Zero-Error Information and Applications in Cryptography

Stefan Wolf
Département d'Informatique et de R.O.
Université de Montréal
E-mail: wolf@iro.umontreal.ca

Jürg Wullschleger
Département d'Informatique et de R.O.
Université de Montréal
E-mail: wullschj@iro.umontreal.ca

Abstract — **In analogy to the zero-error variant of the channel capacity, the *zero-error information* between two random variables is defined. We show that our definition is natural in the sense that the representation of the channel capacity with respect to mutual information carries over to the zero-error variants of the quantities. It is shown that the new notion, together with two operators introduced in the same context, namely the *common random variable* of two random variables and the *dependent part* of a random variable with respect to another, is useful for giving characterizations of the possibility of realizing cryptographic tasks—such as bit commitment, coin tossing, or oblivious transfer—from correlated pieces of information.**

I. Introduction

In cryptography, there are mainly two types of security: *Computational security* is based on the hardness of certain computational problems such as integer factoring or computing discrete logarithms and, hence, inherently dependent on assumptions on an adversary's capabilities as well as, up to now, on the hardness of the underlying problem. *Information-theoretic* or *unconditional security*, on the other hand, does not depend on unproven assumptions. However, certain impossibility results suggest that this type of security is generally less practical: Shannon [17] proved that perfectly secure encryption is possible only between parties sharing a secret key that is as long as the message, and important cryptographic functionalities such as coin tossing, bit commitment, oblivious transfer, or broadcast cannot be achieved in an unconditionally secure way from scratch [5, 12]. This pessimism can, however, often be relativized by showing that cryptographic tasks *can* be realized—in an unconditionally secure way—from information theoretic primitives as simple as a noisy communication channel or correlated pieces of information [13, 3, 7, 6].

In this paper, we consider the scenario where two parties know random variables X and Y, respectively, and ask ourselves under what conditions on P_{XY} this allows for achieving cryptographic goals such as bit commitment. As a preparation, we introduce a number of tools and notions for analyzing the distribution P_{XY} in this context, but that are also of independent interest. These are the *zero-error information* between two random variables, which relates to the "normal" mutual information in exactly the same way as the zero-error capacity to the usual channel capacity, and the *common random variable* between two random variables, as well as the *dependent part* of a random variable with respect to another.

II. Common Random Variables

The *common random variable* of X and Y is the largest random variable that two players Alice and Bob, knowing the random variables X and Y, respectively, can both generate.

Definition 1. Let X and Y be random variables with (disjoint) ranges $\mathcal{X}$ and $\mathcal{Y}$, distributed according to P_{XY}. Then $X \wedge Y$, the *common random variable* of X and Y, is constructed in the following way:

- Let G be the bipartite graph with vertex set $\mathcal{X} \cup \mathcal{Y}$, and such that two vertices $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ are connected by an edge if $P_{XY}(x, y) > 0$ holds.
- Let $f_X : \mathcal{X} \to 2^{\mathcal{X} \cup \mathcal{Y}}$ be the function that maps a vertex $v \in \mathcal{X}$ of G to the set of vertices in the connected component of G containing v. Let $f_Y : \mathcal{Y} \to 2^{\mathcal{X} \cup \mathcal{Y}}$ be defined analogously for vertices $w \in \mathcal{Y}$ of G.
- $X \wedge Y := f_X(X) = f_Y(Y)$.

Note that $X \wedge Y$ is symmetric—i.e., $X \wedge Y \equiv Y \wedge X$ [1]. Because of $X \wedge Y = f_X(X) = f_Y(Y)$, $X \wedge Y$ can be calculated both from X and from Y.

Example 1. Let A, B, and C be independent, and let $X = [A, B]$ and $Y = [B, C]$. Then $X \wedge Y \equiv B$.

Example 2. Let X and Y be two binary random variables such that $X \not\equiv Y$. Then $X \wedge Y$ is a constant.

Lemma 1 shows that $X \wedge Y$ is the "biggest" random variable that can be extracted both from X and from Y.

