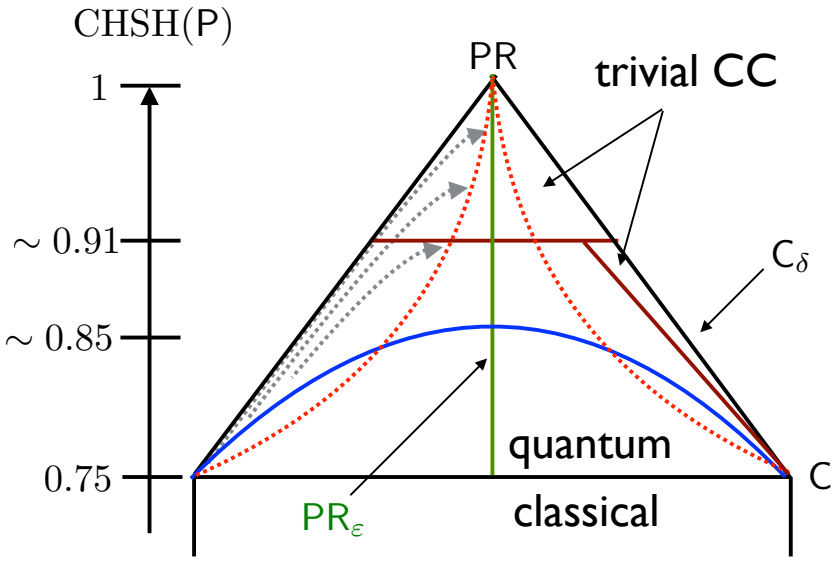

Figure 1.1. Schematic view on a slice of the no-signalling polytope and overview of previous results regarding distillation of nonlocality: The outer sides of the triangle mark the bounds that no-signalling conditions (1.1) *place on a box* $\mathsf{P}(ab\,|\,xy)$*. No-signalling conditions are, as well as positivity and normalisation constraints, linear constraints in the entries* $\mathsf{P}(ab\,|\,xy)$ *and, thus, the set of no-signalling boxes* $\mathsf{P}(ab\,|\,xy)$ *forms a polytope, i.e., the no-signalling polytope. The scale on the left side indicates the degree of nonlocality where the classical limit* (1.7)*, the quantum limit* (1.9)*, the limit beyond which communication complexity becomes trivial, and finally the algebraic maximum are displayed from bottom to top. Below the triangle the set of classical boxes is indicated. The blue curve marks the limit of the set of quantum boxes, i.e., boxes* $\mathsf{P}(ab\,|\,xy)$ *that can be generated by the Born rule* (1.8)*. Above the blue curve are the super-quantum boxes, where the two straight dark red lines, the horizontal one at* $\mathrm{CHSH}(\mathsf{P}) \approx 0.91$ *and the tilted one close to the right edge of the triangle, indicate the limit above which communication complexity has been shown to become trivial [BBL*$^+$*06], [BS09]. The right edge of the triangle is the set of correlated boxes* C_δ*, a mixture of the perfect* PR *and two perfectly random correlated bits which correspond to the bottom right corner of the triangle, where the* δ *indicates the weight of the* PR *box in the composition. A box* C_δ *can be distilled along this edge (almost) up the perfect* PR *box. Boxes above the two dotted red curves have shown to be distillable, however, the the grey curved arrows on the left indicate the direction in which they are mapped by (repeated) application of known distillation protocols [ABL*$^+$*09]. The vertical green line in the middle of the triangle represents the* PR_ε *boxes, which are proven not to be distillable (only) above the quantum bound [BG14]. Below the quantum bound distillation using* PR_ε *boxes is at best very limited [DW08].*

PR_ε are super-quantum, *i.e.*, they violate (1.9). If the players are also allowed to access classical shared randomness during the distillation protocol, their proof requires that a certain quantity named *maximal correlation* takes its minimum on PR_ε for all boxes P with $\mathrm{CHSH}(\mathsf{P}) = 1-\varepsilon$. Numerical evidence indicates that this is not the case when the PR_ε boxes are quantum. A complementary impossibility theorem for the quantum region is still due. Furthermore, it is desirable to find a proof that applies to the quantum *and* the super-quantum region *and* is preferably based on a minimal formal apparatus — according to the quote of Hermann Weyl in the beginning of the introduction.

For general distillation protocols of quantum *and* super-quantum correlations, using n PR_ε boxes as resource, a bound $\mathrm{CHSH}(\hat{\mathsf{P}}) \leq 1 - \theta(\varepsilon^{-n/2})$ can be derived by considering the so-called Elitzur-Popescu-Rohlich decomposition [EPR92] of the resource. The idea is to probabilistically decompose the resource into a nonlocal part and a local part, *i.e.*, a distribution $\mathsf{P}(ab\,|\,xy)$ satisfying (1.7), the weight of the latter being maximal. Consequently, $\hat{\mathsf{P}}$ must be local and satisfy (1.7) with the same probability weight. The drawback of this approach is that it cannot yield stronger bounds; the weight of the local part of n PR_ε is exactly of the order of $\theta(\varepsilon^{-n/2})$ [FHSW10].

We introduce a novel way to derive bounds on distillation protocols by regarding nonlocality distillation as a cryptographic game. We crucially exploit another fundamental feature of nonlocality, which we discuss in more detail in the second part of the introduction, *i.e.*, that *nonlocality provides secrecy against an outside third party*. The stronger the nonlocality $\mathrm{CHSH}(\mathsf{P})$ of a box $\mathsf{P}(ab\,|\,xy)$, the more constrained is the maximal predictive power an outside observer can have about the outputs a and b. We construct adversaries who (statistically) attack the resource boxes P of the distillation protocol and obtain a certain degree of knowledge about the output a of the distillation protocol. Thus, we obtain bounds on the degree of nonlocality $\mathrm{CHSH}(\hat{\mathsf{P}})$ of the distillate.

So far, our method has not yielded optimal bounds, such as [BG14] in the super-quantum region, but compared to other impossibility results [DW08], [BG14] our argument is simpler: Our sole formal ingredient consists in extending the resource distribution with an additional party. A main achievement of this thesis is to connect the type of distillation protocol, specifically the type of interaction of the two players with the resources, with the constraints on the attack on the resources. It turns out that a so-called *time-ordered no-signalling adversary* [AFTS12], who must respect additional no-signalling constrains between the resource systems, limits the degree of nonlocality generated by general distillation protocols. A stronger so-called *Alice-Bob no-signalling adversary*, which does not have to respect these additional constrains, limits the degree of nonlocality generated by *non-adaptive* distil-

lation protocols, where Alice and Bob interact with each of their individual resource boxes P independently. Through this stronger adversary we are able to show that for infinitely many values of ε, non-adaptive distillation using PR_ε is virtually impossible. The remainder of this thesis is dedicated to providing new constructions and more intuition for especially the time-ordered no-signalling adversary, whose study we would like to motivate now from a different, a *cryptographic* perspective.

1.2 Nonlocality and foundations of cryptography

In today's age of information, where information can be of high value, we have the need to *privately* communicate massive amounts of data all over the earth. In classical cryptography Shannon's converse theorem [Sha49] states that, if no further assumptions are made, *e.g.*, on the computational power of the adversary and the complexity of calculating certain functions, then the length of the key used to encode a message must be at least equal to the entropy of the message in order to guarantee perfect secrecy. Thus, the parties who wish to communicate secretly necessarily need a secure channel in advance in order to establish the secret key — which requires a *trusted* physical carrier of information [2] and becomes in general *infeasible* on the large scales of today's communication.

Quantum theory in turn makes the problem of large scale secret key distribution *feasible*. Only a small shared secret is needed to authenticate a classical channel between the parties with a message authentication scheme. Most of the secure communication channel can be replaced with a *completely insecure quantum communication channel*, and thus *untrusted* physical carriers of (quantum) information such as light-particles can be used to establish a long secret key. Such *quantum cryptography* goes back to the celebrated seminal work by Bennett and Brassard in 1984 [BB84]. They devised a protocol based on the exchange of single quantum bits, *e.g.*, encoded into the polarisation of single photons. The security of the protocol depends on the following assumptions (see also Figure 1.2):

- (1) The *laboratories* of the legitimate partners do not leak any information except for the communication specified by the protocol,
- (2) there is *free randomness*[3] to which the partners locally have access,

[2] In order to secure communication via *the red telephone* during cold war times the United States and the Soviet Union used diplomats with suitcases bearing magnetic tapes on which the secret keys were stored.

[3] Again, we refer to the notion of free randomness used by Colbeck and Renner in [CR11].

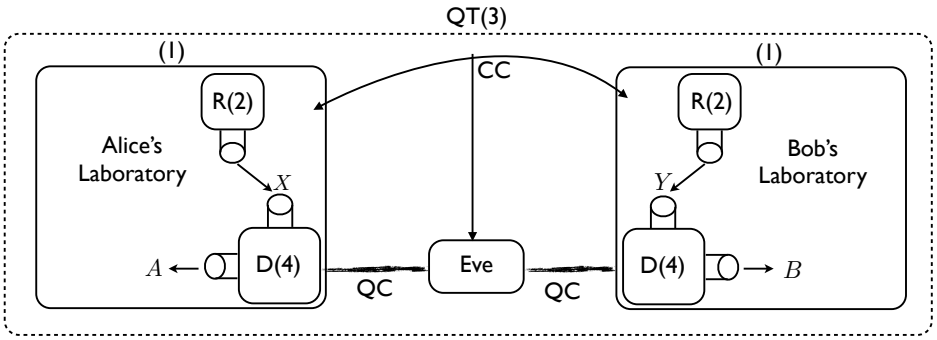

Figure 1.2. Schematic setup of Quantum Key Distribution scenarios with assumptions (1)-(4). The boxes around the legitimate parties' laboratories indicate protection against unwanted information leakage (1). The R's are the sources of free randomness (2) used as the inputs (x, y) to the devices D, which work according to their specification (4). CC refers to a classical insecure (but authenticated) channel to which the adversary Eve also has access. QC is a completely insecure quantum channel with which Eve may interfere to an unspecified extent. The dotted box indicates that the protocol takes place within the rules of quantum theory (3).

- (3) our world behaves according to *quantum theory*, and
- (4) the *devices* generate, and operate on, the specified quantum systems.

It is in the spirit of cryptography to ask for reducing the assumptions under which security can be proven. In the physics community, quantum key distribution became prominent and popular through the work of Artur Ekert [Eke91], who presented a protocol based on *entangled* pairs of quantum bits, and on the phenomenon of *nonlocality* [Bel64] introduced in the previous section. The rationale of Ekert's method is as follows (see Figure 1.3): If, after exchange and measurement on the two parts of the entangled pair, respectively, (1.7) is violated virtually up to the Cirelson's bound (1.9), then the shared state must be (close to) a maximally entangled pair of quantum bits. Furthermore, (the completeness of) quantum theory implies that the outcomes when such a *singlet* state is measured are (a) perfectly correlated with each other yet at the same time (b) completely *un*correlated with any (classical or quantum) information *outside* the two laboratories (and, hence, potentially under an adversary's control): the latter follows from a state violating maximally (1.7) necessarily being *pure*.

Ekert's result (and [MY98] when dealing with noise) has been a big step towards *device-independent* security and the possibility of dropping assumption (4) (see Fig-

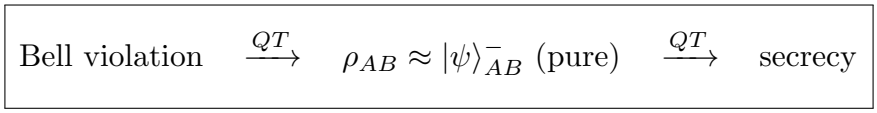

Figure 1.3. Ekert's reasoning: If a system violates the CHSH inequality virtually up to Cirelson's bound (1.9)*, then the framework of quantum theory implies that the state of the system must be close to a maximally entangled and, hence,* pure *state, a Bell state. The purity of the entangled state implies the secrecy of the local measurement outcomes. This reasoning is strongly based on the formalism of quantum theory.*

ure 1.2). Such full device-independent security [ABG+07] was finally achieved by Vazirani and Vidick [VV14] who devised a scheme, similar to Ekert's, which is not only robust against a certain level of noise, but also feasible in the terms of implementation. The key feature of their setup, also roughly depicted in Figure 1.2, is that both parties can *reuse* a single (untrusted) device to produce the raw key. We briefly sketch the rough structure of, and the intuition behind, their protocol (which is typical to quantum key distribution protocols):

1. Repeatedly, the devices and the insecure quantum channel are used to produce, distribute and measure quantum systems, *e.g.*, entangled photons, where for the choice of local measurement settings the sources of randomness are used. The values of the inputs x_i and measurement results a_i for Alice and, respectively, y_i and b_i for Bob in the i-th run are recorded and form the set $\mathcal{M} = \{(a_i, b_i, x_i, y_i)\}$.

2. Then, in a *parameter estimation* phase, the parties compute an estimated CHSH value for a single run on a representative sample: One party uses her source of randomness to randomly choose a subset $\mathcal{A}$ of the measurement statistics $\mathcal{M}$ whose size is a constant fraction of $\mathcal{M}$. Then both parties publicly compute the frequency of winning the game presented in the beginning of Section 1.1 on this subset, *i.e.*,
$$\overline{\text{CHSH}} = \frac{|\{i : a_i \oplus b_i = x_i \cdot y_i\}|}{|\mathcal{A}|} \,. \tag{1.12}$$

3. If the statistics are sufficiently non-local, *i.e.*, if the value $\overline{\text{CHSH}}$ violates (1.7), then the players can conclude that the complete set of outputs $\{(a_i, b_i)\}$ of the remaining statistics $\mathcal{M}\backslash\mathcal{A}$ cannot have been predetermined by an inner mechanism (programmed by on outside party) of the devices (following the argumentation of Section 1.1). Similar to Ekert's reasoning (see Figure 1.3),

one can conclude that it must contain some *partial* secrecy against an outside party.[4] The parties then apply a *privacy-amplification protocol*[5] using standard techniques [BBR88], [BBCM95], [HILL99] to extract an highly secure final key. In a privacy-amplification protocol, *e.g.*, Alice randomly chooses a function from a specific set that maps the partially secure string of outputs $\{a_i\}$ to a shorter, highly secure, string, which we may here assume to be a single bit. She must communicate the choice of the function to Bob such that he can apply it also to his string of outputs $\{b_i\}$ and, thus, this choice eventually becomes also known to the adversary.

However, even Viddick and Vazirani's advanced security proof, like Ekert's, rests on the validity of the entire Hilbert-space formalism of quantum theory. On the other hand, *if* we believe that quantum theory is complete, then one of the central reasons for this is, again, *nonlocality* because, as we argued in the previous section, values displaying non-local correlations are incompatible with predetermined hidden variables if (2) holds. The question is, therefore, a natural one whether it is possible to derive security of the final key *directly and only* from the (extent of) nonlocality of the generated values (see Figure 1.4), together with the assumption (1), that no hidden communication has taken place between the laboratories nor, subsequently, to the adversary. Barrett, Hardy, and Kent [BHK05] have shown that in principle, the answer is *yes*: They presented a protocol generating a secret key based only on

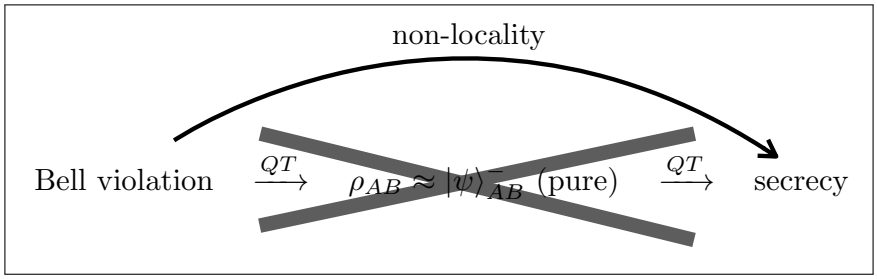

Figure 1.4. Barrett, Hardy, and Kent's reasoning: A Bell-inequality violation indicates a non-local correlation that **directly** *implies a constraint on the predictive power of any external piece of information (such as, e.g., Eve's entire knowledge) about Alice and Bob's measurement outcomes. This reasoning is completely independent of quantum theory.*

[4]This statement holds only in a sense of inference: If the protocol does not abort with non-negligible probability, then the players can conclude that with high probability their remaining statistics $\mathcal{M}\backslash\mathcal{A}$ are at least partially secret.

[5]combined with an error-correction protocol to reduce the noise in the statistics.

assumptions (1) and (2) and not on the validity of quantum theory. Note that both assumptions can (arguably) be also considered as *necessary*. If (1) does not hold, the players do not have access to secure laboratories, then any secret key can eventually leak to the adversary. If (2) does not hold, the players do not have access to private randomness, then the whole protocol is deterministic from the point of view of the adversary, and he can produce the key himself.

Barrett, Hardy, and Kent's work was a proof of principle, their protocol was neither efficient nor practical as it did not tolerate any noise. Several authors have worked on developing protocols that are based on the violation of the CHSH inequality (1.7) (instead of the chained Bell inequality [BC89] used by Barrett, Hardy, and Kent), like the above described protocol of Vazirani and Viddick, to which they are very similar. They exploit a simple relation between the violation of the CHSH inequality and the secrecy towards an outside party, *i.e.*, if $\mathrm{CHSH}(\mathsf{P}) = 1 - \varepsilon$, then *any* outside third party, bounded by quantum theory or not, can guess, *e.g.*, the output a of Alice at best with probability $1/2 + 2\varepsilon$ — which is a non-trivial bound as soon as P violates (1.7), *i.e.*, if $\varepsilon < 1/4$.

However, besides the no-signalling assumption *between* the parties, the protocols' security proofs must be based on the same condition *within* their laboratories in order to perform the final privacy-amplification step. These no-signalling conditions can only be guaranteed if both parties, Alice and Bob, produce their measurement statistics $\mathcal{M} = \{(a_i, b_i, x_i, y_i)\}$ in parallel, *i.e.*, *not* by reusing each a single device. Rather, each tuple (a_i, b_i, x_i, y_i) must be obtained from a separate device isolated in a sub-laboratory for which assumption (1) must hold individually. In today's world it seems to take a large effort to completely secure a space for a laboratory against unwanted information leakage.[6] In their protocols, the number of required sub-laboratories is proportional to the negative logarithm of the estimated ε. Thus, on one hand their protocols prove that the implementation of secret-key distribution based on only most minimalistic assumptions *is* possible, but on the other hand, they become *infeasible* when high security is required since unwanted information leakage has to be guaranteed for many sub-laboratories.

Hänggi, Renner, and Wolf showed that privacy amplification is impossible if absolutely no additional no-signalling conditions are assumed than the one between Alice and Bob [HRW13]. Yet, if Alice and Bob reuse their devices, then *previously obtained outputs cannot depend on future inputs*, since the latter are assumed to be chosen freely; the corresponding additional conditions are termed *time-ordered no-signalling* (TONS) conditions. Arnon-Friedman and Ta-Shma showed that under

[6] In the course of the Snowdon affair the German government, an entity of considerable resources, had to acknowledge that not even its chancellery was protected against unwanted information leakage.

these conditions, super-linear privacy amplification is impossible [AFTS12]: If n is the length of the input to the privacy-amplification protocol, *i.e.*, the size of the set $\mathcal{M}\backslash\mathcal{A}$, then the adversary's knowledge about the output is at least of order $\Omega(1/n)$. No stronger lower bounds on the adversary, which we simply call a *TONS adversary* if he is only bound by time-ordered no-signalling conditions, are known, and linear privacy amplification based on TONS conditions remains possible. With today's technology, quantum systems, such as light-particles, can be created, sent over large distances, and measured within fractions of a second. Therefore, it seems feasible to perform an arbitrary number of repetitions of quantum measurements and to produce a large set of measurement statistics $\mathcal{M}$ in the above described setup of Vazirani and Viddick in a short time. If linear privacy amplification based on TONS conditions can be conducted then one could possibly devise a *feasible* key distribution protocol based only on the minimal conditions (1) and (2).

In the main part of this thesis, Chapter 3, we inspect the power of a TONS adversary. We provide some evidence that there is no trivial connection between privacy amplification based on TONS conditions and a comparable scenario of classical privacy amplification where the privacy-amplification function is chosen deterministically. We show when intuitive reductions from the former to the latter case are possible, and why they are in general *not*. The key contribution of this work is a novel way to construct TONS adversaries: We introduce a set of purely classical games that enables us to analyse the adversaries possibilities with techniques from the field of analysis of Boolean functions. This permits to derive a novel lower bound of $\Omega(\log(n)/n)$ on the adversary's knowledge if *monotonic* functions are used for privacy amplification. Furthermore, finding a constant lower bound on the adversaries knowledge on *random* functions we conclude that *almost all* functions are useless for TONS privacy amplification. The class of attacks generated by our technique is considerably more powerful than the one presented by Arnon-Friedman and Ta-Shma [AFTS12], as we will show by the example of privacy amplification with *majority* functions: their attack provides with only a lower bound of $\Omega(1/\sqrt{n})$ while we obtain a constant lower bound on the adversaries knowledge (independent of n), which implies that privacy amplification cannot be achieved by such a function. We also provide evidence that the class of attacks our construction yields is sufficiently general to show the impossibility of privacy amplification based (only) on TONS conditions — *if* that truly holds: We present a completely analogous construction for setting when only no-signalling conditions between Alice and Bob are assumed and retrieve the result from [HRW13] that privacy amplification is impossible in this scenario.

Even though no stronger general lower bounds on the adversaries knowledge could be derived, the author conjectures that privacy amplification based only on

time-ordered no-signalling conditions is impossible.

Chapter 2

Preliminaries

2.1 Notation

We refer to a system as a black box with an interface consisting of an input and an output. If a system A is shared between m parties, each holding n marginal systems, then we denote the interface of the i-th marginal system held by party j by A_i^j. In the case of three parties we identify the parties with Alice, Bob and Eve ($A^1 = A, A^2 = B, A^3 = E$). Usually we denote sets in the font $\mathcal{S}$, but we also use the shorthand notation $[n] := \{1, 2, ..., n\}$ for the set of the first n natural numbers. We also use contracted indices and define the shorthand notations $A_{\leq i} := A_1 A_2 \dots A_i$ or $A_{\mathcal{S}} := A_{i_1} A_{i_2} \cdots A_{i_s}$ for $\mathcal{S} \subseteq [n]$ and $s := |\mathcal{S}|$. Also for summations we use contracted notation, *e.g.*, $\sum_{i \in \mathcal{S}}$ denotes summation over the set of indices $\{i_1, i_2, ..., i_s\} = \mathcal{S}$. The complement set of $\mathcal{S}$ is denoted $\overline{\mathcal{S}}$, when $\mathcal{S} \subseteq [n]$ then this complement is taken with respect to the set $[n]$, *i.e.*, $\overline{\mathcal{S}} := [n]/\mathcal{S}$.

We identify boxes or systems with conditional probability distributions $\mathsf{P}(ab \,|\, xy)$. For two systems A and B with inputs $x, y \in \mathcal{X} \times \mathcal{Y}$ and outputs $a, b \in \mathcal{A} \times \mathcal{B}$, $\mathsf{P}(ab \,|\, xy)$ is the probability of obtaining output (a, b) if the inputs are (x, y). The whole table of probabilities $\mathsf{P}(ab \,|\, xy)$ specifies thus the complete input-output behaviour of the systems A and B. When considering a more complicated event, *e.g.*, $f(a) = e$ on the outputs of a system AE, we define $\mathsf{P}(f(a) = e) = \sum_{a,e:f(a)=e} \mathsf{P}(ae)$.

2.2 No-signalling conditions

Intuitively, no-signalling conditions between different systems simply mean that the input one party inserts into her system does not affect the output the other party obtains from his system;

Definition 1 (m-Party no-signalling)**.** An m-system box

$$\mathsf{P}(a_1 \dots a_m \,|\, x_1 \dots x_m)$$

is *m-party no-signalling* if no subset of parties, $I^1 \subseteq [m]$, can signal to any other (disjoint) subset of parties. Defining I^2 to be the complementary set to I^1 we have formally

$$\sum_{a_{I^1}} \mathsf{P}(a_{I^1} a_{I^2} \,|\, x_{I^1} x_{I^2}) = \sum_{a_{I^1}} \mathsf{P}(a_{I^1} a_{I^2} \,|\, x'_{I^1} x_{I^2}) \quad \forall I^1, a_{I^2}, x_{I^1}, x'_{I^1}, x_{I^2} \,. \tag{2.1}$$

We will also introduce another short-hand notation and write $A_{I^1} \overset{ns}{\to} A_{I^2}$ if the systems A_{I^1} does not signal to the systems A_{I_2}, *i.e.*, they satisfy (2.1).

Definition 2 (Marginal system)**.** When (2.1) holds, it is possible to define a valid *marginal system* on the systems A_{I^2} that is independent of the inputs chosen by the parties in I^1. In that case, we denote the marginal box simply as $\mathsf{P}(a_{I^2} \,|\, x_{I^2})$.

Definition 3 (No-Signalling extension)**.** A *no-signalling extension* of a given system A (possibly consisting of arbitrarily many subsystems), identified with $\mathsf{P}(a \,|\, x)$, is any joint system AE, identified with $\mathsf{P}'(ae \,|\, xu)$, such that $A \overset{ns}{\leftrightarrow} E$ and the marginals on A coïncide, *i.e.*, $\mathsf{P}'(a \,|\, x) = \mathsf{P}(a \,|\, x)$.

Definition 4 (ABNS)**.** A $(2n+1)$-system

$$\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n} u)$$

is *Alice-Bob no-signalling (ABNS)* if, for the grouping of systems

$$\begin{aligned} A &= A_1 \cup A_2 \cup \dots \cup A_n \quad \text{and} \\ B &= B_1 \cup B_2 \cup \dots \cup B_n \,, \end{aligned} \tag{2.2}$$

the systems ABE is 3-party no-signalling.

Note that in an ABNS system $\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n} u)$, a bit a_i may not only depend on the input x_i but can also depend on all other inputs x_j, $j \neq i$. However, if Alice and Bob observe the systems $A_i B_i$ *consecutively* as results of measurements, where the choices of measurement settings x_i and y_i are free, all previous outcomes $a_{<i}$ and $b_{<i}$ must be independent of x_i and y_i, as, otherwise, x_i and y_i could be predicted. This motivates a stronger set of no-signalling constraints.

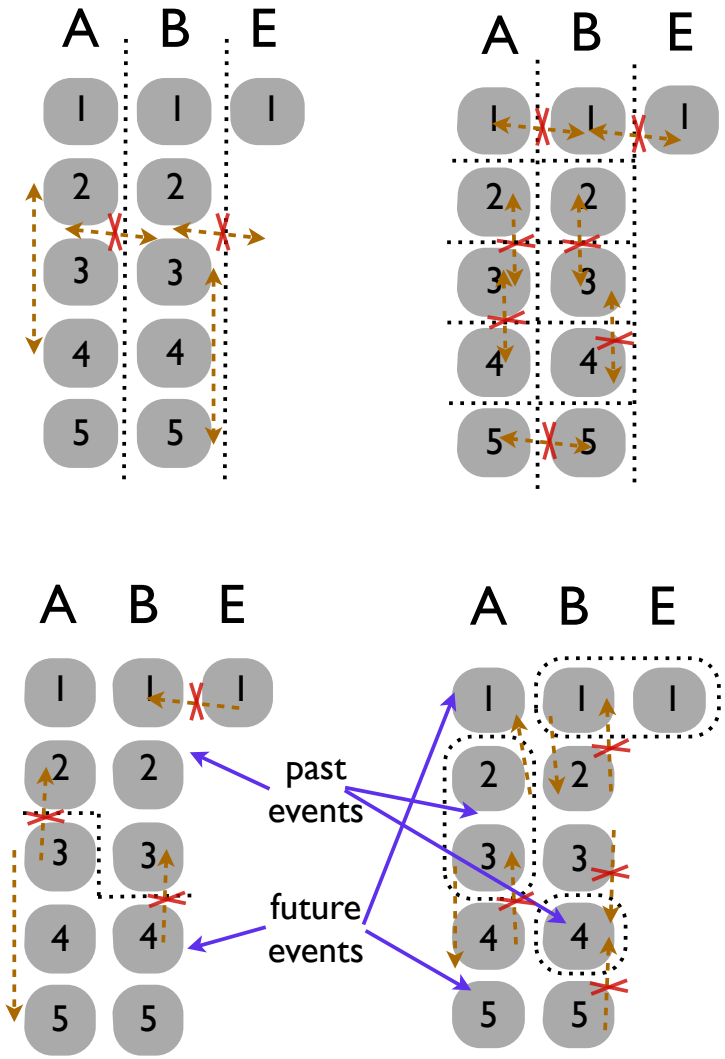

Figure 2.1. Schematic representation of various no-signalling conditions: Alice and Bob each hold $n = 5$ marginal systems, Eve a single one. Allowed directions of signalling are marked with arrows, forbidden directions of signalling with crossed arrows. On the top left are Alice-Bob no-signalling (ABNS) conditions depicted, see Definition 4, and on the top right fully no-signalling conditions, see Definition 7. In the bottom, we have two examples of time-ordered no-signalling conditions (TONS) depicted. On the left we chose $(i_A, i_B, i_E) = (2, 3, 0)$ for an explicit TONS condition. If such a condition is fulfilled, the joint output-distribution of the systems above the dotted lines is independent of the inputs to the systems below the dotted lines. On the right, we chose $(i_A, i_B, i_E) = (2, 2, 1)$ for an explicit dynamic TONS condition. Here the joint distribution of the union of systems inside the dotted lines is independent of the inputs to the systems outside the dotted lines.

Definition 5 (TONS)**.** A $(2n+1)$-system

$$\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}u)$$

is *time-ordered no-signalling (TONS)* if no subset of marginal systems can signal to systems outside its causal future. We define $A_{\leq i}$, with $0 \leq i \leq n$, to be the union of the first i marginal systems held by Alice, and $A_{>i}$ is the union of the last $n-i$ marginal systems held by Alice (and similarly for Bob an Eve). Time-ordered no-signalling is equivalent to any union of past marginal systems $A_{\leq i} \cup B_{\leq j} \cup E_{\leq k}$, with $k \in \{0,1\}$, forming a valid marginal (see Figure 2.1). Formally, this translates to the equations

$$\begin{aligned}
&\sum_{a_{>i}b_{>j}e_{\leq k}} \mathsf{P}(a_{\leq i}a_{>i}b_{\leq j}b_{>j}e_{\leq k} \,|\, x_{\leq i}x_{>i}y_{\leq j}y_{>j}u_{>k}) \\
&= \sum_{a_{>i}b_{>j}e_{\leq k}} \mathsf{P}(a_{\leq i}a_{>i}b_{\leq j}b_{>j}e_{\leq k} \,|\, x_{\leq i}x'_{>i}y_{\leq j}y'_{>j}u'_{>k}) \\
&\forall (a_{\leq i}, b_{\leq j}, x_{\leq i}, y_{\leq j}, e_{\leq k}), (x_{>i}, y_{>j}, u_{>k}), (x'_{>i}, y'_{>j}, u'_{>k}), 0 \leq i, j \leq n, k \in \{0,1\}\,. \quad (2.3)
\end{aligned}$$

Let us now consider a generalisation of the TONS conditions which we call *dynamic* TONS conditions. We relax the condition that Alice and Bob use their n systems in standard order $1, 2, ..., n$ but may use them in *any* order. This order may depend also on previously obtained outputs, *i.e.*, it may vary *dynamically*. We denote j_i (k_i) the next box Alice (Bob) uses, *i.e.*, $j_i = j_i(a_{j_1}, ..., a_{j_{i-1}})$ ($k_i = k_i(b_{k_1}, ..., b_{k_{i-1}})$). Following the short-hand notation introduced in Definition 5, we will use the contracted indices $j_{\leq i} := (j_1, ..., j_i)$ and $a_{j_{\leq i}} = (a_{j_1}, a_{j_2}, ..., a_{j_i})$. Therefore, a dynamic order is uniquely defined by the functions $\{j_1, j_2(a_{j_1}), ..., j_n(a_{j_{<n}})\}$ and $\{k_1, k_2(y, b_{k_1}), ..., k_n(b_{k_{<n}})\}$. For simplicity, we will simply write $\{j_i\}$ and $\{k_i\}$ to indicate these two sets.

Definition 6 (Dynamic TONS)**.** Let $\{j_i\}$ and $\{k_i\}$ correspond to an adaptively chosen order; A $(2n+1)$-system

$$\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}u)$$

is *dynamic time-ordered no-signalling* if no *dynamically chosen* subset of marginal systems can signal to marginal systems in its causal past. We define $A_{j_{\leq i^A}}$, with $0 \leq i^A \leq n$, to be the union of the first i^A adaptively chosen marginal systems used by Alice, *i.e.*, the systems with indices $j_1, j_2(a_{j_1}), ..., j_{i_A}(a_{j_{<i_A}})$ and $A_{j_{>i^A}}$ is the union of the last $n - i^A$ adaptively chosen marginal systems held by Alice (and similarly for Bob). Then it must hold that

$$\sum_{a_{j_{>i_A}} b_{k_{>i_B}} e_{\leq l}} \mathsf{P}(a_{j_{\leq i_A}} a_{j_{>i_A}} b_{k_{\leq i_B}} b_{k_{>i_B}} e_{\leq l} \,|\, x_{j_{\leq i_A}} x_{j_{>i_A}} y_{k_{\leq i_B}} y_{k_{>i_B}} u_{>l})$$
$$= \sum_{a_{j_{>i_A}} b_{k_{>i_B}} e_{\leq l}} \mathsf{P}(a_{j_{\leq i_A}} a_{j_{>i_A}} b_{k_{\leq i_B}} b_{k_{>i_B}} e_{\leq l} \,|\, x_{j_{\leq i_A}} x'_{j_{>i_A}} y_{k_{\leq i_B}} y'_{k_{>i_B}} u'_{>l})$$
$$\forall (a_{j_{\leq i_A}}, b_{k_{\leq i_B}}, x_{j_{\leq i_A}}, y_{k_{\leq i_A}} e_{\leq l}),$$
$$(x_{j_{>i_A}}, y_{k_{>i_B}}, u_{\leq l}), (x'_{j_{>i_A}}, y'_{k_{>i_B}} u'_{>l}), 0 \leq i_A, i_B \leq n, l \in \{0, 1\} \,. \tag{2.4}$$

Definition 7 (Fully NS)**.** A $(2n + 1)$-system

$$\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n} u)$$

is *fully no-signalling*, exactly if it is $(2n + 1)$-party no-signalling for the parties $A_1 A_2 ... A_n B_1 B_2 ... B_n E$.

It is easy to see that Definition 4, 5, 6, and 7 imply a hierarchy of no-signalling constraints; if a $(2n + 1)$-system $\mathsf{P}(a_{\leq n} b_{\leq n} \,|\, x_{\leq n} y_{\leq n})$ is fully no-signalling, then it is dynamic TONS as well as ABNS. Similarly, any (dynamic) TONS system is ABNS. Note that all no-signalling conditions are linear equations in the probabilities $\mathsf{P}(abe \,|\, xyu)$. If we interpret the box P as a vector with entries $\mathsf{P}(abe \,|\, xyu)$, we can identify a set of equations with a matrix M and a vector V. If P satisfies the set of equations, then $\mathrm{M}\mathsf{P} = \mathrm{V}$. Consequently, if a no-signalling condition is fulfilled for two boxes P^1 and P^2, then any convex combination of the two boxes $\mathsf{P} = p\mathsf{P}^1 + (1 - p)\mathsf{P}^2$ satisfies the same no-signalling condition,

$$\begin{aligned} \mathrm{M}\mathsf{P} &= \mathrm{M}(p\mathsf{P}^1 + (1 - p)\mathsf{P}^2) \\ &= p\mathrm{M}\mathsf{P}^1 + (1 - p)\mathrm{M}\mathsf{P}^2 \\ &= p\mathrm{V} + (1 - p)\mathrm{V} = \mathrm{V} \,. \end{aligned} \tag{2.5}$$

2.3 Some explicit no-signalling distributions

In this section we will simply define explicit no-signalling distributions that we make use of in the remainder of the text.

- Denote $\mathsf{D}_{a'}(a \,|\, x)$ as the box that deterministically outputs the bit a'

$$\mathsf{D}_{a'}(a \,|\, x) := \begin{cases} 1 & \text{if } a = a' \\ 0 & \text{otherwise} \,. \end{cases} \tag{2.6}$$

- Denote $\mathsf{U}(a\,|\,x)$ a box that outputs a uniformly random element of the output alphabet $\mathcal{A}$

$$\mathsf{U}(a\,|\,x) := \frac{1}{|\mathcal{A}|} \quad \forall a, x\,. \tag{2.7}$$

- Denote $\mathsf{SR}(ab\,|\,xy)$, with $\mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Y} = \{0, 1\}$, a box that outputs a uniformly random shared bit

$$\mathsf{SR}(ab\,|\,xy) := \begin{cases} \frac{1}{2} & \text{if } a \oplus b = 0 \\ 0 & \text{otherwise}\,. \end{cases} \tag{2.8}$$

- For completion, denote $\mathsf{PR}(ab\,|\,xy)$, with $\mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Y} = \{0, 1\}$, as a box with probabilities

$$\mathsf{PR}(ab\,|\,xy) := \begin{cases} \frac{1}{2} & \text{if } a \oplus b = x \cdot y \\ 0 & \text{otherwise}\,. \end{cases} \tag{2.9}$$

- Denote C_δ, called *correlated boxes*, as

$$\mathsf{C}_\delta := \delta\mathsf{PR} + (1 - \delta)\,\mathsf{SR}\,. \tag{2.10}$$

- Denote $\mathsf{V}(ab\,|\,xy)$, with $\mathcal{A} = \{0, 1\}$ and unspecified alphabets $\mathcal{B}$, $\mathcal{X}$, and $\mathcal{Y}$, as an arbitrary box that satisfies the no-signalling conditions (2.1) and has a uniform marginal on A,

$$\sum_b \mathsf{V}(ab\,|\,xy) = \frac{1}{2} \quad \forall a, x, y\,. \tag{2.11}$$

  An example for this type of boxes is the PR box, or the boxes corresponding to the chained Bell inequalities [BC89] considered in [AFTS12], but also, the system B is not specified and can be composed of an arbitrary number of subsystems, boxes corresponding to the *Guess Your Neighbours Input*-game [ABB$^+$10].

- Denote the noisy version $\mathsf{P}_\varepsilon(ab\,|\,xy)$ of an arbitrary box $\mathsf{P}(ab\,|\,xy)$ as the box with probabilities [1]

$$\mathsf{P}_\varepsilon(ab\,|\,xy) := (1 - 2\varepsilon)\,\mathsf{P}(ab\,|\,xy) + 2\varepsilon\,\mathsf{U}(ab\,|\,xy)\,. \tag{2.12}$$

[1] We chose this decomposition to be conform with the usual definition of PR_ε when P corresponds to the PR box introduced in the introduction and originally by Popescu and Rohrlich in [PR94].

