Information Indivisibility Results

Sayir, Jossy; Land, Ingmar; grant, Alex

Published in:
IEEE International Zurich Seminar on Communications, 2008

DOI (link to publication from Publisher):
10.1109/IZS.2008.4497279

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Information Indivisibility Results

Jossy Sayir*, Ingmar Land†‡, and Alex Grant†
Emails: sayir@ftw.at, {ingmar.land,alex.grant}@unisa.edu.au
* Forschungszentrum Telekommunikation Wien (ftw.),
Donau-City-Str. 1, A1220 Vienna, Austria
† Institute for Telecommunications Research,
University of South Australia, Mawson Lakes, South Australia
‡ Department of Electronic Systems,
Aalborg University, Denmark

***Abstract*—We derive a theorem stating that it is not possible to generate two independent observations of a binary random variable such that each observation gives only partial information about the binary variable, but the two observations jointly determine the binary variable exactly. We illustrate this theorem with examples and elaborate on applications of the theorem to cryptography and to determining the EXIT function of a zero-error code.**

I. Introduction

The following cryptographic scenario motivates the results presented in this paper. Despite never having met in person and always communicating through secret channels, Alice and Bob eventually became good friends and, together, discovered the answer to the fundamental question of information theory, cryptography, and everything. The exact nature of this question is outside the scope of this paper, but for the sake of the story let us assume that its answer is simply 0 or 1. Alice and Bob agreed to keep this answer secret and never to divulge it to anyone during their lifetime. On their deathbeds, Alice and Bob gave their disciples Alfred and Beatrix a one thousand bit clue each. Each clue gives partial information about the precious answer they had discovered, but the exact answer can only be reconstituted by combining the two clues. Note that Alice and Bob lived at opposite ends of the globe, had disabled any secret channels between them at this stage and thus could not collaborate in any way to generate the two clues, nor had they had the foresight of generating common randomness in advance to help in the process. The present paper claims a fundamental flaw in Alice and Bob's plan: according to the results that will be presented, re-constituting the exact answer based on the combined clues is a mathematical impossibility, unless the exact answer is deducible from at least one of the two clues individually.

In Section II, we will introduce the main result of the paper, a theorem stating that it is not possible to generate independent observations of a binary random variable that each give only partial information but together essentially determine the variable. Section III will provide a proof of this result. In Section IV, we will show examples and discuss applications of the result, in particular its use for determining part of the EXIT curve of a zero-error channel code.

II. The Bit Indivisibility Theorem

Let the random variable X carry the binary secret and Y_1 and Y_2 be the two clues that grant access to the secret. By requiring that Y_1 and Y_2 be generated independently, we mean that the two variables give no information about each other when X is known, or in other words that the conditional mutual information $I(Y_1;Y_2|X)$ be zero. This is equivalent to the condition that Y_1 and Y_2 be independent given X, or that Y_1-X-Y_2 form a Markov chain. Generating two independent clues that jointly grant access to a binary secret is an instance of the general case treated in the following theorem:

Theorem 1 (Bit Indivisibility Theorem): Let the random variables $Y_1 - X - Y_2$ form a Markov chain, where X is defined over the alphabet $\{0,1\}$. Then

$$H(X|Y_1Y_2)=0 \iff H(X|Y_1)=0 \text{ OR } H(X|Y_2)=0. \tag{1}$$

In other words, it is not possible to divide the information about a binary random variable X into two independent observations Y_1 and Y_2, such that each observation gives only partial information about X but the two observations jointly give full information about X.

Note that although the theorem is stated for two clues/observations Y_1 and Y_2, either of these could be a vector, so the theorem can be applied recursively to tackle any collection of independent observations $Y_1, Y_2, Y_3, \ldots$.

In the next section, we will provide a proof of this result, then show examples and applications in the following section.

III. Proof of the bit indivisibility theorem

The "$\Leftarrow$" of Theorem 1 can be proved using the chain rule of entropies,

$$H(X|Y_1Y_2) \le H(X|Y_1) \text{ and } H(X|Y_1Y_2) \le H(X|Y_2), \tag{2}$$

and therefore $H(X|Y_1Y_2)=0$ if either $H(X|Y_1)=0$ or $H(X|Y_2)=0$. The proof of the "$\Rightarrow$" is considerably longer and will occupy the rest of this section.