Lemma 1. *For all X, Y, and $\overline{C}$ for which there exist functions $\overline{f}_X$ and $\overline{f}_Y$ such that $\overline{C} = \overline{f}_X(X) = \overline{f}_Y(Y)$ holds, there exists a function g with $\overline{C} = g(X \wedge Y)$.*

Proof. Let us assume that such a function g does not exist. Then there must exist values x_1 and x_2 with $\overline{f}_X(x_1) \neq \overline{f}_X(x_2)$ but $f_X(x_1) = f_X(x_2)$. Hence, x_1 and x_2 are in the same connected component of the graph G from Definition 1. We can therefore find values x_1', x_2', and y with $\overline{f}_X(x_1') \neq \overline{f}_X(x_2')$, $P_{XY}(x_1', y) > 0$, and $P_{XY}(x_2', y) > 0$. This implies that there cannot exist a function $\overline{f}_Y$ with $\overline{C} = \overline{f}_X(X) = \overline{f}_Y(Y)$. □

[1] We say that two random variables A and B are *equivalent*, denoted by $A \equiv B$, if there exists a bijective function $g : \mathcal{A} \to \mathcal{B}$ such that $B = g(A)$ holds.

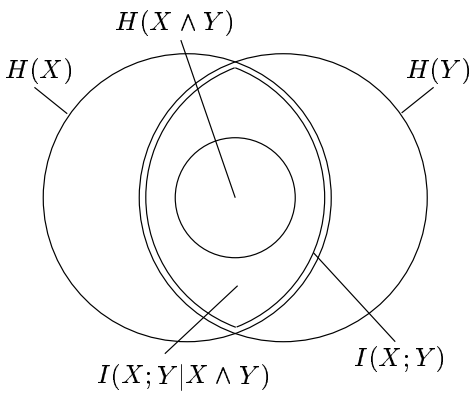

Figure 1: $I(X;Y) = H(X \wedge Y) + I(X;Y|X \wedge Y)$.

Lemma 2. *For all X and Y, we have*

$$I(X;Y) = H(X \wedge Y) + I(X;Y|X \wedge Y) \ .$$

Proof. Let $C = X \wedge Y$. Since $C = f_X(X) = f_Y(Y)$, we have $H(XC) = H(X)$ and $H(X|YC) = H(X|Y)$. Hence,

$$\begin{aligned} H(C) + I(X;Y|C) &= H(XC) - H(X|C) \\ &\quad + H(X|C) - H(X|YC) \\ &= H(X) - H(X|Y) \\ &= I(X;Y) \ . \end{aligned}$$

□

Corollary 1. *For all X and Y, we have*

$$H(X \wedge Y) \leq I(X;Y) \ .$$

Equality holds if and only if $I(X;Y|X \wedge Y) = 0$.

Lemma 3. *Let (X_1, Y_1) and (X_2, Y_2) be independent. Then*

$$(X_1X_2) \wedge (Y_1Y_2) \equiv (X_1 \wedge Y_1)(X_2 \wedge Y_2) \ .$$

Proof. We have $P_{X_1X_2Y_1Y_2}(x_1, x_2, y_1, y_2) > 0$ if and only if $P_{X_1Y_1}(x_1, y_1) > 0$ and $P_{X_2Y_2}(x_2, y_2) > 0$ because of $P_{X_1X_2Y_1Y_2} = P_{X_1Y_1}P_{X_2Y_2}$. Hence, we have $f_{X_1X_2}(x_1, x_2) = f_{X_1X_2}(x_1', x_2')$ if and only if $f_{X_1}(x_1) = f_{X_1}(x_1')$ and $f_{X_2}(x_2) = f_{X_2}(x_2')$. □

III. Zero-Error Information

In contrast to the Shannon information $I(X;Y)$, which is the information that X carries over Y (and *vice versa*) on average *with an arbitrarily small error*, the entropy of $X \wedge Y$ is the information that X has over Y (and Y over X) on average *without any error*. We define the zero-error information between X and Y.

Definition 2. Let X and Y be two random variables, distributed according to the joint distribution P_{XY}. The *zero-error information* between X and Y, denoted by $I_0(X;Y)$, is defined as

$$I_0(X;Y) := H(X \wedge Y) \ .$$

In the following we will show the connection between the zero-error information and the *zero-error capacity* of a channel.

Definition 3. [18, 11] Let $W = P_{Y|X}$ be a channel. Then

$$C_0(W) := \limsup_{n \to \infty} \frac{1}{n} \log_2 N(W, n)$$

is the *zero-error capacity* of W, where $N(W, n)$ stands for the maximal cardinality of a code of length n that is decodable without any error after having been transmitted over the channel W.