Note that (2.6) and (2.7) trivially imply the relation

$$\frac{1}{|\mathcal{A}|}\sum_{a'\in\mathcal{A}}\mathsf{D}_{a'}(a\,|\,x)=\mathsf{U}(a\,|\,x)\,, \tag{2.13}$$

that (2.7) implies a factorisation property for a combined system AB

$$\mathsf{U}(ab\,|\,xy)=\mathsf{U}(a\,|\,x)\,\mathsf{U}(b\,|\,y)\,, \tag{2.14}$$

and that both marginal systems of $\mathsf{PR}(ab\,|\,xy)$ are uniform

$$\sum_{a}\mathsf{PR}(ab\,|\,xy)=\frac{1}{2}=\sum_{b}\mathsf{PR}(ab\,|\,xy)\,. \tag{2.15}$$

Chapter 3

No-Signalling Attacks

We explore the power of a no-signalling adversary, in particular, the time-ordered no-signalling (TONS) adversary. In Section 3.1, we formally define a no-signalling adversary, then, through an example of an attack on a single $\mathsf{PR}_\varepsilon(ab\,|\,xy)$ we provide intuition of the principles behind a no-signalling attack. We close Section 3.1 by illustrating the connection between nonlocality and no-signalling attacks and the limit the former puts on the latter. In Section 3.2, we introduce no-signalling privacy amplification and restate previous results regarding attacks on no-signalling privacy amplification. In Section 3.3, we first make a reference to a somewhat comparable task, deterministic privacy amplification on classical so-called ε-Santha-Vazirani distributions. In this scenario powerful attacks on privacy amplification exist, but we show that these attacks cannot be straightforwardly carried over into its no-signalling counterpart. In Section 3.4, we first introduce a novel way how to construct TONS attacks, which is a central result of this work, via Theorem 3.4.2 and Theorem 3.4.3. The construction is based on a classical privacy-amplification game that shows strong similarities to privacy amplification on ε-Santha-Vazirani distributions. These similarities and further numerical evidence lead us to believe that it is possible to use our construction to extend ε-Santha-Vazirani distributions to TONS attacks, see Conjecture 3.4.1. Then we show, as intuition suspects, that unbalanced functions do not provide more secrecy against a TONS adversary than balanced functions. In Section 3.5, we analyse our classical privacy-amplification game by means of Boolean analysis. First we show how to exclude the possibility of TONS privacy amplification by *linear hashing* (which is possible in the classical, quantum, and fully no-signalling case). We extend this result to almost all functions, by showing impossibility of TONS privacy amplification using *random* functions in Section 3.5.2. We show that our framework comprises previously known TONS attacks on privacy-amplification protocols [AFTS12], we call them *prefix-code at-*

tacks, and derive a stronger lower bound on the adversaries knowledge for these attacks for monotonic privacy-amplification functions. Finally, on the example of the *majority* functions we show that our technique can yield much stronger attacks than prefix-code attacks with a relative increase of knowledge of $\theta(\sqrt{n})$. In Section 3.6, we show how the TONS attacks constructed by our method can be generalised to dynamic TONS attacks. This will be crucial for Chapter 4, where we use dynamic TONS attacks for deriving bounds on general nonlocality distillation protocols. In Section 3.7, we turn to a stronger adversary, the Alice-Bob no-signalling (ABNS) adversary — although it was already shown that ABNS privacy amplification is impossible when the players use $\mathsf{PR}_\varepsilon(ab\,|\,xy)$ [HRW13]. We derive a construction of ABNS attacks via a classical privacy-amplification game, analogous to our construction of TONS attacks, and compare the two corresponding classical games. Then we show that this construction of ABNS attacks retrieves the strong impossibility results from [HRW13], which is an indication that *if* privacy amplification is indeed impossible (as conjectured by the author), then our construction of TONS attacks via a classical privacy-amplification game is also powerful enough to encapsulate this impossibility result. Our construction of TONS attacks and ABNS attacks on $\mathsf{PR}_\varepsilon(ab\,|\,xy)$ boxes via classical privacy-amplification games, played only on the marginal systems of Alice and Eve, also allows a straightforward generalisation to attacks on $\mathsf{V}_\varepsilon(ab\,|\,xy)$, where the B system can be arbitrary, in Section 3.8.

3.1 The no-signalling adversary and nonlocality

3.1.1 Definition of a no-signalling adversary

Assume that Alice and Bob hold a box $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$ and Alice outputs a Boolean function $f(a_{\leq n})$. To analyse the privacy of such a bit $f(a_{\leq n})$ against a no-signalling adversary, one considers, in analogy to the quantum case, an adversary Eve that holds a "no-signalling purifying marginal system" E with input U. The adversary is described by a no-signalling extension, see Definition 3, $A_{\leq n}B_{\leq n}E$ of the systems $A_{\leq n}B_{\leq n}$ where the type of no-signalling adversary is specified by the no-signalling conditions.

Definition 8 (No-signalling attack)**.** The box

$$\mathsf{P}'(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n}u)$$

is a *no-signalling attack* on in the box $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$ if $\mathsf{P}'(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n}u)$ is a no-signalling extension of $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$ and $\mathsf{P}'(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n}u)$ satisfies a

certain set of no-signalling conditions. Here, these conditions are either according to Definitions 4, 5, 6, or 7 defining an ABNS-attack, TONS or dynamic TONS-attack, or a fully NS-attack, respectively. If $n = 1$ and $\mathsf{P}'(abe\,|\,xyu)$ is no-signalling between Alice and Bob, we will just speak of a no-signalling attack.

Note that, *e.g.*, a TONS attack on a system $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$ always implies an ABNS attack on the system, since the TONS conditions imply the ABNS conditions. Obviously, the more strict no-signalling conditions are, the more constrained the adversary is. In the no-signalling attacks that we present the system E has a well-defined marginal distribution $\mathsf{P}(e\,|\,u)$ always by construction, which implies $A_{\leq n}B_{\leq n} \overset{ns}{\to} E$. The requirement that $\mathsf{P}'(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n}u)$ is an extension of $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$ implies that

$$\sum_e \mathsf{P}'(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n}u) = \mathsf{P}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n}) \quad \forall a_{\leq n}, b_{\leq n}, x_{\leq n}, y_{\leq n}, u\,, \tag{3.1}$$

and, therefore, automatically that $E \overset{ns}{\to} A_{\leq n}B_{\leq n}$. If the system $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$ itself already satisfies the respective no-signalling conditions 4, 5, 6, or 7, then the respective conditions only need to be proven on the conditional distribution $\mathsf{P}'(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n}u)$, *e.g.*, for the TONS conditions this is explicitly

$$\begin{aligned}&\sum_{a_{>i}b_{>j}} \mathsf{P}(a_{\leq i}a_{>i}b_{\leq j}b_{>j}\,|\,ex_{\leq i}x_{>i}y_{\leq j}y_{>j}u)\\ &= \sum_{a_{>i}b_{>j}} \mathsf{P}(a_{\leq i}a_{>i}b_{\leq j}b_{>j}\,|\,ex_{\leq i}x'_{>i}y_{\leq j}y'_{>j}u)\\ &\forall (a_{\leq i}, b_{\leq j}, x_{\leq i}, y_{\leq j}, u), (x_{>i}, y_{>j}), (x'_{>i}, y'_{>j}), 0 \leq i, j \leq n\,,\end{aligned} \tag{3.2}$$

and for the ABNS conditions when we restrict $i, j \in \{0, n\}$.

For no-signalling privacy amplification, we restrict Eve to insert always the same input $U = u$ and, for simplicity, suppress her input completely and consider as adversary the extension $\mathsf{P}'(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$, see Section 3.2.

3.1.2 Example - attacking a single PR_ε

Let us assume Alice and Bob share a $\mathsf{PR}_\varepsilon(ab\,|\,xy)$ box. A no-signalling adversary Eve wants to guess the output a of Alice, without ever knowing the inputs (x, y) of Alice and Bob. Our attack, a distribution $\mathsf{P}(abe\,|\,xy)$ (as usually in the following, we restrict Eve to a single possible input) on the (binary) systems ABE, is described by a marginal distribution on Eve's system and a distribution on Alice's and Bob's marginal systems, conditioned on the output of Eve:

$$\mathsf{P}(e) = \frac{1}{2} \qquad \text{and} \tag{3.3}$$

$$\mathsf{P}(ab\,|\,exy) = (1-2\varepsilon)\,\mathsf{PR}(ab\,|\,xy) + 2\varepsilon\,\mathsf{D}_e(a\,|\,x)\mathsf{U}(b\,|\,y)\,. \tag{3.4}$$

In order to show that the construction (3.3) and (3.4) is a valid no-signalling attack we need to show two properties; first that the box $\mathsf{P}(abe\,|\,xy)$ is an extension of $\mathsf{PR}_\varepsilon(ab\,|\,xy)$; second, that $\mathsf{P}(ab\,|\,exy)$ is no-signalling between Alice and Bob. Note that $AB \overset{ns}{\to} E$ follows trivially by construction (3.3). For the first part, we obtain

$$\begin{aligned}
\sum_e \mathsf{P}(abe\,|\,xy) &= \sum_e \mathsf{P}(e)\,\mathsf{P}(ab\,|\,exy) \\
&= \frac{1}{2}((1-2\varepsilon)\,\mathsf{PR}(ab\,|\,xy) + 2\varepsilon\,\mathsf{D}_0(a\,|\,x)\,\mathsf{U}(b\,|\,y)) \\
&\quad + \frac{1}{2}((1-2\varepsilon)\,\mathsf{PR}(ab\,|\,xy) + 2\varepsilon\,\mathsf{D}_1(a\,|\,x)\,\mathsf{U}(b\,|\,y)) \\
&= (1-2\varepsilon)\,\mathsf{PR}(ab\,|\,xy) + \varepsilon\,(\mathsf{D}_0(a\,|\,x) + \mathsf{D}_1(a\,|\,x))\,\mathsf{U}(b\,|\,y) \\
&= (1-2\varepsilon)\,\mathsf{PR}(ab\,|\,xy) + 2\varepsilon\,\mathsf{U}(a\,|\,x)\,\mathsf{U}(b\,|\,y) \\
&= (1-2\varepsilon)\,\mathsf{PR}(ab\,|\,xy) + 2\varepsilon\,\mathsf{U}(ab\,|\,xy) \\
&= \mathsf{PR}_\varepsilon(ab\,|\,xy)\,,
\end{aligned} \tag{3.5}$$

using (3.3) and the definitions of $\mathsf{PR}_\varepsilon(ab\,|\,xy)$, $U(a\,|\,x)$ and $D_{a'}(a\,|\,x)$ boxes in Section 2.3. For the second part, note that $\mathsf{P}(e)$ is by construction independent of (x, y). It remains to show that $\mathsf{P}(ab\,|\,exy)$ is no-signalling between Alice and Bob, *i.e.*, it has well-defined marginal distributions $\mathsf{P}(a\,|\,ex)$ and $\mathsf{P}(b\,|\,ey)$. We have

$$\begin{aligned}
\sum_b \mathsf{P}(ab\,|\,exy) &= \sum_b ((1-2\varepsilon)\,\mathsf{PR}(ab\,|\,xy) + 2\varepsilon\,\mathsf{D}_e(a\,|\,x)\mathsf{U}(b\,|\,y)) \\
&= (1-2\varepsilon)\frac{1}{2} + 2\varepsilon\,\delta(e,a) \\
&= \frac{1}{2} + \varepsilon\cdot(-1)^{e\oplus a} \\
&=: \mathsf{P}(a\,|\,ex)\,,
\end{aligned} \tag{3.6}$$

and

$$\begin{aligned}\sum_a \mathsf{P}(ab\,|\,exy) &= \sum_a \left((1-2\varepsilon)\,\mathsf{PR}(ab\,|\,xy) + 2\varepsilon\,\mathsf{D}_e(a\,|\,x)\mathsf{U}(b\,|\,y)\right)\\ &= (1-2\varepsilon)\frac{1}{2} + 2\varepsilon\frac{1}{2}\\ &= \frac{1}{2} =: \mathsf{P}(b\,|\,ey)\,.\end{aligned}$$

Furthermore, the attack yields

$$\begin{aligned}\mathsf{P}(a=e\,|\,x) &= \sum_e \mathsf{P}(e)\mathsf{P}(a=e\,|\,ex)\\ &= \frac{1}{2}\left(\mathsf{P}(a=0\,|\,e=0,x) + \mathsf{P}(a=1\,|\,e=1,x)\right)\\ &\overset{(3.6)}{=} \frac{1}{2} + \varepsilon \quad \forall x\,. \end{aligned} \tag{3.7}$$

In Section 3.1.3 we prove Theorem 3.1.3, which states that the construction (3.3) and (3.4) is also the optimal attack when the inputs (x, y) remain unknown to the adversary.

3.1.3 Limits of no-signalling attacks from nonlocality

In this section we turn around and prove upper bounds on the degree of knowledge that an adversary can obtain as a function of the nonlocality the box $\mathsf{P}(ab\,|\,xy)$ displays. We will now prove a bound on the no-signalling adversary's guessing probability of Alice's output bit of a box P with $\mathrm{CHSH}(\mathsf{P}) = 1-\varepsilon$ (note that $\mathrm{CHSH}(\mathsf{PR}_\varepsilon) = 1-\varepsilon$). This can be seen as a quantitative version of Bell's theorem, relating the strength of nonlocality to the maximal predictive power of a hidden variable (in the hand of Eve) on the outcome of the measurement. This feature is not only at the heart of secret-key-distribution protocols, but also *randomness-expansion-* and *randomness-amplification* protocols based only on no-signalling assumptions [Col06], [PAM+10], [CR12], [GMDLT+13].

Lemma 3.1.1. *If* $\mathrm{CHSH}(\mathsf{P}) \geq 1-\varepsilon$*, then the guessing probability of a no-signalling adversary, who eventually learns x, of the bit a for any input x is bounded by* $\frac{1}{2}+2\varepsilon$*. Formally, for any box* $\mathsf{P}(abe\,|\,xy)$ *which is fully no-signalling between Alice, Bob and Eve and satisfies*

$$\mathrm{CHSH}\left(\sum_e \mathsf{P}(abe\,|\,xy)\right) = 1-\varepsilon\,, \tag{3.8}$$

it holds that

$$\sum_e \mathsf{P}(e) \max_{a'} [\mathsf{P}(a = a' \,|\, ex)] \leq \frac{1}{2} + 2\varepsilon \qquad \forall x \,. \tag{3.9}$$

The proof is essentially a more explicit version of the one in [HRW13] and only added here for self-containedness.

Proof. We will first show that, for any no-signalling box $\mathsf{P}(ab \,|\, xy)$, $a, b, x, y \in \{0, 1\}$ with

$$\epsilon_{xy} := 1 - \mathsf{P}(a \oplus b = xy \,|\, xy) \,, \tag{3.10}$$

it follows that

$$\mathsf{P}(a \,|\, x) \leq \frac{1}{2} + \frac{1}{2}(\varepsilon_{00} + \varepsilon_{01} + \varepsilon_{10} + \varepsilon_{11}) \qquad \forall a, x \,. \tag{3.11}$$

By the definition of the CHSH value, see (1.2), we have

$$\mathrm{CHSH}(\mathsf{P}) = 1 - \frac{1}{4}(\varepsilon_{00} - \varepsilon_{01} - \varepsilon_{10} - \varepsilon_{11}) \,, \tag{3.12}$$

we argue that Lemma 3.1.1 follows by linearity for any box $\mathsf{P}'(abe \,|\, xy)$ that satisfies (3.8).
Let $\mathsf{P}(a = 0 \,|\, x = 0) = q$. Then it follows with Bayes' rule and no-signalling that

$$\begin{aligned} \mathsf{P}(a = 1, b = 1 \,|\, x = 0, y = 0) &= \mathsf{P}(a = 1 \,|\, x = 0) \cdot \mathsf{P}(b = 1 \,|\, a = 1, x = 0, y = 0) \\ &\leq \mathsf{P}(a = 1 \,|\, x = 0) \\ &\leq 1 - q \,, \end{aligned} \tag{3.13}$$

as the probability $\mathsf{P}(b = 1 \,|\, a = 1, x = 0, y = 0)$ is between 0 and 1. By (3.10) we have $\mathsf{P}(a = b \,|\, x = 0, y = 0) = (1 - \varepsilon_{00})$ and, therefore,

$$\begin{aligned} \mathsf{P}(a = 0, b = 0 \,|\, x = 0, y = 0) &= 1 - \varepsilon_{00} - \mathsf{P}(a = 1, b = 1 \,|\, x = 0, y = 0) \\ &\geq q - \varepsilon_{00} \\ \Rightarrow \quad \mathsf{P}(b = 0 \,|\, y = 0) &= \mathsf{P}(a = 0, b = 0 \,|\, x = 0, y = 0) \\ &\quad + \mathsf{P}(a = 1, b = 0 \,|\, x = 0, y = 0) \\ &\geq \mathsf{P}(a = 0, b = 0 \,|\, x = 0, y = 0) \\ &\geq q - \varepsilon_{00} \,. \end{aligned} \tag{3.14}$$

Repeating the same steps for $y = 1$ we obtain the bound $\mathsf{P}(b = 0\,|\,y = 1) \geq q - \varepsilon_{01}$ (note that for the existence of the marginal $\mathsf{P}(b\,|\,y)$ we implicitly assume $A \overset{ns}{\leftrightarrow} B$). Using

$$\begin{aligned}\mathsf{P}(b = 1\,|\,y = 0) &= 1 - \mathsf{P}(b = 0\,|\,y = 0)\\ &\leq 1 - q + \varepsilon_{00}\,,\end{aligned} \tag{3.15}$$

and (3.10), we obtain

$$\begin{aligned}\mathsf{P}(a = 0, b = 0\,|\,x = 1, y = 0) &= 1 - \varepsilon_{10} - \mathsf{P}(a = 1, b = 1\,|\,x = 1, y = 0)\\ &\geq 1 - \varepsilon_{10} - \mathsf{P}(b = 1\,|\,y = 0)\\ &\geq 1 - \varepsilon_{10} - 1 + q - \varepsilon_{00}\\ &= q - \varepsilon_{00} - \varepsilon_{10}\\ \Rightarrow \quad \mathsf{P}(a = 0\,|\,x = 1) &\geq q - \varepsilon_{00} - \varepsilon_{10}\,.\end{aligned} \tag{3.16}$$

Now we combine

$$\mathsf{P}(b = 1\,|\,y = 1) = 1 - \mathsf{P}(b = 0\,|\,y = 1) \leq 1 - q + \varepsilon_{01} \qquad \text{and} \tag{3.17}$$

$$\mathsf{P}(a = 1\,|\,x = 1) = 1 - \mathsf{P}(a = 0\,|\,x = 1) \leq 1 - q + \varepsilon_{00} + \varepsilon_{10}\,, \tag{3.18}$$

to conclude

$$\begin{aligned}1 - \varepsilon_{11} &= \mathsf{P}(a = 0, b = 1\,|\,x = 1, y = 1) + \mathsf{P}(a = 1, b = 0\,|\,x = 1, y = 1)\\ &\leq \mathsf{P}(b = 1\,|\,y = 1) + \mathsf{P}(a = 1\,|\,x = 1)\\ &\leq 1 - q + \varepsilon_{01} + 1 - q + \varepsilon_{00} + \varepsilon_{10}\\ \Rightarrow \quad q &\leq \frac{1}{2}(1 + \varepsilon_{00} + \varepsilon_{10} + \varepsilon_{01} + \varepsilon_{11})\,.\end{aligned} \tag{3.19}$$

The same bound can be derived by analogous reasoning for the other entries $\mathsf{P}(a\,|\,x)$. We apply the above reasoning to extensions of $\mathsf{P}(ab\,|\,xy)$:

Let E be a "no-signalling purifying system" to $\mathsf{P}(ab\,|\,xy)$, *i.e.*, a $\mathsf{P}(abe\,|\,xy)$ box that is no-signalling between A, B, and E and satisfies

$$\sum_e \mathsf{P}(e)\,\mathsf{P}(ab\,|\,xye) = \mathsf{P}(ab\,|\,xy)\,. \tag{3.20}$$

Furthermore, let us define

$$\varepsilon_{xye} := 1 - \mathsf{P}(a \oplus b = x \cdot y\,|\,exy) \qquad \text{and} \tag{3.21}$$

$$\varepsilon_e := \frac{1}{4}\sum_{xy} \varepsilon_{xye}\,. \tag{3.22}$$

From the discussion above it follows for $\mathsf{P}(ab\,|\,exy)$ that $\mathsf{P}(a\,|\,ex) \leq 1/2 + 2\varepsilon_e$. In addition, from (3.20) if follows that $\sum_e \mathsf{P}(e)\,\varepsilon_e = \varepsilon$ since for $\mathrm{CHSH}(\mathsf{P}) = 1 - \varepsilon$ it follows $\sum_{x,y} \varepsilon_{xy} = 4\varepsilon$. This implies

$$\sum_e \mathsf{P}(e) \max_a [\mathsf{P}(a\,|\,ex)] \leq \frac{1}{2} + 2\varepsilon\,. \tag{3.23}$$

which completes the proof. □

The contrapositive of Lemma 3.1.1 is the key tool we use in Chapter 4 to derive bounds on distillation protocols that produce a distribution $\mathsf{P}(ab\,|\,xy)$.

Corollary 3.1.2. *Let* $\mathsf{P}'(abe\,|\,xy)$ *be a no-signalling attack on* $\mathsf{P}(ab\,|\,xy)$*, i.e.,*

1. $\sum_e \mathsf{P}'(abe\,|\,xy) = \mathsf{P}(ab\,|\,xy)$,
2. $\mathsf{P}(abe\,|\,xy) = \mathsf{P}(e)\,\mathsf{P}(ab\,|\,exy)$ *where* $\mathsf{P}(ab\,|\,exy)$ *is no-signalling.*

If there exists an x *such that* $\mathsf{P}'(a = e\,|\,x) \geq 1/2 + 2\varepsilon$ *then* $\mathrm{CHSH}(\mathsf{P}) \leq 1 - \varepsilon$.

Note that, in (quantum) key distribution protocols it is usually assumed that (at least one of) the inputs x and y become known to the adversary. Without going into details, in the case of *randomness amplification* and *randomness expansion* protocols this can, but not necessarily have to be, the case (see, *e.g.*, [CR12] and [GMDLT+13]). Hence, we would like to finalise with an analogue to Lemma 3.1.1 for the case that the input x of Alice remains unknown to Eve but this time Alice and Bob share a PR_ε box.

Theorem 3.1.3. *Let Alice and Bob share a box* $\mathsf{PR}_\varepsilon(ab\,|\,xy)$*. Then the guessing probability, uniformly averaged over the the inputs x, of a no-signalling adversary of the bit a is bounded by* $1/2 + \varepsilon$*. Formally, we have for* any *no-signalling attack* $\mathsf{P}(abe\,|\,xy)$ *on* $\mathsf{PR}_\varepsilon(ab\,|\,xy)$ *that*

$$\sum_e \mathsf{P}(e) \max_{a'} \left[\frac{1}{2}\mathsf{P}(a = a'\,|\,e, x = 0) + \frac{1}{2}\mathsf{P}(a = a'\,|\,e, x = 1)\right] \leq \frac{1}{2} + \varepsilon\,. \tag{3.24}$$

Proof. The proof is similar to the proof of the previous Lemma 3.1.1. Assume that $\mathsf{P}(a = 0\,|\,x = 0) = q$. We show that, for any no-signalling box $\mathsf{P}(ab\,|\,xy)$, $a, b, x, y \in \{0, 1\}$ with $\epsilon_{xy} := 1 - \mathsf{P}(a = b \oplus xy\,|\,xy)$, it follows that

$$\frac{1}{2}\mathsf{P}(a = a'\,|\,x = 0) + \frac{1}{2}\mathsf{P}(a = a'\,|\,x = 1) \leq \frac{1}{2}(1 + \varepsilon_{01} + \varepsilon_{11})\,. \tag{3.25}$$

With (3.21) and

$$\sum_e \mathsf{P}(e)\,\mathsf{P}(ab\,|\,exy) = \mathsf{PR}_\varepsilon(ab\,|\,xy)\,, \tag{3.26}$$

it follows that

$$\sum_e \mathsf{P}(e)\varepsilon_{xye} = \varepsilon \quad \forall x, y, \tag{3.27}$$

and, therefore, (3.24). Analogous to the discussion around (3.13) and (3.14), we argue that

$$\begin{aligned} & \mathsf{P}(a = 0\,|\,x = 0) = q \\ \Rightarrow \quad & \mathsf{P}(b = 0\,|\,y = 1) \geq q - \varepsilon_{01}\,. \end{aligned} \tag{3.28}$$

Then we proceed with the simple implications

$$\begin{aligned} & \mathsf{P}(b = 1\,|\,y = 1) \leq 1 - q + \varepsilon_{01} \\ \Rightarrow \quad & \mathsf{P}(a = 0, b = 1\,|\,x = 1, y = 1) \leq 1 - q + \varepsilon_{01} \\ \overset{(3.10)}{\Rightarrow} \quad & \mathsf{P}(a = 1, b = 0\,|\,x = 1, y = 1) \geq 1 - \varepsilon_{11} - (1 - q + \varepsilon_{01}) \\ & \qquad = q - \varepsilon_{01} - \varepsilon_{11} \\ \Rightarrow \quad & \mathsf{P}(a = 1\,|\,x = 1) \geq q - \varepsilon_{01} - \varepsilon_{11} \\ \Rightarrow \quad & \mathsf{P}(a = 0\,|\,x = 1) \leq 1 - q + \varepsilon_{01} + \varepsilon_{11} \\ \Rightarrow \quad & \frac{1}{2}(\mathsf{P}(a = 0\,|\,x = 0) + \mathsf{P}(a = 0\,|\,x = 1)) \leq \frac{1}{2}(1 + \varepsilon_{01} + \varepsilon_{11})\,. \end{aligned} \tag{3.29}$$

Similarly, we obtain

$$\frac{1}{2}(\mathsf{P}(a = 1\,|\,x = 0) + \mathsf{P}(a = 1\,|\,x = 1)) \leq \frac{1}{2}(1 + \varepsilon_{01} + \varepsilon_{11})\,, \tag{3.30}$$

which completes the proof. □

3.2 No-signalling attacks on privacy-amplification protocols

3.2.1 The task of privacy amplification

The task of privacy amplification can be seen as follows. Suppose an adversary holding some system E can guess a single bit a_i with probability $1/2 + 2\varepsilon$, but a complete

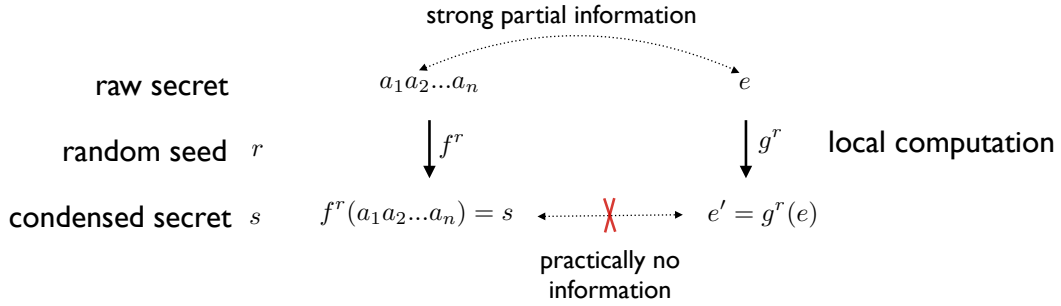

Figure 3.1. Classical privacy amplification: Alice has a string of n bits $a_1a_2...a_n$ about which adversary Eve has some partial information in form of the variable e: Eve can guess parts of the string $a_1a_2...a_n$, e.g., a single bit a_i, with probability $1/2+2\varepsilon$, but the probability to guess the whole string $a_1a_2...a_n$ is exponentially small. If she uses a random seed to choose a function randomly from a suitable set, Alice can amplify the partial privacy *of the string $a_1a_2...a_n$ and condense it into a secret bit s: even the combined knowledge about the string $a_1a_2...a_n$ and the random seed (and, therefore, also about the function Alice uses) does not allow her to compute a random variable e' that contains more than a negligible amount of information about $s = f^r(a_1a_2...a_n)$.*

bit-string $a_1 \dots a_n$ with exponentially small probability, let us say with probability at most $(1/2 + 2\varepsilon)^n$. Usually, in a privacy-amplification protocol (see Figure 3.1) one applies a randomly chosen function f^r, where r denotes the random choice, to obtain a shorter bit-string $s = f^r(a_1 \dots a_n)$ (think of a single bit) that cannot be guessed except with probability close to $1/2$. It is known that if the adversary E is governed by classical or quantum theory, it is possible to generate a single bit s that is $O((1/2 + 2\varepsilon)^{n/2})$ (*i.e.*, exponentially) -close to uniform if the function f^r is chosen uniformly amongst all linear functions [BBR88], [BBCM95], [HILL99], [Ren08]. Observe that it is possible to make this security parameter as small as we wish for any $\varepsilon < 1/4$ by increasing n.

Here we consider the same scenario with the system E being limited only by no-signalling assumptions. If a no-signalling adversary Eve attacks a single $\mathsf{PR}_\varepsilon(ab \,|\, xy)$ box, the probability to guess the output a of Alice is at best $1/2 + 2\varepsilon$, see Lemma 3.1.1. If a no-signalling adversary Eve attacks $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$, then her guessing probability for the complete string $a_1 \dots a_n$ is exponentially small, *i.e.*, of order $O((1/2 + 2\varepsilon)^{n/2})$ (we ignore a polynomial prefactor here). This follows, even for an ABNS adversary, from a threshold theorem on the parallel repetition of no-signalling games [AFRV14]. We study privacy amplification in the context of

secret-key distribution, hence Alice must communicate her choice r of the privacy-amplification function eventually to Bob such that they can arrive at a shared secret key in the end of the protocol. Since we assume that Eve can wiretap the classical communication between Alice and Bob and learn the value r, she can wait to use her system E until that happens and choose her input as a function of r, $u(r)$, accordingly. Her actions are completely specified by the box $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}u)$, which is only constrained by the no-signalling conditions: we can assume without loss of generality that any potential classical (or quantum) processing is encoded into the inner workings of the box. For any set of functions $\mathcal{F}$, with $|\mathcal{F}| = r$, we are interested in a lower bound on the maximal probability $\mathsf{P}(f^r(a_{\leq n}) = e \,|\, x_{\leq n}, u(r))$. One condition on a no-signalling attack is that the systems $A_{\leq n}B_{\leq n}$ held by Alice and Bob have the marginal distribution $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ that, in particular, must be independent of $u(r)$. Hence, each choice of r can be investigated independently and we can confine our analysis on attacks $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on a deterministically chosen function $f(a_{\leq n})$ where E has no input.

3.2.2 Previous results on no-signalling privacy amplification

In this section we present the results of Hänggi *et al.* about privacy amplification against an ABNS and a fully no-signalling adversary, as well as the result from Arnon-Friedman *et al.* against a TONS adversary.

Theorem 3.2.1. *[HRW13] Assume that Alice and Bob share* $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$. *Then, for any function* $f(a_{\leq n})$, *there exists an ABNS-attack* $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ *on the boxes* $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$, *such that*

$$\mathsf{P}(f(a_{\leq n}) = e \,|\, x) \geq \frac{1}{2} + \frac{-1 + \sqrt{1 + 64\varepsilon^2}}{32\varepsilon} \geq \frac{1}{2} + \frac{\varepsilon}{2} \,. \tag{3.31}$$

Thus, (more than constant) privacy amplification for PR_ε against an ABNS-adversary is impossible. On the other hand, for a no-signalling adversary which is much more constrained, *i.e.*, the fully no-signalling-adversary, privacy amplification is possible. If we assume that the systems $A_1A_2...A_nB_1...B_n$ must be $2n$-party no-signalling, see (2.1), then privacy amplification becomes possible even using a certain deterministic function, the parity function.

Theorem 3.2.2. *[HRW10] Assume that Alice and Bob share* $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$. *Then, for any fully NS-attack* $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ *on* $\mathsf{PR}_\varepsilon^{\otimes n}$, *it holds that*

$$\mathsf{P}\left(XOR(a_1, ..., a_n) = e \,|\, x_{\leq n}\right) \leq \frac{1}{2} + (2\varepsilon)^n \qquad \forall x_{\leq n} \,. \tag{3.32}$$

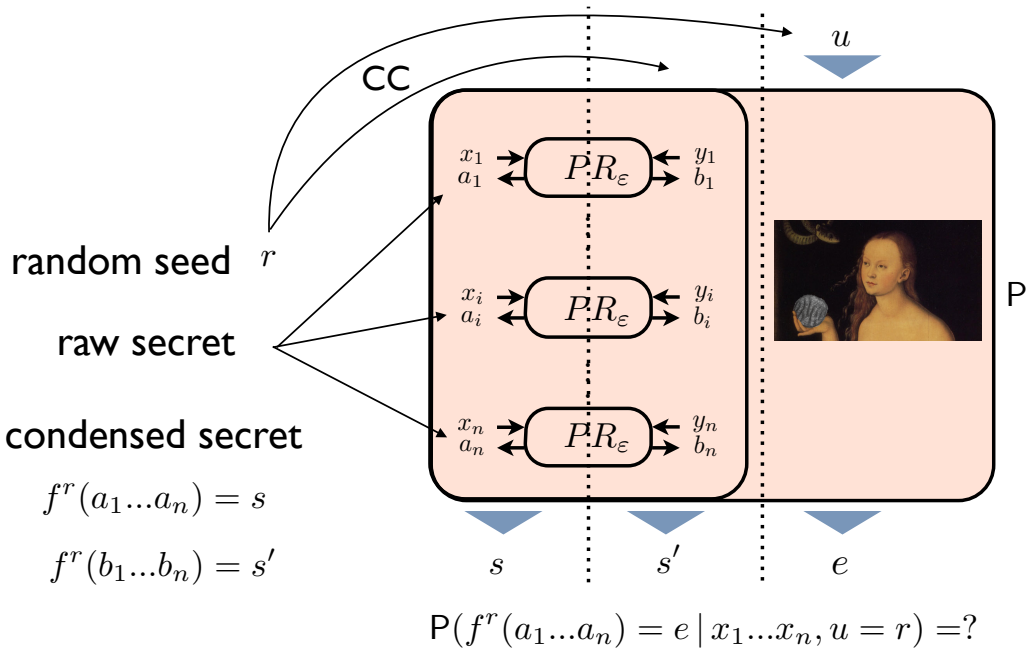

Figure 3.2. No-signalling privacy amplification: We study no-signalling privacy amplification for the case when Alice and Bob hold n PR_ε boxes. Due to the nonlocality of PR_ε, the string of outputs of the n PR_ε is partially secret for no-signalling adversary Eve. In the context of a secret-key distribution protocol Alice needs to communicate the (random) choice r of the privacy-amplification function eventually to Bob. Eve can wait to access her marginal system E until Alice communicates the choice of the privacy-amplification function r and choose her input u accordingly. The output e of the system represents her knowledge about the output of the key-distribution protocol: the final secret key $s = f^r(a_1...a_n)$.

Therefore, (exponential) fully no-signalling privacy amplification is possible, and even achieved by a deterministic function, *i.e.*, the parity of the outputs. Theorems 3.2.1 and 3.2.2 are somewhat the extreme cases of no-signalling privacy amplification. The security offered by Theorem 3.2.2 allows the construction of efficient key distribution protocols based solely on the minimal assumptions (1) and (2) (see Figure 1.2 in the Introduction) [HRW10, Mas09]. However, this comes at the price of ensuring no-signalling conditions between all systems $A_1...A_nB_1...B_n$, *i.e.*, it must be ensured that no information can be exchanged between the measurement devices during the protocol. The number of conditions grows logarithmicly with the security parameter of the protocol; this fact limits the practical feasibility of strong security parameters. On the other hand, the ABNS-attack assumes absolutely no additional no-signalling beside 3-party no-signalling between Alice, Bob and the adversary Eve. Yet, if Alice and Bob reuse their devices to produce the systems A_1B_1 to A_nB_n consecutively, then previously obtained outputs a_i cannot depend on future inputs x_j, for $j > i$, as this would contradict that x_j is freely chosen. This motivates the study of an intermediate adversary, the so-called time-ordered no-signalling (TONS) adversary.

Theorem 3.2.3. *[AFTS12] Assume that Alice and Bob share* $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$. *Then, for any function* $f(a_{\leq n})$*, there exists a TONS-attack* $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ *on* $\mathsf{PR}_\varepsilon^{\otimes n}$*, such that*

$$\mathsf{P}(f(a_{\leq n}) = e \,|\, x_{\leq n}) \geq \frac{1}{2} + \frac{\varepsilon}{2n} \qquad \forall x_{\leq n} \,. \tag{3.33}$$

In Section 3.4, Section 3.5 and Section 3.8 we aim to strengthen and generalise this result.

3.3 Time-ordered no-signalling attacks by extension of Santha-Vazirani distributions

Before discussing TONS privacy amplification, we present in Section 3.3.1 the comparable "classical" case introduced by Santa and Vazirani [SV84]: deterministic privacy amplification against a so-called Santha-Vazirani source. We do so for two reasons: First for the mere sake of comparison to inspect differences and similarities; second, as there are indications that the classical attacks in the Santha-Vazirani setting can be extended to no-signalling setting, which we discuss in Section 3.4.1. However, in Section 3.3.2 we argue that these extensions cannot be constructed in a trivial manner, yet show that they are possible for certain special cases.

3.3.1 Impossibility of deterministic (classical) privacy amplification on Santha-Vazirani distributions

We phrase classical deterministic privacy amplification on Santha-Vazirani distribution as a game of two players, Alice and Eve, which we call a *Santha-Vazirani game*. Let us define first

Definition 9 (Santha-Vazirani distribution)**.** A distribution $\mathsf{Q}(a_{\leq n}e)$ is an ε-*Santha-Vazirani distribution* if it satisfies the two properties

$$\sum_{e} \mathsf{Q}(a_{\leq n}e) = 2^{-n} \,, \tag{3.34}$$

and the so-called *"Santha-Vazirani"-condition*

$$\mathsf{Q}(a_i \,|\, a_{<i}e) \leq \frac{1}{2} + \varepsilon \qquad \forall a_{<i}, a_i, e \,. \tag{3.35}$$

From now on we denote an ε-Santha-Vazirani distribution $\mathsf{Q}(a_{\leq n}e)$ with an extra subscript and write $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ whenever (3.34) and (3.35) are satisfied. Let us state the game:

1. Alice first chooses a natural number n and a function $f(a_{\leq n}) : \{0,1\}^{\otimes n} \to \{0,1\}$ and hands it to Eve,

2. then, upon receiving $f(a_{\leq n})$, Eve chooses an ε-Santha-Vazirani distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e) : \{0,1\}^{\otimes n+1} \to [0,1]$,

3. Eve wins if $f(a_{\leq n}) = e$, Alice wins otherwise, with the winning probabilities with respect to $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$.