Let X, Y_1, Y_2 be random variables as specified by the theorem. We have

$$H(X|Y_1Y_2)=\sum_{y_1y_2} P(y_1y_2)H(X|Y_1Y_2=y_1y_2) \tag{3}$$

so that $H(X|Y_1Y_2)$ can be zero only if $H(X|Y_1Y_2 = y_1y_2) = 0$ for all y_1, y_2 such that $P(y_1y_2) > 0$. Note that we can write $P(y_1y_2) = P(y_1)P(y_2|y_1) = P(y_2)P(y_1|y_2)$ so that $P(y_1y_2) > 0$ implies $P(y_1) > 0$ and $P(y_2) > 0$.

Using $h(.)$ to denote the binary entropy function, we can now proceed as described in Equations 4 to 7 below, where we have used the Markov chain property in the last step. In order for $H(X|Y_1Y_2 = y_1y_2)$ to be zero, the argument of $h(.)$ must be either zero or one, which is only possible if

$$P(X = 0)P(y_1|X = 0)P(y_2|X = 0) = 0 \tag{8}$$

or if

$$P(X = 1)P(y_1|X = 1)P(y_2|X = 1) = 0. \tag{9}$$

If $P(X = 0) = 0$ or $P(X = 1) = 0$, then $H(X) = 0$, which implies that $H(X|Y_1Y_2)$, $H(X|Y_1)$ and $H(X|Y_2)$ are all zero, and the theorem holds. Otherwise, we see that either $P(y_1|X = 0)$, $P(y_2|X = 0)$, $P(y_1|X = 1)$, or $P(y_2|X = 1)$ must be zero. Without loss of generality, let us assume that the first of these probabilities is zero. Then,

$$\begin{aligned} H(X|Y_1 = y_1) &= h(P(X = 0|y_1)) \\ &= h\left(\tfrac{P(X=0)P(y_1|X=0)}{P(X=0)P(y_1|X=0)+P(X=1)P(y_1|X=1)}\right) \\ &= 0. \end{aligned} \tag{10}$$

The same applies to the other three probabilities, so we conclude that $H(X|Y_1Y_2 = y_1y_2) = 0$ if and only if $H(X) = 0$, $H(X|Y_1 = y_1) = 0$ or $H(X|Y_2 = y_2) = 0$.

We already showed that the theorem holds if $H(X) = 0$. If $H(X|Y_1 = y_1) = 0$ for all y_1 such that $P(y_1) > 0$, then $H(X|Y_1) = 0$ and the theorem holds as well. The same is true if $H(X|Y_2 = y_2) = 0$ for all y_2 such that $P(y_2) > 0$. On the other hand, if there exists a y_1 such that $H(X|Y_1 = y_1) > 0$ and a y_2 such that $H(X|Y_2 = y_2) > 0$, then $H(X|Y_1Y_2 = y_1y_2) > 0$. In other words, we cannot zero $H(X|Y_1Y_2)$ by setting $H(X|Y_1 = y_1) = 0$ for some y_1 and setting $H(X|Y_2 = y_2) = 0$ for some y_2. □

We show an example that illustrates the cases considered at the end of the proof. Let X, Y_1 and Y_2 be connected through two Z-channels as illustrated in Figure 1. In this case, $H(X|Y_1 = 1) = H(X|Y_2 = 0) = 0$ but $H(X|Y_1 = 0) > 0$ and $H(X|Y_2 = 1) > 0$. We have $P_{Y_1Y_2}(1, 0) = 0$. But $P_{Y_1Y_2}(0, 1) \neq 0$ and $H(X|Y_1Y_2)$ can only be zero if $H(X|Y_1Y_2 = 01) = 0$, which in turn implies that one of the diagonal connections of the Z-channels must have probability zero, effectively turning it into a noiseless channel as stipulated by the theorem. Interestingly, the joint channel from X to Y_1Y_2 can be seen as an erasure channel, which is symmetric if the probabilities on the diagonal links of the Z-channels are equal, and noiseless if and only if the probability of an erasure given at least one of the input symbols is zero, as illustrated in Figure 2.