Theorem 1 shows that the zero-error capacity can be expressed in a natural way using zero-error information.

Theorem 1. *For all channels $W = P_{Y|X}$, we have*

$$C_0(W) = \lim_{n \to \infty} \max_{P_{X^n}} \frac{1}{n} I_0(X^n; Y^n) \ .$$

Proof. Let X^n be distributed such that $I_0(X^n; Y^n)$ is maximal, and let $C^n = X^n \wedge Y^n$. There exists $\overline{X}^n$ such that $\overline{X}^n \wedge Y^n \equiv C^n \equiv \overline{X}^n$ holds since X^n can be modified such that for every c^n there exists exactly one x^n with positive probability. We have $H(\overline{X}^n) = H(C^n) = I_0(X^n; Y^n)$. Because $H(\overline{X}^n)$ is maximal, all the $\overline{x}^n$ with positive probability have the same probability, hence, $H(\overline{X}^n) = H_0(\overline{X}^n)$. Because both the sender and the receiver are able to calculate $C^n \equiv \overline{X}^n$ and because $H(C^n)$ is maximal, $\overline{X}^n$ forms a code with maximal cardinality that the receiver can decode without error. We have $H_0(\overline{X}^n) = \log_2 N(W, n)$. Therefore, $\max_{P_{X^n}} I_0(X^n; Y^n)/n$ approaches $C_0(W)$ as n goes to infinity. □

Note that the (normal) capacity of a noisy channel can be written in the exact same way, using the mutual information instead of the zero-error information:

$$C(W) = \max_{P_X} I(X;Y) = \lim_{n \to \infty} \max_{P_{X^n}} \frac{1}{n} I(X^n; Y^n) \ .$$

IV. Dependent Parts

In this section we give the definition of the dependent part of a random variable with respect to another. The notion has already been introduced in [6] and independently in [7].

Definition 4. [6] Let X and Y be two random variables, and let $f(x) = P_{Y|X=x}$. The *dependent part of X from Y* is defined as $X \searrow Y := f(X)$.

The random variable $X \searrow Y$ is a function of X and takes on the value of the conditional probability distribution $P_{Y|X=x}$. Lemma 4 shows that all of X that is dependent on Y is included in $X \searrow Y$, i.e., more formally, $I(X;Y|X \searrow Y) = 0$ holds or, equivalently, X, $X \searrow Y$, and Y form a Markov chain[2].

Lemma 4. [6] *For all X and Y, we have*

$$X \longleftrightarrow (X \searrow Y) \longleftrightarrow Y \ .$$

Proof. Let $K = f(X) = X \searrow Y$. For all $x \in \mathcal{X}$ and $k = f(x)$, we have $P_{Y|X=x,K=k} = P_{Y|K=k}$. □

On the other hand, there does not exist a random variable with the same properties that is "smaller" than $X \searrow Y$.

Lemma 5. *Let X, Y, and $\overline{K}$ be random variables such that there exists a function $\overline{f}$ with $\overline{K} = \overline{f}(X)$ and $X \longleftrightarrow \overline{K} \longleftrightarrow Y$. Then there exists a function g with $X \searrow Y = g(\overline{K})$.*

Proof. $X \longleftrightarrow \overline{K} \longleftrightarrow Y$ implies $P_{Y|X\overline{K}} = P_{Y|\overline{K}}$. Because of $\overline{K} = \overline{f}(X)$, we have $P_{Y|X=x} = P_{Y|\overline{K}=\overline{k}}$ for all x and $\overline{k} = \overline{f}(x)$. Hence, we have $P_{Y|X=x_1} = P_{Y|X=x_2}$ for all x_1 and x_2 with $\overline{f}(x_1) = \overline{f}(x_2)$, and therefore there exists a function g such that $X \searrow Y = g(\overline{K})$. □

Lemma 6. *For random variables X and Y, we have*

$$H(Y|X \searrow Y) = H(Y|X) \ .$$

Proof. Let $K = X \searrow Y$. Because of $K = f(X)$ and of $X \longleftrightarrow K \longleftrightarrow Y$, we have $H(Y|X) = H(Y|XK) = H(Y|K)$. □

Corollary 2. *For all X and Y, we have*

$$I(X;Y) = H(X \searrow Y) - H(X \searrow Y|Y) = I(X \searrow Y;Y) \ .$$

Proof. We have $I(X \searrow Y;Y) = H(Y) - H(Y|X \searrow Y) = H(Y) - H(Y|X) = I(X;Y)$. □

Corollary 3 follows immediately from Corollary 2.