Note that this game is indeed comparable with no-signalling privacy amplification on $\mathsf{PR}_{\varepsilon}^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ (see Section 3.2.1). In no-signalling privacy amplification we can analyse the attack $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ also for deterministic functions $f(a_{\leq n})$. Since the marginal $\mathsf{PR}_{\varepsilon}(a_i \,|\, x_i)$ is uniform (for any x_i), the marginal of Alice systems are also uniform

$$\sum_{b_{\leq n}e} \mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = 2^{-n} \quad \forall a_{\leq n}, x_{\leq n} \,. \tag{3.36}$$

And, finally, if the input $x_{\leq n}$ remains unknown to Eve, then her knowledge about each bit is bounded, see (3.24), by

$$\mathsf{P}(e = a_i \,|\, x_{\leq n}) \leq \frac{1}{2} + \varepsilon \,. \tag{3.37}$$

The definition of a Santha-Vazirani distribution and a Santha-Vazirani game we introduce here differs from the original in [SV84]: There, Santa and Vazirani consider distributions on n bits $\mathsf{Q}(a_{\leq n})$ that satisfy

$$\mathsf{Q}(a_i \,|\, a_{<i}) \leq \frac{1}{2} + \varepsilon \quad \forall a_{<i}, a_i \,, \tag{3.38}$$

but are otherwise unconstrained. The game then has two steps; First Alice chooses the function $f(a_{\leq n})$ and communicates her choice to Eve. Then Eve chooses a bit e and a distribution $\mathsf{Q}(a_{\leq n})$. Eve wins if $\mathsf{Q}_{\varepsilon-sv}(f(a_{\leq n}) = e)$ and Alice otherwise. Clearly, the original scenario is more advantageous to Eve since in our game (3.34) has to be satisfied, but also in our game privacy amplification is impossible. By the condition (3.35) it is clear that Alice can always win the game with probability $1/2 + \varepsilon$; she chooses simply $n = 1$ and $f(a_1) = a_1$. However, this is also her optimal strategy. To see this let us present an elegant argument for *balanced* functions $f(a_{\leq n})$ by Reingold *et al.* [RVW].[1]

Definition 10 (Balanced function)**.** A function $f(a_{\leq n}) : \{0,1\}^n \to \{0,1\}$ is *balanced* if

$$|\{a_{\leq n} : f(a_{\leq n}) = 0\}| = 2^{n-1} \,. \tag{3.39}$$

Reingold *et al.* construct the distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ as

$$\mathsf{Q}_{\varepsilon-sv}(e) = \frac{1}{2} \qquad \text{and} \tag{3.40}$$

$$\mathsf{Q}_{\varepsilon-sv}(a_{\leq n} \,|\, e) = 2^{-n}\left(1 + 2\varepsilon(-1)^{(f(a_{\leq n})\oplus e}\right) \,. \tag{3.41}$$

The equations (3.40) and (3.41) define an ε-Santha-Vazirani distribution, which we refer to as the *Reingold-distribution*. (3.34) follows directly for balanced functions, to prove (3.35) notice that

$$\frac{1-2\varepsilon}{1+2\varepsilon} \leq \frac{\mathsf{Q}_{\varepsilon-sv}(a_{\leq n} \,|\, e)}{\mathsf{Q}_{\varepsilon-sv}(a'_{\leq n} \,|\, e)} \leq \frac{1+2\varepsilon}{1-2\varepsilon} \quad \forall a_{\leq n}, a'_{\leq n} \,. \tag{3.42}$$

Since $\mathsf{Q}_{\varepsilon-sv}(a_{\leq i} \,|\, e) = \sum_{a_{>i}} \mathsf{Q}_{\varepsilon-sv}(a_{\leq i}a_{>i} \,|\, e)$ this implies also

$$\frac{1-2\varepsilon}{1+2\varepsilon} \leq \frac{\mathsf{Q}_{\varepsilon-sv}(a_{\leq i} \,|\, e)}{\mathsf{Q}_{\varepsilon-sv}(a'_{\leq i} \,|\, e)} \leq \frac{1+2\varepsilon}{1-2\varepsilon} \quad \forall a_{\leq i}, a'_{\leq i} \,, \tag{3.43}$$

[1] Through an analogous construction to the one we use in the proof of Theorem 3.4.4, one can show that using an *unbalanced* function cannot provide any advantage to Alice in the above game.

and, therefore,

$$\begin{aligned}\mathsf{Q}_{\varepsilon-sv}(a_i \,|\, a_{>i}e) &= \frac{\mathsf{Q}_{\varepsilon-sv}(a_{>i}a_i \,|\, e)}{\mathsf{Q}_{\varepsilon-sv}(a_{>i} \,|\, e)} \\ &= \frac{\mathsf{Q}_{\varepsilon-sv}(a_{>i}a_i \,|\, e)}{\mathsf{Q}_{\varepsilon-sv}(a_{>i}a_i \,|\, e) + \mathsf{Q}_{\varepsilon-sv}(a_{>i}\overline{a}_i \,|\, e)} \\ &\leq \frac{1+2\varepsilon}{1+2\varepsilon+1-2\varepsilon} = \frac{1}{2} + \varepsilon \,. \end{aligned} \tag{3.44}$$

Obviously, $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ satisfies for *any* balanced function

$$\mathsf{Q}_{\varepsilon-sv}(f(a_{\leq n}) = e) = \frac{1}{2} + \varepsilon \,. \tag{3.45}$$

Consequently, this construction proves that deterministic privacy amplification on ε-Santha-Vazirani-distributions is impossible. One reason why we present the distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ defined by construction (3.40) and (3.41) in such detail, is that there is evidence, which we will discuss subsequently in this Section 3.4.1, that such a distribution can be extended to a TONS-attack $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on $\mathsf{PR}_{\varepsilon}^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$, *i.e.*,

Conjecture 3.3.1. *For any balanced function $f(a_{\leq n})$ there exists a $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$*

1. *that is a TONS attack on $\mathsf{PR}_{\varepsilon}^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$, and*
2. *has the distribution $\mathsf{Q}_{\varepsilon-sv}$ defined by* (3.40),(3.41) *as a marginal on $A_{\leq n}E$, i.e.,*

$$\mathsf{P}(a_{\leq n}e \,|\, x_{\leq n}) = \mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e) \quad \forall x_{\leq n} \,. \tag{3.46}$$

Corollary 3.3.2. *If Conjecture 3.3.1 holds then privacy amplification against a time-ordered no-signalling adversary is impossible.*

In Section 3.3.2 we discuss two instructive but unsuccessful attempts to prove Conjecture 3.3.1. In Section 3.4.1 we give indications on how such a proof might work, see Conjecture 3.4.1.

3.3.2 Limits of straightforward extensions of ε-Santha-Vazirani distributions to time-ordered no-signalling attacks

We show that (in most cases) there is no straightforward way to extend an adversary attacking an ε-Santha-Vazirani distribution to a TONS-attack on $\mathsf{PR}_{\varepsilon}^{\otimes n}$, *i.e.*, to construct an extension $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ to the ε-Santha-Vazirani distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ which is a TONS-attack on $\mathsf{PR}_{\varepsilon}^{\otimes n}$. The rough idea behind such an extension would be to combine distributions consecutively such that

1. that $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ is naturally embedded in $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$,

2. the distribution on the i-th systems A_iB_i of Alice and Bob is composed of no-signalling distributions, to ensure no-signalling between Alice and Bob, and

3. the composition of the distribution on the i-th systems A_iB_i depends only on past inputs and outputs, to ensure time-ordered no-signalling constraints from future to past systems.

Fix an ε-Santha-Vazirani distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$, *e.g.*, defined by (3.40), (3.41). Now define the sets of functions $\{t(a_{<i}, e)\}$ and $\{\tau(a_{<i}, e)\}$ by the equations

$$\mathsf{Q}_{\varepsilon-sv}(a_i \,|\, a_{<i}e) = \frac{1}{2} + (-1)^{a_i \oplus t(a_{<i},e)}\tau(a_{<i}, e) \,. \tag{3.47}$$

Since $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ satisfies (3.34) and (3.35), this definition implies the following properties for $\{t(a_{<i}, e)\}$ and $\{\tau(a_{<i}, e)\}$

$$\overset{(3.35)}{\Rightarrow} \quad \tau(a_{<i}, e) \leq \varepsilon \qquad \forall a_{<i}, e \,, \tag{3.48}$$

$$\overset{(3.34)}{\Rightarrow} \quad t(a_{<i}, e) = t(a_{<i}, \overline{e}) \oplus 1 \qquad \forall a_{<i} \,, \tag{3.49}$$

$$\overset{(3.34)}{\Rightarrow} \quad \tau(a_{<i}, e)\, \mathsf{Q}_{\varepsilon-sv}(a_{<i}e) = \tau(a_{<i}, \overline{e})\, \mathsf{Q}_{\varepsilon-sv}(a_{<i}\overline{e}) \qquad \forall a_{<i} \,. \tag{3.50}$$

Now we proceed with the first attempt to extend the distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ to a TONS attack. We define the distribution $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ as

$$\mathsf{P}(e) = \mathsf{Q}_{\varepsilon-sv}(e) \tag{3.51}$$

$$\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n}) = \prod_{i=1}^{n} \mathsf{P}(a_ib_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) \qquad \text{with} \tag{3.52}$$

$$\begin{aligned}\mathsf{P}(a_ib_i \,|\, x_iy_ia_{<i}b_{<i}e) = {} & (1 - 2\varepsilon)\, \mathsf{PR}(a_ib_i \,|\, x_iy_i) + 2(\varepsilon - \tau(a_{<i}, e))\, \mathsf{U}(a_ib_i) \\ & + 2\tau(a_{<i}, e)\, \mathsf{D}_{t(a_{<i},e)}(a_i)\, \mathsf{U}(b_i) \,.\end{aligned} \tag{3.53}$$

Theorem 3.3.3. *The box* $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ *defined by* (3.51)*,* (3.52)*, and* (3.53) *has the marginal* $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ *on systems* $A_{\leq n}E$*. Furthermore,* $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ *is a valid time-ordered no-signalling attack on* $\mathsf{PR}_{\varepsilon}^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ *if and only if for each string* $a_{\leq n}$ *there exists only a single prefix* $a_{<i}$ *such that* $\tau(a_{<i}, e) \neq 0$*.*

Proof. First of all, note that $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ is indeed a valid probability distribution; it is normalised since (3.53) is a normalised composition of again normalised probability distributions and every entry is positive due to equation (3.48). In Theorem 3.3.4 we prove that there exists a TONS attack $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on $\mathsf{PR}_\varepsilon^{\otimes n}$ with the correct marginal $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ on the systems $A_{\leq n}E$ if the condition on the $\tau(a_{<i}, e)$ holds in a more general context. Therefore, we will only briefly sketch the proof that $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ satisfies the TONS conditions (2.3) and has indeed the correct marginal on the systems $A_{\leq n}E$, *i.e.*,

$$\sum_{b_{\leq n}} \mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e) \quad \forall x_{\leq n}, y_{\leq n} \,. \tag{3.54}$$

From (3.53) we obtain by the definition of the boxes PR, U, and D in Section 2.3

$$\begin{aligned}\sum_{b_i} \mathsf{P}(a_ib_i \,|\, x_iy_ia_{<i}b_{<i}e) &= \sum_{b_i} \Big((1-2\varepsilon)\,\mathsf{PR}(a_ib_i \,|\, x_iy_i) + 2(\varepsilon - \tau(a_{<i}, e))\,\mathsf{U}(a_ib_i) \\ &\qquad + 2\tau(a_{<i}, e)\,\mathsf{D}_{t(a_{<i},e)}(a_i)\,\mathsf{U}(b_i) \Big) \\ &= (1-2\varepsilon)\frac{1}{2} + 2(\varepsilon - \tau(a_{<i}, e))\frac{1}{2} + 2\tau(a_{<i}, e)\,\delta(t(a_{<i}, e), a_i) \\ &= \frac{1}{2} + (-1)^{a_i \oplus t(a_{<i},e)}\,\tau(a_{<i}, e) \,, \end{aligned} \tag{3.55}$$

as well as

$$\begin{aligned}\sum_{a_i} \mathsf{P}(a_ib_i \,|\, x_iy_ia_{<i}b_{<i}e) &= \sum_{a_i} \Big((1-2\varepsilon)\,\mathsf{PR}(a_ib_i \,|\, x_iy_i) + 2(\varepsilon - \tau(a_{<i}, e))\,\mathsf{U}(a_ib_i) \\ &\qquad + 2\tau(a_{<i}, e)\,\mathsf{D}_{t(a_{<i},e)}(a_i)\,\mathsf{U}(b_i) \Big) \\ &= (1-2\varepsilon)\frac{1}{2} + 2(\varepsilon - \tau(a_{<i}, e))\frac{1}{2} + 2\tau(a_{<i}, e)\frac{1}{2} \\ &= \frac{1}{2} \,. \end{aligned} \tag{3.56}$$

From (3.55) and (3.56) together with the recursive construction of $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ (3.52) one can easily derive (3.54) and the TONS conditions for the distribution conditioned on system E, (3.2), by straightforwardly carrying out the respective summations. We discuss in more detail why $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ has the marginal $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ on systems $A_{\leq n}B_{\leq n}$ if and only if for each string $a_{\leq n}$ there exists at most one prefix $a_{<i}$ such that $\tau(a_{<i}, e) \neq 0$. The problem arises when we try to show that the construction (3.51) to (3.53) is an extension of the distribution $\mathsf{PR}_\varepsilon^{\otimes n}$ on the systems $A_{\leq n}B_{\leq n}$, *i.e.*,

$$\sum_{e} \mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}) \,. \tag{3.57}$$

Through a recursive argument one can see that (3.57) is true if and only if for any $1 \leq i \leq n$ it holds that

$$\sum_e \mathsf{P}(a_i b_i e \,|\, a_{<i} b_{<i} x_{<i} x_i y_{<i} y_i) = \mathsf{PR}_\varepsilon(a_i b_i \,|\, x_i y_i) \qquad \forall a_{<i}, b_{<i}, x_{<i}, y_{<i} \,. \tag{3.58}$$

For the left side of (3.58) we have

$$\begin{aligned}
&\sum_e \mathsf{P}(a_i b_i e \,|\, a_{<i} b_{<i} x_{<i} x_i y_{<i} y_i) \\
=&\sum_e \mathsf{P}(e \,|\, a_{<i} b_{<i} x_{\leq i} y_{\leq i}) \, \mathsf{P}(a_i b_i e \,|\, a_{<i} b_{<i} x_{<i} x_i y_{<i} y_i) \\
=&\sum_e \mathsf{P}(e \,|\, a_{<i} b_{<i} x_{\leq i} y_{\leq i}) \Big((1-2\varepsilon)\, \mathsf{PR}(a_i b_i \,|\, x_i y_i) + 2(\varepsilon - \tau(a_{<i}, e))\, \mathsf{U}(a_i b_i) \\
&\quad + 2\tau(a_{<i}, e)\, \mathsf{D}_{t(a_{<i},e)}(a_i)\, \mathsf{U}(b_i) \Big) \\
=&\mathsf{PR}_\varepsilon(a_i b_i \,|\, x_i y_i) + \sum_e \mathsf{P}(e \,|\, a_{<i} b_{<i} x_{\leq i} y_{\leq i})\, \tau(a_{<i}, e)\, \left(\mathsf{D}_{t(a_{<i},e)}(a_i) - \mathsf{U}(a_i)\right) \,,
\end{aligned} \tag{3.59}$$

which is equal to the right side of (3.58) exactly when the sum vanishes. Using that $\mathsf{U}(a_i) = 1/2 = 1/2(\mathsf{D}_0(a_i) + \mathsf{D}_1(a_i))$ and (3.49), we conclude that for the sum to vanish it must hold that

$$\mathsf{P}(e \,|\, a_{<i} b_{<i} x_{\leq i} y_{\leq i})\, \tau(a_{<i}, e) = \mathsf{P}(\overline{e} \,|\, a_{<i} b_{<i} x_{\leq i} y_{\leq i})\, \tau(a_{<i}, \overline{e}) \,. \tag{3.60}$$

By the recursive construction of $\mathsf{P}(a_{\leq n} b_{\leq n} | e x_{\leq n} y_{\leq n})$ (3.52), we can conclude that

$$\begin{aligned}
\mathsf{P}(e \,|\, a_{<i} b_{<i} x_{\leq i} y_{\leq i})\, \mathsf{P}(a_{<i} b_{<i} \,|\, x_{\leq i} y_{\leq i}) &= \mathsf{P}(a_{<i} b_{<i} e \,|\, x_{\leq i} y_{\leq i}) \\
&= \mathsf{P}(a_{<i} b_{<i} e \,|\, x_{<i} y_{<i}) \,,
\end{aligned} \tag{3.61}$$

and (3.60) becomes equivalent to

$$\mathsf{P}(a_{<i} b_{<i} e \,|\, x_{<i} y_{<i})\, \tau(a_{<i}, e) = \mathsf{P}(a_{<i} b_{<i} \overline{e} \,|\, x_{<i} y_{<i})\, \tau(a_{<i}, \overline{e}) \,. \tag{3.62}$$

If $\tau(a_{<i}, e) \neq 0$, then using (3.54) and (3.50), (3.62) is equivalent to

$$\frac{\mathsf{P}(a_{<i} b_{<i} e \,|\, x_{<i} y_{<i})}{\mathsf{P}(a_{<i} b_{<i} \overline{e} \,|\, x_{<i} y_{<i})} = \frac{\mathsf{P}(a_{<i} e \,|\, x_{<i})}{\mathsf{P}(a_{<i} \overline{e} \,|\, x_{<i})} = \frac{\mathsf{Q}_{\varepsilon-sv}(a_{<i} e)}{\mathsf{Q}_{\varepsilon-sv}(a_{<i} \overline{e})} \,. \tag{3.63}$$

If there exists a prefix $a_{<j}$ of $a_{<i}$ such that $\tau(a_{<j}, e) \neq 0$ then one can always find values for $b_{<i}, x_{<i}$ and $y_{<i}$ such that (3.63) does not hold: assume without loss of

generality that $a_{<j}$ is the *only* prefix of $a_{<i}$ such that $\tau(a_{<j}, e) \neq 0$, then using the construction (3.51)-(3.53) we obtain

$$\begin{aligned}
&\frac{\mathsf{P}(a_{<i}b_{<i}e \,|\, x_{<i}y_{<i})}{\mathsf{P}(a_{<i}b_{<i}\overline{e} \,|\, x_{<i}y_{<i})} \\
&= \frac{\mathsf{P}(e)\,\mathsf{P}(a_{<i}b_{<i} \,|\, e x_{<i}y_{<i})}{\mathsf{P}(\overline{e})\,\mathsf{P}(a_{<i}b_{<i} \,|\, \overline{e} x_{<i}y_{<i})} \\
&= \frac{\mathsf{P}(a_jb_j \,|\, e a_{<j}b_{<j}x_{\leq j}y_{\leq j}) \prod_{k=1,k\neq j}^{i-1} \mathsf{PR}_\varepsilon(a_kb_k \,|\, x_ky_k)}{\mathsf{P}(a_jb_j \,|\, \overline{e} a_{<j}b_{<j}x_{\leq j}y_{\leq j}) \prod_{k=1,k\neq j}^{i-1} \mathsf{PR}_\varepsilon(a_kb_k \,|\, x_ky_k)} \\
&= \frac{\mathsf{P}(a_jb_j \,|\, e a_{<j}b_{<j}x_{\leq j}y_{\leq j})}{\mathsf{P}(a_jb_j \,|\, \overline{e} a_{<j}b_{<j}x_{\leq j}y_{\leq j})} \\
&= \frac{(1-2\varepsilon)\,\mathsf{PR}(a_jb_j \,|\, x_jy_j) + 2(\varepsilon - \tau(a_{<j}, e))\,\mathsf{U}(a_jb_j) + 2\tau(a_{<j}, e)\,\mathsf{D}_{t(a_{<j},e)}(a_j)\,\mathsf{U}(b_j)}{(1-2\varepsilon)\,\mathsf{PR}(a_jb_j \,|\, x_jy_j) + 2(\varepsilon - \tau(a_{<j}, \overline{e}))\,\mathsf{U}(a_jb_j) + 2\tau(a_{<j}, \overline{e})\,\mathsf{D}_{t(a_{<j},\overline{e})}(a_j)\,\mathsf{U}(b_j)}
\end{aligned} \tag{3.64}$$

By assumption, we have

$$\mathsf{Q}_{\varepsilon-sv}(a_{<j}e) = 2^{-j} = \mathsf{Q}_{\varepsilon-sv}(a_{<j}\overline{e})\,, \tag{3.65}$$

and thus $t(a_{<j}, e) = \bar{t}(a_{<j}, \overline{e})$. Due to the $\mathsf{PR}(a_jb_j \,|\, x_jy_j)$ part, the fraction in (3.64) must take two different values depending on whether $a_j = x_j{\cdot}y_j \oplus b_j$ or $a_j \neq x_j{\cdot}y_j \oplus b_j$ and is in contradiction to (3.63).

On the other hand, if each string of bits $a_{\leq n}$ has only a *single* prefix $a_{<i}$ such that $\tau(a_{<i}e) \neq 0$, then (3.60) is true trivially for any $j \neq i$. And, since $\tau(a_{<j}, e) = 0$ for any prefix of $a_{<j}$ of $a_{<i}$ by construction (3.51)-(3.53) we obtain

$$\begin{aligned}
\mathsf{P}(a_{<i}b_{<i}e \,|\, x_{<i}y_{<i}) &= \mathsf{P}(e)\,\mathsf{PR}_\varepsilon^{\otimes i-1}(a_{<i}b_{<i} \,|\, x_{<i}y_{<i}) \\
&= \mathsf{P}(\overline{e})\,\mathsf{PR}_\varepsilon^{\otimes i-1}(a_{<i}b_{<i} \,|\, x_{<i}y_{<i}) \\
&= \mathsf{P}(a_{<i}b_{<i}\overline{e} \,|\, x_{<i}y_{<i}) \\
\Rightarrow \quad \mathsf{Q}(a_{<i}e) &= \mathsf{P}(a_{<i}e \,|\, x_{<i}) \\
&= \sum_{b_{<i}} \mathsf{P}(a_{<i}b_{<i}e \,|\, x_{<i}y_{<i}) \\
&= \sum_{b_{<i}} \mathsf{P}(a_{<i}b_{<i}\overline{e} \,|\, x_{<i}y_{<i}) \\
&= \mathsf{P}(a_{<i}\overline{e} \,|\, x_{<i}) \\
&= \mathsf{Q}(a_{<i}\overline{e})\,,
\end{aligned} \tag{3.66}$$

and, therefore, also equation (3.63). □

The $\tau(a_{<i}, e)$ satisfy (3.62) in the average over $b_{\leq n}$, see (3.50) and (3.50). This inspires another idea to extend an ε-Santha-Vazirani distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ in an analogous fashion to (3.51) to (3.53), however, using rescaled values of $\tau(a_{<i}e)$, namely $\tilde{\tau}(a_{<i}b_{<i}ex_{<i}y_{<i})$ in construction (3.67)-(3.70). For such a construction to work, such rescaled values $\{\tilde{\tau}(a_{<i}b_{<i}ex_{<i}y_{<i})\}$ should be chosen such that

1. $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ satisfies (3.60) or, equivalently, one of the equations (3.62) or (3.63), in order to guarantee that it is an extension of $\mathsf{PR}_{\varepsilon}^{\otimes n}$,

2. in the average over $b_k, ..., b_{i-1}$ the value $\tilde{\tau}(a_{<i}b_{<i}ex_{<i}y_{<i})$ becomes independent of $y_k, ..., y_{i-1}$, to ensure TONS conditions (2.3) (since $\tilde{\tau}(a_{<i}b_{<i}ex_{<i}y_{<i})$ influences the marginal distribution of a_i, see (3.55))

3. in the average over $b_{\leq n}$ the value $\tilde{\tau}(a_{<i}b_{<i}ex_{<i}y_{<i})$ equals $\tau(a_{<i}e)$ to ensure that (3.54) is satisfied.

This wish list is quite demanding, but surprisingly, there exists a set $\{\tilde{\tau}(a_{<i}b_{<i}ex_{<i}y_{<i})\}$ such that all three conditions are satisfied. These lead to the following construction for $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$:

$$\mathsf{P}(e) = \mathsf{Q}_{\varepsilon-sv}(e) \tag{3.67}$$

$$\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n}) = \prod_{i=1}^{n} \mathsf{P}(a_ib_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) \tag{3.68}$$

$$\begin{aligned}\mathsf{P}(a_ib_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) &= (1-2\varepsilon)\,\mathsf{PR}(a_ib_i \,|\, x_iy_i) + 2(\varepsilon - \tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e))\,\mathsf{U}(a_ib_i) \\ &\quad + 2\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e)\,\mathsf{D}_{t(a_{<i},e)}(a_i)\,\mathsf{U}(b_i)\end{aligned} \tag{3.69}$$

$$\text{with} \quad \tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e) := \tau(a_{<i}, e) \cdot \frac{\mathsf{Q}_{\varepsilon-sv}(e \,|\, a_{<i})}{\mathsf{P}(e \,|\, a_{<i}b_{<i}x_{<i}y_{<i})}\,. \tag{3.70}$$

The only problem of construction (3.67)-(3.70) is, that it does not necessary hold that $\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e) \leq \varepsilon$ for general functions $f(a_{\leq n})$, which implies negative entries in $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$. This leads us to the following Theorem:

Theorem 3.3.4. *Let the functions $\{\tau(a_{<i}e)\}$ and $\{t(a_{<i}e)\}$ be derived from an ε-Santha-Vazirani distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ through equation* (3.47)*. Then the distribution $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ defined by* (3.67),(3.68)*, and* (3.69) *has the marginal $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ on systems $A_{\leq n}E$. Furthermore, the distribution $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ is a valid time-ordered no-signalling attack on $\mathsf{PR}_{\varepsilon}^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ if (and only if)*

$$|\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e)| \leq \varepsilon \qquad \forall a_{<i}, b_{<i}, x_{<i}, y_{<i}, e, \tag{3.71}$$

for the function $f(a_{\leq n})$.

Proof. The proof is divided into three steps, essentially working through the three points of wish list we stated above the construction. If these three conditions hold, then the distribution $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ is a valid no-signalling attack exactly if it is a valid probability distribution, *i.e.*, it is normalised and with non-negative entries. By construction (3.67) to (3.69) $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ is normalised as it is a composed of normalised probability distributions. Therefore, $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ is a valid probability distribution exactly if it is positive, which is the case if and only if $|\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e)| \leq \varepsilon$. The latter is easy to see by choosing $a_i \oplus b_i \neq x_i y_i$ in (3.69).

1. We start by showing that $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ is an extension of $\mathsf{PR}_{\varepsilon}^{\otimes n}$, *i.e.*, (3.57). Following the same reasoning as between the equations (3.58) and (3.60), (3.57) is equivalent to

$$\mathsf{P}(e \,|\, a_{<i}b_{<i}x_{\leq i}y_{\leq i})\, \tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e) = \mathsf{P}(\overline{e} \,|\, a_{<i}b_{<i}x_{\leq i}y_{\leq i})\, \tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, \overline{e}) \,, \quad (3.72)$$

which holds by (3.70), together with (3.50).

2. We address the TONS conditions (3.2), *i.e.*, we show that $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ has a well-defined marginal distribution for all systems A_iB_j, for any $0 \leq i, j \leq n$. Analogous to obtaining (3.55) and (3.56) from (3.53), (3.69) implies

$$\sum_{b_i} \mathsf{P}(a_i b_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) = \frac{1}{2} + (-1)^{a_i \oplus t(a_{<i}, e)}\, \tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e) \,, \quad (3.73)$$

as well as

$$\sum_{a_i} \mathsf{P}(a_i b_i \,|\, x_i y_i a_{<i}b_{<i}e) = \frac{1}{2} \,, \quad (3.74)$$

and, of course,

$$\sum_{a_i b_i} \mathsf{P}(a_i b_i \,|\, x_i y_i a_{<i}b_{<i}e) = 1 \,. \quad (3.75)$$

The latter directly implies that $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ has a well-defined marginal on $A_{\leq j}B_{\leq j}$ with

$$\begin{aligned} \sum_{a_{>j}b_{>j}} \mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n}) &= \prod_{i=1}^{n} \mathsf{P}(a_i b_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) \\ &= \prod_{i=1}^{j} \mathsf{P}(a_i b_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) \\ &=: \mathsf{P}(a_{\leq j}b_{\leq j} \,|\, ex_{\leq j}y_{\leq j}) \,. \end{aligned} \quad (3.76)$$

To prove that $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ has a well-defined marginal on $A_{\leq i}B_{\leq j}$ for $i < j$, we use (3.74) and obtain

$$\begin{aligned}\sum_{a_{i+1}\ldots a_j} \mathsf{P}(a_{\leq j}b_{\leq j} \,|\, ex_{\leq j}y_{\leq j}) &= \mathsf{P}(a_{\leq i}b_{\leq i} \,|\, ex_{\leq i}y_{\leq i}) \prod_{k=i+1}^{j} \sum_{a_k} \mathsf{P}(a_k b_k \,|\, a_{<k}b_{<k}ex_{\leq k}y_{\leq k}) \\ &= 2^{-(j-i)}\, \mathsf{P}(a_{\leq i}b_{\leq i} \,|\, ex_{\leq i}y_{\leq i}) \\ &=: \mathsf{P}(a_{\leq i}b_{\leq j} \,|\, ex_{\leq i}y_{\leq j})\,. \end{aligned} \tag{3.77}$$

Finally, we will come to the case of $i > j$, which is a little more intricate to prove. To facilitate this, we need some small preparations. Using Bayes' rule and that $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}) = 2^{-n}$, see (3.34), we obtain

$$\begin{aligned}\mathsf{Q}_{\varepsilon-sv}(e \,|\, a_{<i}) &= \frac{\mathsf{Q}_{\varepsilon-sv}(a_{<i}e)}{\mathsf{Q}_{\varepsilon-sv}(a_{<i})} \\ &= 2^{i-1}\, \mathsf{Q}_{\varepsilon-sv}(a_{<i}e)\,. \end{aligned} \tag{3.78}$$

Since $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ is an extension of $\mathsf{PR}_\varepsilon^{\otimes n}$ we conclude, again with Bayes' help, that

$$\begin{aligned}\mathsf{P}(e \,|\, a_{<i}b_{<i}x_{<i}y_{<i}) &= \frac{\mathsf{P}(a_{<i}b_{<i}e \,|\, x_{<i}y_{<i})}{\mathsf{P}(a_{<i}b_{<i} \,|\, x_{<i}y_{<i})} \\ &= \mathsf{P}(e \,|\, a_{\leq j}b_{\leq j}x_{\leq j}y_{\leq j})\, \frac{\mathsf{P}(a_{j+1}\ldots a_{i-1}b_{j+1}\ldots b_{i-1} \,|\, a_{\leq j}b_{\leq j}ex_{<i}y_{<i})}{\mathsf{P}(a_{j+1}\ldots a_{i-1}b_{j+1}\ldots b_{i-1} \,|\, a_{\leq j}b_{\leq j}x_{\leq i}y_{\leq i})} \\ &= \mathsf{P}(e \,|\, a_{\leq j}b_{\leq j}x_{\leq j}y_{\leq j})\, \frac{\mathsf{P}(a_{j+1}\ldots a_{i-1}b_{j+1}\ldots b_{i-1} \,|\, a_{\leq j}b_{\leq j}ex_{<i}y_{<i})}{\prod_{k=j+1}^{i-1} \mathsf{PR}_\varepsilon(a_k b_k \,|\, x_k y_k)}\,. \end{aligned} \tag{3.79}$$

Using $\sum_{b_k} \mathsf{PR}_\varepsilon(a_k b_k \,|\, x_k y_k) = 1/2$, we obtain

$$\begin{aligned}&\sum_{b_{j+1}\ldots b_{i-1}} \mathsf{P}(a_{j+1}\ldots a_{i-1}b_{j+1}\ldots b_{i-1} \,|\, a_{\leq j}b_{\leq j}ex_{<i}y_{<i})\, \tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e) \\ \overset{(3.70)}{\Rightarrow} &= \sum_{b_{j+1}\ldots b_{i-1}} \mathsf{P}(a_{j+1}\ldots a_{i-1}b_{j+1}\ldots b_{i-1} \,|\, a_{\leq j}b_{\leq j}ex_{<i}y_{<i})\, \tau(a_{<i}e)\, \frac{\mathsf{Q}_{\varepsilon-sv}(e \,|\, a_{<i})}{\mathsf{P}(e \,|\, a_{<i}b_{<i}x_{<i}y_{<i})} \\ \overset{(3.78),(3.79)}{\Rightarrow} &= \mathsf{Q}_{\varepsilon-sv}(a_{<i}e)\, \tau(a_{<i}e)\, \frac{2^{i-1}}{\mathsf{P}(e \,|\, a_{\leq j}b_{\leq j}x_{\leq j}y_{\leq j})} \sum_{b_{j+1}\ldots b_{i-1}} \prod_{k=j+1}^{i-1} \mathsf{PR}_\varepsilon(a_k b_k \,|\, x_k y_k) \\ &= \mathsf{Q}_{\varepsilon-sv}(a_{<i}e)\, \tau(a_{<i}e)\, \frac{2^{j}}{\mathsf{P}(e \,|\, a_{\leq j}b_{\leq j}x_{\leq j}y_{\leq j})}\,. \end{aligned} \tag{3.80}$$

Now we can compute the marginal of $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ on systems A_iB_j for $i > j$ by using (3.80) consecutively $(i-j)$ times:

$$
\begin{aligned}
&\sum_{b_{j+1}\ldots b_i} \mathsf{P}(a_{\leq i}b_{\leq i} \,|\, ex_{\leq i}y_{\leq i}) \\
&= \mathsf{P}(a_{\leq j}b_{\leq j} \,|\, ex_{\leq j}y_{\leq j}) \sum_{b_{j+1}\ldots b_i} \mathsf{P}(a_{j+1}\ldots a_i b_{j+1}\ldots b_i \,|\, a_{\leq j}b_{\leq j}ex_{\leq i}y_{\leq i}) \qquad (3.81) \\
&= \mathsf{P}(a_{\leq j}b_{\leq j} \,|\, ex_{\leq j}y_{\leq j}) \sum_{b_{j+1}\ldots b_{i-1}} \Bigg(\mathsf{P}(a_{j+1}\ldots a_{i-1}b_{j+1}\ldots b_{i-1} \,|\, a_{\leq j}b_{\leq j}ex_{\leq i}y_{\leq i}) \\
&\quad \left(\frac{1}{2} + (-1)^{t(a_{<i}e)}\,\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e)\right)\Bigg) \\
&= \mathsf{P}(a_{\leq j}b_{\leq j} \,|\, ex_{\leq j}y_{\leq j}) \cdot \Bigg(\frac{1}{2} \sum_{b_{j+1}\ldots b_{i-1}} \mathsf{P}(a_{j+1}\ldots a_{i-1}b_{j+1}\ldots b_{i-1} \,|\, a_{\leq j}b_{\leq j}ex_{\leq i}y_{\leq i}) \\
&\quad + (-1)^{t(a_{<i}e)}\,\mathsf{Q}_{\varepsilon-sv}(a_{<i}e)\,\tau(a_{<i}e)\,\frac{2^j}{\mathsf{P}(e \,|\, a_{\leq j}b_{\leq j}x_{\leq j}y_{\leq j})}\Bigg) \\
&= \\
&\vdots \\
&= \mathsf{P}(a_{\leq j}b_{\leq j} \,|\, ex_{\leq j}y_{\leq j})\,\frac{2^j}{\mathsf{P}(e \,|\, a_{\leq j}b_{\leq j}x_{\leq j}y_{\leq j})} \cdot \sum_{k=j+1}^{i} 2^{-i+k}\,\mathsf{Q}_{\varepsilon-sv}(a_{<k}e)\,(-1)^{t(a_{<i}e)}\,\tau(a_{<k}e) \\
&=: \mathsf{P}(a_{\leq i}b_{\leq j} \,|\, ex_{\leq i}y_{\leq j})\,, \qquad (3.82)
\end{aligned}
$$

and, thus, the TONS conditions are satisfied for all $0 \leq i, j \leq n$.

3. Finally, we show that $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ has the marginal $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ on systems $A_{\leq n}E$:

$$
\begin{aligned}
\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e) &= \mathsf{Q}_{\varepsilon-sv}(a_{<n}e)\,\mathsf{Q}_{\varepsilon-sv}(a_n \,|\, a_{<n}e) \\
&= \mathsf{Q}_{\varepsilon-sv}(a_{<n}e)\,\frac{1}{2} + \mathsf{Q}_{\varepsilon-sv}(a_{<n}e)\,(-1)^{t(a_{<n}e)}\,\tau(a_{<k}e) \\
&\;\;\vdots \\
&= \sum_{k=1}^{n} \left(\frac{1}{2}\right)^{n-k} \mathsf{Q}_{\varepsilon-sv}(a_{<k}e)\,(-1)^{t(a_{<k}e)}\,\tau(a_{<k}e) \\
&= \mathsf{P}(e)\,\mathsf{P}(e)^{-1} \sum_{k=1}^{n} \left(\frac{1}{2}\right)^{n-k} \mathsf{Q}_{\varepsilon-sv}(a_{<k}e)\,(-1)^{t(a_{<k}e)}\,\tau(a_{<k}e) \\
&= \mathsf{P}(a_{\leq n}e)\,, \qquad (3.83)
\end{aligned}
$$

where the last equation follows by setting $i = n$ and $j = 0$ in (3.81). □

Unfortunately, $|\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}e)| \leq \varepsilon$ is in general not satisfied if we use construction (3.67)-(3.70) to extend the distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ constructed in (3.40), (3.41). One might wonder if construction (3.67)-(3.70) can be used to prove a constant lower bound on the knowledge of a TONS adversary by shrinking ε in (3.40), (3.41) by a constant factor, *i.e.*, by substitution ε with $\varepsilon' = c_1 \cdot \varepsilon$ for an arbitrarily small constant $c > 0$:

$$\mathsf{Q}_{\varepsilon'-sv}(e) = \frac{1}{2} \quad \text{and} \tag{3.84}$$

$$\mathsf{Q}_{\varepsilon'-sv}(a_{\leq n} \,|\, e) = 2^{-n}\left(1 + 2c_1 \cdot \varepsilon\,(-1)^{(f(a_{\leq n})\oplus e}\right) . \tag{3.85}$$

This implies that $\tau(a_{\leq n}e) \leq \varepsilon'$ and, therefore, that also $|\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}e)|$ could remain small enough. In the remainder of the section we will show that this is not possible. We present an example where $|\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}e)|$ becomes arbitrarily larger than ε', the factor growing by $\sqrt{n}$.

We define the majority of n bits (n odd) in the usual way

$$\mathrm{Maj}_n(a_{\leq n}) := \begin{cases} 0 & H(a_{\leq n}) \leq \frac{n-1}{2} \\ 1 & \text{otherwise} , \end{cases} \tag{3.86}$$

where $H(a_{\leq n})$ is the *Hamming weight* of the string $a_{\leq n}$.

Theorem 3.3.5. *Let $\varepsilon' = c_1\varepsilon$, and use the Rheingold construction* (3.84) *and* (3.85) *for $Maj_n(a_{\leq n})$ to construct the ε'-Santha-Vazirani distribution $\mathsf{Q}_{\varepsilon'-sv}(a_{\leq n}e)$. For sufficiently large (odd) n there always exist inputs and outputs $a_{<i}, b_{<i}, e, x_{<i}, y_{<i}$ such that*

$$\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e) > \frac{\sqrt{n}}{c_2}\,\varepsilon' = \frac{c_1}{c_2}\,\sqrt{n}\,\varepsilon, \tag{3.87}$$

for some positive constant c_2.

For the proof Theorem 3.3.5 we use the following technical Lemma:

Lemma 3.3.6. *For the construction* (3.67) *to* (3.69) *of $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$, it holds that*

$$\mathsf{P}(e \,|\, a_{\leq i}b_{\leq i}x_{\leq i}y_{\leq i}) = \mathsf{P}(e) + \frac{1}{4}\sum_{j=1}^{i} \frac{\mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{\leq j}) - \mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{<j})}{\mathsf{PR}_{\varepsilon}(a_jb_j \,|\, x_jy_j)} . \tag{3.88}$$

Proof. We rewrite (3.69) using the definition of PR_ε and that $\mathsf{U}(a_ib_i) = 1/4$ and obtain

$$\begin{aligned}\mathsf{P}(a_ib_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) &= (1-2\varepsilon)\,\mathsf{PR}(a_ib_i \,|\, x_iy_i) + 2(\varepsilon - \tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e))\,\mathsf{U}(a_ib_i) \\ &\quad + 2\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e)\,\mathsf{D}_{t(a_{<i}e)}(a_i)\,\mathsf{U}(b_i) \\ &= \mathsf{PR}_\varepsilon(a_ib_i \,|\, x_iy_i) + \frac{1}{2}(-1)^{t(a_{<i}e)\oplus a_i}\,\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e)\,. \qquad (3.89)\end{aligned}$$

With Bayes' rule and (3.47) and $\mathsf{Q}_{\varepsilon'-sv}(a_i) = 1/2$ we conclude that

$$\begin{aligned}\mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{\leq i}) &= \frac{\mathsf{Q}_{\varepsilon'-sv}(a_{\leq i}e)}{\mathsf{Q}_{\varepsilon'-sv}(a_{\leq i})} \\ &= \mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{<i})\,\frac{\mathsf{Q}_{\varepsilon'-sv}(a_i \,|\, a_{<i}e)}{\mathsf{Q}_{\varepsilon'-sv}(a_i)} \\ &= \mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{<i}) \cdot \Big(1 + (-1)^{t(a_{<i}e)\oplus a_i}\,2\tau(a_{<i}e)\Big) \\ \Rightarrow \quad \frac{1}{2}(\mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{\leq i}) - \mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{<i})) &= \tau(a_{<i}e)\,\mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{<i})\,(-1)^{t(a_{<i}e)\oplus a_i}\,. \qquad (3.90)\end{aligned}$$

With the definition of $\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}e)$,

$$\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e) = \tau(a_{<i}, e)\,\frac{\mathsf{Q}_{\varepsilon-sv}(e \,|\, a_{<i})}{\mathsf{P}(e \,|\, a_{<i}b_{<i}x_{<i}y_{<i})}\,, \qquad (3.91)$$

and (3.89) and (3.90) we obtain

$$\begin{aligned}\mathsf{P}(e \,|\, a_{\leq i}b_{\leq i}x_{\leq i}y_{\leq i}) &= \frac{\mathsf{P}(a_{\leq i}b_{\leq i}e \,|\, x_{\leq i}y_{\leq i})}{\mathsf{P}(a_{\leq i}b_{\leq i} \,|\, x_{\leq i}y_{\leq i})} \\ &= \mathsf{P}(e \,|\, a_{<i}b_{<i}x_{<i}y_{<i})\,\frac{\mathsf{P}(a_ib_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i})}{\mathsf{P}(a_ib_i \,|\, x_iy_i)} \\ &= \mathsf{P}(e \,|\, a_{<i}b_{<i}x_{<i}y_{<i}\,)\,\frac{\mathsf{PR}_\varepsilon(a_ib_i \,|\, x_iy_i) + \frac{1}{2}(-1)^{t(a_{<i}e)\oplus a_i}\,\tilde{\tau}(a_{<i}b_{<i}x_{<i}y_{<i}, e)}{\mathsf{PR}_\varepsilon(a_ib_i \,|\, x_iy_i)} \\ &= \mathsf{P}(e \,|\, a_{<i}b_{<i}x_{<i}y_{<i}) + \frac{1}{2}\frac{(-1)^{t(a_{<i}e)\oplus a_i}\,\tau(a_{<i}e)\,\mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{<i})}{\mathsf{PR}_\varepsilon(a_ib_i \,|\, x_iy_i)} \\ &\;\;\vdots \\ &= \mathsf{P}(e) + \frac{1}{4}\sum_{j=1}^{i}\frac{\mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{\leq j}) - \mathsf{Q}_{\varepsilon'-sv}(e \,|\, a_{<j})}{\mathsf{PR}_\varepsilon(a_jb_j \,|\, x_jy_j)}\,, \qquad (3.92)\end{aligned}$$

which completes the proof of Lemma 3.3.6. □

Now we will return to the proof of Theorem 3.3.5.