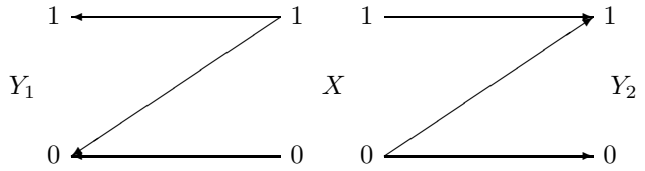

Fig. 1. Two parallel Z-channels

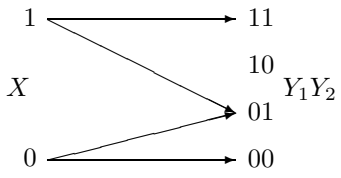

Fig. 2. Equivalent channel for two parallel Z channels

IV. Discussion, examples and applications

If the Markov condition is lifted, it is easy to see that there is a trivial solution to Alice and Bob's problem: let Y_1 be a uniformly distributed binary random variable independent of X and let $Y_2 = X + Y_1$ where the addition is taken modulo 2, i.e., in GF(2). Then we have $H(X|Y_1) = H(X|Y_2) = 1$ but $H(X|Y_1Y_2) = 0$, i.e., the two one-bit clues Y_1 and Y_2 give no information at all about X individually, but together they determine X exactly. However, Bob needs to know the Y_1 generated by Alice in order to generate the corresponding Y_2, so the two keys are not generated independently in this scenario.

A simple example also shows that the theorem does not apply to non-binary X in its current form. Consider the case of a uniformly distributed quaternary random variable X. The four possible values of X can be written as two binary digits. Let Y_1 be the first of these digits, transmitted through a

---

$$\begin{aligned} H(X|Y_1Y_2 = y_1y_2) &= h\left(P(X = 0|y_1y_2)\right) && (4)\\ &= h\left(\frac{P(X = 0, y_1, y_2)}{P(y_1y_2)}\right) && (5)\\ &= h\left(\frac{P(X = 0)P(y_1|X = 0)P(y_2|y_1, X = 0)}{\sum_x P(xy_1y_2)}\right) && (6)\\ &= h\left(\frac{P(X = 0)P(y_1|X = 0)P(y_2|X = 0)}{P(X = 0)P(y_1|X = 0)P(y_2|X = 0) + P(X = 1)P(y_1|X = 1)P(y_2|X = 1)}\right), && (7) \end{aligned}$$

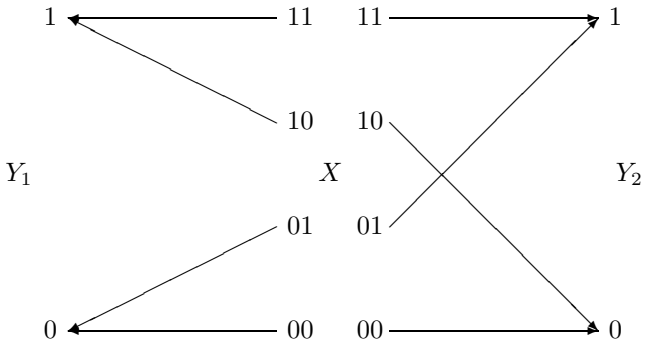

Fig. 3. Quaternary divisibility scenario

deterministic noiseless channel, and Y_2 be the second digit, as illustrated in Figure 3. Then we have

$$I(Y_1; Y_2|X) = H(Y_1|X) - H(Y_1|XY_2) = 0 - 0 = 0,$$

so $Y_1 - X - Y_2$ form a Markov chain as required by the theorem. However, $H(X|Y_1) = 1$ bit, $H(X|Y_2) = 1$ bit, while $H(X|Y_1Y_2) = 0$, which shows that an equivalent theorem would not hold for this non-binary random variable X.

Beyond the anecdotal application to cryptography described in the introduction, the theorem has consequences for a variety of disciplines. In physics, the theorem implies that when measuring a binary quantity (e.g., the spin of an electron), repeated independent measurements can only determine the quantity of interest beyond doubt if one of the measurements does so on its own. In decision theory, the theorem implies that a binary decision (e.g., whether to "buy" or to "sell" a financial product on a stock exchange) based on a collection of independent criteria can only be unambiguous if one of the criteria determines the decision without ambiguity.