Corollary 3. *For all X and Y, we have*

$$I(X;Y) \leq H(X \searrow Y) \ .$$

Equality holds if and only if $H(X \searrow Y|Y) = 0$.

Corollary 4. *For all X and Y, we have*

$$H(X|Y) = H(X|X \searrow Y) + H(X \searrow Y|Y) \ .$$

[2] A sequence of three random variables A, B, C forms a *Markov chain*, denoted by $A \longleftrightarrow B \longleftrightarrow C$, if $I(A;C|B) = 0$ holds or, equivalently, if we have $P_{C|AB}(c,a,b) = P_{C|B}(c,b)$ for all $(a,b,c) \in \mathcal{A} \times \mathcal{B} \times \mathcal{C}$.

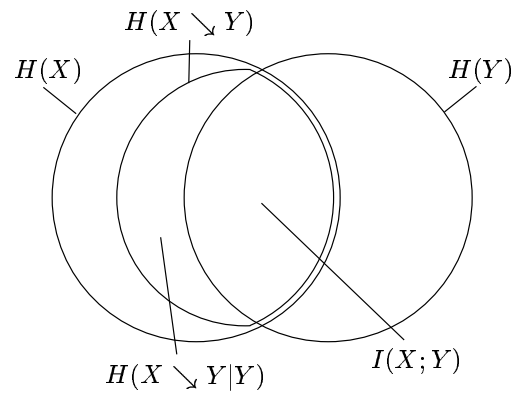

Figure 2: $H(X \searrow Y) = I(X;Y) + H(X \searrow Y|Y)$.

Proof. Let $K = X \searrow Y$. From Corollary 2 we get

$$\begin{aligned} H(X|Y) &= H(X) - I(X;Y) \\ &= H(XK) - H(K) + H(K|Y) \\ &= H(X|K) + H(K|Y) \ . \end{aligned}$$

□

Lemma 7. *Let (X_1, Y_1) and (X_2, Y_2) be independent. Then*

$$(X_1X_2) \searrow (Y_1Y_2) \equiv (X_1 \searrow Y_1)(X_2 \searrow Y_2) \ .$$

Proof. We have $P_{Y_1Y_2|(X_1,X_2)=(x_1,x_2)} = P_{Y_1|X_1=x_1} P_{Y_2|X_2=x_2}$. Hence, we have $P_{Y_1Y_2|(X_1,X_2)=(x_1,x_2)} \neq P_{Y_1Y_2|(X_1,X_2)=(x_1',x_2')}$ if and only if either $P_{Y_1|X_1=x_1} \neq P_{Y_1|X_1=x_1'}$ or $P_{Y_2|X_2=x_2} \neq P_{Y_2|X_2=x_2'}$ holds. □

The random variable $K = X \searrow Y$ is the part of X that someone who knows Y can verify to be correct. This was shown in [6] and used for deriving the exact condition under which so-called *pseudo-signatures* and *broadcast* among three players are possible among parties sharing correlated randomness. Lemma 8 shows that every random variable $\overline{K}$ that a player knowing X can generate and that has the same joint distribution with Y as the actual K must in fact be *identical* with K. Lemma 9 shows that, on the other hand, from K, a random variable $\overline{X}$ can be constructed which has the same joint distribution with Y as X.