Proof. By assumption $\mathsf{Q}_{\varepsilon'-sv}(a_{\leq n}e)$ is computed as

$$\begin{aligned}\mathsf{Q}_{\varepsilon'-sv}(e) &= \frac{1}{2} \qquad \text{and}\\ \mathsf{Q}_{\varepsilon'-sv}(a_{\leq n}\,|\,e) &= 2^{-n}\left(1+2\varepsilon'(-1)^{f(a_{\leq n})\oplus e}\right)\,,\end{aligned} \tag{3.93}$$

which implies

$$\begin{aligned}\mathsf{Q}_{\varepsilon'-sv}(a_{\leq i}\,|\,e) &= \sum_{a_{>i}} 2^{-n}\left(1+2\varepsilon'(-1)^{f(a_{\leq n})\oplus e}\right)\\ &= 2^{-i} + 2^{-n}\cdot 2\varepsilon'\big(|\{a_{>i} : f(a_{\leq i}, a_{>i}) = e\}| - |\{a_{>i} : f(a_{\leq i}, a_{>i}) = \overline{e}\}|\big)\,.\end{aligned} \tag{3.94}$$

Then we have for the conditioned probability $\mathsf{Q}_{\varepsilon'-sv}(e\,|\,a_{\leq i})$

$$\begin{aligned}\mathsf{Q}_{\varepsilon'-sv}(e\,|\,a_{\leq i}) &= \frac{\mathsf{Q}_{\varepsilon'-sv}(e)}{\mathsf{Q}_{\varepsilon'-sv}(a_{\leq i})}\,\mathsf{Q}_{\varepsilon'-sv}(a_{\leq i}\,|\,e)\\ &= 2^{i-1}\,\mathsf{Q}_{\varepsilon'-sv}(a_{\leq i}\,|\,e)\\ &= \frac{1}{2}\left(1+2^{-n+i}\cdot 2\varepsilon'\big(|\{a_{>i} : f(a_{\leq i}, a_{>i}) = e\}| - |\{a_{>i} : f(a_{\leq i}, a_{>i}) = \overline{e}\}|\big)\right)\,.\end{aligned} \tag{3.95}$$

Consider the string $a'_{\leq n} = 1(01)^{\frac{n-1}{2}}$ where here $(01)^{\frac{n-1}{2}}$ is be interpreted as the $\frac{n-1}{2}$-fold concatenation of the string 01. We have for even i that $H(a'_{\leq i}) = i/2$, and for $f(a_{\leq n}) = \mathrm{Maj}_n(a_{\leq n})$ that

$$\begin{aligned}&|\{a_{>i} : \mathrm{Maj}_n(a'_{\leq i}, a_{>i}) = 0\}| - |\{a_{>i} : \mathrm{Maj}_n(a'_{\leq i}, a_{>i}) = 1\}|\\ =&|\{a_{>i} : H(a_{>i}) \leq \frac{n-i-1}{2}\}| - |\{a_{>i} : H(a_{>i}) \geq \frac{n-i+1}{2}\}|\\ =&\sum_{k=0}^{\frac{n-i-1}{2}}\binom{n-i}{k} - \sum_{k=\frac{n-i-1}{2}}^{n-i}\binom{n-i}{k}\\ =&0\,,\end{aligned} \tag{3.96}$$

using $\binom{n}{k} = \binom{n}{n-k}$. And if i is odd, we obtain $H(a'_{\leq i}) = (i+1)/2$ and

$$\begin{aligned}&|\{a_{>i} : \mathrm{Maj}_n(a'_{\leq i}, a_{>i}) = 0\}| - |\{a_{>i} : \mathrm{Maj}_n(a'_{\leq i}, a_{>i}) = 1\}|\\ =&\sum_{k=0}^{\frac{n-i}{2}-1}\binom{n-i}{k} - \sum_{k=\frac{n-i}{2}}^{n-i}\binom{n-i}{k}\\ =&(-1)\binom{n-i}{\frac{n-i}{2}}\,.\end{aligned} \tag{3.97}$$

We insert (3.96) and (3.97) into (3.95) and conclude

$$\mathsf{Q}_{\varepsilon'-sv}(e=0\,|\,a'_{\leq i}) = \begin{cases} \frac{1}{2} & i \text{ even} \\ \frac{1}{2}\left(1-2^{-n+i}\cdot 2\varepsilon'^{\binom{n-i}{\frac{n-i}{2}}}\right) & i \text{ odd} \end{cases}$$

$$\Downarrow \tag{3.98}$$

$$\mathsf{Q}_{\varepsilon'-sv}(e=0\,|\,a'_{\leq i}) - \mathsf{Q}_{\varepsilon-sv}(e=0\,|\,a'_{<i}) = \varepsilon' \cdot \begin{cases} 2^{-n+i-1}\binom{n-i+1}{\frac{n-i+1}{2}} & i \text{ even} \\ -2^{-n+i}\binom{n-i}{\frac{n-i}{2}} & i \text{ odd}\,. \end{cases} \tag{3.99}$$

We compute $\tau(a'_{<n}, e=0)$ with (3.90), (3.98), and (3.99)

$$\begin{aligned} \tau(a'_{<n}, e=0) &= \frac{(-1)^{t(a_{<n},e=0)\oplus a'_n}\left(\mathsf{Q}_{\varepsilon'-sv}(e=0\,|\,a'_{\leq n}) - \mathsf{Q}_{\varepsilon'-sv}(e=0\,|\,a'_{<n})\right)}{2\mathsf{Q}_{\varepsilon'-sv}(e=0\,|\,a'_{<n})} \\ &= \varepsilon'\,, \end{aligned} \tag{3.100}$$

which implies for $\tilde{\tau}(a'_{<n}, b_{<n}, x_{<n}, y_{<n})$, defined in (3.70),

$$\begin{aligned} \tilde{\tau}(a'_{<n}, b_{<n}, x_{<n}, y_{<n}, e=0) &= \tau(a'_{<n}, e)\,\frac{\mathsf{Q}_{\varepsilon'-sv}(e=0\,|\,a'_{<n})}{\mathsf{P}(e=0\,|\,a'_{<n}b_{<n}x_{<n}y_{<n})} \\ &= \frac{\varepsilon'}{2\mathsf{P}(e=0\,|\,a'_{<n}b_{<n}x_{<n}y_{<n})}\,. \end{aligned} \tag{3.101}$$

We show now that we can choose $b'_{\leq i}, x'_{\leq i}, y'_{\leq i}$ such that $|\mathsf{P}(e=0\,|\,a'_{<i}b'_{<i}x'_{<i}y'_{<i})|$ becomes arbitrarily small. We choose

$$\begin{aligned} a'_j \oplus b'_j &= x'_j \cdot y'_j \oplus 1 \quad \text{for } j \text{ odd, and} \\ a'_j \oplus b'_j &= x'_j \cdot y'_j \oplus \alpha_j \quad \text{for } j \text{ even}\,, \end{aligned} \tag{3.102}$$

where the set of bits $\{\alpha_j\}$ is to be determined later. With this choice of $b'_{<n}, x'_{<n}, y'_{<n}$,

and Lemma 3.3.6 we compute $\mathsf{P}(e = 0 \,|\, a'_{<n} b'_{<n} x'_{<n} y'_{<n})$ as

$$\begin{aligned}
&\mathsf{P}(e = 0 \,|\, a'_{<n} b'_{<n} x'_{<n} y'_{<n}) \\
&= \mathsf{P}(e = 0) + \frac{1}{4} \sum_{j=1}^{n-1} \frac{\mathsf{Q}_{\varepsilon'-sv}(e = 0 \,|\, a'_{\leq j}) - \mathsf{Q}(e = 0 \,|\, a'_{\leq j})}{\mathsf{PR}_{\varepsilon}(a'_j b'_j \,|\, x'_j y'_j)} \\
&\overset{(3.99)}{\Rightarrow} = \frac{1}{2} + \frac{1}{4} \sum_{j \text{ even}}^{n-1} \frac{\varepsilon' \cdot 2^{-n+j-1} \binom{n-j+1}{\frac{n-j+1}{2}}}{\mathsf{PR}_{\varepsilon}(a'_j b'_j \,|\, x'_j y'_j)} - \frac{1}{4} \sum_{j \text{ odd}}^{n-1} \frac{\varepsilon' \cdot 2^{-n+j} \binom{n-j}{\frac{n-j}{2}}}{\mathsf{PR}_{\varepsilon}(a'_j b'_j \,|\, x'_j y'_j)} \\
&\overset{(3.102)}{\Rightarrow} = \frac{1}{2} + \frac{1}{2} \sum_{j \text{ even}}^{n-1} \frac{\varepsilon' \cdot 2^{-n+j-1} \binom{n-j+1}{\frac{n-j+1}{2}}}{\delta(\alpha_j, 0) \cdot (1-\varepsilon) + \delta(\alpha_j, 1) \cdot \varepsilon} - \frac{1}{4} \sum_{j \text{ odd}}^{n-1} \frac{\varepsilon' \cdot 2^{-n+j} \binom{n-j}{\frac{n-j}{2}}}{\varepsilon} \\
&= \frac{1}{2} + \frac{1}{2} \sum_{j \text{ odd}}^{n-1} c_1 \cdot \delta(c_{j-1}, 0) \cdot \left(\frac{\varepsilon}{1-\varepsilon} - 1\right) \cdot \binom{n-j}{\frac{n-j}{2}} \cdot 2^{-n+j} \\
&\quad + \frac{c_1 \varepsilon}{2\Big(\delta(\alpha_{n-1}, 0) \cdot (1-\varepsilon) + \delta(\alpha_{n-1}, 1) \cdot \varepsilon\Big)} \cdot \binom{n-1}{\frac{n-1}{2}} \cdot 2^{-1} \,, \qquad (3.103)
\end{aligned}$$

where the last equation was obtained by a substitution $j \to j+1$ in the left sum and $\varepsilon' = c_1 n$. For j being smaller than a constant fraction of n, *i.e.*, $j < c_2 \cdot n$ for some constant $0 < c_2 < 1$, we can use Stirlings approximation for large n and obtain

$$\binom{n-j}{\frac{n-j}{2}} \cdot 2^{-n+j} \approx \frac{2^{n-j}}{\sqrt{n-j}} 2^{-n+j} = \frac{1}{\sqrt{n-j}} \,. \qquad (3.104)$$

For values of $j \approx n$ the term $\binom{n-j}{\frac{n-j}{2}} \cdot 2^{-n+j}$ becomes obviously constant. Thus, for $i = n-1$, then there must exist a choice for set $\{\alpha_0, ..., \alpha_{n-3}\}$ and some constant c_2, such that

$$|\mathsf{P}(e = 0 \,|\, a'_{<n} b'_{<n} x'_{<n} y'_{<n})| \leq \frac{c_2}{\sqrt{n}} \,. \qquad (3.105)$$

With this choice, we have

$$\begin{aligned}
\tilde{\tau}(a'_{<n-1} b'_{<n-1} x'_{<n-1} y'_{<n-1}, e = 0) &= \tau(a'_{<n-1}, e = 0) \, \frac{\mathsf{Q}_{\varepsilon'-sv}(a'_{<n-1} \,|\, e = 0)}{\mathsf{P}(e = 0 \,|\, a'_{<i} b'_{<i} x'_{<i} y'_{<i})} \\
&= \frac{\varepsilon'}{2\mathsf{P}(e = 0 \,|\, a'_{<i} b'_{<i} x'_{<i} y'_{<i})} \\
&\geq \sqrt{n} \, \frac{c_1}{2c_2} \, \varepsilon >> \varepsilon \,. \qquad (3.106)
\end{aligned}$$

□

3.4 TONS attacks via another classical game

3.4.1 From a classical game over a weighted set of distributions to TONS attacks

In this section we present a novel Ansatz to construct TONS-attacks on the boxes $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$ from special distributions $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$. We construct a set of classical distributions $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$, parametrised by a set $\mathcal{S} \in \mathcal{P}([n])$. Each distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ can be identified with a privacy-amplification game between Alice and Eve that is similar to the Santha-Vazirani game we presented in Section 3.3.1. However, in this game Eve can obtain *perfect* information of the bits $a_i, i \in \mathcal{S}$ but *no* information about the bits $a_i, i \in \overline{\mathcal{S}}$.

In Theorem 3.4.2 we will show that each of the distributions $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ can be extended to a TONS attack on a product distribution of noise boxes $\mathsf{U}(a_ib_i\,|\,x_iy_i)$ for $i \in \mathcal{S}$ and perfect $\mathsf{PR}(a_ib_i\,|\,x_iy_i)$ boxes for $i \in \overline{\mathcal{S}}$. A weighted sum of these product distributions corresponds to $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$. We define $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ as the sum over the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ with the same weights and, thus, the distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ induces a TONS attack on $\mathsf{PR}_\varepsilon^n$ (see Figure 3.3). Let us define the distributions $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ and $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$.

Definition 11 (Ordered $\mathcal{S}$-influenceable distributions)**.** For a set $\mathcal{S} \in \mathcal{P}([n])$ we define an *ordered $\mathcal{S}$-influenceable distribution* $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ as a probability distribution that satisfies uniformity on $a_{\leq n}$

$$\sum_e \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e) = 2^{-n} \quad \forall a_{\leq n} \quad \text{and} \tag{3.107}$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_i\,|\,a_{<i}e) = \frac{1}{2} \quad \forall a_{\leq i}, e, \quad \text{and} \quad i \in \overline{\mathcal{S}}\,. \tag{3.108}$$

We call the distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ ordered because the number of conditions that (3.108) imposes on $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ grows exponentially in i for each $i \in \overline{\mathcal{S}}$. In Section 3.7.1 we define a set of distributions $\{\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)\}$ in the context of an analogous construction of ABNS attacks where the respective conditions to (3.108) are symmetric for each i, see (3.231).

Definition 12 (Ordered $(\varepsilon, \mathcal{S})$-divisible distribution)**.** Fix a given set of ordered $\mathcal{S}$-influenceable distributions $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$. Define an *ordered $(\varepsilon, \mathcal{S})$-divisible distribution* $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ as

$$\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e) := \sum_{\mathcal{S}\in\mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\,, \tag{3.109}$$

with weights

$$\omega(\mathcal{S}, n, \varepsilon) := (1-2\varepsilon)^{n-|S|}\,(2\varepsilon)^{|S|}\,. \tag{3.110}$$

Let us consider the following privacy-amplification game between Alice and Eve. First Alice chooses a function $f(a_{\leq n})$ and hands it to Eve. Then Eve constructs an ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$. Draw the string $a_{\leq n}e$ according to the distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$. Eve wins if $f(a_{\leq n}) = e$; Alice wins otherwise.

Note that implicitly Eve must construct the set of distributions $\{\mathsf{Q}_{o-\mathcal{S}}\}$. The trivial strategy for Alice is to let the function be equal to a single bit, *e.g.*, $f(a_{\leq n}) = a_1$. For this strategy the optimal winning probability for Eve is the same as in the Santha-Vazirani game from Section 3.3.1, *i.e.*, $\mathsf{Q}_{o-\varepsilon}(a_1 = e) = 1/2 + \varepsilon$: If $\{1\} \in \mathcal{S}$, the optimal strategy for Eve is to set $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e) = 2^{-n}\delta(a_1, e)$ and win with probability 1. If $\{1\} \in \overline{\mathcal{S}}$, then (3.108) implies that $\mathsf{Q}_{o-\mathcal{S}}(a_1 = e) = 1/2$, independently from how Eve constructs $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$. By the definition of $\omega(\mathcal{S}, n, \varepsilon)$ (3.110) the weight of the sets $\mathcal{S}$ that contain $\{1\}$ is 2ε, and thus Eve's winning probability is

$$\mathsf{Q}_{o-\varepsilon}(a_1 = e) = 2\varepsilon \cdot 1 + (1-2\varepsilon)\cdot\frac{1}{2} = \frac{1}{2} + \varepsilon. \tag{3.111}$$

For Santha-Vazirani games, we saw that this trivial strategy is the best that Alice can do: Eve can always achieve a winning probability of $1/2 + \varepsilon$ due to the Reingold construction, (3.40) and (3.41). However, the best general lower bound for optimal strategies $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ that we obtain is $\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e) \geq 1/2 + \varepsilon/n$. Note that the two games are quite similar: both $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ and $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ must satisfy that the marginal distribution on Alice's bits is uniform, *i.e.*,

$$\sum_e \mathsf{Q}_{o-\varepsilon}(a_{\leq n}e) = 2^{-n} = \sum_e \mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)\,. \tag{3.112}$$

Furthermore, one can easily show that Eve's knowledge about a single bit a_i is bounded by the same value in both games, *i.e.*,

$$\sum_e \mathsf{Q}_{o-\varepsilon}(e)\,\max_{a'}\mathsf{Q}_{o-\varepsilon}(a_i = a' \,|\, e) \leq \frac{1}{2} + \varepsilon \tag{3.113}$$

$$\sum_e \mathsf{Q}_{\varepsilon-sv}(e)\,\max_{a'}\mathsf{Q}_{\varepsilon-sv}(a_i = a' \,|\, e) \leq \frac{1}{2} + \varepsilon\,. \tag{3.114}$$

However, $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ does not have to satisfy the Santa-Vazirani condition (3.35) in general, *i.e.*, one can construct distributions $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$, such that

$$\mathsf{Q}_{o-\varepsilon}(a_i \,|\, a_{<i}e) > \frac{1}{2} + \varepsilon\,. \tag{3.115}$$

It would be surprising (at least to the author) if the $\mathcal{S}$-divisibility condition (3.109) is so restrictive that there is a large gap of the order of $\theta(n)$ in between the optimal winning probabilities for Eve in both games. In addition, numerical analysis suggests that there is a deeper connection between ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ and the Reingold distributions $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$, defined by (3.40) and(3.41): For randomly chosen balanced functions $f(a_{\leq n})$ up to $n = 8$, each corresponding Reingold distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ could be shown to be also an ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$.

Furthermore, if $f(a_{\leq n})$ is a prefix-coded function, see Definition 16, the construction (3.170) and (3.171) for the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ yields exactly an $\mathcal{S}$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ which is equal to the $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ from the Reingold construction (3.40) and (3.41). This leads us to the following conjecture

Conjecture 3.4.1. *For any balanced function $f(a_{\leq n})$, the ε-Santha-Vazirani distribution $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ defined by Reingold construction* (3.40) *and* (3.41) *is also an ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$, i.e., there exists a set $\{\mathsf{Q}_{\mathcal{S}}\}$ such that for the corresponding $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$, see* (3.109)*, it holds that*

$$\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e) = \mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e) \,. \tag{3.116}$$

If Conjecture 3.4.1 is true, then Theorem 3.4.3 implies that TONS privacy amplification is impossible.

We now show that any ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ induces a TONS attack on a privacy amplification protocol on $\mathsf{PR}_{\varepsilon}^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$.

Theorem 3.4.2. *Any ordered $\mathcal{S}$-influenceable distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ can be extended to a TONS-attack $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on the systems $A_{\leq n}B_{\leq n}$ with marginal distribution*

$$\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}) := \prod_{i\in\mathcal{S}} \mathsf{U}(a_ib_i \,|\, x_iy_i) \prod_{i\in\overline{\mathcal{S}}} \mathsf{PR}(a_ib_i \,|\, x_iy_i) \,. \tag{3.117}$$

Proof. The proof consists of an explicit construction of $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$:

$$\mathsf{P}_{o-\mathcal{S}}(e) = \mathsf{Q}_{o-\mathcal{S}}(e) \tag{3.118}$$

$$\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}e) = \prod_{i=1}^{n} \mathsf{P}_{o-\mathcal{S}}(a_ib_i \,|\, a_{<i}b_{<i}x_{\leq n}y_{\leq n}e) \tag{3.119}$$

$$\mathsf{P}_{o-\mathcal{S}}(a_ib_i \,|\, a_{<i}b_{<i}x_{\leq n}y_{\leq n}e) = \begin{cases} \mathsf{U}(b_i \,|\, y_i)\, \mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) & i \in \mathcal{S} \\ \mathsf{PR}(a_ib_i \,|\, x_iy_i) & \text{otherwise} \,. \end{cases} \tag{3.120}$$

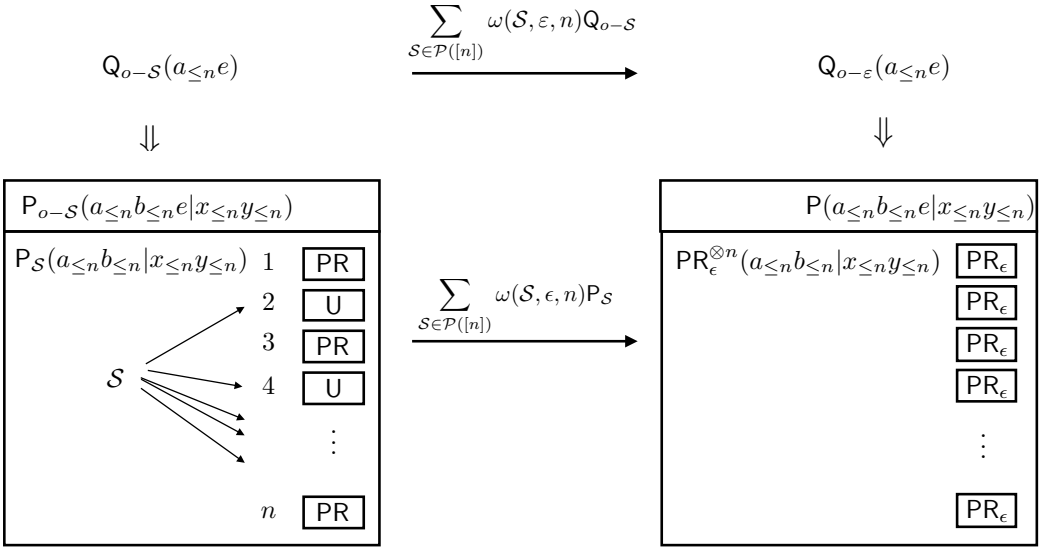

Figure 3.3. Schematic view on Theorem 3.4.2 and Theorem 3.4.3. In Theorem 3.4.2 we show that a distribution $\mathsf{Q}_{o-\mathcal{S}}$ can be always be extended to a TONS attack on a box $\mathsf{P}_{\mathcal{S}}$ consisting of a product of perfect PR*-boxes and boxes* U *which just output uniform bits. Taking a weighted sum of the $\mathsf{Q}_{o-\mathcal{S}}$ distributions implies Theorem 3.4.3, which states that a $\mathsf{Q}_{o-\varepsilon}$ distribution can alway be extended to a TONS attack on $\mathsf{PR}_{\varepsilon}^{\otimes n}$.*

We have to show that (3.118)-(3.120) implies that $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$

1. satisfies the TONS-conditions (3.2),
2. has the correct marginal on systems $A_{\leq n}B_{\leq n}$:
$$\sum_e \mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \prod_{i\in\mathcal{S}} \mathsf{U}(a_ib_i \,|\, x_iy_i) \prod_{i\in\overline{\mathcal{S}}} \mathsf{PR}(a_ib_i \,|\, x_iy_i)\,, \tag{3.121}$$
3. and has the correct marginal on systems $A_{\leq n}E$ (which must be independent also of $x_{\leq n}$):
$$\sum_{b_{\leq n}} \mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\,. \tag{3.122}$$

We start with the TONS conditions (2.3): We need to show for all $0 \leq i_A, i_B \leq n$, that the systems $A_{\leq i_A}B_{\leq i_B}$ have a well-defined marginal distribution $\mathsf{P}_{o-\mathcal{S}}(a_{\leq i_A}b_{\leq i_B} \,|\, ex_{\leq i_A}y_{\leq i_B})$: First note that from (3.120) it follows directly $\forall a_{<i}, b_{\leq i}e, x_{\leq i}, y_{\leq i}$ that

$$\sum_{a_ib_i} \mathsf{P}_{o-\mathcal{S}}(a_ib_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) = 1 \tag{3.123}$$

$$\sum_{a_i} \mathsf{P}_{o-\mathcal{S}}(a_ib_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) = \mathsf{U}(b_i \,|\, y_i)\,, \tag{3.124}$$

and also

$$\sum_{b_i} \mathsf{P}_{o-\mathcal{S}}(a_ib_i \,|\, a_{<i}b_{<i}ex_{\leq i}y_{\leq i}) = \begin{cases} \mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) & i \in \mathcal{S} \\ \frac{1}{2} = \mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) = \mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}\overline{e}) & \text{otherwise}\,. \end{cases} \tag{3.125}$$

We show that $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ has a well-defined marginal first on systems $A_{\leq i}B_{\leq i}$. Then we show this for the systems $A_{\leq i}B_{\leq k}$ first for $i < k$ and then for $i > k$, and thus prove that satisfies the TONS $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ conditions. From the second equation in (3.123), it follows

$$\begin{aligned}
&\sum_{a_{>i}b_{>i}} \mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n}) \\
&= \sum_{a_{>i}b_{>i}} \prod_{j=1}^{n} \mathsf{P}_{o-\mathcal{S}}(a_jb_j \,|\, a_{<j}b_{<j}ex_{\leq j}y_{\leq j}) \\
&= \prod_{j=1}^{i} \mathsf{P}_{o-\mathcal{S}}(a_jb_j \,|\, a_{<j}b_{<j}ex_{\leq j}y_{\leq j}) \prod_{j=i+1}^{n} \sum_{a_jb_j} \mathsf{P}_{o-\mathcal{S}}(a_jb_j \,|\, a_{<j}b_{<j}ex_{\leq j}y_{\leq j}) \\
&= \prod_{j=1}^{i} \mathsf{P}_{o-\mathcal{S}}(a_jb_j \,|\, a_{<j}b_{<j}ex_{\leq j}y_{\leq j}) \cdot 1 \\
&=: \mathsf{P}_{o-\mathcal{S}}(a_{\leq i}b_{\leq i} \,|\, ex_{\leq i}y_{\leq i})\,,
\end{aligned} \tag{3.126}$$

where the expression in second to last line is independent of $x_{>i}y_{>i}$ and, therefore, is the well-defined marginal on $A_{\leq i}B_{\leq i}$. Now for $i < k$, we have

$$\begin{aligned}
&\sum_{a_{i+1}\ldots a_k} \mathsf{P}_{o-\mathcal{S}}(a_{\leq k}b_{\leq k} \,|\, ex_{\leq k}y_{\leq k}) \\
&= \mathsf{P}_{o-\mathcal{S}}(a_{\leq i}b_{\leq i} \,|\, ex_{\leq i}y_{\leq i}) \sum_{a_{i+1}\ldots a_k} \prod_{j=i+1}^{k} \mathsf{P}_{o-\mathcal{S}}(a_j b_j \,|\, a_{<j}b_{<j}ex_{\leq j}y_{\leq j}) \\
&= \mathsf{P}_{o-\mathcal{S}}(a_{\leq i}b_{\leq i} \,|\, ex_{\leq i}y_{\leq i}) \prod_{j=i+1}^{k} \mathsf{U}(b_i \,|\, y_i) \\
&=: \mathsf{P}_{o-\mathcal{S}}(a_{\leq i}b_{\leq k} \,|\, ex_{\leq i}y_{\leq k}) \,, \qquad (3.127)
\end{aligned}$$

where we used for the second equality (3.124) $k - i$ times. For $i > k$, we obtain similarly

$$\begin{aligned}
&\sum_{b_{k+1}\ldots b_i} \mathsf{P}_{o-\mathcal{S}}(a_{\leq i}b_{\leq i} \,|\, ex_{\leq i}y_{\leq i}) \\
&= \mathsf{P}_{o-\mathcal{S}}(a_{\leq k}b_{\leq k} \,|\, ex_{\leq k}y_{\leq k}) \sum_{b_{k+1}\ldots b_i} \prod_{j=k+1}^{i} \mathsf{P}_{o-\mathcal{S}}(a_j b_j \,|\, a_{<j}b_{<j}ex_{\leq j}y_{\leq j}) \\
&= \mathsf{P}_{o-\mathcal{S}}(a_{\leq i}b_{\leq i} \,|\, ex_{\leq i}y_{\leq i}) \prod_{j=k+1}^{i} \mathsf{Q}_{o-\mathcal{S}}(a_j \,|\, a_{<j}e) \\
&=: \mathsf{P}_{o-\mathcal{S}}(a_{\leq i}b_{\leq k} \,|\, ex_{\leq i}y_{\leq k}) \,. \qquad (3.128)
\end{aligned}$$

Thus, by (3.126),(3.127) and (3.128) all TONS conditions are satisfied.

In order to show that our construction yields the correct marginal distribution on $A_{\leq n}E$ we simply set $i = n$ and $k = 0$ in (3.128) and directly obtain (3.122). Finally, we need to prove the correct marginal distribution on the systems $A_{\leq n}B_{\leq n}$, *i.e.*, that

(3.121) is satisfied. Using the construction rules (3.118) to (3.120) we conclude that

$$
\begin{aligned}
&\sum_e \mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n}) \\
&= \sum_e \mathsf{P}(e) \prod_{i=1}^{n} \mathsf{P}(a_i b_i \,|\, a_{<i} b_{<i} e x_{\leq i} y_{\leq i}) \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i) \cdot \Big(\sum_e \prod_{i \in \mathcal{S}} \mathsf{U}(b_i \,|\, y_i)\, \mathsf{Q}_{o-\mathcal{S}}(e)\, \mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i} e) \Big) \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i) \cdot \prod_{i \in \mathcal{S}} \mathsf{U}(b_i \,|\, y_i) \\
&\quad \cdot \Big(2^{n-|\mathcal{S}|}\, \mathsf{Q}_{o-\mathcal{S}}(e)\, \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} \,|\, e) + 2^{n-|\mathcal{S}|}\, \mathsf{Q}_{o-\mathcal{S}}(\overline{e})\, \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} \,|\, \overline{e}) \Big) \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i) \cdot \prod_{i \in \mathcal{S}} \frac{1}{2} \mathsf{U}(b_i \,|\, y_i) \cdot 2^n \Big(\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} e) + \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} \overline{e}) \Big) \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i) \prod_{i \in \mathcal{S}} \mathsf{U}(a_i \,|\, x_i)\, \mathsf{U}(b_i \,|\, y_i) \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i) \prod_{i \in \mathcal{S}} \mathsf{U}(a_i b_i \,|\, x_i y_i) \,, \qquad (3.129)
\end{aligned}
$$

where we used (3.108) in the third equation, (3.107) for the fifth equation, and that $\mathsf{U}(a_i b_i \,|\, x_i y_i) = \mathsf{U}(a_i \,|\, x_i)\, \mathsf{U}(b_i \,|\, y_i)$ for the last equation. □

Theorem 3.4.3. *For any ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n} e)$, there exists a TONS-attack $\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n})$ on $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n} b_{\leq n} \,|\, x_{\leq n} y_{\leq n})$ such that*

$$
\sum_{b_{\leq n}} \mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n}) = \mathsf{Q}_{o-\varepsilon}(a_{\leq n} e) \qquad \forall x_{\leq n}, y_{\leq n} \qquad (3.130)
$$

Proof. Theorem 3.4.3 follows from Definition 12, Theorem 3.4.2 and the fact that

$$
\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n} b_{\leq n} \,|\, x_{\leq n} y_{\leq n}) = \sum_{\mathcal{S} \in \mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{P}_{o-\mathcal{S}}(a_{\leq n} b_{\leq n} \,|\, x_{\leq n} y_{\leq n}) \,. \qquad (3.131)
$$

□

In the remainder of Section 3.4 we derive TONS attacks from the ordered $(\varepsilon, \mathcal{S})$-divisible distributions $\mathsf{Q}_{o-\varepsilon}(a_{\leq n} e)$. Consequently, we refer to a $\mathsf{Q}_{o-\varepsilon}(a_{\leq n} e)$ that is specifically constructed to ensure a high level of knowledge $\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e)$ for a certain function $f(a_{\leq n})$ simply as an "attack" or "TONS attack" on $f(a_{\leq n})$.

3.4.2 Unbalanced functions do not provide more secrecy than balanced functions

Theorem 3.4.4. *For any function* $f(a_{\leq n}) : \{0,1\}^n \to \{0,1\}$ *there exists a balanced function* $f'(a_0 a_{\leq n}) : \{0,1\}^{n+1} \to \{0,1\}$ *such that for any* $\mathsf{Q}'_{o-\varepsilon}(a_0 a_{\leq n} e)$ *there exists a* $\mathsf{Q}_{o-\varepsilon}(a_{\leq n} e)$ *such that*

$$\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e) = \mathsf{Q}'_{o-\varepsilon}(f'(a_0 a_{\leq n}) = e) \,. \tag{3.132}$$

Proof. For a given function $f(a_{\leq n})$ define

$$f'(a_0 a_{\leq n}) := a_0 \oplus f(a_{\leq n}) \,. \tag{3.133}$$

Let $\mathsf{Q}'_{o-\varepsilon}(a_0 a_{\leq n} e)$ be an attack on the function $f'(a_0 a_{\leq n})$ derived from the set of ordered $\mathcal{T}$-influenceable distributions for $\mathcal{T} \in \mathcal{P}(n+1)$

$$\{\mathsf{Q}'_{o-\mathcal{T}}(a_0 a_{\leq n} e)\} = \{\mathsf{Q}'_{o-\mathcal{S}}(a_0 a_{\leq n} e)\} \cup \{\mathsf{Q}'_{o-\mathcal{S}\cup\{0\}}(a_0 a_{\leq n} e)\} \quad \text{with } \mathcal{S} \in \mathcal{P}([n]) \,. \tag{3.134}$$

Then define the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} e)\}$ with elements

$$\boxed{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} e) := (1-2\varepsilon) \sum_{a_0} \mathsf{Q}'_{o-\mathcal{S}}(a_0 a_{\leq n}, a_0 \oplus e) + 2\varepsilon \sum_{a_0} \mathsf{Q}'_{o-\mathcal{S}\cup\{0\}}(a_0 a_{\leq n}, a_0 \oplus e) \,.} \tag{3.135}$$

Now we need to show that the definition (3.135) implies (3.107) and (3.108) for $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} e)$. These follow from the same properties for $\mathsf{Q}'_{o-\mathcal{T}}(a_0 a_{\leq n} e)$; for (3.107) we have

$$\begin{aligned}
&\sum_e \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} e) \\
&= (1-2\varepsilon) \sum_e \sum_{a_0} \mathsf{Q}'_{o-\mathcal{S}}(a_0 a_{\leq n}, a_0 \oplus e) + 2\varepsilon \sum_e \sum_{a_0} \mathsf{Q}'_{o-\mathcal{S}\cup\{0\}}(a_0 a_{\leq n}, a_0 \oplus e) \\
&= (1-2\varepsilon) \sum_{a_0} \mathsf{Q}'_{o-\mathcal{S}}(a_0 a_{\leq n}) + 2\varepsilon \sum_{a_0} \mathsf{Q}'_{o-\mathcal{S}\cup\{0\}}(a_0 a_{\leq n}) \\
&= (1-2\varepsilon) \sum_{a_0} 2^{-(n+1)} + 2\varepsilon \sum_{a_0} 2^{-(n+1)} \\
&= 2^{-n} \,.
\end{aligned} \tag{3.136}$$

Note that (3.108) implies for $\mathsf{Q}'_{o-\mathcal{T}}(a_0 a_{\leq n} e)$ that

$$\begin{aligned}
\mathsf{Q}'_{o-\mathcal{T}}(a_i \,|\, a_0 a_{<i} e) &= \frac{\mathsf{Q}'_{o-\mathcal{T}}(a_0 a_{<i} \,|\, e)}{\mathsf{Q}'_{o-\mathcal{T}}(a_0 a_{<i} a_i \,|\, e)} \\
&= \frac{\alpha}{2\alpha} \qquad \forall i \notin \mathcal{T} \,,
\end{aligned} \tag{3.137}$$

and thus (3.108) follows also for $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$

$$\begin{aligned}\mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) &= \frac{\mathsf{Q}_{o-\mathcal{S}}(a_{<i}a_i \,|\, e)}{\mathsf{Q}_{o-\mathcal{S}}(a_{<i} \,|\, e)} \\ &= \frac{(1-2\varepsilon)\sum_{a_0}\mathsf{Q}'_{o-\mathcal{S}}(a_0a_{<i}a_i \,|\, a_0 \oplus e) + 2\varepsilon\sum_{a_0}\mathsf{Q}'_{o-\mathcal{S}\cup 0}(a_0a_{<i}a_i \,|\, a_0 \oplus e)}{(1-2\varepsilon)\sum_{a_0}\mathsf{Q}'_{o-\mathcal{S}}(a_0a_{<i} \,|\, a_0 \oplus e) + 2\varepsilon\sum_{a_0}\mathsf{Q}'_{o-\mathcal{S}\cup 0}(a_0a_{<i} \,|\, a_0 \oplus e)} \\ &= \frac{(1-2\varepsilon)(\alpha_1+\alpha_2) + 2\varepsilon(\alpha_3+\alpha_4)}{(1-2\varepsilon)(2\alpha_1+2\alpha_2) + 2\varepsilon(2\alpha_3+2\alpha_4)} \\ &= \frac{1}{2}\,. \end{aligned} \tag{3.138}$$

We complete the proof by showing (3.132): note that

$$\begin{aligned}\mathsf{Q}'_{o-\varepsilon}(f'(a_0a_{\leq n}) = e) = \sum_{\mathcal{S}\in\mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\Big((1-2\varepsilon)\,\mathsf{Q}'_{o-\mathcal{S}}(f'(a_0a_{\leq n}) = e) \\ + 2\varepsilon\,\mathsf{Q}'_{o-\mathcal{S}\cup 0}(f'(a_0a_{\leq n}) = e)\Big)\,,\end{aligned} \tag{3.139}$$

which together with

$$\begin{aligned}&(1-2\varepsilon)\,\mathsf{Q}'_{o-\mathcal{S}}(f'(a_0a_{\leq n}) = e) + 2\varepsilon\,\mathsf{Q}'_{o-\mathcal{S}\cup 0}(f'(a_0a_{\leq n}) = e) \\ &= (1-2\varepsilon)\,\mathsf{Q}'_{o-\mathcal{S}}(a_0 \oplus f(a_{\leq n}) = e) + 2\varepsilon\,\mathsf{Q}'_{o-\mathcal{S}\cup 0}(a_0 \oplus f(a_{\leq n}) = e) \\ &= (1-2\varepsilon)\,\mathsf{Q}'_{o-\mathcal{S}}(f(a_{\leq n}) = a_0 \oplus e) + 2\varepsilon\,\mathsf{Q}'_{o-\mathcal{S}\cup 0}(f(a_{\leq n}) = a_0 \oplus e) \\ &= \mathsf{Q}_{o-\mathcal{S}}(f(a_{\leq n}) = e)\,,\end{aligned} \tag{3.140}$$

implies (3.132). □

3.5 Classical analysis of ordered $(\varepsilon, \mathcal{S})$-divisible distributions $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$

In the following, if not specified otherwise we will assume that probabilities $\mathsf{Q}(a_{\leq n})$ are uniformly distributed.

3.5.1 Attacking linear functions

The first candidates as privacy-amplification functions are the set of parities $\bigoplus_{i\in\mathcal{S}} a_i$ for $\mathcal{S} \in \mathcal{P}([n])$. Against both, classical and quantum adversaries, privacy amplification using *2-universal hashing* (using a randomly chosen parity) is optimal. In the setting of the *fully no-signalling adversary* studied in [HRW10] privacy amplification is even possible when the parity is chosen deterministically, see Theorem 3.2.2.

Using our framework, we will show that privacy amplification against a TONS adversary is impossible using linear functions.