A further telecommunications-related application of the theorem is to determine the EXtrinsic Information Transfer (EXIT) function [1] of a zero-error code. "Zero-error" here refers to the ability of the code to provide a vanishing probability of error for channels whose capacity lies above a certain threshold. This can be interpreted in the asymptotic regime, in which case we are considering the limiting case of a "good code", or a "capacity-achieving" family of codes. EXIT charts are normally used for code design, where the EXIT functions of code components (check nodes and variable nodes for LDPC codes, or component convolutional codes for turbo codes) are designed to optimize the performance of the code. If a zero-error code is to be used itself as a component within an iterative setup, for example within a turbo equalization scheme, then the overall EXIT function of the code becomes relevant. There have been claims (e.g., [2]) that this EXIT function is a step function as indicated in Figure IV, whose value remains zero up to an a-priori mutual information corresponding to the threshold/capacity of the code, and becomes 1 beyond that threshold. While it is not trivial to prove that the mutual information is zero below

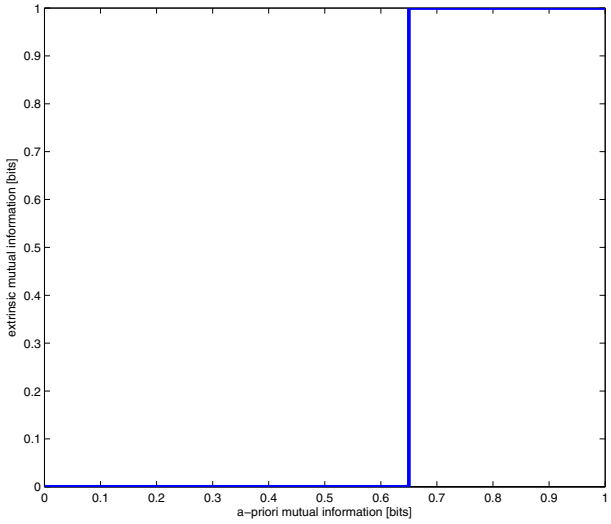

Fig. 4. Claimed EXIT function of a capacity-achieving code

the threshold, we can easily show that the mutual information is 1 above the threshold using the bit indivisibility theorem.

Let $X_1, X_2, \ldots$ be the sequence of code digits and $Y_1, Y_2, \ldots$ be the sequence of associated channel output observations, where the channel is assumed to be memoryless. We write $Y_{\setminus i}$ for the sequence of all but the i-th observation. Due to the memoryless nature of the channel, $Y_{\setminus i} - X_i - Y_i$ forms a Markov chain. We assume a binary code, so X_i is a binary random variable. Since the code is error-free when we are operating above the threshold, we know that $I(X_i; Y_1Y_2\ldots) = 1$, or in other words $H(X_i|Y_1Y_2\ldots) = 0$, i.e., the sequence of observations essentially determines every code digit. Therefore, the conditions of the theorem apply, and unless the channel is noiseless, we know that $H(X_i|Y_i) \neq 0$, which in turn implies that $H(X_i|Y_{\setminus i}) = 0$ and

$$I(X_i; Y_{\setminus i}) = 1, \tag{11}$$

which is the extrinsic mutual information plotted in the EXIT function above the threshold.

ACKNOWLEDGMENT

Part of this work was performed during a visit sponsored by the European Network of Excellence NEWCOM and by the Australian Research Council's (ARC) research network ACoRN (RN0459498). The work was also supported by the ARC Discovery Grant DP0663567 and by NEWCOM++. ftw. is a research center within the Austrian goverment's COMET funding scheme.

REFERENCES

[1] S. ten Brink, *Convergence behavior of iteratively decoded parallel concatenated codes,* IEEE Trans. Commun., vol. 49, no. 10, pp. 1727-1737, Oct. 2001.

[2] M. Peleg, A. Sanderovich and S. Shamai (Shitz), *On Extrinsic Information of Good Binary Codes Operating on Gaussian Channels*, European Trans. on Telecom. (ETT), Vol. 18, No. 2, pp. 133-139, 2007.