Lemma 8. [6] *Let X, K, $\overline{K}$, and Y be random variables with $K = X \searrow Y$, $Y \longleftrightarrow X \longleftrightarrow \overline{K}$, and $P_{KY} = P_{\overline{K}Y}$. Then we have $\overline{K} = K$.*

Lemma 9. [6] *Let X and Y be random variables, and let $K = X \searrow Y$. Then there exists a channel $P_{\overline{X}|K}$—which is equal to $P_{X|K}$—such that $P_{XY} = P_{\overline{X}Y}$ holds, where $P_{\overline{X}Y} = \sum_k P_{KY} P_{\overline{X}|K}$.*

V. The Connection Between $X \wedge Y$ and $X \searrow Y$

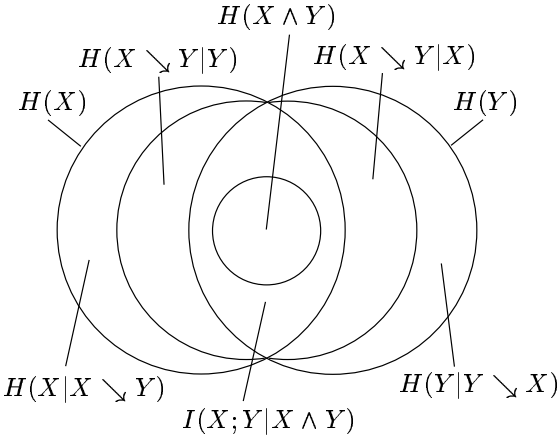

Figure 3: $H(XY)$ can be split into six regions.

Corollary 5. *For all X and Y, we have*

$$\begin{aligned} H(XY) &= H(X \wedge Y) + I(X;Y|X \wedge Y) \\ &\quad + H(X \searrow Y|Y) + H(Y \searrow X|X) \\ &\quad + H(X|X \searrow Y) + H(Y|Y \searrow X)\,. \end{aligned}$$

Proof. This follows from $H(XY) = I(X;Y) + H(X|Y) + H(Y|X)$ and from Lemma 2 and Corollary 4. □

Lemma 10. *For all X and Y there exists a function g such that $X \wedge Y = g(X \searrow Y)$ holds.*

Proof. Let $C = X \wedge Y = f_X(X) = f_Y(Y)$ and $K = X \searrow Y = f(X)$. Let x_1 and x_2 be two values X can take on with $f(x_1) = f(x_2)$. We have $P_{Y|X=x_1} = P_{Y|X=x_2}$, which means that from Y one cannot distinguish these two cases, and therefore we must have $f_X(x_1) = f_X(x_2)$. It follows that there exists a function g with $X \wedge Y = g(X \searrow Y)$. □

Theorem 2 gives alternative characterizations of the fact that the entire mutual information between two random variables is *noiseless*. As a preparation, we prove two lemmas.

Lemma 11. *For all X and Y, we have $I(X;Y) = H(X \wedge Y)$ if and only if $X \wedge Y \equiv X \searrow Y$.*

Proof. Because of Corollary 1 it follows from $I(X;Y) = H(X \wedge Y)$ that $I(X;Y|X \wedge Y) = 0$ holds, hence, $X \longleftrightarrow X \wedge Y \longleftrightarrow Y$. Using Lemmas 5 and 10, we get $X \wedge Y \equiv X \searrow Y$. Using Corollaries 1 and 3, it follows directly from $X \wedge Y \equiv X \searrow Y$ that $I(X;Y) = H(X \wedge Y)$ holds. □

Lemma 12. *For all X and Y, we have $I(X;Y) = H(X \searrow Y)$ if and only if $X \wedge Y \equiv X \searrow Y$.*

Proof. Because of Corollary 3, $I(X;Y) = H(X \searrow Y)$ implies $H(X \searrow Y|Y) = 0$, hence, $X \searrow Y$ is a function of Y. Using Lemmas 1 and 10, we get $X \wedge Y \equiv X \searrow Y$. Using Corollaries 1 and 3, it follows directly from $X \wedge Y \equiv X \searrow Y$ that $I(X;Y) = H(X \searrow Y)$ holds. □

Theorem 2. *For two random variables X and Y, the following properties are equivalent.*

1) $I(X;Y) = I_0(X;Y)$

2) $H(X \wedge Y) = I(X;Y)$

3) $I(X;Y|X \wedge Y) = 0$

4) $H(X \searrow Y) = I(X;Y)$

5) $H(Y \searrow X) = I(X;Y)$

6) $H(X \searrow Y|Y) = 0$

7) $H(Y \searrow X|X) = 0$

8) $X \wedge Y \equiv X \searrow Y$

9) $X \wedge Y \equiv Y \searrow X$

Proof. Follows directly from Definition 2, Corollaries 1 and 3, and Lemmas 11 and 12. □

VI. Bit Commitment

Bit commitment was introduced in [2]—together with distributed coin-flipping among two players (the topic of Section VII).