Theorem 3.5.1. *No privacy amplification with linear (parity) functions. For any $\mathcal{T} \in \mathcal{P}([n])$ there exists an ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ such that*

$$\mathsf{Q}_{o-\varepsilon}(\bigoplus_{i\in\mathcal{T}} a_i = e) = \frac{1}{2} + \varepsilon \tag{3.141}$$

Proof. Intuitively, we prove Theorem 3.5.1 by constructing an adversary $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ who influences the bit a_{i_t} where $i_t := \max_i[i \in \mathcal{T}]$ by a value of ε into a preferred direction depending of the parity of the previous bits $a_{i_1}, ..., a_{i_{t-1}}$, $i_j \in \mathcal{T}$. Formally the construction is as follows: For all $\mathcal{S}$ such that $i_t \in \mathcal{S}$ let $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ be defined as

$$\mathsf{Q}_{o-\mathcal{S}}(e) = \frac{1}{2}\,, \tag{3.142}$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) = \begin{cases} \delta(a_i, e \oplus \bigoplus_{j\in\mathcal{T}/i} a_i) & i = i_t \\ \frac{1}{2} & \text{otherwise}\,. \end{cases} \tag{3.143}$$

It is easy to see that (3.108) is satisfied by construction. Note that (3.143) is equivalent to

$$\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} \,|\, e) = 2^{-n+1}\,\delta(a_{i_t}, e \oplus \bigoplus_{j\in\mathcal{T}/i} a_i)\,, \tag{3.144}$$

which yields (3.107) when averaged over $\mathsf{Q}_{o-\mathcal{S}}(e)$. For all $\mathcal{S}$ such that $i_t \notin \mathcal{S}$ set

$$\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e) = 2^{-(n+1)}\,. \tag{3.145}$$

Here (3.107) and (3.108) are satisfied trivially. The constructions (3.143) and (3.145) for the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ induce a $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ with

$$\begin{aligned} \mathsf{Q}_{o-\varepsilon}(\bigoplus_{i\in\mathcal{T}} a_i = e) &= \sum_{\mathcal{S}\in\mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\,\mathsf{Q}_{o-\mathcal{S}}(\bigoplus_{i\in\mathcal{T}} a_i = e) \\ &= \sum_{\mathcal{S}\ni i_t} \omega(\mathcal{S}, n, \varepsilon)\,\mathsf{Q}_{o-\mathcal{S}}(\bigoplus_{i\in\mathcal{T}} a_i = e) \\ &\quad + \sum_{\mathcal{S}\not\ni i_t} \omega(\mathcal{S}, n, \varepsilon)\,\mathsf{Q}_{o-\mathcal{S}}(\bigoplus_{i\in\mathcal{T}} a_i = e) \\ &= 2\varepsilon \cdot 1 + (1 - 2\varepsilon) \cdot \frac{1}{2} = \frac{1}{2} + \varepsilon\,, \end{aligned} \tag{3.146}$$

where we used that

$$\sum_{\mathcal{S}\ni i_t} \omega(\mathcal{S}, n, \varepsilon) = 2\varepsilon \; . \tag{3.147}$$

□

Note that Theorem 3.5.1 directly extends to the case of *affine functions* $f(a_{\leq n})$. A function $f(a_{\leq n})$ is called *affine* if there exists some set $\mathcal{T} \in \mathcal{P}([n])$ and a bit $b \in \{0, 1\}$ such that

$$f(a_{\leq n}) = b \oplus \bigoplus_{i\in\mathcal{T}} a_i \; . \tag{3.148}$$

To extend Theorem 3.5.1 to affine functions, one just needs to insert the bit b into second argument of the δ-function in (3.143).

3.5.2 Attacking random functions — bias the last bit

An important class of functions are *random functions* since, roughly speaking, most functions share the properties of random functions. We show that random functions can be attacked with a relatively simple attack; analogous to the case when the privacy-amplification function is, *e.g.*, the parity of all n bits, the adversary simply biases the last bit a_n by a value of ε into the preferred direction. The intuition behind the attack is that a random function non-trivially depends on the the last bit a_n with probability $1/2$.

Definition 13 (Random function)**.** We define a *random function* $f_r : \{0, 1\}^n \to \{0, 1\}$ as a function which is drawn according to the uniform measure from the set $\mathcal{F}_n$ of all Boolean functions of n bits.

Note that if we denote the uniform measure over the set $\mathcal{F}_n$ as $\mathsf{Q}_{\mathcal{F}_n}$ then a random function satisfies

$$\mathsf{Q}_{\mathcal{F}_n}(f_r(a_{\leq n}) = 0) = \frac{1}{2} = \mathsf{Q}_{\mathcal{F}_n}(f_r(a_{\leq n}) = 1) \; . \tag{3.149}$$

Theorem 3.5.2. *For a random function $f_r(a_{\leq n})$ there exists an ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ such that*

$$\mathsf{Q}_{o-\varepsilon}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e) \geq \frac{1}{2} + \frac{\varepsilon}{2} \; . \tag{3.150}$$

Note that we use the notation $\mathsf{Q}_{o-\varepsilon}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e)$ to indicate that first the function f_r is drawn according to the uniform distribution from the set of all functions $f(a_{\leq n})$ and then the distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ is constructed accordingly.

Proof. For a function $f(a_{\leq n})$ define the set

$$\mathcal{D}_f := \{a_{<n} : f(a_{<n}, 0) \neq f(a_{<n}, 1)\} . \tag{3.151}$$

We construct the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ as follows:

$$\mathsf{Q}_{o-\mathcal{S}}(e) = \frac{1}{2} , \tag{3.152}$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_{<n} \,|\, e) = 2^{-n+1} \tag{3.153}$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_n \,|\, a_{<n}e) = \begin{cases} \delta(e, f(a_{\leq n})) & n \in \mathcal{S} \,\cap\, a_{<n} \in \mathcal{D}_f \\ \frac{1}{2} & \text{otherwise} . \end{cases} \tag{3.154}$$

It is easy to see that (3.108) is satisfied by construction through (3.154). Note that when $n \in \mathcal{S} \,\cap\, a_{<n} \in \mathcal{D}_f$, then (3.153) and (3.154) are equivalent to

$$\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} \,|\, e) = 2^{-n+1}\, \delta(e, f(a_{\leq n})) , \tag{3.155}$$

which, in the average over $\mathsf{Q}_{o-\mathcal{S}}(e)$, yields (3.107) through the condition $a_{<n} \in \mathcal{D}_f$. In the second case of (3.154), (3.107) is satisfied trivially.

It is left to prove (3.150). To analyse $\mathsf{Q}_{o-\varepsilon}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e)$ we distinguish between two cases:

1. $n \notin \mathcal{S}$: (3.152)–(3.154) implies that, independently of the choice of function f_r, we have

$$\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e) = 2^{-(n+1)} . \tag{3.156}$$

Consequently, we have that

$$\begin{aligned} \mathsf{Q}_{o-\mathcal{S}}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e) &= \sum_{a_{\leq n}} \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n})\, \mathsf{Q}_{o-\mathcal{S}}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e \,|\, a_{\leq n}) \\ &= \sum_{a_{\leq n}} 2^{-n}\, \mathsf{Q}_{o-\mathcal{S}}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e \,|\, a_{\leq n}) \\ &= \sum_{a_{\leq n}} 2^{-n}\, \frac{1}{2} \\ &= \frac{1}{2}, \end{aligned} \tag{3.157}$$

where we use that, independently of the value $f_r(a_{\leq n})$, the bit e is uniformly random in order to calculate $\mathsf{Q}_{o-\mathcal{S}}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e \,|\, a_{\leq n})$

2. $n \in \mathcal{S}$: From (3.152) and (3.153) it follows that $\mathsf{Q}_{o-\mathcal{S}}(a_{<n}e) = 2^{-n}$, *i.e.*, $a_1 ... a_{n-1}e$ is a string of uniform independent bits that, furthermore, is independent of the choice of the function f_r. From (3.149) also follows that for $(f_r(a_{<n}, 0), f_r(a_{<n}, 1))$ there are the four (equally likely) possibilities $(0,0)$, $(1,1)$, $(0,1)$ and $(1,0)$. In the two cases $(0,0)$ and $(1,1)$ we have that $a_{<n} \notin \mathcal{D}_f$, and thus

$$\mathsf{Q}_{o-\mathcal{S}}(f_r(a_{\leq n}) = e \,|\, a_{<n}) = \frac{1}{2}\,, \tag{3.158}$$

since by (3.154) the bit e is uniform conditioned on $a_{<n}$. In the cases $(0,1)$ and $(1,0)$, we have

$$\mathsf{Q}_{o-\mathcal{S}}(f_r(a_{\leq n}) = e \,|\, a_{<n}) = 1\,, \tag{3.159}$$

due to (3.154), since then $a_{<n} \in \mathcal{D}_f$.

We conclude that the constructions (3.152)–(3.154) for the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ thus induce a $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ with

$$\begin{aligned}
\mathsf{Q}_{o-\varepsilon}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e) &= \sum_{\mathcal{S}\in\mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{o-\mathcal{S}}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e) \\
&= \sum_{\mathcal{S}\ni n} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{o-\mathcal{S}}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e) \\
&\quad + \sum_{\mathcal{S}\not\ni n} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{o-\mathcal{S}}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e) \\
&= \sum_{\mathcal{S}\ni n} \omega(\mathcal{S}, n, \varepsilon) \sum_{a_{<n}} \mathsf{Q}_{o-\mathcal{S}}(a_{<n}) \mathsf{Q}_{o-\mathcal{S}}(\mathsf{Q}_{\mathcal{F}_n}(f_r)(a_{\leq n}) = e \,|\, a_{<n}) \\
&\quad + (1-2\varepsilon)\,\frac{1}{2} \\
&= 2\varepsilon \left(\frac{1}{2}\,\frac{1}{2} + \frac{1}{2}\,1\right) + (1-2\varepsilon)\,\frac{1}{2} = \frac{1}{2} + \frac{\varepsilon}{2}\,.
\end{aligned} \tag{3.160}$$

We use that

$$\sum_{\mathcal{S}\ni n} \omega(\mathcal{S}, n, \varepsilon) = 2\varepsilon\,, \tag{3.161}$$

and for the fourth equation that (3.158) and (3.159) each happen with probability $1/2$ (for each prefix $a_{<n}$) for random functions. □

Now we quantify the intuition that most functions share the properties of random functions and a refined analysis of the construction (3.152)–(3.154) allows us to conclude that for *almost all* functions the same attack allows the adversary to gain a (maybe small but) constant knowledge about the output of Alice.

Theorem 3.5.3. *For any given n and* $0 < \gamma_1 < 1$*, there exists for a fraction* $1 - \gamma_2$ *of the set* $\mathcal{F}_n$ *of all Boolean functions* $f : \{0,1\}^n \to \{0,1\}$ *an ordered* $(\varepsilon, \mathcal{S})$*-divisible distribution* $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ *such that*

$$\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e) \geq \frac{1}{2} + \frac{(1-\gamma_1)\,\varepsilon}{2}\,, \tag{3.162}$$

with

$$\gamma_2 = e^{-\gamma_1^2 2^{n-3}}\,. \tag{3.163}$$

Proof. The proof follows from the attack (3.152)–(3.154) on random functions via an application of a Chernoff bound [Che52]. In Case 2 of the proof of Theorem 3.5.2 the probability $\mathsf{Q}_{o-\mathcal{S}}(f_r(a_{\leq n}) = e \,|\, a_{<n})$ depends on whether $(f_r(a_{<n}, 0), f_r(a_{<n}, 1))$ is either in the cases $(0,0)$ and $(1,1)$, then (3.158) holds, or in the cases $(0,1)$ and $(1,0)$ and then (3.159) holds; in other words if $a_{<n} \in \mathcal{D}_f$ or not. For a specific function $f(a_{\leq n})$ the knowledge of the adversary implied by the attack (3.152)–(3.154) only depends the size of the set $\mathcal{D}_f$. With respect to the (uniform) average over all functions $\mathsf{Q}_{\mathcal{F}_n}(f_r)$ the probability that $a_{<n} \in \mathcal{D}_f$ is $1/2$, as we argued in the fourth equation of (3.160), and thus $\mathcal{D}_f$ is on average half the size of the set of all $a_{<n}$, *i.e.*, $|\mathcal{D}_f| = 2^{n-2}$. Now we use a particular Chernoff bound from [AV79] to provide an (exponentially small) upper bound on the probability (with respect to the uniform measure $\mathsf{Q}_{\mathcal{F}_n}$) that $|\mathcal{D}_f|$ is below a constant fraction of its expectation value

$$\mathsf{Q}_{\mathcal{F}_n}\left(|\mathcal{D}_{f_r}| \leq (1-\gamma_1)2^{n-2}\right) \leq e^{-\gamma_1^2 \frac{2^{n-2}}{2}}\,. \tag{3.164}$$

Consequently, with probability $1 - e^{-\gamma_1^2 \frac{2^{n-2}}{2}}$ (with respect to the uniform measure $\mathsf{Q}_{\mathcal{F}_n}(f_r)$) we obtain (3.162), following the steps of (3.160). □

3.5.3 Prefix-code attacks and their limits

In this section we construct a set of distributions $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ that we call "prefix-code attacks". This is the set of attacks introduced in [AFTS12], but our presentation is simpler and allows for a simple generalisation to the *dynamic* TONS adversary as well as systems with more parties, higher dimensions and arbitrary input alphabets. Similar to the attack on linear functions in Section 3.5.1, the idea is that for every string $a_{\leq n}$ the adversary biases a certain bit a_i in his preferred direction by a value of ε. In the attacks in Section 3.5.1 for the function $f(a_{\leq n}) = \bigoplus_{i\in\mathcal{T}} a_i$ the bit a_{i_t} was biased by a value of ε and the direction of the bias depended on the party of the previously obtained bits $a_{i_1}, ..., a_{i_t}$. Here, also the choice of which bit a_i will depend on the previously obtained bits $a_1, a_2, ..., a_{i-1}$. Note that as a consequence

of Theorem 3.4.4, we may, without loss of generality, restrict Alice to the use of balanced functions $f(a_{\leq n})$ only.

Definition 14 (Prefix-code)**.** We define a *prefix-code* C to be a set of codewords $C = \{c_1, c_2, ..., c_k\}$ with $c_m \in \{0,1\}^\ell, 0 \leq \ell \leq n-1, 1 \leq m \leq k$ such that for any $a_{\leq n}$ the code C contains a unique $c_m \in C$ that is a prefix of $a_{\leq n}$.

Note that such a code saturates the *Kraft inequality* [Kra49], *i.e.*,

$$\sum_{m=1}^{k} 2^{-|c_m|} = 1 \,. \tag{3.165}$$

Definition 15 (Influence)**.** We define the *influence* $\Delta^f(a_{<i})$ *of* a_i *given the prefix* $a_{<i}$ *on the function* $f(a_{\leq n})$ as

$$\Delta^f(a_{<i}) := \frac{1}{2}\Big(\mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}, a_i = 0) - \mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}, a_i = 1)\Big) \,, \tag{3.166}$$

where $\mathsf{Q}(a_{\leq n}) = 2^{-n}$.

Note that for the uniform distribution $\mathsf{Q}(a_{\leq n})$ we have

$$\mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}) = \frac{1}{2}(\mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}, a_i = 0) + \mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}, a_i = 1)) \,. \tag{3.167}$$

and, therefore,

$$\mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}, a_i) = \mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}) + (-1)^{a_i}\Delta^f(a_{<i}) \,. \tag{3.168}$$

Theorem 3.5.4. *For any prefix code* C *and any balanced function* $f(a_{\leq n})$ *there exists an ordered* $(\varepsilon, \mathcal{S})$*-divisible distribution* $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ *such that*

$$\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e) = \frac{1}{2} + 2\varepsilon \cdot \sum_m 2^{-|c_m|}|\Delta^f(c_m)| \,. \tag{3.169}$$

Proof. We construct the set of ordered $\mathcal{S}$-influenceable distributions $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$. For any $\mathcal{S} \in \mathcal{P}([n])$ let $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ be defined as

$$\mathsf{Q}_{o-\mathcal{S}}(e) = \frac{1}{2} \,, \tag{3.170}$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) = \begin{cases} \frac{1}{2}\left(1 + \mathrm{sign}(\Delta^f(c_m))(-1)^{e\oplus a_i}\right) & \text{if } \exists m: \; a_{<i} = c_m \cap i \in \mathcal{S} \\ \frac{1}{2} & \text{otherwise} \,, \end{cases} \tag{3.171}$$

where $\mathrm{sign}(x)$ is the signum function, defined as

$$\mathrm{sign}(x) := \begin{cases} -1 & x < 0 \\ 1 & \text{otherwise} . \end{cases} \tag{3.172}$$

Conditions (3.107) and (3.108) can be easily verified. (3.171) is equivalent to

$$\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} \,|\, e) = \begin{cases} 2^{-n} \cdot (1 + \mathrm{sign}(\Delta^f(c_m))(-1)^{e \oplus a_i}) & \exists i, m : \ i \in \mathcal{S} \cap a_{<i} = c_m \\ 2^{-n} & \text{otherwise} , \end{cases} \tag{3.173}$$

and (3.107) follows by averaging (3.173) over e. Equation (3.108) holds directly by construction (3.171). In order to complete the proof we need to show that our construction (3.170) and (3.171) for the set $\{\mathsf{Q}_{o-\mathcal{S}}\}$ for $\mathcal{S} \in \mathcal{P}([n])$ satisfies (3.169). The argument is analogous to the one in Section 3.5.1 for the equations (3.146) and (3.147). For the $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ that satisfy the condition $i \in \mathcal{S}$ in (3.171) we use (3.168) and obtain

$$\mathsf{Q}_{o-\mathcal{S}}(f(a_{\leq n}) = 0 \,|\, a_{<i} = c_m, e) = \mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}) + (-1)^e |\Delta^f(c_m)| , \tag{3.174}$$

for the uniform distribution $\mathsf{Q}(a_{\leq n})$. In the decomposition

$$\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e) = \sum_{\mathcal{S} \in \mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e) , \tag{3.175}$$

the distributions $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ that satisfy the condition $i \in \mathcal{S}$ in (3.171) have weight exactly 2ε. We conclude that for balanced functions $f(a_{\leq n})$ we have

$$\begin{aligned} \mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = 0 \,|\, e) &= \sum_{m=1}^{k} \mathsf{Q}_{o-\varepsilon}(a_{<i} = c_m | e)\, \mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = 0 \,|\, a_{<i} = c_m, e) \\ &= \sum_{m=1}^{k} 2^{-c_m}\, \mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = 0 \,|\, a_{<i} = c_m, e) \\ &= \sum_{m=1}^{k} 2^{-c_m} \Big((1 - 2\varepsilon)\, \mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}) \\ &\quad + 2\varepsilon \left(\mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}) + (-1)^e |\Delta^f(c_m)| \right) \Big) \\ &= \frac{1}{2} + 2\varepsilon \sum_{m=1}^{k} 2^{-c_m} (-1)^e \, |\Delta^f(c_m)| , \end{aligned} \tag{3.176}$$

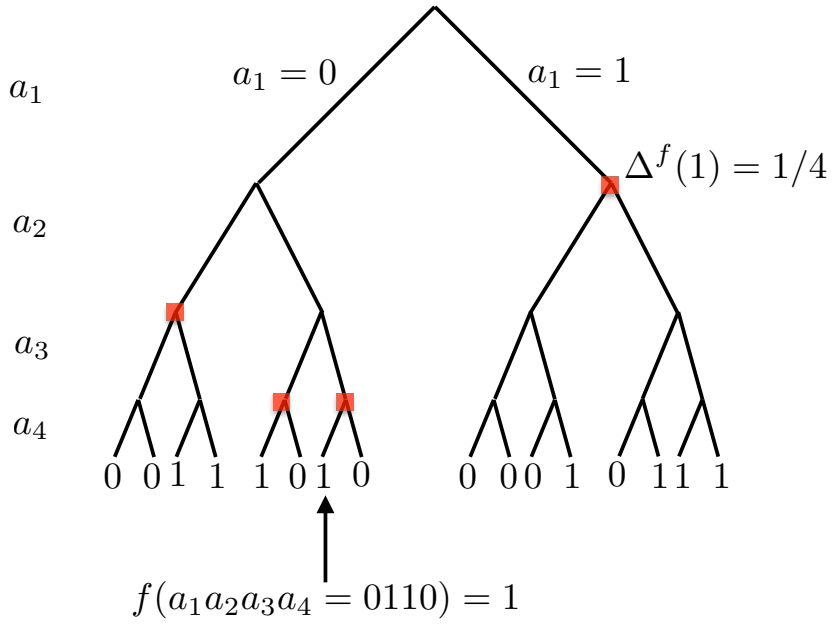

Figure 3.4. Example of an unbalanced function f. The branches of the tree represent the string inputs $a_1a_2a_3a_4$ the leaves of the tree the values of the function $f(a_1a_2a_3a_4)$. The red marks indicate (one possibility of) an optimal prefix-code C with codewords $\{00, 010, 011, 1\}$ that maximises the expectation value of $|\Delta^f(c_m)|$.

and, finally, that

$$\begin{aligned}\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e) &= \frac{1}{2}\Big(\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = 0 \,|\, e = 0) + 1 - \mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = 0 \,|\, e = 1)\Big) \\ &= \frac{1}{2} + 2\varepsilon \sum_{m=1}^{k} 2^{-c_m} |\Delta^f(c_m)| \,. \end{aligned} \tag{3.177}$$

□

We will continue with the argument presented in [AFTS12], which implies the impossibility of super-linear privacy amplification against a TONS-adversary.

Theorem 3.5.5. *[AFTS12] Assume that $f(a_{\leq n}) : \{0,1\}^n \to \{0,1\}$ is balanced, i.e., $|\{a_{\leq n} : f(a_{\leq n}) = 0\}| = 2^{n-1}$, then for every $a_{\leq n}$ there exists a prefix $a_{<i}$ with $i \in [n]$, such that*

$$|\Delta^f(a_{<i})| \geq \frac{1}{2n} \,. \tag{3.178}$$

Proof. For balanced functions, we have $\mathsf{Q}(f(a_{\leq n}) = 0) = 1/2$ and, by applying (3.168) recursively i times, we obtain

$$\begin{aligned}\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = 0 \,|\, a_{\leq i}) &= \mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = 0 \,|\, a_{<i}) + (-1)^{a_i}\Delta^f(a_{<i}) \\ &= \frac{1}{2} + \sum_{j=1}^{i}(-1)^{a_j}\Delta(a_{<j}) \,. \end{aligned} \tag{3.179}$$

For all $a_{\leq n}$, we have trivially $\mathsf{Q}(f(a_{\leq n}) = 0) \,|\, a_{\leq n}) = 0$ (or $= 1$), which implies

$$\begin{aligned}\frac{1}{2} &= \left|\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = 0 \,|\, a_{\leq n}) - \frac{1}{2}\right| \\ &= \left|\sum_{i=1}^{n}(-1)^{a_i}\Delta^f(a_{<i})\right| \quad \forall a_{\leq n} \,. \end{aligned} \tag{3.180}$$

□

Corollary 3.5.6. *For any balanced function $f(a_{\leq n})$, by Theorem 3.5.5 there must exist a, not necessarily unique, prefix code $C^* = \{c_1, c_2, ..., c_k\}$, such that*

$$\frac{1}{2}\sum_{m=1}^{k} 2^{-|c_m|}|\Delta^f(c_m)| \geq \frac{1}{2n} \,. \tag{3.181}$$

Note that for any prefix code $\sum_{m=1}^{k} 2^{-|c_m|} = 1$.

Theorem 3.5.7. *For* any *function $f(a_{\leq n})$ there exists an ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$, such that*

$$\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e) \geq \frac{1}{2} + \frac{\varepsilon}{n} \,. \tag{3.182}$$

Proof. Theorem 3.5.7 follows from Corollary 3.5.6 and Theorem 3.5.4 for balanced functions, and thus via Theorem 3.4.4 also for unbalanced functions. □

Another direct consequence of Theorem 3.5.4 is that privacy amplification is impossible using functions that are composed of affine functions glued together by prefix-codes.

Definition 16 (Prefix-coded functions)**.** A function $f(a_{\leq n}) : \{0,1\}^n :\to \{0,1\}$ is called a *prefix-coded affine function* if there exists a prefix code $C = \{c_1, c_2, ..., c_k\}$ and corresponding set of bits $\mathcal{B} = \{b_1, b_2, ..., b_k\}$ such that

$$f(a_{\leq n}) = \sum_{m=1}^{k} \delta(a_{\leq |c_m|}, c_m) \cdot (a_{|c_m+1} \oplus b_m) \,. \tag{3.183}$$

In the definition of a prefix-coded function (3.183) implies that

$$\sum_{m=1}^{k} 2^{-|c_m|} |\, \Delta^f(c_m)| = \sum_{m=1}^{k} 2^{-|c_m|} \frac{1}{2} = \frac{1}{2}\,, \tag{3.184}$$

and consequently also prefix-coded functions, which are a generalisation of affine functions, cannot be used for privacy amplification:

Corollary 3.5.8. *Impossibility of privacy amplification using prefix-coded affine functions. Let $f(a_{\leq n}) : \{0,1\}^n :\to \{0,1\}$ be a prefix-coded function, then there exists a $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ such that*

$$\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e) = \frac{1}{2} + \varepsilon\,. \tag{3.185}$$

We want to end this section by proving a stronger bound than the one from Theorem 3.5.7 that for a large class of boolean functions, the *monotonic* functions.

Definition 17 (Monotonic functions)**.** A function $f(a_{\leq n})$ is *monotonic* if for any $i \in [n]$ it holds that

$$f(a_{<i}, a_i = 0, a_{>i}) \leq f(a_{<i}, a_i = 1, a_{>i}) \quad \forall a_{<i}, a_{>i}\,. \tag{3.186}$$

For monotonic functions we can make use of the so-called *KKL Theorem* by Kahn, Kalai and Linial [KKL88], [O'D14] on influences of individual variables on boolean functions. In the present context a simplified version of the theorem states that

Theorem 3.5.9. *Let $f(a_{\leq n})$ be a monotonic balanced function. Then there must exist at least one $i \in [n]$ such that*

$$2^{-i+1} \sum_{a_{<i}} \Delta^f(a_{<i}) \geq c\, \frac{\log(n)}{n}\,, \tag{3.187}$$

for some constant $c > 0$.

Together with Theorem 3.5.4 this yields the following theorem:

Theorem 3.5.10. *For any monotonic balanced function $f(a_{\leq n})$ there exists an ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$, such that*

$$\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e) \geq \frac{1}{2} + 2\,c\,\varepsilon\,\frac{\log(n)}{n}\,, \tag{3.188}$$

where the constant c is the one from Theorem 3.5.9.

3.5.4 Majority and "prefix-code" attacks vs. the "maximum-likelihood" attack.

In this section we investigate in more detail the case of a specific set of functions, the majority functions, defined in (3.86) (for odd n). The majority functions are, due to the high degree of symmetry (relatively) easy to analyse as they are only functions of the Hamming weight of $a_{\leq n}$. First we will inspect the performance of "prefix-code" attacks introduced in Section 3.5.3. We will see that for large n this performance scales with $\theta(\varepsilon/\sqrt{n})$. Then we will introduce another construction of the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$, which we denote as a "maximum-likelihood" attack. Intuitively, for each $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$, the adversary approximates the majority function on the whole string $a_{\leq n}$ with the (partial) majority on the substring $a_{\mathcal{S}}$. This attack performs far better and proves that with majority functions no privacy amplification against a TONS adversary can be achieved at all.

Theorem 3.5.11. *Let the distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ be a "prefix-code attack" constructed as in the proof of Theorem 3.5.4 for the majority function $Maj_n(a_{\leq n})$. Then, for the optimal choice of a prefix-code $C = \{c_1, ..., c_k\}$ and bits $\{a^*(1), ..., a^*(k)\}$, the performance of this attack is*

$$\mathsf{Q}_{o-\varepsilon}(Maj_n(a_{\leq n}) = e) = \frac{1}{2} + \theta\left(\frac{\varepsilon}{\sqrt{n}}\right) . \tag{3.189}$$

Proof. Independent of the choice of prefix-code $C = \{c_1, c_2, ..., c_k\}$, the optimal choice of bits $\{a^*(1), ..., a^*(k)\}$ is always $a^*(m) = 0$ since

$$\Delta^{\mathrm{Maj}_n}(a_{<i}) \geq 0 \quad \forall a_{<i}, 1 \leq i \leq n . \tag{3.190}$$

The distributions $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n})$ constructed in Theorem 3.5.4 by

$$\mathsf{Q}_{o-\mathcal{S}}(e) = \frac{1}{2} , \tag{3.191}$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) = \begin{cases} \frac{1}{2}\left(1 + \mathrm{sign}(\Delta^f(c_m))\,(-1)^{e\oplus a_i}\right) & \text{if } \exists m:\ a_{<i} = c_m \cap i \in \mathcal{S} \\ \frac{1}{2} & \text{otherwise} , \end{cases} \tag{3.192}$$

have the following property: for any $i \in [n]$ the conditioned distribution of a single bit $\mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e)$ is uniform, except if $a_{<i} \in C \cap i \in \mathcal{S}$, then $\mathsf{Q}_{o-\mathcal{S}}(a_i = e \,|\, a_{<i}e) = 1$. For each $a_{\leq n}$ the condition $a_{<i} \in C \cap i \in \mathcal{S}$ holds in the composition

$$\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e) = \sum_{\mathcal{S}} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e) , \tag{3.193}$$

with probability 2ε, and with probability $(1 - 2\varepsilon)$ the string $a_{\leq n}$ is uniform, *i.e.*, $\mathsf{Q}_{o-\varepsilon}(a_{\leq n} \,|\, e) = 2^{-n}$. Thus, in $\mathsf{Q}_{o-\varepsilon}(a_{\leq n} \,|\, e)$ with probability 2ε a single bit a_i is equal to e and all other bits $a_{\neq i}$ are uniformly distributed, while, otherwise, the whole distribution is uniform. Thus,

$$\begin{aligned}
\mathsf{Q}_{o-\varepsilon}(\mathrm{Maj}_n(a_{\leq n}) = e) &= \sum_e \mathsf{Q}_{o-\varepsilon}(e)\, \mathsf{Q}_{o-\varepsilon}(\mathrm{Maj}_n(a_{\leq n}) = e \,|\, e) \\
&= (2\varepsilon) \cdot 2^{-n+1} \sum_{i=0}^{\frac{n-1}{2}} \binom{n-1}{i} + (1-2\varepsilon)\frac{1}{2} \\
&= (2\varepsilon)\left(\frac{1}{2} + 2^{-n}\binom{n-1}{\frac{n-1}{2}}\right) + (1-2\varepsilon)\frac{1}{2} \\
&= \frac{1}{2} + \varepsilon \cdot 2^{-n+1}\binom{n-1}{\frac{n-1}{2}} \\
n \to \infty \quad &\approx \frac{1}{2} + \frac{\varepsilon}{\sqrt{n}}\,, \qquad (3.194)
\end{aligned}$$

where the last line follows from Stirling's approximation. □

Now we derive another type of attacks on majority functions that we refer to as "maximum-likelihood" attacks. In case of the majority, this essentially boils down to constructing a distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ that, out of the view of the adversary, approximates $\mathrm{Maj}_n(a_{\leq n})$ with $\mathrm{Maj}_s(a_{\mathcal{S}})$.

Theorem 3.5.12. *For any odd n and any $\mathcal{S} \in \mathcal{P}([n])$ where $s = |\mathcal{S}|$ is odd, there exists an ordered $\mathcal{S}$-influenceable distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$, see* (3.107) *and* (3.108)*, with*

$$\begin{aligned}
\mathsf{Q}_{o-\mathcal{S}}(Maj_n(a_{\leq n}) = e) &= \mathsf{Q}_{o-\mathcal{S}}(Maj_n(a_{\leq n}) = Maj_s(a_{\mathcal{S}})) \\
&= 1 - \sum_{h_1 = \max(0, \frac{1-n}{2}+s)}^{\frac{s-1}{2}} \binom{s}{h_1} \sum_{h_2 = \frac{n-1}{2} - h_1}^{n-s} \binom{n-s}{h_2}. \qquad (3.195)
\end{aligned}$$

Proof. We prove this by construction of the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ as

$$\mathsf{Q}_{o-\mathcal{S}}(e) = \frac{1}{2} \qquad (3.196)$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n} \,|\, e) = 2^{-n+1} \cdot \delta(\mathrm{Maj}_{\mathcal{S}}(a_{\mathcal{S}}), e)\,. \qquad (3.197)$$

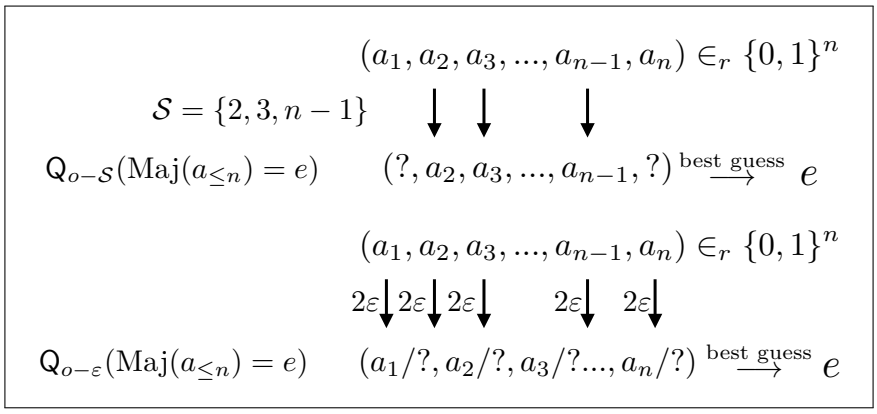

Figure 3.5. Intuition behind the majority attacks $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ constructed in Theorem 3.5.12. In the upper scenario we depict the knowledge the adversary obtains in the attack (3.196) *and* (3.197) *on $Maj_n(a_{\leq n})$ for $\mathcal{S} = 2, 3, n-1$. The string $a_{\leq n}$ is drawn uniformly at random and then the bits in $a_{\mathcal{S}}$ are given to the adversary. His guess for $Maj_n(a_{\leq n})$ is the majority of the bits $a_{\mathcal{S}}$ that he receives $Maj_s(a_{\mathcal{S}})$, see* (3.197). *The lower scenario depicts the intuition behind the attack $\mathsf{Q}_{o-\varepsilon}$ on the majority of n bits which is derived from the set $\{\mathsf{Q}_{o-\mathcal{S}}\}$. Here again $a_{\leq n}$ is drawn uniformly at random and but this time each bit a_i are given to the adversary independently with probability 2ε.*

Uniformity of Alice's string $a_{\leq n}$ (3.107) follows directly from (3.196), (3.197), and the fact that $\mathrm{Maj}_{\mathcal{S}}(a_{\mathcal{S}})$ is a balanced function. Equation (3.108) follows from (3.197) and that if $i \notin \mathcal{S}$, then $\mathrm{Maj}_{\mathcal{S}}(a_{\mathcal{S}})$ is (by definition) independent of a_i. Therefore, we have

$$\begin{aligned}\mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) &= \frac{\sum_{a_{>i}} \mathsf{Q}_{o-\mathcal{S}}(a_{\leq i}, a_{>i} \,|\, e)}{\sum_{a_{\geq i}} \mathsf{Q}_{o-\mathcal{S}}(a_{<i}, a_{\geq i} \,|\, e)} \\ &= \frac{|\{a_{>i} : \mathrm{Maj}_{\mathcal{S}}(a_{\leq i}, a_{>i}) = e\}|}{|\{a_{\geq i} : \mathrm{Maj}_{\mathcal{S}}(a_{<i}, a_{\geq i}) = e\}|} \\ &= \frac{1}{2}\,. \end{aligned} \tag{3.198}$$

We calculate Eve's guessing probability as

$$\begin{aligned}\mathsf{Q}_{o-\mathcal{S}}(\mathrm{Maj}_n(a_{\leq n}) = e) &= 1 - \mathsf{Q}_{o-\mathcal{S}}(\mathrm{Maj}_n(a_{\leq n}) \neq e) \\ &= 1 - \sum_e \mathsf{Q}_{o-\mathcal{S}}(e)\, \mathsf{Q}_{o-\mathcal{S}}(\mathrm{Maj}_n(a_{\leq n}) \neq e \,|\, e) \\ &\overset{(3.196),(3.197)}{\Rightarrow} \quad = 1 - \sum_e 2^{-n} \sum_{a_{\mathcal{S}}} |\{a_{\overline{\mathcal{S}}} : \mathrm{Maj}_s(a_{\mathcal{S}}) = e \ \cap\ \mathrm{Maj}_n(a_{\mathcal{S}}, a_{\overline{\mathcal{S}}}) = \overline{e}\}| \\ &= 1 - 2^{-n+1} \sum_{a_{\mathcal{S}}} |\{a_{\overline{\mathcal{S}}} : \mathrm{Maj}_s(a_{\mathcal{S}}) = 0 \ \cap\ \mathrm{Maj}_n(a_{\mathcal{S}}, a_{\overline{\mathcal{S}}}) = 1\}| \\ &= 1 - 2^{-n+1} \sum_{h_1 = \max(0, s - \frac{n-1}{2})}^{\frac{s-1}{2}} \binom{s}{h_1} \sum_{h_2 = \frac{n+1}{2} - h_1}^{n-s} \binom{n-s}{h_2}, \end{aligned} \tag{3.199}$$

where we used for the fourth equality the fact that the majority function is odd. The idea behind the last equation is to count how many strings $a_{\mathcal{S}}$ have Hamming weight h_1 and then count how many strings $a_{\overline{\mathcal{S}}}$ have Hamming weight $h_2 \geq (n+1)/2 - h_1$. Note that the latter is only possible if

$$n - s \geq \frac{n+1}{2} - h_1 \quad \Leftrightarrow \quad h_1 \geq s - \frac{n-1}{2}\,. \tag{3.200}$$

□

The probability $\mathsf{Q}_{o-\mathcal{S}}(\mathrm{Maj}_n = e)$ in the attack presented in Theorem 3.5.12 only depends on the size of $\mathcal{S}$, *i.e.*, $s := |\mathcal{S}|$. For an ordered $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ we have

$$\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e) = \sum_{\mathcal{S} \in \mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\,. \tag{3.201}$$

For large n, almost all of the probability weight in the composition (3.109) of $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ is concentrated on $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ with $s \in [(1-\delta)\cdot\varepsilon n, (1+\delta)\cdot\varepsilon n])$ for $\delta \in \theta(1/\sqrt{n})$. For s odd we let $\mathsf{Q}_{o-\mathcal{S}}$ be as in Theorem 3.5.12 and for s even we let $\mathsf{Q}_{o-\mathcal{S}}$ be as an $\mathsf{Q}_{o-\mathcal{S}\setminus\{i\}}$ in Theorem 3.5.12 for some arbitrary $i \in \mathcal{S}$. In the limit of large n we have then (through the concentration of measure induced by the CLT)

$$\begin{aligned}\mathsf{Q}_{o-\varepsilon}(Maj_n(a_{\leq n}) = e) &= \sum_{\mathcal{S}\in\mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{o-\mathcal{S}}(Maj_n(a_{\leq n}) = e) \\ &\overset{n\to\infty}{\geq} \mathsf{Q}_{o-\mathcal{S}'}(Maj_n(a_{\leq n}) = e)\,, \end{aligned} \tag{3.202}$$

for some $\mathcal{S}'$ with $s' = (\varepsilon - \delta)\cdot n$ for any $\delta > 0$.

Theorem 3.5.13. *Let $|\mathcal{S}| = s = c\,n$ for some constant $0 < c < 1$ such that s is odd. Then there exists a series of ordered $\mathcal{S}$-influenceable distributions $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ such that*

$$\mathsf{Q}_{o-\mathcal{S}}(Maj_n(a_{\leq n}) = e) \overset{n\to\infty}{=} 1 - \frac{arctan\left(\sqrt{\frac{1-c}{c}}\right)}{\pi} \tag{3.203}$$

Proof. Let the series of $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ be defined by (3.196) and (3.197) for increasing n. By (3.199) we have

$$\begin{aligned}\mathsf{Q}_{o-\mathcal{S}}(\mathrm{Maj}_n(a_{\leq n}) = e) &= 1 - 2 \sum_{h_1=\max(0,s-\frac{n-1}{2})}^{\frac{s-1}{2}} 2^{-s}\binom{s}{h_1} \sum_{h_2=\frac{n+1}{2}-h_1}^{n-s} 2^{-n+s}\binom{n-s}{h_2} \\ &= 1 - 2 \sum_{h_1=0}^{\mathrm{Min}(\frac{s-1}{2},\frac{n-s}{2}-1)} 2^{-s}\binom{s}{\frac{s-1}{2}-h_1} \sum_{h_2=h_1+1}^{n-s} 2^{-n+s}\binom{n-s}{\frac{-n+s}{2}+h_2}, \end{aligned} \tag{3.204}$$

where in the last line h_1 and h_2 resemble the distance in Hamming weight of the strings $a_{\mathcal{S}}$ and $a_{\overline{\mathcal{S}}}$ from the string with half the bits 1

$$h_1 := -\frac{s-1}{2} + \sum_{i=1}^{s} a_{j_i} \qquad j_i \in \mathcal{S} \tag{3.205}$$

$$h_2 := -\frac{n-s}{2} + \sum_{i=}^{n-s} a_{k_i} \qquad k_i \in \overline{\mathcal{S}}\,. \tag{3.206}$$

We define $h_1' := (h_1 + 1/2)/\sqrt{s}$ and $h_2' := h_2/\sqrt{n-s}$ and if we set $s = c\,n$ and

$h_2 = h_1 + 1$, we obtain

$$\begin{aligned} h_2' &= \frac{h_2}{\sqrt{(1-c)\,n}} \\ &= \frac{h_1 + 1}{\sqrt{(1-c)\,n}} \\ &= \sqrt{\frac{c}{(1-c)}} h_1' + \frac{1}{2\sqrt{(1-c)\,n}} \,. \end{aligned} \tag{3.207}$$

Then, the central limit theorem (CLT) and (3.204) imply

$$\begin{aligned} \mathsf{Q}_{o-\mathcal{S}}(\mathrm{Maj}_n(a_{\leq n}) = e) =& 1 - 2 \sum_{h_1=0}^{\mathrm{Min}(\frac{s-1}{2}, \frac{n-s}{2}-1)} 2^{-s} \binom{s}{\frac{s-1}{2} - h_1} \sum_{h_2=h_1+1}^{n-s} 2^{-n+s} \binom{n-s}{\frac{-n+s}{2} + h_2} \\ =& 1 - 2 \sum_{h_1=0}^{\mathrm{Min}(\frac{c\cdot n-1}{2}, \frac{(1-c)n}{2}-1)} 2^{-c\cdot n} \binom{c\cdot n}{\frac{c\cdot n-1}{2} - h_1} \sum_{h_2=h_1+1}^{n-s} 2^{-(1-c)n} \binom{(1-c)n}{\frac{-(1-c)n}{2} + h_2} \\ \overset{CLT}{\Rightarrow} \quad \overset{n\to\infty}{=}& 1 - 2 \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \int_{\sqrt{\frac{c}{1-c}}x_1}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}} dx_2\, dx_1 \\ =& 1 - \frac{1}{\pi} \int_0^\infty e^{-\frac{x_1^2}{2}} \int_{\sqrt{\frac{c}{1-c}}x_1}^\infty e^{-\frac{x_2^2}{2}} dx_2\, dx_1 \\ =& 1 - \frac{\arctan\left(\sqrt{\frac{1-c}{c}}\right)}{\pi} \,. \end{aligned} \tag{3.208}$$

□

A direct consequence of Theorem 3.5.13 and (3.202) is

Theorem 3.5.14. *For any $\delta > 0$, there exists a series of $\{\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)\}$ such that*

$$\mathsf{Q}_{o-\varepsilon}(Maj_n(a_{\leq n}) = e) \overset{n\to\infty}{\geq} 1 - \frac{arctan\left(\sqrt{\frac{1-(\varepsilon-\delta)}{(\varepsilon-\delta)}}\right)}{\pi} \,. \tag{3.209}$$

3.6 Generalisation to a dynamic TONS adversary

In this section we show that the construction of TONS attacks from classical distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ in Section 3.4.1 can be generalised to dynamic TONS attacks, see (2.3) and (2.4) for a comparison of the two. This will become a crucial step to derive bounds on distillation protocols in Chapter 4. We proceed in two steps. First, we

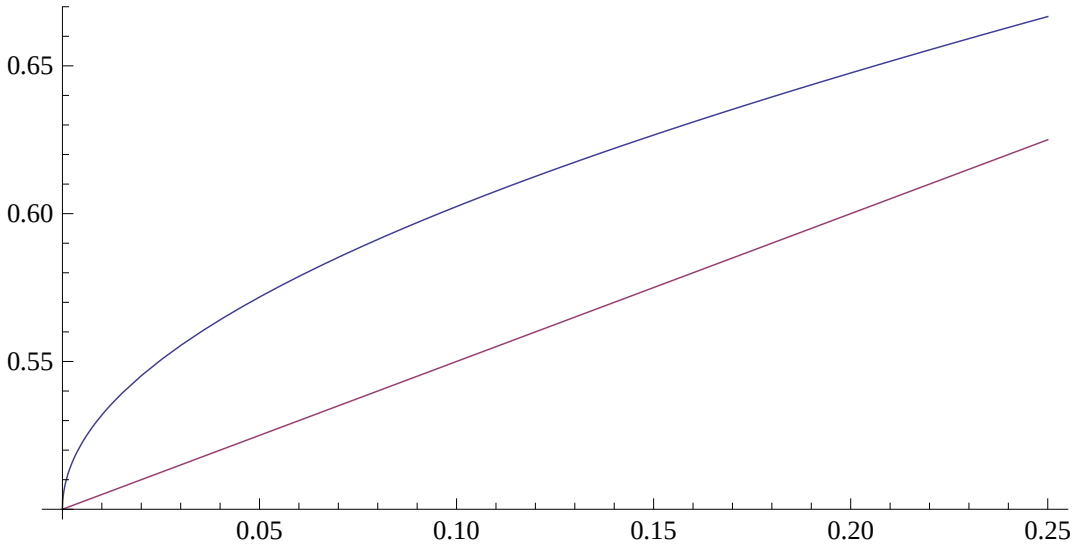

Figure 3.6. Comparison of the knowledge of the adversary about majority in the limit of large n (the upper line) with his knowledge about a single bit a_i (the lower line). The upper line represents the function $1 - \arctan(\sqrt{(1-\varepsilon)/\varepsilon})/\pi$, see (3.203), and the lower line $1/2 + \varepsilon$, see Theorem 3.1.3. The x-axis corresponds to the values for ε between 0 and $1/4$. We observe that the majority function is not useful for TONS privacy amplification: the knowledge of the adversary about $\mathrm{Maj}_n(a_{\leq n})$ is larger than about the individual bits a_i.

show in Theorem 3.6.1 that the distributions $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ we constructed in Theorem 3.4.2 from classical distributions $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ satisfy more than just the TONS conditions but hold for *any* dynamic order $\{k_i\}$ on Bob's side. Then, we show in Theorem 3.6.2 that for any function $f(a_{\leq n})$ a dynamic TONS attack can be constructed from a *standard* TONS attack on a function $\tilde{f}$, which is constructed from $f(a_{\leq n})$ via suitable permutations of the input strings.

Theorem 3.6.1. *Fix an arbitrary dynamic order $\{k_i\}$ for Bob, i.e., let the functions $k_i(b_{k_{<i}})$ be arbitrary, and fix Alice dynamic order to be trivial, i.e., for the set $\{j_i\}$ let*

$$j_i(a_{j_{<i}}) = i \quad \forall a_{j_{<i}}, i \in [n]\,. \tag{3.210}$$

For any set $\mathcal{S} \subseteq [n]$ the distributions $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ constructed in Theorem 3.4.2 by (3.118)-(3.120) *satisfy the dynamic time-ordered no-signalling conditions with respect to the dynamic orders $\{j_i\}$ and $\{k_i\}$.*

Proof. We need to show that $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n})$ has a well-defined marginal distribution $\mathsf{P}_{o-\mathcal{S}}(a_{\leq i_A}b_{k_{\leq i_B}}\,|\,ex_{\leq i_A}y_{k_{\leq i_B}})$ on subsystems $A_{\leq i_A}B_{k_{\leq i_B}}$ for any $0 \leq i_A, i_B \leq n$. To prevent us from entering a horrible index jungle, let us define some sets. Define the set $\mathcal{T}_{b_{k_{<i_B}}}$ for the string $b_{k_{<i_B}} = b_{k_1}...b_{k_{i_B-1}}$ as the indices of the "past" i_B systems of Bob:

$$\mathcal{T}_{b_{<i_B}} := \{i : i \in \{k_1, k_2(b_{k_1}), ..., k_{i_B}(b_{k_{<i_B}})\}\}\,, \tag{3.211}$$

and $\overline{\mathcal{T}}_{b_{k_{<i_B}}}$ as the "future" $n - i_B$ systems

$$\overline{\mathcal{T}}_{b_{<i_B}} := [n]\backslash\mathcal{T}_{b_{<i_B}}\,. \tag{3.212}$$

We divide both sets $\mathcal{T}_{b_{<i_B}}$ and $\overline{\mathcal{T}}_{b_{<i_B}}$ again into two subsets, the intersection with the set $[i_A]$ and its complement

$$\mathcal{T}^1 := \mathcal{T}_{b_{<i_B}} \cap [i_A] \tag{3.213}$$

$$\mathcal{T}^2 := \overline{\mathcal{T}}_{b_{<i_B}} \cap [i_A] \tag{3.214}$$

$$\mathcal{T}^3 := \mathcal{T}_{b_{<i_B}} \cap ([n]\backslash[i_A]) \tag{3.215}$$

$$\mathcal{T}^4 := \overline{\mathcal{T}}_{b_{<i_B}} \cap ([n]\backslash[i_A])\,. \tag{3.216}$$

Recall the construction (3.118) to (3.120) of $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ in Theorem 3.4.2 from the distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e\,|\,x_{\leq n})$ and take a close look at (3.120)

$$\mathsf{P}_{o-\mathcal{S}}(a_ib_i\,|\,a_{<i}b_{<i}x_{\leq n}y_{\leq n}e) = \begin{cases} \mathsf{U}(b_i\,|\,y_i)\,\mathsf{Q}_{o-\mathcal{S}}(a_i\,|\,a_{<i}e) & i \in \mathcal{S} \\ \mathsf{PR}(a_ib_i\,|\,x_iy_i) & \text{otherwise}\,. \end{cases} \tag{3.217}$$

The right-hand side is completely independent of $b_{<i}$, $x_{\neq i}$, and $y_{\neq i}$. Hence, we do not need to condition on these variables

$$\mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} b_{<i} x_{\leq n} y_{\leq n} e) = \mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} x_i y_i e) \,. \tag{3.218}$$

Let us also recall from the proof of Theorem 3.4.2 that

$$\sum_{a_i b_i} \mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} x_i y_i e) = 1 \tag{3.219}$$

$$\sum_{a_i} \mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} x_i y_i e) = \mathsf{U}(b_i \,|\, x_i) \tag{3.220}$$

$$\sum_{b_i} \mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} x_i y_i e) = \mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i} e) \,. \tag{3.221}$$

We are ready to show that $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n} b_{\leq n} \,|\, e x_{\leq n} y_{\leq n})$ has a well-defined marginal distribution $\mathsf{P}_{o-\mathcal{S}}(a_{\leq i_A} b_{k_{\leq i_B}} \,|\, e x_{\leq i_A} y_{k_{\leq i_B}})$ on subsystems $A_{\leq i_A} B_{k_{\leq i_B}}$ for any $0 \leq i_A, i_B \leq n$:

$$\begin{aligned}
&\sum_{a_{>i_A} b_{k_{>i_B}}} \mathsf{P}(a_{\leq n} b_{\leq n} \,|\, e x_{\leq n} y_{\leq n}) \\
&= \sum_{a_{>i_A} b_{k_{>i_B}}} \prod_{j \in [n]} \mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} x_i y_i e) \\
&= \sum_{b_{k_{>i_B}}} \prod_{j \in [i_A]} \mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} x_i y_i e) \prod_{j \in [n]/[i_A]} \mathsf{U}(b_i \,|\, y_i) \\
&= \prod_{j \in \mathcal{T}^1} \mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} x_i y_i e) \cdot \Big(\sum_{b_{\mathcal{T}^2}} \prod_{j \in \mathcal{T}^2} \mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} x_i y_i e) \Big) \\
&\qquad \cdot \Big(\sum_{b_{\mathcal{T}^4}} \prod_{j \in \mathcal{T}^3 \cap \mathcal{T}^4} \mathsf{U}(b_i \,|\, y_i) \Big) \\
&= \prod_{j \in \mathcal{T}^1} \mathsf{P}_{o-\mathcal{S}}(a_i b_i \,|\, a_{<i} x_i y_i e) \cdot \Big(\prod_{j \in \mathcal{T}^2} \mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i} e) \Big) \cdot \prod_{j \in \mathcal{T}^3} \mathsf{U}(b_i \,|\, y_i) \\
&=: \mathsf{P}_{o-\mathcal{S}}(a_{\leq i_A} b_{k_{\leq i_B}} \,|\, e x_{\leq i_A} y_{k_{\leq i_B}}) \,,
\end{aligned} \tag{3.222}$$

where we use (3.220) for the second equation and the definition of the sets $\mathcal{T}^1$ to $\mathcal{T}^4$ (3.213) to (3.216) for the third equation. For the fourth equation we use (3.219) and (3.221) and, finally, since the right-hand side of the fourth equation is independent of $x_{>i_A}$ and $y_{k_{>i_B}}$ it forms a well-defined marginal on systems $A_{\leq i_A} B_{k_{\leq i_B}}$. □

Theorem 3.6.2. *Let $\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n})$ be a TONS attack constructed according to Theorem 3.4.2 on a function $\tilde{f}(a_{\leq n})$ that we will specify below. For any choice of*

dynamic orders $\{j_i\}$ *and* $\{k_i\}$ *and any function* $f(a_{\leq n})$ *there exists a dynamic TONS attack* $\mathsf{P}'(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ *such that*

$$\mathsf{P}'(f(a_{\leq n}) = e) = \mathsf{P}(\tilde{f}(a_{\leq n}) = e)\,. \tag{3.223}$$

The function $\tilde{f}$ *is defined by*

$$\tilde{f}(a_1, a_2, ..., a_n) := f(a'_1, a'_2, ..., a'_n) \tag{3.224}$$

with $a'_i = a_{j_i}$.

Note that $\tilde{f}(a_1, a_2, ..., a_n)$ is well-defined as for any $a_{\leq n}$, the set $\{j_1, j_2(a_{j_1}), ..., j_n(a_{j_{<n}})\}$ forms a permutation of the set $\{1, 2, ..., n\}$.

Proof. The proof follows from Theorem 3.6.1 and the construction

$$\mathsf{P}'(e) = \mathsf{P}(e) \tag{3.225}$$

$$\mathsf{P}'(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n}) = \prod_i \mathsf{P}'(a_{j_i}b_{j_i}\,|\,a_{j_{<i}}ex_{j_i}y_{j_i}) \tag{3.226}$$

$$\mathsf{P}'(a_{j_i}b_{j_i}\,|\,a_{j_{<i}}ex_{j_i}y_{j_i}) = \mathsf{P}(a'_ib'_i\,|\,a'_{<i}ex'_iy'_i)\,, \tag{3.227}$$

again with $a'_i = a_{j_i}$. Note that for the construction (3.225) to (3.227) the dynamic TONS conditions (2.4) for the dynamic orders $\{j_i\}$ and $\{k_i\}$ follow for the distribution $\mathsf{P}'(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ from Theorem 3.6.1. Furthermore, note that

$$\mathsf{P}'(a_{j_{\leq i}}b_{j_{\leq i}}\,|\,ex_{j_{\leq i}}y_{j_{\leq i}}) = \mathsf{P}(a'_{\leq i}b'_{\leq i}\,|\,ex'_{\leq i}y'_{\leq i})\,, \tag{3.228}$$

with $a_{j_i} = a'_i$ and, similarly, for the other entries. As $\mathsf{P}(a_{\leq i}b_{\leq i}\,|\,ex_{\leq i}y_{\leq i})$ has a well-defined marginal on the systems $A_{\leq i_A}B_{k_{\leq i_B}}$ for any $i_A, i_B \in [n]$ and arbitrary dynamic order $\{k_i\}$, so does $\mathsf{P}'(a_{\leq i}b_{\leq i}\,|\,ex_{\leq i}y_{\leq i})$ on the systems $A_{\leq j_{i_A}}B_{k_{\leq i_B}}$. The proof of (3.223) follows from the construction (3.225) to (3.227) and the definition of $\tilde{f}$ (3.224)

$$\begin{aligned}\mathsf{P}'(f(a_{\leq n}) = e) &= \mathsf{P}'(e)\,\mathsf{P}'(f(a_{\leq n}) = e\,|\,e)\\ &= \mathsf{P}(e)\,\mathsf{P}(f(a'_{\leq n}) = e\,|\,e)\\ &= \mathsf{P}(e)\,\mathsf{P}(\tilde{f}(a_{\leq n}) = e\,|\,e)\\ &= \mathsf{P}(\tilde{f}(a_{\leq n}) = e)\,,\end{aligned} \tag{3.229}$$

which completes the proof. □

3.7 An analogous construction of ABNS attacks from a classical game

Although it was already known that privacy amplification against an ABNS adversary is impossible we present, for the sake of comparison, a construction of ABNS attacks from a classical game which is analogous to the construction of TONS attacks in Section 3.4.1. Again, we construct classical distributions $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}n)$ that induce ABNS attacks on product distributions of $\mathsf{U}(a_ib_i \,|\, x_iy_i)$ for $i \in \mathcal{S}$ and perfect $\mathsf{PR}(a_ib_i \,|\, x_iy_i)$-boxes for $i \in \overline{\mathcal{S}}$. Then, a weighted sum $\mathsf{Q}_{\varepsilon}(a_{\leq n}e)$ of the distributions $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}n)$ induces an ABNS-attack on $\mathsf{PR}^n_{\varepsilon}(a_{\leq_n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$.

As the TONS conditions are stronger no-signalling conditions than the ABNS conditions (compare Definitions 4 and 5) the conditions on the distributions $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}n)$ are also weaker (see Figure 3.7). In Section 3.7.3 we show that this construction of ABNS attacks retrieves the impossibility result for ABNS privacy amplification of Hänggi, Renner and Wolf. We argue that this is a strong indication that, since the analogous construction for ABNS attacks is optimal (up to a constant factor), our construction of TONS attacks should be also optimal. Consequently, *if* TONS privacy amplification is impossible, we conjecture that our construction of TONS attacks from Section 3.4.1 is general enough to retrieve this result.

3.7.1 From a classical game over a weighted set of distributions to ABNS attacks

Consider a distribution $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)$ that has to satisfy two properties, the marginal distribution on $a_{\leq n}$ must be uniform and only the bits $a_{\mathcal{S}}$ can be influenced (*i.e.*, made non-uniform) when conditioned on Eve's bit e. On the other side, the marginal distribution on the bits $a_{\overline{\mathcal{S}}}$ must remain uniform even when conditioned on e. More formally, we have:

Definition 18 ($\mathcal{S}$-influenceable distribution)**.** For the subset $\mathcal{S} \in \mathcal{P}([n])$ we define an *$\mathcal{S}$-influenceable distribution* $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e) : \{0,1\}^{n+1} \to \mathbb{R}$ as a probability distribution that satisfies uniformity on $a_{\leq n}$

$$\sum_e \mathsf{Q}_{\mathcal{S}}(a_{\leq n}e) = 2^{-n} \quad \forall a_{\leq n} \,, \tag{3.230}$$

and

$$\mathsf{Q}_{\mathcal{S}}(a_{\overline{\mathcal{S}}} \,|\, e) = 2^{-|\overline{\mathcal{S}}|} \quad \forall a_{\overline{\mathcal{S}}}, e \,. \tag{3.231}$$

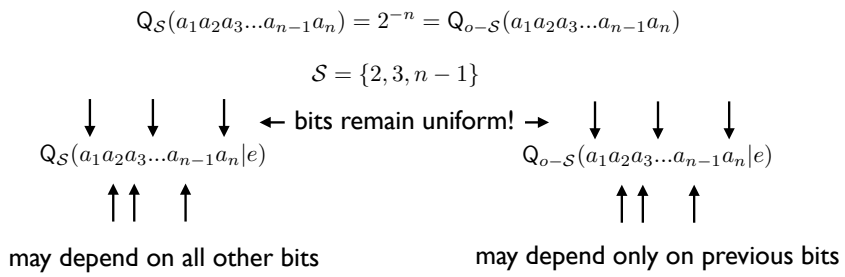

Figure 3.7. Comparison of $\mathcal{S}$-influenceable distributions $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)$, which induce ABNS attacks, with ordered $\mathcal{S}$-influenceable distributions $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$, which induce TONS attacks, for the example $\mathcal{S} = \{2, 3, n-1\}$. For $\mathcal{S}$-influenceable distributions $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)$, each of the bits a_i, $i \in \mathcal{S}$, can be influenced depending on all other bits $a_{\bar{i}}$. However, for ordered $\mathcal{S}$-influenceable distributions $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$, each of the bits a_i, $i \in \mathcal{S}$, can be influenced depending only on the bits $a_{<i}$.

Analogue to Definition 12 we define the weighted sum over $\mathcal{S}$-influanceable distributions as as an $(\varepsilon, \mathcal{S})$-*divisible distribution*.

Definition 19 ($(\varepsilon, \mathcal{S})$-divisible distribution)**.** Define an $(\varepsilon, \mathcal{S})$-*divisible distribution* as a distribution $\mathsf{Q}_{\varepsilon}(a_{\leq n}e) : \{0,1\}^{n+1} \to \mathbb{R}$ if there exists as set $\{\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)\}$ containing each $\mathcal{S} \in \mathcal{P}([n])$, such that

$$\mathsf{Q}_{\varepsilon}(a_{\leq n}e) = \sum_{\mathcal{S}\in\mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)\,, \tag{3.232}$$

with weights

$$\omega(\mathcal{S}, n, \varepsilon) := (1-2\varepsilon)^{n-|\mathcal{S}|}(2\varepsilon)^{|\mathcal{S}|}\,. \tag{3.233}$$

Now let us consider the following privacy amplification game between Alice and Eve. First let Alice choose a function $f(a_{\leq n})$ and hand it to Eve. Then Eve constructs an $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{\varepsilon}(a_{\leq n}e)$. Then draw the string $a_{\leq n}e$ according to the distribution $\mathsf{Q}_{\varepsilon}(a_{\leq n}e)$. Eve wins if $f(a_{\leq n}) = e$, Alice wins otherwise. In Section 3.7.3 we show that Eve can always win the game with probability at least $1/2+\varepsilon/2$, a considerable improvement to the best-known lower bound $1/2 + \varepsilon/n$ for the analogous game for ordered $(\varepsilon, \mathcal{S})$-divisible distributions $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$.

Theorem 3.7.1. *Any $\mathcal{S}$-influenceable distribution $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)$ can be extended to an ABNS-attack $\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on the systems $A_{\leq n}B_{\leq n}$ with distribution*

$$\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}) := \prod_{i\in\mathcal{S}} \mathsf{U}(a_ib_i \,|\, x_iy_i) \prod_{i\in\overline{\mathcal{S}}} \mathsf{PR}(a_ib_i \,|\, x_iy_i)\,. \tag{3.234}$$

Proof. We construct $\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ as:

$$\mathsf{P}_{\mathcal{S}}(e) = \mathsf{Q}_{\mathcal{S}}(e)\,, \tag{3.235}$$

$$\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}e) = \prod_{i\in\overline{\mathcal{S}}} \mathsf{PR}(a_ib_i \,|\, x_iy_i)\, \mathsf{P}_{\mathcal{S}}(a_{\mathcal{S}}b_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}b_{\overline{\mathcal{S}}}x_{\leq n}y_{\leq n}e)\,, \tag{3.236}$$

$$\mathsf{P}_{\mathcal{S}}(a_{\mathcal{S}}b_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}b_{\overline{\mathcal{S}}}x_{\leq n}y_{\leq n}e) = \mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}e) \prod_{i\in\mathcal{S}} \mathsf{U}(b_i \,|\, y_i)\,. \tag{3.237}$$

We have to show that (3.235)-(3.237) implies that $\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$

1. satisfies the ABNS-conditions,

2. has the correct marginal on systems $A_{\leq n}B_{\leq n}$:

$$\sum_e \mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \prod_{i\in\mathcal{S}} \mathsf{U}(a_ib_i \,|\, x_iy_i) \prod_{i\in\overline{\mathcal{S}}} \mathsf{PR}(a_ib_i \,|\, x_iy_i)\,, \tag{3.238}$$

3. and has the correct marginal on systems $A_{\leq n}E$ (which must be independent also of $x_{\leq n}$):

$$\sum_{b_{\leq n}} \mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)\,. \tag{3.239}$$

We start with the ABNS conditions, see Definition 4. We need to show that the systems $A_{\leq n}$ and $B_{\leq n}$ have well-defined marginal distributions $\mathsf{P}_{\mathcal{S}}(a_{\leq n} \,|\, ex_{\leq n})$ and $\mathsf{P}_{\mathcal{S}}(b_{\leq n} \,|\, ey_{\leq n})$, respectively: From (3.237), we obtain

$$\begin{aligned}\sum_{b_{\mathcal{S}}} \mathsf{P}_{\mathcal{S}}(a_{\mathcal{S}}b_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}b_{\overline{\mathcal{S}}}x_{\leq n}y_{\leq n}e) &= \sum_{b_{\mathcal{S}}} \prod_{i\in\mathcal{S}} \mathsf{U}(b_i \,|\, y_i)\, \mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}e) \\ &= \mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}e)\,,\end{aligned} \tag{3.240}$$

which, together with (3.236), implies no-signalling from Bob to Alice

$$\begin{aligned}\sum_{b_{\leq n}} \mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}e) &= \sum_{b_{\overline{\mathcal{S}}}} \prod_{i\in\overline{\mathcal{S}}} \mathsf{PR}(a_ib_i \,|\, x_iy_i)\, \mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}e) \\ &= 2^{-n+|\mathcal{S}|}\mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}e) \\ &= \mathsf{Q}_{\mathcal{S}}(a_{\leq n} \,|\, e) \\ &=: \mathsf{P}_{\mathcal{S}}(a_{\leq n} \,|\, ex_{\leq n})\,.\end{aligned} \tag{3.241}$$

Note that (3.241) and (3.235) already impliy the correct marginal $\mathsf{Q}_(a_{\leq n}e)$ on the systems $A_{\leq n}E$:

$$\begin{aligned}\sum_{b_{\leq n}} \mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) &= \sum_{b_{\leq n}} \mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}e)\,\mathsf{P}(e) \\ &= \mathsf{Q}_{\mathcal{S}}(a_{\leq n} \,|\, e)\,\mathsf{Q}_{\mathcal{S}}(e) \\ &= \mathsf{Q}_{\mathcal{S}}(a_{\leq n}e) \qquad \forall x_{\leq n}, y_{\leq n}\,. \end{aligned} \tag{3.242}$$

No-signalling from Alice to Bob follows directly by construction (3.236) and (3.237):

$$\begin{aligned}\sum_{a_{\leq n}} \mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}e) &= \sum_{a_{\leq n}} \prod_{i\in\overline{\mathcal{S}}} \mathsf{PR}(a_ib_i \,|\, x_iy_i)\,\mathsf{P}_{\mathcal{S}}(a_{\mathcal{S}}b_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}b_{\overline{\mathcal{S}}}x_{\leq n}y_{\leq n}e) \\ &= \sum_{a_{\overline{\mathcal{S}}}} \prod_{i\in\overline{\mathcal{S}}} \mathsf{PR}(a_ib_i \,|\, x_iy_i) \sum_{a_{\mathcal{S}}} \mathsf{P}_{\mathcal{S}}(a_{\mathcal{S}}b_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}b_{\overline{\mathcal{S}}}x_{\leq n}y_{\leq n}e) \\ &= 2^{-n+s} \sum_{a_{\mathcal{S}}} \mathsf{U}(b_{\mathcal{S}} \,|\, y_{\mathcal{S}})\,\mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}e) \\ &= 2^{-n}\,. \end{aligned} \tag{3.243}$$

To finish the proof, we need to show (3.238): First note that (3.230) and (3.231) together imply

$$\begin{aligned}\sum_e \mathsf{Q}_{\mathcal{S}}(a_{\leq n}e) &= \sum_e \mathsf{Q}_{\mathcal{S}}(e)\,\mathsf{Q}_{\mathcal{S}}(a_{\overline{\mathcal{S}}} \,|\, e)\,\mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}e) \\ &= 2^{-n+s} \sum_e \mathsf{Q}_{\mathcal{S}}(e)\,\mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}e) \\ &= 2^{-n} \\ \Rightarrow \quad \sum_e \mathsf{Q}_{\mathcal{S}}(e)\,\mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}}e) &= 2^{-s}\,. \end{aligned} \tag{3.244}$$

Together with the construction rules (3.235) to (3.237), we obtain

$$\begin{aligned}
\sum_e \mathsf{P}_{\mathcal{S}}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n}) &= \sum_e \mathsf{P}_{\mathcal{S}}(e)\, \mathsf{P}_{\mathcal{S}}(a_{\leq n} b_{\leq n} \,|\, e x_{\leq n} y_{\leq n}) \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i) \sum_e \mathsf{Q}_{\mathcal{S}}(e)\, \mathsf{P}_{\mathcal{S}}(a_{\mathcal{S}} b_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}} b_{\overline{\mathcal{S}}} x_{\leq n} y_{\leq n} e) \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i)\, \mathsf{U}(b_{\mathcal{S}} \,|\, y_{\mathcal{S}}) \sum_e \mathsf{Q}_{\mathcal{S}}(e)\, \mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}} \,|\, a_{\overline{\mathcal{S}}} e) \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i)\, \mathsf{U}(b_{\mathcal{S}} \,|\, y_{\mathcal{S}})\, 2^{-s} \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i)\, \mathsf{U}(b_{\mathcal{S}} \,|\, y_{\mathcal{S}})\, \mathsf{U}(a_{\mathcal{S}} \,|\, x_{\mathcal{S}}) \\
&= \prod_{i \in \overline{\mathcal{S}}} \mathsf{PR}(a_i b_i \,|\, x_i y_i)\, \mathsf{U}(a_{\mathcal{S}} b_{\mathcal{S}} \,|\, x_{\mathcal{S}} y_{\mathcal{S}}) \,. \qquad (3.245)
\end{aligned}$$

□

Theorem 3.7.2. *For any $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_\varepsilon(a_{\leq n} e)$, there exists an ABNS-attack $\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n})$ on $\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n} b_{\leq n} \,|\, x_{\leq n} y_{\leq n})$ such that*

$$\sum_{b_{\leq n}} \mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n}) = \mathsf{Q}_\varepsilon(a_{\leq n} e) \quad \forall x_{\leq n}, y_{\leq n} \,. \qquad (3.246)$$

Proof. Theorem 3.7.2 follows from Definition 19, Theorem 3.7.1 and the fact that

$$\mathsf{PR}_\varepsilon^{\otimes n}(a_{\leq n} b_{\leq n} \,|\, x_{\leq n} y_{\leq n}) = \sum_{\mathcal{S} \in \mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{P}_{\mathcal{S}}(a_{\leq n} b_{\leq n} \,|\, x_{\leq n} y_{\leq n}) \,. \qquad (3.247)$$

□

3.7.2 Attacking linear functions

In this section we show that, in the limit of large n, an ABNS adversary can gain complete knowledge if $f(a_{\leq n})$ is the parity of the string $a_{\leq n}$, no matter how small the noise parameter ε (and consequently the secrecy of the individual bits a_i) is.

Theorem 3.7.3. *Linear functions reduce secrecy. For any $\mathcal{T} \in \mathcal{P}([n])$ there exists an $(\varepsilon - \mathcal{S})$-divisible distribution $\mathsf{Q}_\varepsilon(a_{\leq n} e)$ such that*

$$\mathsf{Q}_\varepsilon\Big(\bigoplus_{i \in \mathcal{T}} a_i = e\Big) = 1 - (1 - 2\varepsilon)^t \,, \qquad (3.248)$$

where $t = |\mathcal{T}|$.

Proof. In the proof we use that in $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)$ a bit a_i may be influenced as a function of all other bits $a_{\bar{i}}$ (when conditioned on e). Thus the adversary simply chooses one bit a_i and influences it towards or against the parity of the other bits $a_{\bar{i}}$. Formally the construction of $\{\mathsf{Q}_{\mathcal{S}}\}$ is as follows: If $\mathcal{S} \cap \mathcal{T} = \emptyset$ then let $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)$ be the trivial distribution

$$\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e) = 2^{-(n+1)} \,. \tag{3.249}$$

For all $\mathcal{S}$ such that $\mathcal{S} \cap \mathcal{T} \neq \emptyset$ choose an $i^* \in \mathcal{S} \cap \mathcal{T}$. Then let $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)$ be defined as

$$\mathsf{Q}_{\mathcal{S}}(a_{\bar{i}}e) = 2^{-n} \quad \text{and} \tag{3.250}$$

$$\mathsf{Q}_{\mathcal{S}}(a_i \,|\, a_{\bar{i}}e) = \delta(a_i, e \oplus \bigoplus_{j \in \mathcal{T}/i} a_j) \quad \text{for } i = i^* \,. \tag{3.251}$$

Note that (3.250) is equivalent to setting $\mathsf{Q}_{\mathcal{S}}(e) = 1/2$ and $\mathsf{Q}_{\mathcal{S}}(a_{\bar{i}} \,|\, e) = 2^{-(n-1)}$. Therefore,

$$\mathsf{Q}_{\mathcal{S}}(a_i a_{\bar{i}} \,|\, e) = 2^{-(n-1)}\, \delta(a_i, e \oplus \bigoplus_{j \in \mathcal{T}/i} a_j) \,, \tag{3.252}$$

which yields (3.230) where averaged with $\mathsf{Q}_{\mathcal{S}}(e) = 1/2$. Furthermore, $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)$ satisfies (3.231) since $\overline{\mathcal{S}} \in \bar{i}$. We complete the proof by showing (3.248). From (3.249)-(3.251) it follows that

$$\mathsf{Q}_{\mathcal{S}}(\bigoplus_{i \in \mathcal{T}} a_i = e) = \begin{cases} \frac{1}{2} & \mathcal{S} \cap \mathcal{T} = \emptyset \\ 1 & \text{otherwise} \,. \end{cases} \tag{3.253}$$

Furthermore, notice that by the definition of $\omega(\mathcal{S}, n, \varepsilon)$ we have

$$\sum_{\mathcal{S} \in \mathcal{P}([n]) : i \in \overline{\mathcal{S}}} \omega(\mathcal{S}, n, \varepsilon) = (1 - 2\varepsilon) \,, \tag{3.254}$$

and, for, $\mathcal{T} = \{i_1, i_2, ..., i_t\}$

$$\begin{aligned} \mathsf{Q}_{\varepsilon}(\bigoplus_{i \in \mathcal{T}} a_i = e) &= \sum_{\mathcal{S} \in \mathcal{P}([n])} \omega(\mathcal{S}, n, \varepsilon)\, \mathsf{Q}_{\mathcal{S}}(\bigoplus_{i \in \mathcal{T}} a_i = e) \\ &= \sum_{\mathcal{S} : \bigcap_j^t i_j \in \overline{\mathcal{S}}} \omega(\mathcal{S}, n, \varepsilon) \cdot \frac{1}{2} + \sum_{\mathcal{S} : \exists j : i_j \in \mathcal{S}} \omega(\mathcal{S}, n, \varepsilon) \cdot 1 \\ &= (1 - 2\varepsilon)^t \cdot \frac{1}{2} + (1 - (1 - 2\varepsilon)^t) \cdot 1 \\ &= 1 - \frac{1}{2}(1 - 2\varepsilon)^t \end{aligned} \tag{3.255}$$

□

3.7.3 Impossibility of ABNS privacy amplification from $(\varepsilon, \mathcal{S})$-divisible distributions $\mathsf{Q}_\varepsilon(a_{\leq n}e)$

Theorem 3.7.4. *For any function $f(a_{\leq n})$ there exists an $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_\varepsilon(a_{\leq n}e)$ such that*

$$\mathsf{Q}_\varepsilon(f(a_{\leq n}) = e) \geq \frac{1}{2} + \frac{\varepsilon}{2} \,. \tag{3.256}$$

The proof of Theorem 3.7.4 follows below Theorem 3.7.9.

Theorem 3.7.5. *For any function $f(a_{\leq n})$ and any $\mathcal{S} \in \mathcal{P}([n])$ there exists a $\mathsf{Q}_\mathcal{S}(a_{\leq n}e)$ such that*

$$\mathsf{Q}_\mathcal{S}(f(a_{\leq n}) = e) = \frac{1}{2} + 1 - 2^{-n} \sum_{a_{\overline{\mathcal{S}}}} \max\left(|\mathcal{A}_0(a_{\overline{\mathcal{S}}})|, |\mathcal{A}_1(a_{\overline{\mathcal{S}}})|\right) \tag{3.257}$$

Proof. Fix a function $f(a_{\leq n})$ and let us define the sets $\mathcal{A}_e(a_{\overline{\mathcal{S}}})$ as

$$\mathcal{A}_e(a_{\overline{\mathcal{S}}}) := \{a_\mathcal{S} : f(a_\mathcal{S}, a_{\overline{\mathcal{S}}}) = e\} \,, \tag{3.258}$$

and the number $d(a_{\overline{\mathcal{S}}})$ as

$$d(a_{\overline{\mathcal{S}}}) := \min(|\mathcal{A}_0(a_{\overline{\mathcal{S}}})|, |\mathcal{A}_1(a_{\overline{\mathcal{S}}})|) \,. \tag{3.259}$$

We choose $d(a_{\overline{\mathcal{S}}})$ strings $a_\mathcal{S} \in \mathcal{A}_e(a_{\overline{\mathcal{S}}})$ arbitrarily (this can be the whole set $\mathcal{A}_e(a_{\overline{\mathcal{S}}})$) to form the set $\mathcal{A}'_e(a_{\overline{\mathcal{S}}})$. We construct the $\mathsf{Q}_\mathcal{S}(a_\mathcal{S}a_{\overline{\mathcal{S}}}e)$ in the following way.

$$\mathsf{Q}_\mathcal{S}(e) = \frac{1}{2} \quad \text{and,} \tag{3.260}$$

$$\mathsf{Q}_\mathcal{S}(a_{\overline{\mathcal{S}}} \,|\, e) = 2^{-|\overline{\mathcal{S}}|} \tag{3.261}$$

$$\mathsf{Q}_\mathcal{S}(a_\mathcal{S} \,|\, a_{\overline{\mathcal{S}}}e) = \begin{cases} 2^{-|\mathcal{S}|+1} & \text{if } a_\mathcal{S} \in \mathcal{A}'_e(a_{\overline{\mathcal{S}}}) \\ 2^{-|\mathcal{S}|} & \text{if } a_\mathcal{S} \in \mathcal{A}_e(a_{\overline{\mathcal{S}}})/\mathcal{A}'_e(a_{\overline{\mathcal{S}}}) \\ 2^{-|\mathcal{S}|} & \text{if } a_\mathcal{S} \in \mathcal{A}_{\bar{e}}(a_{\overline{\mathcal{S}}})/\mathcal{A}'_{\bar{e}}(a_{\overline{\mathcal{S}}}) \\ 0 & \text{if } a_\mathcal{S} \in \mathcal{A}'_{\bar{e}}(a_{\overline{\mathcal{S}}}) \,. \end{cases} \tag{3.262}$$

Note that equations (3.260)-(3.262) are directly implying the defining conditions of an $\mathcal{S}$-influenceable distribution (3.230) and (3.231). We analyse the efficiency of

this attack:

$$\begin{aligned}
\mathsf{Q}_{\mathcal{S}}(f(a_{\leq n}=e)) &= \sum_e \mathsf{Q}_{\mathcal{S}}(e)\,\mathsf{Q}_{\mathcal{S}}(f(a_{\leq n}=e)\,|\,e) \\
&= \sum_e \mathsf{Q}_{\mathcal{S}}(e) \sum_{a_{\leq n}:f(a_{\leq n})=e} \mathsf{Q}_{\mathcal{S}}(a_{\leq n}\,|\,e) \\
\overset{(3.260),(3.261)}{\Rightarrow} \quad &= \frac{1}{2}\sum_e \sum_{a_{\overline{\mathcal{S}}}} 2^{-|\overline{\mathcal{S}}|} \sum_{a_{\overline{\mathcal{S}}}:a_{\overline{\mathcal{S}}}\in\mathcal{A}_e} \mathsf{Q}_{\mathcal{S}}(a_{\mathcal{S}}a_{\overline{\mathcal{S}}}\,|\,e) \\
\overset{(3.262)}{\Rightarrow} \quad &= \frac{1}{2}\sum_e \sum_{a_{\overline{\mathcal{S}}}} 2^{-|\overline{\mathcal{S}}|} \sum_{a_{\overline{\mathcal{S}}}:a_{\overline{\mathcal{S}}}\in\mathcal{A}_e} 2^{-|\mathcal{S}|}\left(|\mathcal{A}_e(a_{\overline{\mathcal{S}}})| + |\mathcal{A}'_e(a_{\overline{\mathcal{S}}})|\right) \\
&= \frac{1}{2}\sum_e \sum_{a_{\overline{\mathcal{S}}}} 2^{-|\overline{\mathcal{S}}|} \sum_{a_{\overline{\mathcal{S}}}:a_{\overline{\mathcal{S}}}\in\mathcal{A}_e} 2^{-|\mathcal{S}|}\left(|\mathcal{A}_e(a_{\overline{\mathcal{S}}})| + d(a_{\overline{\mathcal{S}}})\right) \\
\overset{(3.259)}{\Rightarrow} \quad &= 2^{-n-1}\sum_e |\mathcal{A}_e| + 2^{-|\overline{\mathcal{S}}|}\sum_{a_{\overline{\mathcal{S}}}} 2^{-|\mathcal{S}|}\min\left[|\mathcal{A}_0(a_{\overline{\mathcal{S}}})|, |\mathcal{A}_1(a_{\overline{\mathcal{S}}})|\right] \\
&= \frac{1}{2} + 1 - 2^{-|\overline{\mathcal{S}}|}\sum_{a_{\overline{\mathcal{S}}}} 2^{-|\mathcal{S}|}\max\left[|\mathcal{A}_0(a_{\overline{\mathcal{S}}})|, |\mathcal{A}_1(a_{\overline{\mathcal{S}}})|\right] \\
&= \frac{1}{2} + 1 - 2^{-n}\sum_{a_{\overline{\mathcal{S}}}} \max\left[|\mathcal{A}_0(a_{\overline{\mathcal{S}}})|, |\mathcal{A}_1(a_{\overline{\mathcal{S}}})|\right], \qquad (3.263)
\end{aligned}$$

where we use the facts that $\sum_{a_{\overline{\mathcal{S}}}} \mathcal{A}_e(a_{\overline{\mathcal{S}}}) = \mathcal{A}_e$ and that $\sum_e |\mathcal{A}_e| = 2^n$. □

Intuitively, the sum on the second to last line is the winning probability of someone guessing $f(a_{\leq n})$ after he has received the (uniformly distributed string) $a_{\overline{\mathcal{S}}}$. We will formalise this intuition now.