Bit commitment is a cryptographic primitive where at some point Alice has to commit herself to a value of her choice that Bob does not get to know until later, when Alice opens her commitment to Bob. On the other hand, it is guaranteed that Alice cannot reveal any other value than the committed one. Bit commitments are used in identification schemes, zero-knowledge proofs, and general multi-party computation.

Bit commitment based on common randomness was introduced in [15]. In [7], the *commitment capacity* of two correlated random variables is defined. It is the supremum of all rates (bits per instance) that can be achieved with an arbitrarily small error. Note that in that model, *string commitment* is considered: All bits are committed to and opened *simultaneously*.

Theorem 3. [7] *The commitment capacity of X and Y, where the commiter holds X and the verifier holds Y, is $H(X \searrow Y|Y)$.*

Corollary 6. *Bit commitment between two parties, where the commiter holds X and the verifier holds Y, is possible if and only if $I(X;Y) > I_0(X;Y)$.*

The algorithm of [7] is based on a code that has been introduced in [19]. We will present here—briefly and without any proofs—a simpler protocol which only relies on standard coding techniques and privacy amplification [1].

Let Alice and Bob know $X^n = X_1, \ldots, X_n$ and $Y^n = Y_1, \ldots, Y_n$, respectively. Assume that Alice wants to commit to some string D.

To construct the commitment to D, Alice uses the part of her X^n that Bob can verify. Lemmas 8 and 9 tell us that this part

is $K^n = X^n \searrow Y^n$. In order to ensure that Bob has no information about Alice's secret, she applies privacy amplification on K^n, using some additional randomness R, and gets a key L about which Bob has no information. Her commitment consists of $C := L \oplus D$ and the additional randomness R.

In the opening phase, Alice has to send K^n to Bob. Bob checks whether the values (K^n, Y^n) form a typical sequence. However, a dishonest Alice might still be able to change a sub-linear amount of values in K^n. To ensure that this is not possible, she has to send to Bob some additional parity-checks P in the commitment phase. The linear code for the parity-checks must have a minimal distance which is such that there do not exist two sequences k_1 and k_2 that have the same parity-checks and for which there exists y such that both (k_1, y) and (k_2, y) are typical. Bob checks whether P are valid parity-checks for K^n. If so, he knows that K^n is valid, and he can extract L using R and get $D = L \oplus C$.

Of course, these additional parity-check bits lower the amount of extractable randomness and, therefore, the commitment rate. However, this loss can be made arbitrarily small, since Alice is only able to cheat for a sub-linear amount of values. Hence, this scheme approaches the rate $H(X \searrow Y|Y)$ as n goes to infinity.

Instead of the parity-checks of a linear code, *universal hashing* can be used to ensure that Alice cannot change a sub-linear amount of values in K^n: Bob randomly chooses an element h of a universal class of hash functions $\mathcal{H}$ and sends it to Alice. Instead of the parity-checks P of K^n, Alice sends the value $h(K^n)$ to Bob. Alice's cheating probability is exponentially small in the length of the hash value.

VII. Distributed Coin Tossing and Coin Extraction

In distributed coin-flipping, Alice and Bob have to agree on two equal random bit-strings such that neither Alice nor Bob can bias the probability distribution if the other party plays honestly. Distributed coin-flipping is used in any sort of two-player games.

For finding the number of coins that can be tossed based on some additional randomness, we need to distinguish between two different kinds of "coins": Coins that come *directly* from the randomness—we will call this *coin extraction*—, and coins that can be *tossed later*, during the execution of a protocol—here, we will talk about *coin tossing*. It depends on the particular application whether extracted coins are sufficient or whether the coins need to be tossed at a specific time (and should not be known beforehand to any of the parties). Furthermore, we also need to distinguish between coins *without any error*, i.e., bias, and coins with an arbitrarily small error $\varepsilon > 0$.

Lemma 13. *Any protocol for coin extraction or coin tossing without any error—based only on noiseless communication and additional randomness—can be transformed into a "protocol" not using any communication.*

Proof. (Sketch) Let us assume that we have an interactive protocol between Alice and Bob such that both Alice and Bob get a coin without any error. Let Alice be the last to send a message. Let us assume that there exists an execution of the protocol—where this last message is m and the shared coin is x—such that there exists a message m' that Alice could send to Bob who would then end up with the coin flip $x' \neq x$. This means that Alice is able to bias the coin by guessing m' and sending it to Bob. Since this must be impossible, for no possible executions does there exist such a message m', which means that Bob knows x already before receiving the last message from Alice. The same argument can now be repeated for all messages sent. □

Corollary 7. *Coin tossing without any error based only on noiseless communication and additional randomness is impossible.*

Proof. Otherwise, Lemma 13 would imply that a protocol without any communication would exist, which is obviously impossible. □

Theorem 4. *The rate at which coins without any error can be extracted from X and Y is $I_0(X, Y)$.*

Proof. Both Alice and Bob can calculate $X \wedge Y$. It was shown in [9] that random coins without any error can be extracted from $X \wedge Y$ with a rate approaching $H(X \wedge Y) = I_0(X; Y)$.

Lemma 13 implies that a protocol for coin extraction without any error can be assumed not to use any communication. Lemma 1 implies that for any n, not more than $H(X^n \wedge Y^n) = nH(X \wedge Y)$ bits can be extracted. Hence, the rate cannot exceed $I_0(X; Y)$. □

If we allow an arbitrarily small error, any bit-commitment scheme can be used for coin-tossing [2]. Recently, it has been shown that coin tosses can be multiplied, i.e., if a few coin tosses are possible, than any amount of coins can be tossed [7]. Note that for the realization and multiplication of coin tosses, *extracted coins* are *not* sufficient.

Theorem 5. [7] *Coin tossing with an arbitrarily small error is possible if and only if $H(X \searrow Y|Y) > 0$ holds. Then the rate at which such coin tosses can be generated is infinite.*

Corollary 8. *Coin tossing with an arbitrarily small error is possible if and only if $I(X; Y) > I_0(X; Y)$ holds. Then the rate is infinite.*

Theorem 6. *The rate at which coins can be extracted is equal to $I(X, Y)$ if $I(X, Y) = I_0(X; Y)$, and infinite otherwise.*

Proof. If $I_0(X, Y) = I(X; Y)$, no coin-tossing is possible. Theorem 4 implies that coins can be extracted at a rate approaching $I(X; Y)$, which is obviously also an upper bound.

If $I_0(X, Y) < I(X; Y)$ holds, it follows from Theorem 2 that $H(X \searrow Y|Y) > 0$, hence, coin-tossing is possible at an infinite rate. □

Corollary 9. *Coin extraction with communication is possible if and only if $I(X, Y) > 0$.*

VIII. Oblivious Transfer

Oblivious transfer goes back to [14]. It is a primitive where Alice sends two values of which Bob gets to know one of his choice. The protocol ensures that Bob does not get any information on the other value, and that Alice does not get any information about which value Bob has chosen. It was shown in [10] that with oblivious transfer, *any* computation among two players can be achieved.

In [4] it was shown that almost any noisy channel can be used to implement oblivious transfer. We will use this result to state the exact condition under which oblivious transfer is possible when some correlated randomness is given.

Lemma 14. *Let X and Y be two random variables. We have $H(X \searrow Y|Y) > 0$ if and only if there exist x_1, x_2, and y such that $P_{XY}(x_1, y) > 0$, $P_{XY}(x_2, y) > 0$, and $P_{Y|X=x_1} \neq P_{Y|X=x_2}$ hold.*

Proof. Let $K = X \searrow Y$. We have $H(K|Y) > 0$ if and only if there exists y with $H(K|Y = y) > 0$. It follows that there must exist two values k_1 and k_2 such that $P_{KY}(k_1, y) > 0$ and $P_{KY}(k_2, y) > 0$ hold. Choosing x_1 that is mapped to k_1 and x_2 that is mapped to k_2 concludes the proof. □

Theorem 7. *Oblivious transfer between two players knowing X and Y, respectively, is possible if and only if $I(X;Y) > I_0(X;Y)$ holds.*

Proof. (Sketch) Let Alice have $X_1X_2 \cdots X_n$ and Bob know $Y_1Y_2 \cdots Y_n$. Alice and Bob simulate a channel in the following way: For all i, Alice erases the values X_i with a certain probability, such that all x occur with the probability of the least probable x. On input value x, Alice sends Bob the index i of the first value X_i with $X_i = x$. Bob outputs the value Y_i. Note that the index i does not carry any information about the value x.