Definition 20 (Maximum-likelihood function)**.** The *maximum-likelihood function* $g(b)$ of $f(a)$ is defined as

$$g(b) = \max[c : \mathsf{P}(f(a) = c\,|\,b)]\,. \qquad (3.264)$$

Definition 21 (Correlation)**.** The *correlation* $\mathrm{Cor}_{\mathsf{P}}(f,g)$ between two binary random variables $f(a)$ and $g(b)$ with respect to the joint distribution $\mathsf{P}(a,b)$ is defined as

$$\begin{aligned}
\mathrm{Cor}_{\mathsf{P}}(f,g) &:= \mathsf{P}(f(a) = g(b)) - \mathsf{P}(f(a) \neq g(b)) \\
&= 2\mathsf{P}(f(a) = g(b)) - 1\,. \qquad (3.265)
\end{aligned}$$

Consider two n-bit strings $A_{\leq n}B_{\leq n}$ where $|\overline{\mathcal{S}}|$ bits are perfectly correlated and the

other $|\mathcal{S}|$ are completely uncorrelated

$$\mathsf{P}_{\mathcal{S}}(a_{\mathcal{S}}b_{\mathcal{S}}) = 2^{-2|\mathcal{S}|} \tag{3.266}$$

$$\mathsf{P}_{\mathcal{S}}(a_{\overline{\mathcal{S}}} \,|\, a_{\mathcal{S}}b_{\mathcal{S}}) = 2^{-|\overline{\mathcal{S}}|} \tag{3.267}$$

$$\mathsf{P}_{\mathcal{S}}(b_{\overline{\mathcal{S}}} \,|\, a_{\leq n}b_{\mathcal{S}}) = \delta(b_{\mathcal{S}}, a_{\mathcal{S}}) \,. \tag{3.268}$$

Note that $\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n})$ [2] is symmetric with respect to the exchange of $a_{\leq n}$ and $b_{\leq n}$. Note alsothat with this definition the sum in the last line of (3.263) is exactly the probability that the maximum-likelihood function $g(a_{\leq n})$ guesses $f(a_{\leq n})$ correctly, *i.e.*,

$$\begin{aligned}\mathsf{P}_{\mathcal{S}}(f(a_{\leq n}) = g(b_{\leq n})) &= \sum_{b_{\leq n}} \mathsf{P}_{\mathcal{S}}(b_{\leq n}) \max\left[c : \mathsf{P}(f(a_{\leq n}) = c \,|\, b_{\leq n})\right] \\ &= 2^{-n} \sum_{a_{\overline{\mathcal{S}}}} \max\left[|\mathcal{A}_0(a_{\overline{\mathcal{S}}})|, |\mathcal{A}_1(a_{\overline{\mathcal{S}}})|\right] . \end{aligned} \tag{3.269}$$

This implies that the higher the correlation between the function $f(a_{\leq n})$ and its best guess $g(b_{\leq n})$ with respect to the distribution $\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n})$ is, the lower is the probability $\mathsf{Q}_{\mathcal{S}}(f(a_{\leq n}) = e)$.

Lemma 3.7.6. *Define the distribution* $\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n})$ *as in* (3.266)-(3.268) *and the distribution* $\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)$ *according to* (3.260)-(3.262) *or a function* $f(a_{\leq n})$*. Then*

$$\mathsf{Q}_{\mathcal{S}}(f(a_{\leq n}) = e) = 1 - \frac{1}{2} Cor_{\mathsf{P}_{\mathcal{S}}}(f, g) \,, \tag{3.270}$$

where we denote $g_{\mathcal{S}}(b_{\leq n})$ *as the maximum-likelihood function of* $f(a_{\leq n})$ *with respect to* $\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n})$*. Note that* $g_{\mathcal{S}}(b_{\leq n})$ *only depends on* $b_{\overline{\mathcal{S}}}$*.*

Proof. Lemma 3.7.6 follows directly from (3.263), (3.269), and (3.270). □

The distribution $\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n})$ defined in (3.266)-(3.268) can be interpreted as a *partial erasure channel*; the bits $b_{\overline{\mathcal{S}}}$ are perfectly transmitted from Alice to Bob, but the bits $b_{\mathcal{S}}$ are completely erased (or vice versa). This insight allows us to compute $\mathsf{Q}_{\varepsilon}(a_{\leq n}e)$ via a definition of a probabilistic erasure channel, since

$$\mathsf{Q}_{\varepsilon}(a_{\leq n}e) = \sum_{\mathcal{S}} (1 - 2\varepsilon)^{n-|\mathcal{S}|} (2\varepsilon)^{\mathcal{S}} \, \mathsf{Q}_{\mathcal{S}}(a_{\leq n}e) \,. \tag{3.271}$$

[2] We denote the distribution defined in (3.266)-(3.268) with a subscript $\mathcal{S}$ as it is exactly the same distribution as the conditional distribution in (3.234) for the inputs $x_{\leq n} = y_{\leq n} = 0^n$.

Definition 22 (Probability p-erasure channel)**.** We call the probability distribution $\gamma_p(ab) : \{0,1\} \times \{0,1,\perp\} \to [0,1]$ a *probability p-erasure channel* with uniform input a if $\gamma_p(a) = 1/2$ and

$$\gamma_p(b\,|\,a) = \begin{cases} p & \text{if } b = \perp\,, \\ \delta(b,a)\cdot(1-p) & \text{otherwise}\ . \end{cases} \tag{3.272}$$

Theorem 3.7.7. *Let* $\mathsf{P}_{2\varepsilon}(a_{\leq n}b_{\leq n}) = \gamma^n_{2\varepsilon}(a_{\leq n}b_{\leq n})$*. Let* $g(b_{\leq n}) : \{0,1,\perp\}$ *be the maximum-likelihood guess* $g(b_{\leq n})$ *for the function* $f(a_{\leq n})$ *with respect to the* $\mathsf{P}_{2\varepsilon}(a_{\leq n}b_{\leq n})$*. Construct the set of* $\mathcal{S}$*-influenceable distributions* $\{\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)\}$ *as in Theorem 3.7.5. Then for the derived distribution* $\mathsf{Q}_{\varepsilon-sv}(f(a_{\leq n}) = e)$ *it holds that*

$$\mathsf{Q}_{\varepsilon}(f(a_{\leq n}) = e) = 1 - \frac{1}{2}Cor_{\mathsf{P}_{2\varepsilon}}(f,g)\ . \tag{3.273}$$

Proof. Note that if $b_i = \perp$, which happens with probability 2ε, then $\mathsf{P}_{2\varepsilon}(a_i\,|\,b_i) = \frac{1}{2}$. On the other hand, with probability $(1-2\varepsilon)$ we have $b_i \neq \perp$ and $\mathsf{P}_{2\varepsilon}(a_i\,|\,b_i) = \delta(a_i,b_i)$. Therefore,

$$\mathsf{P}_{2\varepsilon}(b_{\overline{\mathcal{S}}} = a_{\overline{\mathcal{S}}} \cap b_{\mathcal{S}} = \perp_{\mathcal{S}}) = (1-2\varepsilon)^{|\overline{\mathcal{S}}|}\,(2\varepsilon)^{|\mathcal{S}|}\ , \tag{3.274}$$

and the conditional probability probability of the string $a_{\leq n}$ is the same as for the distribution $\mathsf{P}_{\mathcal{S}}$, *i.e.*,

$$\mathsf{P}_{2\varepsilon}(a_{\overline{\mathcal{S}}}a_{\mathcal{S}}\,|\,b_{\leq n} = b_{\overline{\mathcal{S}}}\,\perp_{\mathcal{S}}) = \mathsf{P}_{\mathcal{S}}(a_{\overline{\mathcal{S}}}a_{\mathcal{S}}\,|\,b_{\leq n} = b_{\overline{\mathcal{S}}}b_{\mathcal{S}}). \tag{3.275}$$

Thus, also the maximum-likelihood function $g(b_{\leq n})$ of $f(a_{\leq n})$ with respect to $\mathsf{P}_{2\varepsilon}$ is equivalent to $g_{\mathcal{S}}(b_{\overline{\mathcal{S}}})$ when conditioned on $b_{\leq n} = b_{\overline{\mathcal{S}}}\,\perp_{\mathcal{S}}$:

$$g(b_{\overline{\mathcal{S}}},\perp_{\mathcal{S}}) = g_{\mathcal{S}}(b_{\overline{\mathcal{S}}})\ , \tag{3.276}$$

and we can compute the probability $\mathsf{P}_{2\varepsilon}(f(a_{\leq n}) = g(b_{\leq n})$ as

$$\mathsf{P}_{2\varepsilon}(f(a_{\leq n}) = g(b_{\leq n})) = (1-2\varepsilon)^{|\overline{\mathcal{S}}|}\,(2\varepsilon)^{|\mathcal{S}|}\,\mathsf{P}_{\mathcal{S}}(f(a_{\leq n}) = g_{\mathcal{S}}(b_{\leq n}))\ . \tag{3.277}$$

Using $\mathrm{Cor}_{\mathsf{P}}(f,g) = 2\mathsf{P}(f=g) - 1$ we can complete the proof with

$$\begin{aligned} \mathsf{Q}_{\varepsilon}(f(a_{\leq n}) = e) &= \sum_{\mathcal{S}}(1-2\varepsilon)^{n-|\mathcal{S}|}\,(2\varepsilon)^{|\mathcal{S}|}\,\mathsf{Q}_{\mathcal{S}}(f(a_{\leq n}) = e) \\ &= \sum_{\mathcal{S}}(1-2\varepsilon)^{n-|\mathcal{S}|}\,(2\varepsilon)^{|\mathcal{S}|}\,\mathsf{Q}_{\mathcal{S}}(f(a_{\leq n}) = e) \\ &\overset{\text{Lemma 3.7.6}}{\Rightarrow} = 1 - \frac{1}{2}\sum_{\mathcal{S}}(1-2\varepsilon)^{n-|\mathcal{S}|}\,(2\varepsilon)^{|\mathcal{S}|}\,\mathrm{Cor}_{\mathsf{P}_{\mathcal{S}}}(f,g_{\mathcal{S}}) \\ &= 1 - \frac{1}{2}\mathrm{Cor}_{\mathsf{P}_{2\varepsilon}}(f,g)\ . \end{aligned} \tag{3.278}$$

□

In Theorem 3.7.7 we establish a connection between the success probability of an attack from Eve on $f(a_{\leq n})$ and the correlation of $f(a_{\leq n})$ to its maximum-likelihood guess $g(b_{\leq n})$ where $b_{\leq n}$ is obtained obtained from $a_{\leq n}$ via a probabilistic binary erasure channel. We argue in the remainder of Section 3.7.3 that this correlation cannot be too strong, unless both functions $f(a_{\leq n})$ and $g(b_{\leq n})$ are strongly biased. However, if $f(a_{\leq n})$ is strongly biased, then Eve can guess the output of $f(a_{\leq n})$ anyway (by a trivial attack) with high probability.

Definition 23 (Bias)**.** Define the *bias* $\beta_{\mathsf{P}}(f)$ of the function $f(a_{\leq n})$ with respect to the distribution $\mathsf{P}(a_{\leq n})$ as

$$\beta_{\mathsf{P}}(f) := \mathsf{P}(f(a_{\leq n}) = 0) - \frac{1}{2} \,. \tag{3.279}$$

Lemma 3.7.8. *The correlation $Cor_{\mathsf{P}}(f, g)$ is limited by the difference of the biases of f and g as*

$$Cor_{\mathsf{P}}(f, g) \leq 1 - 2|\beta_{\mathsf{P}}(f) - \beta_{\mathsf{P}}(g)| \,. \tag{3.280}$$

Proof.

$$\begin{aligned}
&\mathrm{Cor}_{\mathsf{P}}(f, g) \\
&= 2\mathsf{P}(f(a_{\leq n}) = g(b_{\leq n})) - 1 \\
&= 2\Big(\mathsf{P}(f(a_{\leq n}) = 0 \cap g(b_{\leq n}) = 0) + \mathsf{P}(f(a_{\leq n}) = 1 \cap g(b_{\leq n}) = 1)\Big) - 1 \\
&\leq 2\Big(\min[\mathsf{P}(f(a_{\leq n}) = 0), \mathsf{P}(g(b_{\leq n}) = 0)] + \min[\mathsf{P}(f(a_{\leq n}) = 1), \mathsf{P}(g(b_{\leq n}) = 1)]\Big) - 1 \\
&= 2\Big(\min[\frac{1}{2} + \beta_{\mathsf{P}}(f), \frac{1}{2} + \beta_{\mathsf{P}}(g)] + \min[\frac{1}{2} - \beta_{\mathsf{P}}(f), \frac{1}{2} - \beta_{\mathsf{P}}(g)]\Big) - 1 \\
&\leq 2\Big(1 - |\beta_{\mathsf{P}}(f) - \beta_{\mathsf{P}}(g)|\Big) - 1 \\
&= 1 - 2|\beta_{\mathsf{P}}(f) - \beta_{\mathsf{P}}(g)| \,.
\end{aligned} \tag{3.281}$$

□

Theorem 3.7.9. *[Yan07, O'D04] Let the strings $a_{\leq n} \in \{0, 1\}^n$ and $b_{\leq n} \in \{0, 1, \perp\}^n$ be drawn according to the joint distribution $\mathsf{P}_{2\varepsilon}(a_{\leq n} b_{\leq n}) = \gamma_{2\varepsilon}^n(a_{\leq n} b_{\leq n})$, i.e., be related via n independent 2ε-erasure channels. Then for any two functions $f(a_{\leq n})$, $g(b_{\leq n})$ their correlation is limited by*

$$Cor_{\mathsf{P}_{2\varepsilon}}(f(a_{\leq n}), g(b_{\leq n})) \leq \sqrt{1 - 2\varepsilon\,(1 - 4\beta_{\mathsf{P}_{2\varepsilon}}(f)\beta_{\mathsf{P}_{2\varepsilon}}(g))} \,. \tag{3.282}$$

We come back to the proof of Theorem 3.7.4. Assume without loss of generality that the function is biased towards 0, *i.e.*,

$$\mathsf{P}_{2\varepsilon}(f(a_{\leq n}) = 0) \geq 0 \,. \tag{3.283}$$

First, assume that $f(a_{\leq n})$ is strongly biased with respect to $\mathsf{P}_{2\varepsilon}$, *i.e.*,

$$\beta_{\mathsf{P}_{2\varepsilon}}(f) \geq \frac{\varepsilon}{2} \,. \tag{3.284}$$

As the marginal distribution $\mathsf{P}_{2\varepsilon}(a_{\leq n})$ is uniform, we can define the $\mathcal{S}$-influenceable distributions $\{\mathsf{Q}_{\mathcal{S}}(a_{\leq n}e)\}$ trivially by $\mathsf{Q}_{\mathcal{S}}(e = 0) = 1$ and $\mathsf{Q}_{\mathcal{S}}(a_{\leq n} \,|\, e = 0) = 2^{-n}$. Obviously, this induces an ε-divisible $\mathsf{Q}_{\varepsilon}(a_{\leq n}e)$ with

$$\mathsf{Q}_{\varepsilon}(f(a_{\leq n}) = e) = \mathsf{Q}_{\varepsilon}(f(a_{\leq n}) = 0) \geq \frac{1+\varepsilon}{2} \,, \tag{3.285}$$

and Theorem 3.7.4 follows. From now on we assume that $\beta_{\mathsf{P}_{2\varepsilon}}(f) < \frac{\varepsilon}{2}$. Let the bias $\beta_{\mathsf{P}_{2\varepsilon}}(g) \geq \varepsilon$, then for any function $g(b_{\leq n})$, we have by Lemma 3.7.8 that

$$\begin{aligned} \mathrm{Cor}_{\mathsf{P}_{2\varepsilon}}(f, g) &\leq 1 - 2|\beta_{\mathsf{P}_{2\varepsilon}}(f) - \beta_{\mathsf{P}_{2\varepsilon}}(g)| \\ &< 1 - \varepsilon \,. \end{aligned} \tag{3.286}$$

If we assume that the maximum-likelihood function $g(b_{\leq n})$ of $f(a_{\leq n})$ with respect to $\mathsf{P}_{2\varepsilon}$ would have such a large bias (which is unlikely) then by Theorem 3.7.7 we obtain Theorem 3.7.4

$$\begin{aligned} \mathsf{Q}_{\varepsilon-sv}(f(a_{\leq n}) = e) &= 1 - \frac{1}{2}\mathrm{Cor}_{\mathsf{P}_{2\varepsilon}}(f, g) \\ &> \frac{1}{2} + \frac{\varepsilon}{2} \,. \end{aligned} \tag{3.287}$$

Thus, we may from now on assume that for the maximum-likelihood function $g(b_{\leq n})$ of $f(a_{\leq n})$ with respect to $\mathsf{P}_{2\varepsilon}$ the bias is also small, *i.e.*, $\beta_{\mathsf{P}_{2\varepsilon}}(g) < \varepsilon$. Then we can conclude by Theorem 3.7.9 that

$$\begin{aligned} \mathrm{Cor}_{\mathsf{P}_{2\varepsilon}}(f(a_{\leq n}), g(b_{\leq n})) &\leq \sqrt{1 - 2\varepsilon\,(1 - 4\beta_{\mathsf{P}_{2\varepsilon}}(f)\beta_{\mathsf{P}_{2\varepsilon}}(g))} \\ &< \sqrt{1 - 2\varepsilon\,(1 - 4\frac{\varepsilon}{2}\varepsilon)} \\ &= \sqrt{1 - 2\varepsilon + 4\varepsilon^3} \\ &\leq \sqrt{1 - 2\varepsilon + \varepsilon^2} \\ &= 1 - \varepsilon \,, \end{aligned} \tag{3.288}$$

where we have used the fact that we are only interested in the parameters $\varepsilon < 1/4$. Theorem 3.7.4 follows via (3.287).

3.8 Application to more general systems

The no-signalling attacks we derived so far for TONS, dynamic TONS, and ABNS privacy-amplification protocols using n PR_ε boxes can be applied also to more general systems with arbitrary system B, *i.e.*, to $\mathsf{V}_\varepsilon(ab\,|\,xy)$, see (2.11). The proofs of Theorems 3.4.2 and 3.7.1 only require that the marginal $\mathsf{PR}(a\,|\,x)$ be uniform.

Theorem 3.8.1. *Any ordered $\mathcal{S}$-influenceable distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ can be extended to a TONS-attack $\mathsf{V}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ on the systems $A_{\leq n}B_{\leq n}$ with distribution*

$$\mathsf{V}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n}) := \prod_{i\in\mathcal{S}}\mathsf{U}(a_ib_i\,|\,x_iy_i)\prod_{i\in\overline{\mathcal{S}}}\mathsf{V}(a_ib_i\,|\,x_iy_i)\,. \tag{3.289}$$

Proof. By a minimal adaptation of construction (3.118)–(3.120) we obtain the generalisation of Theorem 3.4.2 from $\mathsf{PR}_\varepsilon(ab\,|\,xy)$ to $\mathsf{V}_\varepsilon(ab\,|\,xy)$; we substitute the boxes $\mathsf{PR}(a_ib_i\,|\,x_iy_i)$ with $\mathsf{V}(a_ib_i\,|\,x_iy_i)$ and $\mathsf{U}(b_i\,|\,y_i)$ with $\mathsf{V}(b_i\,|\,y_i)$ and obtain

$$\mathsf{V}_{o-\mathcal{S}}(e) = \mathsf{Q}_{o-\mathcal{S}}(e) \tag{3.290}$$

$$\mathsf{V}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n}e) = \prod_{i=1}^{n}\mathsf{V}_{o-\mathcal{S}}(a_ib_i\,|\,a_{<i}b_{<i}x_{\leq n}y_{\leq n}e) \tag{3.291}$$

$$\mathsf{V}_{o-\mathcal{S}}(a_ib_i\,|\,a_{<i}b_{<i}x_{\leq n}y_{\leq n}e) = \begin{cases}\mathsf{V}(b_i\,|\,y_i)\,\mathsf{Q}_{o-\mathcal{S}}(a_i\,|\,a_{<i}e) & i\in\mathcal{S}\\ \mathsf{V}(a_ib_i\,|\,x_iy_i) & \text{otherwise}\,.\end{cases} \tag{3.292}$$

The rest of the proof follows exactly the same steps as the proof of Theorem 3.4.2. □

By Theorem 3.8.1 we obtain also a generalisation of Theorem 3.4.3.

Theorem 3.8.2. *For any ordered $(\varepsilon,\mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$, there exists a TONS-attack $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ on $\mathsf{V}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$ such that*

$$\sum_{b_{\leq n}}\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n}) = \mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)\quad\forall x_{\leq n},y_{\leq n}\,. \tag{3.293}$$

Thus, all bounds we derived in Section 3.5 for classical privacy amplification against $\mathsf{Q}_{o-e}(a_{\leq n}e)$ also hold for TONS privacy-amplification protocols on $\mathsf{V}_\varepsilon(ab\,|\,xy)$ boxes. Similarly, by substituting $\mathsf{PR}(a_ib_i\,|\,x_iy_i)$ with $\mathsf{V}(a_ib_i\,|\,x_iy_i)$ and $\mathsf{U}(b_i\,|\,y_i)$ with $\mathsf{V}(b_i\,|\,y_i)$ in the proof of Theorem 3.6.1 in Section 3.6 we can generalise Theorem 3.8.2 to dynamic TONS attacks. Again by the same substitution into the proof of Theorem 3.7.1 we can conclude analogously on a generalisation of Theorem 3.7.2 to V_ε boxes and obtain the following statement.

Theorem 3.8.3. *For any $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_\varepsilon(a_{\leq n}e)$, there exists an ABNS-attack $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on $\mathsf{V}_\varepsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ such that*

$$\sum_{b_{\leq n}} \mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \mathsf{Q}_\varepsilon(a_{\leq n}e) \quad \forall x_{\leq n}, y_{\leq n} \,. \tag{3.294}$$

This implies by Theorem 3.7.4 that ABNS privacy amplification on $\mathsf{V}_\varepsilon(ab \,|\, xy)$ is impossible.

Chapter 4

Distillation of Nonlocality

4.1 Definition of a nonlocality distillation protocol

We consider bipartite nonlocality distillation protocols. Two players Alice and Bob share n resource boxes denoted R. Conditional probability distributions that are produced via interaction of Alice and Bob with the resource systems are marked with an accent $\hat{\mathsf{P}}$. Also the final output of a nonlocality distillation protocols, *i.e.*, the *distillate* is marked with an accent $\hat{\mathsf{P}}(ab\,|\,xy)$. The inputs and outputs of the resource boxes R are indexed with a subscript, *i.e.*, a_i, b_i, x_i, y_i, in contrast to the inputs and outputs of the protocol a, b, x, y. Without communication, the players goal is to use these boxes to simulate a single box $\hat{\mathsf{P}}(ab\,|\,xy)$ such that $\mathrm{CHSH}(\hat{\mathsf{P}}) > \mathrm{CHSH}(\mathsf{R})$. To encompass the most general case, we allow the players to use their boxes in any given (dynamic) order (see Figure 4.1), which may also depends on x and y.

Definition 24 (Bipartite distillation protocol)**.** A *bipartite nonlocality distillation protocol* using n resource boxes R is defined by the tuple of functions

$$(\{j_i^x\}, \{k_i^y\}, \{x_{j_i}^x\}, \{y_{k_i}^y\}, \{f^x\}, \{g^y\}) \quad \text{for } i \in [n]\,, \tag{4.1}$$

in the following way: Given (x, y), the outputs (a, b) of the box $\hat{\mathsf{P}}(ab\,|\,xy)$ are functions of the outputs $a_{\leq n}, b_{\leq n}$ of the n resource boxes, *i.e.*, $a = f^x(a_{\leq n})$, $b = g^y(b_{\leq n})$. At the i-th step of the protocol, the function $j_i^x = j_i^x(a_{j_{<i}})$ of the previously obtained outputs $a_{j_{<i}}$ determine which is the next box Alice uses and function $x_{j_i}^x(a_{j_{<i}})$ what to input in this box (similarly for Bob k_i^y and $y_{k_i}^y$).[1]

[1] Note that in the present discussion, Alice and Bob do not use shared randomness in addition to their non-local resources. However, the linearity of the CHSH value in the output probabilities of the distillation protocol allows to extend the results in Section 4.3 straightforwardly to distillation protocols with shared randomness.

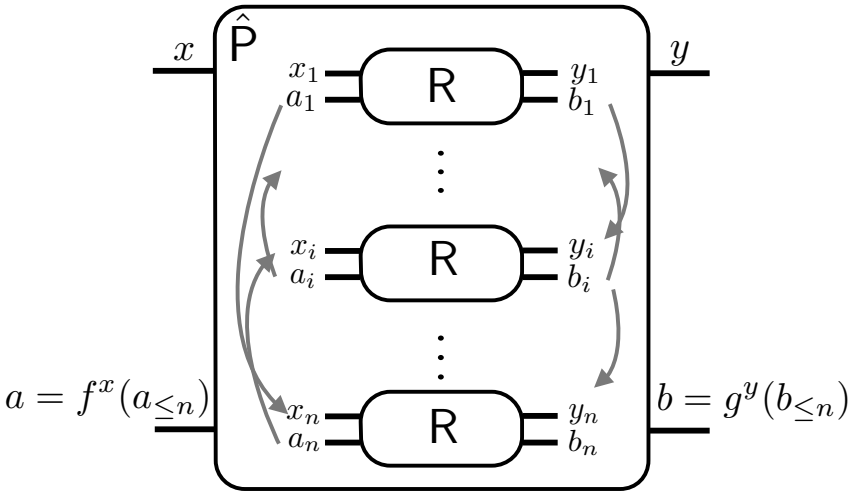

Figure 4.1. Schematic representation of a general distillation protocol. The arrows indicate the orders in which Alice and Bob use their resource boxes. The i-th system of Bob has two outgoing arrows pointing to distinct following systems, which indicates a dynamic order that depends non-trivially on b_i.

The functions $\{j_i^x\}$, $\{k_i^y\}$, $\{x_{j_i}^x\}$, and $\{y_{k_i}^y\}$, in the literature also referred to as the *wiring*, specify the interaction of the players with the resources. They determine the order of usage of and the inputs to the resources at any given step in the protocol and induce a first mapping, the output functions $f^x(a_{\leq n})$, $g^y(b_{\leq n})$ induce a second mapping:

$$\begin{aligned} &\mathsf{R}^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}) \\ &\quad\downarrow \{j_i^x\}, \{k_i^y\}, \{x_{j_i}^x\}, \{y_{k_i}^y\} \\ \hat{\mathsf{P}}(a_{\leq n}b_{\leq n} \,|\, xy) = &\mathsf{R}^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}^x(a_{\leq n})y_{\leq n}^y(b_{\leq n})) \\ &\quad\downarrow \{f^x, g^y\} \\ \hat{\mathsf{P}}(ab \,|\, xy) = &\sum_{\substack{a_{\leq n}: f^x(a_{\leq n})=a \\ b_{\leq n}: g^y(b_{\leq n})=b}} \hat{\mathsf{P}}(a_{\leq n}b_{\leq n} \,|\, xy)\,, \end{aligned} \tag{4.2}$$

where we write $x_{\leq n}^x(a_{\leq n})$ for the vector of functions $x_{j_i}^x(a_{j_{<i}})$. A commonly addressed class of distillation protocols are so-called *non-adaptive* distillation protocols, one example being the first distillation protocol presented by Forster, Winkler, and Wolf [FWW09].

Definition 25 (Non-adaptive distillation)**.** A distillation protocol is *non-adaptive* if the inputs to the resource are chosen independently from any outputs ($a_{\leq n}$, $b_{\leq n}$) of the resources, *i.e.*, if $x_i = x_i(x)$ and $y_i = y_i(y)$.

Note that for non-adaptive protocols there is also no need to specify order functions $\{j_i^x\}$, $\{k_i^y\}$; all inputs can, in principle, be inserted *simultaneously*.

4.2 Examples of distillation protocols

4.2.1 The Forster-Winkler-Wolf non-adaptive protocol

The possibility that it is, in principle, possible to *distill nonlocality* was discovered by Forster, Winkler, and Wolf in 2009 [FWW09]. They presented the first protocol allowed to distill nonlocality from some no-signalling boxes. In their protocol, the parties essentially transfer their inputs (x, y) to the n resource boxes and each party outputs the parity of the outcomes it obtains. Explicitly, the input and output functions are

$$x_i = x_i(x) = x\,, \qquad a^x = \bigoplus_{i=1}^{n} a_i\,, \tag{4.3}$$

$$y_i = y_i(y) = y\,, \qquad b^y = \bigoplus_{i=1}^{n} b_i\,. \tag{4.4}$$

As the protocol is non-adaptive, the inputs x_i and y_i do not depend on previous outputs, we do not need to specify the orders $\{j_i\}$ and $\{k_i\}$ in which the players intact with the resources. We show that the protocol distills nonlocality from correlated boxes $\mathsf{C}_\delta(ab\,|\,xy)$, see (2.10), for any $0 < \delta < 1/2$. A correlated box is defined as a convex combination of a (noiseless) PR and a pair of perfect shared random bits SR, hence,

$$\begin{aligned} \mathrm{CHSH}(\mathsf{C}_\delta) &= \delta\,\mathrm{CHSH}(\mathsf{PR}) + (1-\delta)\,\mathrm{CHSH}(\mathsf{SR}) \\ &= 1\cdot\delta + \frac{3}{4}(1-\delta) = \frac{3+\delta}{4}\,. \end{aligned} \tag{4.5}$$

For values of $\delta < 1/2$ the protocol distills nonlocality of correlated boxes, *i.e.*, for the above distillation protocol we have $\mathrm{CHSH}(\hat{\mathsf{P}}) > \mathrm{CHSH}(\mathsf{C}_\delta)$, with an asymptotic value of $\mathrm{CHSH}(\hat{\mathsf{P}}) = 7/8$ for large n. The rough intuition is the following (for a detailed analysis we refer to [FWW09]): For each input (x, y) the parity of the output is the average parity of the sum of outputs of n copies of a C_δ box. For $(x, y) \neq (1, 1)$ the parity of the outputs of a single copy is with certainty 0, *i.e.*, $\mathsf{C}_\delta(a_i \oplus b_i = 0\,|\,xy) = 1$. Therefore, also the sum of all these parities $\sum_{i=1}^n a_i \oplus b_i$ is with certainty 0. For $(x, y) = (1, 1)$, we have $\mathsf{C}_\delta(a_i \oplus b_i = 0\,|\,xy) = 1 - \delta$ for each

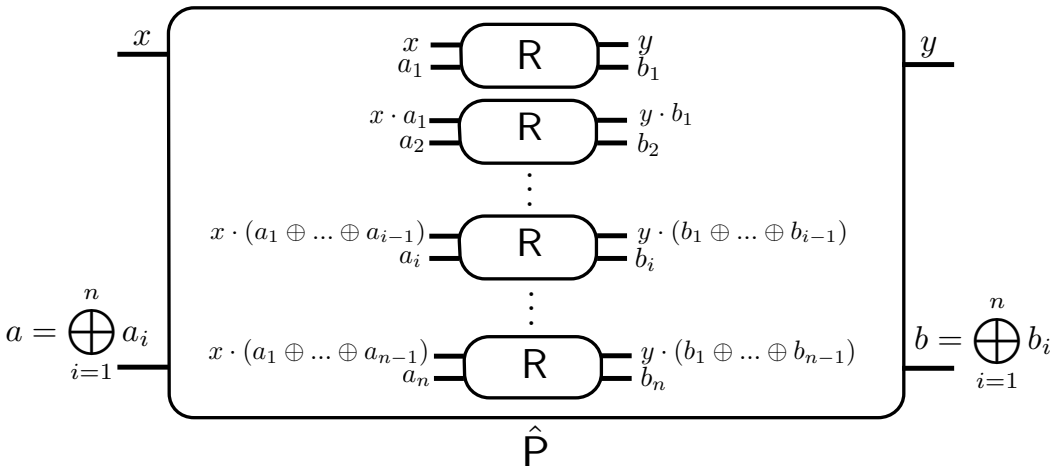

Figure 4.2. The Brunner-Skrzypczyk distillation protocol. If the resources are correlated boxes $\mathsf{R}(a_i b_i \,|\, x_i y_i) = \mathsf{C}_\delta(a_i b_i \,|\, x_i y_i)$*, then for* $n \to \infty$ *the distillate becomes a perfect* PR*, i.e.,* $\hat{\mathsf{P}}(ab \,|\, xy) = \mathsf{PR}(ab \,|\, xy)$*.*

copy independently. Then, for every $\delta > 0$, the probability to have the correct output parity, *i.e.*, $\sum_{i=1}^{n} a_i \oplus b_i = 1$, approaches $1/2$ for large n. Therefore, we have for the distillate $\hat{\mathsf{P}}(ab \,|\, xy)$:

$$\begin{aligned} \mathrm{CHSH}(\hat{\mathsf{P}}) &= \frac{1}{4} \sum_{x,y} \mathsf{P}(a \oplus b = x \cdot y \,|\, xy) \\ &= \frac{1}{4}(1 + 1 + 1 + 0.5) = 0.875 \,. \end{aligned} \tag{4.6}$$

4.2.2 The Brunner-Skrzypczyk adaptive protocol

The second protocol we present is due to Brunner and Skrzypczyk [BS09], in a form generalised to the use of n resource boxes (see Figure 4.2). The protocol was the first *adaptive* protocol: Inputs to resource boxes depend on previously obtained outputs. Alice and Bob use their resources in standard order, *i.e.*, the order functions are trivially $j_i = i = k_i$. The input and output functions for Alice and Bob are given as

$$x_1 = x\,, \qquad x_i = x \cdot \Big(\bigoplus_{j=1}^{i-1} a_j \Big)\,, \qquad a = \bigoplus_{i}^{n} a_i\,, \tag{4.7}$$

$$y_1 = y\,, \qquad y_i = y \cdot \Big(\bigoplus_{j=1}^{i-1} b_j \Big)\,, \qquad b = \bigoplus_{i}^{n} b_i\,. \tag{4.8}$$

This protocol was highly celebrated as it can distill the nonlocality of C_δ boxes even up to an asymptotic value of $\mathrm{CHSH}(\hat{\mathsf{P}}) \to 1$ (in the limit of large n), *i.e.*, the simulated box $\hat{\mathsf{P}}$ becomes equivalent to the PR, for any $\delta > 0$. We again provide the intuition behind this result: First, consider the case where $(x, y) \neq (1, 1)$. Then we have $(x_i, y_i) \neq (1, 1)$ for $1 \leq i \leq n$ and analogously to the previous example we have again $\hat{\mathsf{P}}(a \oplus b = 0 \,|\, xy) = 1$.

Now to the case where $(x, y) = (1, 1)$. In this case the protocol performs a sort of *error correction*: as soon as the parity of the outputs of any first j boxes is odd, *i.e.*, $\bigoplus_{i=1}^{j} a_i \oplus b_i = 1$, then the inputs cannot be both 1 for the rest of the protocol, *i.e.*, $(x_i, y_i) \neq (1, 1)$ for any $i > j$. Hence, once the sum of the outputs has the correct parity, it also remains correct for the rest of the protocol. Until this happens, *i.e.*, as long as $\bigoplus_{i=1}^{j} a_i \oplus b_i = 0$, there is a 50% chance that the next input is $(x_{j+1}, y_{j+1}) = (1, 1)$. Since $\mathsf{C}_\delta(a_{j+1} \oplus b_{j+1} = 1 \,|\, x_{j+1} = 1, y_{j+1} = 1) = \delta > 0$, sooner or later the parity of the outputs becomes odd, and then stays odd for the rest of the protocol. This implies that also that $\hat{\mathsf{P}}(a \oplus b = 1 \,|\, x = 1, y = 1) \to 1$ in the limit of large n and, therefore, nonlocality reaches its algebraic maximum $\mathrm{CHSH}(\hat{\mathsf{P}}) \to 1$.

For further examples of adaptive distillation protocols we refer the reader, *e.g.*, to [Ras12] and [ABL$^+$09].

4.3 Distillation as a cryptographic game

Both protocols presented in Section 4.2 crucially exploit the fact that there is no noise in the output (a_i, b_i) if $(x_i, y_i) \neq (1, 1)$. For PR_ε, which can be decomposed into two parts, a part with maximal nonlocality and the other just (white) noise,

$$\mathsf{PR}_\varepsilon(ab \,|\, xy) = (1 - 2\varepsilon)\, \mathsf{PR}(ab \,|\, xy) + 2\varepsilon\, \mathsf{U}(ab \,|\, xy)\,, \tag{4.9}$$

nonlocality distillation seems to be impossible.[2]

We introduce a novel method to derive bounds on nonlocality distillation protocols. We construct an adversary $\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n})$ who attacks the resource distribution $\mathsf{R}^{\otimes n}(a_{\leq n} b_{\leq n} \,|\, x_{\leq n} y_{\leq n})$ to gain knowledge on Alice's output of the distilla-

[2] As proven by Beigi and Gohari for super-quantum PR_ε boxes [BG14].

tion protocol $f^{x=0}(a_{\leq n})$ (see also Figure 4.3). If the interaction between the players and $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ is well-defined, then (4.2) maps $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ to a box $\hat{\mathsf{P}}(abe \,|\, xy)$ with $\hat{\mathsf{P}}(e \,|\, xy) = \mathsf{P}(e)$ and $\sum_e \hat{\mathsf{P}}(abe \,|\, xy) = \hat{\mathsf{P}}(ab \,|\, xy)$ for the output $\hat{\mathsf{P}}(ab \,|\, xy)$ of the distillation protocol. If $\hat{\mathsf{P}}(ab \,|\, exy)$ is also no-signalling, then we are able to apply Corollary 3.1.2 in order to derive limitations on $\mathrm{CHSH}(\hat{\mathsf{P}})$ from the knowledge of the adversary

$$\hat{\mathsf{P}}(a = e \,|\, x) \geq 1/2 + 2\varepsilon \quad \Rightarrow \quad \mathrm{CHSH}(\mathsf{P}) \leq 1 - \varepsilon \,. \tag{4.10}$$

We derive sufficient conditions on the attack $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ such that $\hat{\mathsf{P}}(ab \,|\, exy)$ is no-signalling. The intuition behind these conditions is that as long as the resource distribution conditioned on Eve $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ respects the orders of use of the distillation protocol the interaction of the players with the box $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ is well-defined and induces a no-signalling distribution $\hat{\mathsf{P}}(ab \,|\, exy)$ on the output of the protocol via the mappings (4.2). In general protocols, *e.g.*, Alice can choose the inputs $x_{j_{>i}}$ as a function of $a_{j_{\leq i}}$, which is why we require the outputs $a_{j_{\leq i}}$ to be independent of the inputs $x_{j_{>i}}$ by enforcing the dynamic TONS conditions on $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$. For non-adaptive protocols the inputs $x_{\leq n}$ are independent of the outputs $a_{\leq n}$ and therefore it is sufficient to enforce the ABNS conditions.

4.3.1 ABNS-attacks induce no-signalling attacks on non-adaptive protocols

As a warm-up for the case of general distillation protocols, we show how to derive bounds on *non-adaptive* protocols from ABNS-attacks.

Theorem 4.3.1. *Let $\hat{\mathsf{P}}$ denote a box generated by a non-adaptive distillation protocol using n R boxes as resource. Let $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ be an ABNS attack on the resource boxes $\mathsf{R}^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$. Then $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ induces a no-signalling attack $\hat{\mathsf{P}}(abe \,|\, xy)$ on the distillate $\hat{\mathsf{P}}(ab \,|\, xy)$.*

Proof. In order to prove the theorem, we need to show that the interaction between Alice and Bob with the resources conditioned on Eve $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ is well-defined, and then that $\hat{\mathsf{P}}(ab \,|\, exy)$ is no-signalling between Alice and Bob. The distribution $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, xy)$ can be calculated from the resource $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}y_{\leq n})$ directly:

$$\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, xy) = \mathsf{P}(a_{\leq n}b_{\leq n} \,|\, ex_{\leq n}(x)y_{\leq n}(y)) \,. \tag{4.11}$$

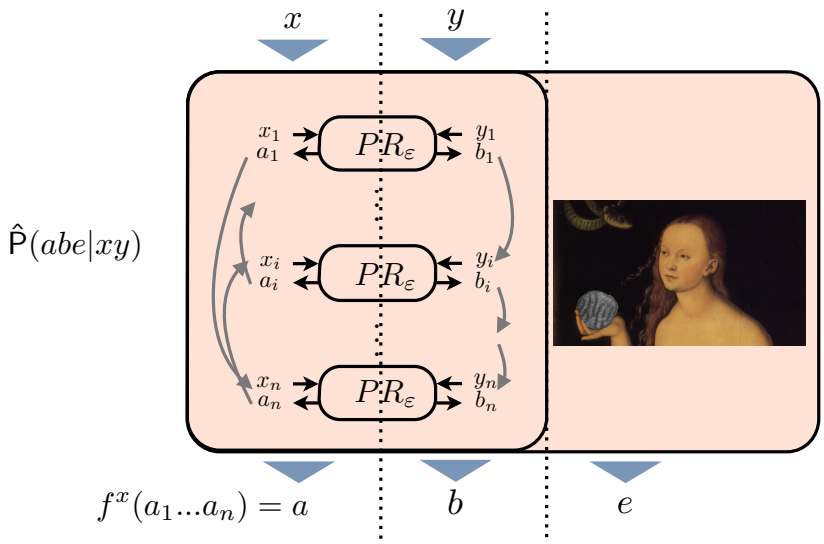

Figure 4.3. Eve's view on distillation: A no-signalling attack on the resources of an adaptive distillation protocol. Similar to the situation in a no-signalling privacy amplification protocol the adversary attacks an output bit that is a function of the outputs of $\mathsf{PR}_\varepsilon^{\otimes n}$, $f^x(a_1, ..., a_n)$. *Via the mappings* (4.2), *the attack* $\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n})$ *on the resources is mapped on* $\hat{\mathsf{P}}(abe \,|\, xy)$. *To apply Corollary 3.1.2 in order to derive bounds on* $\mathrm{CHSH}(\hat{\mathsf{P}})$, *the box* $\hat{\mathsf{P}}(abe \,|\, xy)$ *must be no-signalling. For this it is sufficient that* $\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n})$ *is a dynamic TONS attack on the resources* $\mathsf{PR}_\varepsilon^{\otimes n}$.

For no-signalling from Bob to Alice we have

$$\begin{aligned}\sum_a \mathsf{P}(ab\,|\,exy) &= \sum_{\substack{a_{\leq n}:\ f^x(a_{\leq n})=a \\ b_{\leq n}}} \mathsf{P}(a_{\leq n}b_{\leq n}\,|\,exy) \\ &= \sum_{\substack{a_{\leq n}:\ f^x(a_{\leq n})=a \\ b_{\leq n}}} \mathsf{P}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}(x)y_{\leq n}(y)) \\ &= \sum_{\substack{a_{\leq n}:\ f^x(a_{\leq n})=a \\ b_{\leq n}}} \mathsf{P}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}(x)y_{\leq n}(y')) \\ &= \sum_{\substack{a_{\leq n}:\ f^x(a_{\leq n})=a \\ b_{\leq n}}} \mathsf{P}(a_{\leq n}b_{\leq n}\,|\,exy') \\ &= \sum_a \mathsf{P}(ab\,|\,xy'e)\,, \end{aligned} \tag{4.12}$$

where we use Definition 4 for the second equality. No-signalling from Alice to Bob is proven analogously. □

Using Theorem 4.3.1 we can directly apply the ABNS attack from Theorem 3.2.1 by Hänggi *et al.* [HRW13] (or, equivalently, Theorem 3.7.2 and Theorem 3.7.4 to obtain a slightly weaker bound), to obtain Theorem 4.3.2.

Theorem 4.3.2. *Let $\hat{\mathsf{P}}(ab\,|\,xy)$ be generated by a non-adaptive distillation protocol using n PR_ε boxes as resource. Then its degree of nonlocality is bounded by*

$$\mathrm{CHSH}(\hat{\mathsf{P}}) \leq 1 - \frac{\varepsilon}{4}\,. \tag{4.13}$$

Proof. Theorem 3.2.1 states that for any function $f(a_{\leq n})$, there exists an ABNS attack $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ such that

$$\mathsf{P}(f(a_{\leq n}) = e\,|\,x_{\leq n}) \geq \frac{\varepsilon}{2} \quad \forall x_{\leq n}\,. \tag{4.14}$$

Thus, an ABNS attack $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ that satisfies (4.14) for the output function $f^0(a_{\leq n})$ of the distillation protocol induces via Theorem 4.3.1 an attack $\hat{\mathsf{P}}(abe\,|\,xy)$ with

$$\hat{\mathsf{P}}(a = e\,|\,x = 0) \geq \frac{\varepsilon}{2}\,. \tag{4.15}$$

By Corollary 3.1.2, this implies

$$\mathrm{CHSH}(\hat{\mathsf{P}}) \leq 1 - \frac{\varepsilon}{4}\,, \tag{4.16}$$

which completes the proof. □

By an argument of combining subprotocols we can even derive a stronger bound than (4.13) for many values of ε.

Corollary 4.3.3. *For any $\delta > 0$ and any $\varepsilon < 1/4$, there exists a subset $\mathcal{S} \subseteq [\varepsilon/4, \varepsilon]$ of non-zero measure, such that for any non-adaptive protocol using $\mathsf{PR}_{\varepsilon^*}$, with $\varepsilon^* \in \mathcal{S}$, as resources the nonlocality of the distillate $\hat{\mathsf{P}}$ is bounded by*

$$\mathrm{CHSH}(\hat{\mathsf{P}}) \leq 1 - \varepsilon^* + \delta \,. \tag{4.17}$$

Proof. First note that via use of a depolarisation protocol [MAG06] and three bits of classical shared randomness, any box $\mathsf{P}(ab\,|\,xy)$ with $\mathrm{CHSH}(\mathsf{P}) = 1 - \varepsilon$ can be converted into a PR_ε box without communication. For a given $\varepsilon < 1/4$, define $1-\varepsilon'$ as the supremum of $\mathrm{CHSH}(\hat{\mathsf{P}})$ over all non-adaptive distillation protocols using (arbitrarily many) PR_ε as a resource. Without loss of generality, we can assume that $\varepsilon' < \varepsilon$. Then it must be possible to generate any box $\mathsf{PR}_{\varepsilon''}$ with $\varepsilon' < \varepsilon'' < \varepsilon$ with PR_ε boxes as resource. Hence, for any ε'', $1 - \varepsilon'$ is also the supremum of $\mathrm{CHSH}(\hat{\mathsf{P}})$ over all non-adaptive distillation protocols using $\mathsf{PR}_{\varepsilon''}$ as a resource. Thus, for a given $\delta > 0$ we choose the set $\mathcal{S} = (\varepsilon', \varepsilon' + \delta)$, which completes the proof. □

4.3.2 Sufficient conditions for a no-signalling attack on general distillation protocols

Theorem 4.3.4. *Let $\hat{\mathsf{P}}(ab\,|\,xy)$ denote a box simulated by a general distillation protocol using n R boxes as resource in dynamic orders $\{j_i^x\},\{k_i^y\}$. Let $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ be a dynamic TONS attack on $\mathsf{R}^{\otimes n}(a_{\leq n}b_{\leq n}\,|\,x_{\leq n}y_{\leq n})$ that fulfils* each *of the four dynamic TONS conditions specified by the orders $(\{j_i^x\},\{k_i^y\})$ for $x \in 0,1$ and $y \in 0,1$. Then $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ induces a no-signalling attack $\hat{\mathsf{P}}(abe\,|\,xy)$ to $\hat{\mathsf{P}}(ab\,|\,xy)$.*

Proof. The intuition behind Theorem 4.3.4 is that the interaction between Alice and Bob and the dynamic TONS system $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n})$ is (locally) well-defined for each dynamic order $\{j_i^x\},\{k_i^y\}$ and that, in addition, the distributions that arise on either Alice' or Bob's side are independent of the other players interaction with $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n})$.

We first show that the interaction of the players with the box $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n})$ is well-defined and then we prove no-signalling from Bob to Alice, $B \overset{ns}{\to} A$. No-signalling from Alice to Bob is proven analogously. If the box $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n})$ satisfies the dynamic TONS conditions (2.4) for the order $\{j_i^x\},\{k_i^y\}$ it can be written

as

$$\begin{aligned}\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n}) &= \mathsf{P}(a_{\leq n}\,|\,ex_{\leq n})\,\mathsf{P}(b_{\leq n}\,|\,a_{\leq n}ex_{\leq n}y_{\leq n})\\ &= \prod_{i=1}^{n}\mathsf{P}(a_{j_i^x}\,|\,a_{j^x_{<i}}ex^x_{j^x_{\leq i}})\prod_{i=1}^{n}\mathsf{P}(b_{k_i^y}\,|\,a_{\leq n}b_{k^y_{<i}}ex_{\leq n}y^y_{k^y_{\leq i}})\,. \end{aligned}\tag{4.18}$$

At each step the interaction of Alice and Bob and the system $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n})$ in a distillation protocol with dynamic orders $(\{j_i^x\},\{k_i^y\})$ is well-defined. The complete distribution $\hat{\mathsf{P}}(a_{\leq n}b_{\leq n}\,|\,exy)$ induced by this interaction, see (4.2), is computed as

$$\hat{\mathsf{P}}(a_{\leq n}b_{\leq n}\,|\,exy) = \hat{\mathsf{P}}(a_{\leq n}\,|\,ex)\,\hat{\mathsf{P}}(b_{\leq n}\,|\,a_{\leq n}exy)\tag{4.19}$$

$$\hat{\mathsf{P}}(a_{\leq n}\,|\,ex) = \prod_{i=1}^{n}\mathsf{P}(a_{j_i^x}\,|\,a_{j^x_{<i}}ex^x_{j^x_{\leq i}})\tag{4.20}$$

$$\hat{\mathsf{P}}(b_{\leq n}\,|\,a_{\leq n}exy) = \prod_{i=1}^{n}\mathsf{P}(b_{k_i^y}\,|\,a_{\leq n}b_{k^y_{<i}}ex^x_{j^x_{\leq i}}y^y_{k^y_{\leq i}})\,,\tag{4.21}$$

where $j_i^x = j_i^x(a_{<j_i})$, $k_i^y = k_i^y(b_{<k_i})$ and

$$x^x_{j_{\leq i}}(a_{j_{<i}}) := (x^x_{j_1}, x^x_{j_2}(a_{j_1}, x), ..., x^x_{j_i}(a_{j_{<i}}))\tag{4.22}$$

$$y^y_{k_{\leq i}}(b_{k_{<i}}) := (y^y_{k_1}, y^y_{k_2}(b_{k_1}), ..., y^y_{k_i}(b_{k_{<i}}))\,.\tag{4.23}$$

Note that, using the dynamic TONS conditions (2.4) one can also show that any marginal distribution arising during the protocol, *i.e.*, $\hat{\mathsf{P}}(a_{j^x_{\leq i_A}}b_{k^y_{\leq i_B}}\,|\,exy)$, is well-defined for any $i_A, i_B \in [n]$. It follows directly from (4.21) that

$$\begin{aligned}\sum_{b_{\leq n}}\hat{\mathsf{P}}(b_{\leq n}\,|\,a_{\leq n}exy) &= \prod_{i=1}^{n}\sum_{b_{k_i^y}}\mathsf{P}(b_{k_i^y}\,|\,a_{\leq n}b_{k^y_{<i}}ex^x_{j^x_{\leq i}}y^y_{k_{\leq i}}(b_{k^y_{<i}}))\\ &= 1\\ &= \prod_{i=1}^{n}\sum_{b_{k_i^{y'}}}\mathsf{P}(b_{k_i^{y'}}\,|\,a_{\leq n}b_{k^{y'}_{<i}}ex^x_{j^x_{\leq i}}y^{y'}_{k_{\leq i}}(b_{k^{y'}_{<i}}))\\ &= \sum_{b_{\leq n}}\mathsf{P}(b_{\leq n}\,|\,a_{\leq n}exy')\,,\end{aligned}\tag{4.24}$$

which implies $B \xrightarrow{ns} A$ through

$$\begin{aligned}\sum_b \hat{\mathsf{P}}(ab\,|\,exy) &= \sum_{\substack{a_{\leq n}:f^x(a_{\leq n})=a \\ b_{\leq n}}} \hat{\mathsf{P}}(a_{\leq n}b_{\leq n}\,|\,exy) \\ &= \sum_{\substack{a_{\leq n}:f^x(a_{\leq n})=a \\ b_{\leq n}}} \hat{\mathsf{P}}(a_{\leq n}\,|\,ex)\hat{\mathsf{P}}(b_{\leq n}\,|\,a_{\leq n}exy) \\ &= \sum_{\substack{a_{\leq n}:f^x(a_{\leq n})=a \\ b_{\leq n}}} \hat{\mathsf{P}}(a_{\leq n}\,|\,ex)\hat{\mathsf{P}}(b_{\leq n}\,|\,a_{\leq n}exy') \\ &= \sum_b \hat{\mathsf{P}}(ab\,|\,exy')\,, \end{aligned} \tag{4.25}$$

which completes the proof. □

4.3.3 Application of dynamic TONS attacks to adaptive distillation protocols

When we constructed dynamic TONS attack in Section 3.6 from the distributions $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$, we did so for a *fixed* dynamic order $\{j_i\}$, and only showed that dynamic TONS conditions hold for this choice of dynamic order $\{j_i\}$ (but for all possible dynamic orders of Bob $\{k_i\}$). At any step in the protocol we could have for example $j_i^0(a_{<j_i}) \neq j_i^1(a_{<j_i})$ and the two orders may deviate. If, in this case, an attack $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ satisfies the dynamic TONS conditions only for a *single* dynamic order, *e.g.*, $\{j_i^0\}$, $\{k_i^0\}$, then the proof of Theorem 4.3.4 does not apply anymore.

Thus, in order to apply Theorem 4.3.4, $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ must satisfy the dynamic TONS conditions for *each* set of dynamic orders $\{i_j^x\}$, $\{k_i^y\}$

- $\{i_j^0\}, \{k_i^0\}$
- $\{i_j^0\}, \{k_i^1\}$
- $\{i_j^1\}, \{k_i^0\}$
- $\{i_j^1\}, \{k_i^1\}$.

In Section 3.6, we show how all TONS attacks constructed in Chapter 3 can be generalised to dynamic TONS attacks for an arbitrary order $\{k_i\}$ for Bob, however, a *fix* order $\{j_i\}$ for Alice — it can by *any* order $\{j_i\}$ but the choice *which* order is determined by construction (3.225) to (3.227). If we want to use (3.225) to (3.227) then the problem is that the attack $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ must be constructed *independently* of the inputs to the distillation protocol (x, y), and, hence, in particular of the two

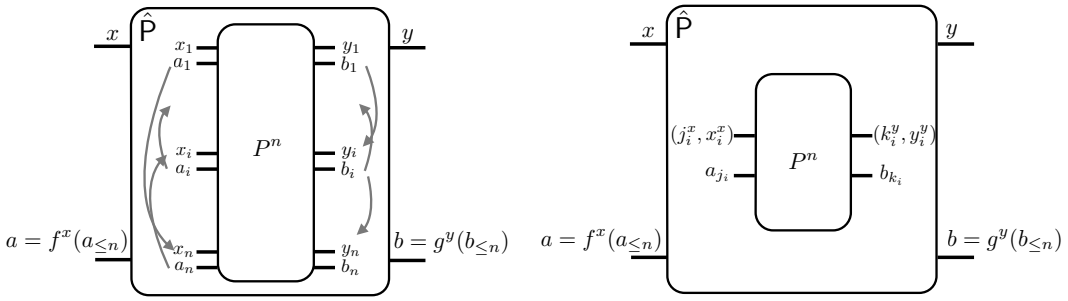

Figure 4.4. Mathematically equivalent view on a distillation protocol. On the left, arrows denote the order in which Alice and Bob use their interfaces. On the right, Alice and Bob repeatedly *use the same interface. If equations* (4.28) *are satisfied, then the distillate* $\hat{\mathsf{P}}(ab\,|\,xy)$ *is identical on both sides.*

orders $\{j_i^0\}$ and $\{j_i^1\}$. We construct an attack independent of (x, y), but not attacking the distillation protocol itself, but rather a *mathematically equivalent version* (see also Figure 4.4).

When a distillation protocol is executed, Alice and Bob may proceed with the next step using arbitrary input slots of their choice. In principle, one should regard the information concerning which box is used in the next step as an *additional* input of each party Alice and Bob to the system $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$. Consider a distillation protocol defined by the tuple of functions

$$\left(\{j_i^x\}, \{k_i^y\}, \{x_{j_i}^x\}, \{y_{k_i}^y\}, \{f^x\}, \{g^y\}\right) . \tag{4.26}$$

Let us define the $i-th$ interaction of Alice with the system $\mathsf{P}(a_{\leq n}b_{\leq n}\,|\,ex_{\leq n}y_{\leq n})$ as inserting the input $x_{j_i}^x$ into the box j_i^x and obtaining the output $a_{j_i^x}$, and similarly for Bob. Alternatively, we could consider another system

$$\mathsf{P}'(a_{\leq n}b_{\leq n}\,|\,e\,j_{\leq n}k_{\leq n}x_{\leq n}y_{\leq n}) , \tag{4.27}$$

which has a *single* interface on Alice's and Bob's side and the $i-th$ interaction of Alice is defined as her inserting input $(j_i^x, x_i^x) \equiv x_{j_i^x}^x$ and obtaining $a_{j_i^x}$ as output (see also Figure 4.4). If we define

$$\mathsf{P}'(a_{j_i^x}\,|\,a_{j_{<i}^x}j_{\leq i}^x x_{\leq i}^x) = \mathsf{P}(a_{j_i^x}\,|\,a_{j_{<i}^x}x_{j_{\leq i}}^x) \quad \text{and} \tag{4.28}$$

$$\mathsf{P}'(b_{k_i^y}\,|\,a_{\leq n}b_{k_{<i}^y}j_{\leq n}^x k_{\leq i}^y x_{\leq n}^x y_{\leq i}^y) = \mathsf{P}(b_{k_i^y}\,|\,a_{\leq n}b_{k_{<i}^y}x_{\leq n}^x y_{k_{\leq i}}^y) , \tag{4.29}$$

then we have

$$\mathsf{P}'(a_{\leq n}b_{\leq n}\,|\,xy) = \mathsf{P}(a_{\leq n}b_{\leq n}\,|\,xy)\ . \tag{4.30}$$

We present the attack $\mathsf{P}'(a_{\leq n}b_{\leq n}\,|\,e\,j_{\leq n}k_{\leq n}x_{\leq n}y_{\leq n})$ on a sequential-input distillation protocol with output function $f^0(a_1,...,a_n)$, similar to the one in Section 3.6.
Let the function $\tilde{f}(a_1,...,a_n)$ be defined for the dynamic order $\{j_i^0\}$ by

$$\tilde{f}(a_1,a_2,...,a_n) = f(a'_1,a'_2,...,a'_n) \tag{4.31}$$

with $a'_i := a_{j_i}$. Now let $\mathsf{P}(a_{\leq n}b_{\leq n}e\,|\,x_{\leq n}y_{\leq n})$ be a TONS attack on the function $\tilde{f}$ and on the boxes $\mathsf{PR}_\varepsilon^{\otimes n}$ constructed from an $(\varepsilon,\mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ according to the proof of Theorem 3.4.3. Then we define $\mathsf{P}'(a_{\leq n}b_{\leq n}\,|\,e\,j_{\leq n}k_{\leq n}x_{\leq n}y_{\leq n})$ as

$$\mathsf{P}'(e) = \mathsf{P}(e) \tag{4.32}$$

$$\mathsf{P}'(a_{\leq n}b_{\leq n}\,|\,e\,x_{\leq n}y_{\leq n}) = \prod_i \mathsf{P}'(a_{j_i}b_{j_i}\,|\,a_{j_{<i}}e\,j_i x_i y_i) \tag{4.33}$$

$$\mathsf{P}'(a_{j_i}b_{j_i}\,|\,a_{j_{<i}}e\,j_i x_i y_i) = \mathsf{P}(a'_i b'_i\,|\,a'_{<i}e\,x'_i y'_i)\ . \tag{4.34}$$

with $a'_i = a_{j_i}$.

Theorem 4.3.5. *Consider a distillation protocol given by the tuple*

$$(\{j_i^x\},\{k_i^y\},\{x_{j_i}^x\},\{y_{k_i}^y\},\{f^x\},\{g^y\})\ ,$$

using $\mathsf{PR}_\varepsilon^{\otimes n}$ *as resources. Then the box* $\mathsf{P}'(a_{\leq n}b_{\leq n}e\,|\,j_{\leq n}k_{\leq n}x_{\leq n}y_{\leq n})$ *constructed in* (4.32)-(4.34) *defines a dynamic TONS attack on* $\mathsf{PR}_\varepsilon^{\otimes n}$ *with respect to all four dynamic orders* $\{j_i^x\}$ *and* $k_i^y\}$*. Furthermore, Eve's guessing probability of Alice's output given* $x = 0$ *is given by*

$$\hat{\mathsf{P}}'(a = e\,|\,x = 0) = \mathsf{P}(\tilde{f}(a_{\leq n}) = e)\ . \tag{4.35}$$

Theorem 4.3.5 follows directly from the proof of Theorem 3.6.2 and the above considerations. Finally we can derive a bound on the nonlocality of the distillate for general distillation protocol.

Theorem 4.3.6. *Let* $\hat{\mathsf{P}}(ab\,|\,xy)$ *be generated by a general distillation protocol using* n PR_ε *as resource. Then its degree of nonlocality is bounded by*

$$\mathrm{CHSH}(\hat{\mathsf{P}}) \leq 1 - \frac{\varepsilon}{2n}\ . \tag{4.36}$$

Proof. Due to Theorem 3.188, for any function $f(a_{\leq n})$ there exists an $(\varepsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\varepsilon}(a_{\leq n}e)$ with

$$\mathsf{Q}_{o-\varepsilon}(f(a_{\leq n}) = e) \geq 1/2 + \varepsilon/n \,. \tag{4.37}$$

Through Theorem 3.4.3 there must also exist a TONS attack $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on $\mathsf{PR}_{\varepsilon}^{\otimes n}$ with $\mathsf{P}(f(a_{\leq n}) = e) \geq 1/2 + \varepsilon/n$ for any function $f(a_{\leq n})$. For the given distillation protocol with output function $f^0(a_{\leq n})$ we use a TONS attack $\mathsf{P}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on $\tilde{f}(a_{\leq n})$ with $\mathsf{P}(\tilde{f}(a_{\leq n}) = e) \geq 1/2 + \varepsilon/n$. Then we construct the attack on the distillation $\mathsf{P}'(a_{\leq n}b_{\leq n} \,|\, e j_{\leq n}k_{\leq n}x_{\leq n}y_{\leq n})$ according to (4.32) to (4.34), which by Theorem 4.3.5 and Theorem 4.3.4 implies a no-signalling attack $\hat{\mathsf{P}}'(abe \,|\, xy)$ on the distillate $\hat{\mathsf{P}}(ab \,|\, xy)$ with

$$\hat{\mathsf{P}}'(a = f^0(a_{\leq n}) = e \,|\, x = 0) \geq 1/2 + \frac{\varepsilon}{n} \,. \tag{4.38}$$

Finally, we apply Corollary 3.1.2 and obtain

$$\mathrm{CHSH}(\hat{\mathsf{P}}) \leq 1 - \frac{\varepsilon}{2n} \,, \tag{4.39}$$

this completes the proof. □

Chapter 5

Summary and Outlook

5.1 Results on no-signalling attacks

In Chapter 3, we study the power of no-signalling attacks on privacy-amplification protocols using $\mathsf{PR}_\varepsilon(ab \,|\, xy)$ boxes. We focus on time-ordered no-signalling (TONS) attacks, which are relevant for the security of *feasible* secret-key distribution protocols based only *provably minimal assumptions*. In Section 3.3, we present a scenario of classical deterministic privacy amplification of ε-Santha-Vazirani-sources, which is comparable to TONS privacy amplification. We show that the impossibility result for classical deterministic privacy amplification of ε-Santha-Vazirani-sources, achieved through Reingold distributions (3.40) and (3.41), *cannot* be easily extended to TONS privacy amplification in Section 3.3.2 . In Section 3.4.1, we present a construction of TONS attacks from another classical privacy-amplification game, see Theorems 3.4.2 and 3.4.3. Numerical evidence, as well as the analysis of these games for, *e.g.*, parity functions, suggest the possibility that the Reingold distributions can be extended to TONS attacks through this construction, see Conjecture 3.4.1, and thus impossibility of TONS privacy amplification would be proven. A particularly simple "bias-the-last-bit" attack provides further strong evidence for the impossibility of TONS privacy amplification; a lower bound of $1/2 + \varepsilon/2$ on the knowledge of the adversary for random functions and a lower bound of $1/2 + (1 - \gamma_1)\varepsilon/2$ for all but an exponentially small fraction (exponentially small in γ_1, doubly exponentially small in n) of all functions $f(a_{\leq n})$, see Theorems 3.5.2 and 3.5.3. Our construction comprises "prefix-code" attacks, used in [AFTS12] to derive the strongest known bound on the adversaries knowledge of $1/2 + \varepsilon/n$ for general privacy-amplification functions. Using a theorem by Kahn, Kalai, and Linial [KKL88] from the field of analysis of Boolean functions, we show that this bound can be strengthened for monotonic functions to $1/2 + \varepsilon \cdot \theta(\log(n)/n)$, see

Theorem 3.5.10. For majority functions, "prefix-code" attacks yield a bound of $1/2 + \varepsilon \cdot \theta(1/\sqrt{n})$. With our construction we derive a much stronger attack, which yields a bound of more than $1/2+\varepsilon$ in the limit of large n, see Theorem 3.5.14; the attack proves that TONS privacy amplification with majority functions is impossible. We show that our construction can also be used to derive dynamic TONS attacks, where the time-ordering of the subsystems of Alice and Bob is permuted, see Theorem 3.6.2. We use these attacks to derive bounds on nonlocality distillation protocols in Chapter 4. In Section 3.7, we present further evidence that *if* TONS privacy amplification is impossible, then our construction should also be powerful enough to prove this: an analogous construction for the stronger ABNS adversary retrieves the fact that privacy amplification is impossible against this adversary [HRW13]. In Section 3.8, we show that all of the above results also hold for privacy-amplification protocols on more general distributions $\mathsf{V}_\varepsilon(ab\,|\,xy)$, which are relevant for secret-ket distribution protocols based on other Bell inequalities than the CHSH inequality.

5.2 Results on nonlocality distillation

In a nonlocality-distillation protocol two players, Alice and Bob, generate, without communication, a box $\hat{\mathsf{P}}(ab\,|\,xy)$ with a higher degree of nonlocality than the resource boxes $\mathsf{P}(ab\,|\,xy)$ they use, *i.e.*, $\mathrm{CHSH}(\hat{\mathsf{P}}) > \mathrm{CHSH}(\mathsf{P})$. We present a novel method to derive bounds on nonlocality distillation protocols using $\mathsf{PR}_\varepsilon(ab\,|\,xy)$ boxes as a resource. We construct a third party Eve who gains knowledge on the output of the distillate $\hat{\mathsf{P}}(ab\,|\,xy)$ through a no-signalling attack on the resource boxes. The degree of this knowledge limits the degree of nonlocality of the distillate $\mathrm{CHSH}(\hat{\mathsf{P}})$ if the no-signalling attack on the resources induces also a no-signalling attack on the distillate. We relate the specifics of the distillation protocol, in particular the interaction of the players with the resources, with the constraints on the no-signalling attack of the resource. We derive sufficient conditions for the no-signalling attack on the resource boxes to induce a no-signalling attack on the distillate: the ABNS conditions for non-adaptive distillation protocols and the dynamic TONS conditions for general distillation protocols. For general nonlocality distillation protocols using PR_ε boxes as resource we obtain the bound $\mathrm{CHSH}(\hat{\mathsf{P}}) \leq 1 - \varepsilon/2n$, see Theorem 4.3.6. For nonadaptive protocols we derive a bound of at most $\mathrm{CHSH}(\hat{\mathsf{P}}) \leq 1-\varepsilon/4$ for all values of ε. Since this bound is constant in n, this implies also the existence of an infinity of values ε, for which distillation is virtually impossible. These limits on nonlocality distillation also hold if the players use the more general boxes $\mathsf{V}_\varepsilon(ab\,|\,xy)$ as a resource, where the B system and the correlation between the A system and the B system are completely arbitrary.

5.3 Outlook

The central open problem of this thesis is the question whether TONS privacy amplification is possible. It is the author's firm belief, confirmed for almost all functions as well as linear and majority functions, that this is not the case. One way to answer the question in the negative would be to prove Conjecture 3.4.1. Intuitively, this means to show that ε-Santha-Vazirani distributions $\mathsf{Q}_{\varepsilon-sv}(a_{\leq n}e)$ arising from the Reingold construction can be *fine-grained* into a set of ordered $\mathcal{S}$-influenceable distributions $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$.

We confined our analysis of TONS privacy amplification to the case of two parties, Alice and Bob, and a privacy-amplification function $f(a_{\leq n})$ that depends only on Alice' systems. If in this case TONS privacy amplification turns out to be impossible, then, in order to investigate if *feasible* key distribution based on minimal assumptions is possible, one needs is to consider *multipartite* time-ordered-no-signalling systems with a *fixed* number of parties and privacy-amplification functions that depend on the outputs of systems of several parties. A no-go theorem for privacy amplification in this multipartite scenario would bear the insight, that in general no-signalling theories the achievable *secrecy* against an outside observer is limited by the *space* the parties control — rather than by the amount of repetitive "measurements", which are just interactions with boxes in the present context. Such a relation between information and space would parallel efforts from, *e.g.*, the field of *black hole thermodynamics*, where the amount of information (entropy) contained in a volume of space is limited by the size of its surface [tH93], [Sus95].

Bibliography

[ABB+10] Mafalda L. Almeida, Jean-Daniel Bancal, Nicolas Brunner, Antonio Acín, Nicolas Gisin, and Stefano Pironio. Guess your neighbor's input: A multipartite nonlocal game with no quantum advantage. *Phys. Rev. Lett.*, 104:230404, Jun 2010.

[ABG+07] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98:230501, Jun 2007.

[ABL+09] Jonathan Allcock, Nicolas Brunner, Noah Linden, Sandu Popescu, Paul Skrzypczyk, and Tamás Vértesi. Closed sets of nonlocal correlations. *Phys. Rev. A*, 80:062107, Dec 2009.

[ADR82] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental Test of Bell's Inequalities Using Time- Varying Analyzers. *Phys. Rev. Lett.*, 49:1804–1807, Dec 1982.

[AFRV14] Rotem Arnon-Friedman, Renato Renner, and Thomas Vidick. Non-signalling parallel repetition using de finetti reductions. *http://arxiv.org/abs/1411.1582*, 2014.

[AFTS12] Rotem Arnon-Friedman and Amnon Ta-Shma. Limits of privacy amplification against nonsignaling memory attacks. *Phys. Rev. A*, 86:062333, Dec 2012.

[AV79] D. Angluin and L.G. Valiant. Fast probabilistic algorithms for hamiltonian circuits and matchings. *Journal of Computer and System Sciences*, 18(2):155 – 193, 1979.

[BB84] Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the Inter-*

national Conference on Computers, Systems and Signal Processing, pages 175–179, 1984.

[BBCM95] Charles H. Bennett, Gilles Brassard, Claude Crepeau, and Ueli M. Maurer. Generalized privacy amplification. *IEEE Trans. Inf. Theor.*, 41(6):1915–1923, Nov 1995.

[BBL+06] Gilles Brassard, Harry Buhrman, Noah Linden, André Allan Méthot, Alain Tapp, and Falk Unger. Limit on nonlocality in any world in which communication complexity is not trivial. *Phys. Rev. Lett.*, 96:250401, Jun 2006.

[BBR88] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy amplification by public discussion. *SIAM J. Comput.*, 17(2):210–229, Apr 1988.

[BC89] Samuel L. Braunstein and Carlton M. Caves. Wringing out better Bell inequalities. *Nuclear Physics B - Proceedings Supplements*, 6(0):211 – 221, 1989.

[Bel64] John S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[BG14] Salman Beigi and Amin Gohari. A monotone measure for non-local correlations. *http://arxiv.org/abs/1409.3665v3*, 2014.

[BHK05] Jonathan Barrett, Lucien Hardy, and Adrian Kent. No signaling and quantum key distribution. *Phys. Rev. Lett.*, 95:010503, Jun 2005.

[BS09] Nicolas Brunner and Paul Skrzypczyk. Nonlocality distillation and postquantum theories with trivial communication complexity. *Phys. Rev. Lett.*, 102:160403, Apr 2009.

[Che52] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23(4):493–507, Dec 1952.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[Cir80] Boris S. Cirel'son. Quantum generalizations of Bell's inequality. *Letter in Mathematical Physics*, 4:93–100, 1980.

[Col06] Roger Colbeck. *Quantum And Relativistic Protocols For Secure Multi-Party Computation*. PhD thesis, University of Cambridge, 2006.

[CR11] Roger Colbeck and Renato Renner. No extension of quantum theory can have improved predictive power. *Nat. Commun.*, 2:411, Aug 2011.

[CR12] Roger Colbeck and Renato Renner. Free randomness can be amplified. *Nat. Phys.*, 8(6):450–453, Jun 2012.

[DLR12] Oscar C. O. Dahlsten, Daniel Lercher, and Renato Renner. Tsirelson's bound from a generalised data processing inequality. *New J. Phys.*, 14(063024), 2012.

[DW08] Dejan Dukaric and Stefan Wolf. A limit on non-locality distillation. *http://arxiv.org/abs/0808.3317*, Aug 2008.

[Eke91] Artur K. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661–663, Aug 1991.

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777, 1935.

[EPR92] Avshalom C. Elitzur, Sandu Popescu, and Daniel Rohrlich. Quantum nonlocality for each pair in an ensemble. *Physics Letters A*, 162(1):25–28, 1992.

[FHSW10] Matthias Fitzi, Esther Hänggi, Valerio Scarani, and Stefan Wolf. The non-locality of n noisy popescu–rohrlich boxes. *Journal of Physics A: Mathematical and Theoretical*, 43(46):465305, 2010.

[For11] Manuel Forster. Bounds for nonlocality distillation protocols. *Phys. Rev. A*, 83:062114, Jun 2011.

[FSA+13] Tobias Fritz, Ana Bélen Sainz, Remigiusz Augusiak, J. Bohr Brask, Rafael Chaves, Anthony Leverrier, and Antonio Acín. Local orthogonality as a multipartite principle for quantum correlations. *Nat Commun*, 4, Aug 2013.

[FWW09] Manuel Forster, Severin Winkler, and Stefan Wolf. Distilling nonlocality. *Phys. Rev. Lett.*, 102:120401, Mar 2009.

[GMDLT+13] Rodrigo Gallego, Lluis Masanes, Gonzalo De La Torre, Chirag Dhara, Leandro Aolita, and Antonio Acín. Full randomness from arbitrarily deterministic events. *Nat Commun*, 4, Oct 2013.

[HILL99] Johan Hastad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, Mar 1999.

[HR10] Peter Hoyer and Jibran Rashid. Optimal protocols for nonlocality distillation. *Phys. Rev. A*, 82:042118, Oct 2010.

[HRW10] Esther Hänggi, Renato Renner, and Stefan Wolf. Efficient device-independent quantum key distribution. In *Proceedings of the 29th Annual International Conference on Theory and Applications of Cryptographic Techniques*, EUROCRYPT'10, pages 216–234, 2010.

[HRW13] Esther Hänggi, Renato Renner, and Stefan Wolf. The impossibility of non-signaling privacy amplification. *Theoretical Computer Science*, 486(0):27–42, 2013.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, SFCS '88, pages 68–80, 1988.

[Kra49] Leon Gordon Kraft. A device for quantizing, grouping, and coding amplitude-modulated pulses. Master's thesis, Massachusetts Institute of Technology, 1949.

[LPSW07] Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter. Quantum nonlocality and beyond: Limits from nonlocal computation. *Phys. Rev. Lett.*, 99:180502, Oct 2007.

[MAG06] Lluis Masanes, Antonio Acín, and Nicolas Gisin. General properties of nonsignaling theories. *Phys. Rev. A*, 73:012112, Jan 2006.

[Mas09] Lluis Masanes. Universally composable privacy amplification from causality constraints. *Phys. Rev. Lett.*, 102:140501, Apr 2009.

[MY98] Dominic Mayers and Andrew Yao. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS '98, page 503, 1998.

[NW10] Miguel Navascués and Harald Wunderlich. A glance beyond the quantum model. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 466(2115):881–890, Jan 2010.

[O'D04] Ryan O'Donnell. Hardness amplification within NP. *Journal of Computer and System Sciences*, 69(1):68–94, Aug 2004.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Number ISBN 978-1-107-03832-5. Cambridge University Press, 2014.

[PAM+10] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe. Random numbers certified by Bell's theorem. *Nature*, 464(7291):1021–1024, Apr 2010.

[PPK+09] Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, Oct 2009.

[PR94] Sandu Popescu and Daniel Rohrlich. Nonlocality as an axiom. *Foundations of Physics*, 24(379), 1994.

[Ras12] Jibran Rashid. *Limits and Consequences of Nonlocality Distillation*. PhD thesis, University of Calgary, Apr 2012.

[Ren08] Renato Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6(01):1–127, 2008.

[RVW] Omar Reingold, Salil Vadhan, and Avi Wigderson. unpublished. http://windowsontheory.org/2012/02/21/no-deterministic-extraction-from-santha-vazirani-sources-a-simple-proof/.

[Sha49] Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[Sho09] Anthony J. Short. No deterministic purification for two copies of a noisy entangled state. *Phys. Rev. Lett.*, 102:180502, May 2009.

[Sus95] Leonard Susskind. The world as a hologram. *Journal of Mathematical Physics*, 36(11):6377–6396, 1995.

[SV84] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from slightly-random sources (extended abstract). In *FOCS*, pages 434–440, 1984.

[tH93] Gerardus 't Hooft. Dimensional reduction in quantum gravity. *arXiv:gr-qc/9310026*, 1993.

[vD99] Wim van Dam. Implausible consequences of superstrong nonlocality. *arXiv:quant-ph/0501159*, 1999.

[VV14] Umesh Vazirani and Thomas Vidick. Fully device-independent quantum key distribution. *Phys. Rev. Lett.*, 113:140501, Sep 2014.

[Yan07] Ke Yang. On the (im)possibility of non-interactive correlation distillation. *Theoretical Computer Science*, 382(2):157 – 166, 2007.