Lemma 14 and Theorem 2 imply that if $I(X;Y) > I_0(X;Y)$ holds, then the resulting channel satisfies the condition stated in [4] to allow for achieving oblivious transfer. □

Note that the same result was found independently in [8].

IX. Concluding Remarks

We have defined new information-theoretic notions such as the *common random variable* of two random variables, the *zero-error information* between them, and the *dependent part* of a random variable with respect to another. An important property is the fact that the "normal" mutual information between two random variables X and Y *exceeds* their zero-error information: In this case, two parties knowing X and Y, respectively, can realize, in an unconditionally secure way, cryptographic tasks such as bit commitment or oblivious transfer. This result is another step towards making unconditional cryptographic security—clearly the most desirable type—more practical.

We suggest as an open problem to find the exact rate at which oblivious transfer can be generated from repeated realizations of random variables X and Y.

References

[1] Charles Bennett, Gilles Brassard, Claude Crepeau, and Ueli Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, Vol. 41, 1995.

[2] Manuel Blum. Coin flipping by telephone: A protocol for solving impossible problems. In *Proceedings of the 24th IEEE Computer Conference*, pages 133–137, 1982.

[3] Claude Cr´epeau. Efficient cryptographic protocols based on noisy channels. In *Advances in Cryptology: EUROCRYPT '97, Lecture Notes in Computer Science*, Springer-Verlag, 1997.

[4] Claude Crepeau, Kirill Morozov, and Stefan Wolf. Oblivious transfer from any noisy channel. To appear in *Proceedings of SCN'04*, Springer-Verlag, 2004.

[5] Ivan Damgård, Joe Kilian, and Louis Salvail. On the (im)possibility of basing oblivious transfer and bit commitment on weakened security assumptions. In *Advances in Cryptology: CRYPTO '99*, volume 1592 of *Lecture Notes in Computer Science*, pages 56–73. Springer-Verlag, 1999.

[6] Matthias Fitzi, Stefan Wolf, and Jürg Wullschleger. Pseudo-signatures, broadcast, and multi-party computation from correlated randomness. In *Advances in Cryptology: CRYPTO '04*, Lecture Notes in Computer Science. Springer-Verlag, 2004.

[7] Hideki Imai, Jörn Müller-Quade, Anderson Nascimento, and Andreas Winter. Rates for bit commitment and coin tossing from noisy correlation. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT '04)*, 2004.

[8] Hideki Imai, Anderson Nascimento, and Andreas Winter. Oblivious transfer from any genuine noise. Unpublished manuscript, 2004.

[9] Ari Juels, Markus Jakobsson, Elizabeth Shriver, and Bruce K. Hillyer. How to turn loaded dice into fair coins. *IEEE Transactions on Information Theory*, Vol. 46, 2000.

[10] Joe Kilian. Founding cryptography on oblivious transfer. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, pages 20–31, 1988.

[11] J´anos Körner and Alon Orlitsky. Zero-error information theory. *IEEE Transactions on Information Theory*, Vol. 44, 1998.

[12] Leslie Lamport, Robert Shostak, and Marshall Pease. The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3):382–401, 1982.

[13] Ueli Maurer. Secret key agreement by public discussion. *IEEE Transaction on Information Theory*, 39(3):733–742, 1993.

[14] Michael O. Rabin. How to exchange secrets by oblivious transfer. Technical Report TR-81, Harvard Aiken Computation Laboratory, 1981.

[15] Ronald L. Rivest. Unconditionally secure commitment and oblivious transfer schemes using private channels and a trusted initializer. Unpublished manuscript, 1999.

[16] Claude E. Shannon. A mathematical theory of communication. *Bell System Tech. Journal*, 27:379–423, 623–656, 1948.

[17] Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4), 1949.

[18] Claude E. Shannon. The zero-error capacity of a noisy channel. *IEEE Transactions on Information Theory*, 1956.

[19] Aaron D. Wyner. The wire-tap channel. *Bell System Tech. Journal*, 54:1355–1387, 1975.