On Oblivious Transfer Capacity

Rudolf Ahlswede* and Imre Csiszár[1],**

[1] Rényi Institute of Mathematics, P.O. Box 127, Budapest, Hungary
csiszar.imre@renyi.mta.hu

Abstract. Upper and lower bounds to the oblivious transfer (OT) capacity of discrete memoryless channels and multiple sources are obtained, for 1 of 2 strings OT with honest but curious participants. The upper bounds hold also for one-string OT. The results provide the exact value of OT capacity for a specified class of models, and the necessary and sufficient condition of its positivity, in general.

Keywords: entropy difference bound, generalized erasure channel, honest but curious, oblivios transfer, one of two strings, secret key, wiretap channel.

This paper is based on the ISIT-07 contribution [2]. The authors did intend to write up a full version and devoted substantial amount of work to that project, but abandoned it as other obligations delayed completion and the elapsed time caused loss of novelty. Still, the second author considers it proper to publish this paper in this volume, paying tribute to the memory of Rudolph Ahlswede. The results in [2] are completed by some previously unpublished ones which originated from the authors' discussions during their work towards a full version of [2].

1 Introduction

Oblivious transfer (OT) is a fundamental concept in cryptography, see for example [9]. The term has been used with different meanings, including a simple transmission over a binary erasure channel. In this paper, unless stated otherwise, OT means "1 out of 2 oblivious string transfer" [9]. Two parties are involved, commonly called Alice and Bob. Alice is initially given two binary strings K_0, K_1 of length k, and Bob is given a single bit Z. An OT protocol performed by Alice and Bob is supposed to let Bob learn K_Z while he remains ignorant of $K_{\overline{Z}}$ ($\overline{Z} = 1 - Z$) and Alice remains ignorant of Z. The Shannon-theoretic approach is used, thus ignorance means negligible amount of information. Formal definitions are in Section 2.

* Rudolf Ahlswede, one of the world's top information theorists at the University of Bielefeld, died December 18, 2010.

** Supported by Hungarian National Foundation for Scientific Research, Grant 76088.

H. Aydinian, F. Cicalese, and C. Deppe (Eds.): Ahlswede Festschrift, LNCS 7777, pp. 145–166, 2013.

Throughout this paper, it is assumed that Alice and Bob may use the following resources for free: (i) unlimited computing power (ii) local randomness provided by random experiments they may perform, independently of each other (iii) a noiseless public channel, available for unlimited communication in any number of rounds. These free resources alone are not sufficient for OT. In this paper, two kinds of models will be considered which involve an additional (non-free) resource, either a discrete memoryless multiple source (DMMS) or a noisy discrete memoryless channel (DMC).

A *source model* is determined by a DMMS with two component sources, i.e., a sequence of i.i.d. repetitions (X_i, Y_i), $i = 1, 2, \ldots$ of a pair (X, Y) of "generic" random variables (RVs) taking values in finite sets $\mathcal{X}, \mathcal{Y}$ called source alphabets. At the ith access to this DMMS, Alice observes X_i and Bob Y_i. A *channel model* is determined by a DMC whose (finite) input and output alphabets are denoted by $\mathcal{X}, \mathcal{Y}$, and the conditional probability of Bob receiving $y \in \mathcal{Y}$ when Alice sends $x \in \mathcal{X}$ is denoted by $W(y|x)$. At the ith access to this DMC, Alice selects an input X_i and Bob observes the corresponding output Y_i. In either model, the cost of one access to the DMMS resp. DMC is one unit. Thus the cost of an OT protocol is the number of accesses to the DMMS resp. DMC.

The OT capacity C_{OT} of a DMMS or DMC is the limit as $n \to \infty$ of $1/n$ times the largest k for which OT is possible with cost n. This concept has been introduced by Nascimento and Winter [11,12] who also proved $C_{\mathrm{OT}} > 0$ under a natural condition. See also Imai et al. [7] who for the binary erasure channel with erasure probability 1/2 proved $C_{\mathrm{OT}} = 1/2$. For previous results showing that a DMMS or DMC makes OT possible for any k (but not that k/n may be bounded away from 0 while the conditions (1)-(3) below are satisfied) see the references in [12]. A related concept of commitment capacity has been introduced and characterized in [15].

In the literature of OT much of the effort is devoted to designing protocols that prevent a malicious Alice from learning Bob's bit Z or a malicious Bob from obtaining information also about $K_{\overline{Z}}$. This issue is not entered here, we assume following [11,12] that Alice and Bob are "honest but curious". This means that they honestly follow the protocol but do not discard any information they get access to in the process, and may use all of it to infer what they are supposed to remain ignorant about. Nevertheless, we will point out that a modification of the basic protocol does provide some protection against cheating, while not decreasing OT capacity.

2 Preliminaries

The basic notation of the book [6] is used, except that source and channel alphabets are denoted by script rather than boldface capitals. In particular, log denotes logarithm to base 2, and a DMC with matrix $W = \{W(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ is referred to as DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ or just $\{W\}$. In order to define admissible OT protocols for source and channel models, general two-party protocols are described first.

A *noiseless protocol*, assuming Alice and Bob have initial knowledge or *view* U and V, is described as follows; here U and V are not necessarily independent RVs. At the beginning of the protocol, both Alice and Bob perform a random experiment to generate RVs M resp. N, where M, N and (U, V) are independent. Then Alice sends Bob over the noiseless public channel a message F_1 which is a function of U and M, and Bob returns Alice a message F_2, a function of V, N and F_1. The formal role of the RVs M, N is to model possible randomization in Alice's choice of F_1 and Bob's choice of F_2, as well as in their actions later on. In following rounds (as many as desired) Alice and Bob alternatingly send messages $F_3, F_4, \ldots, F_{2t}$ which are functions of their instantenous views. In other words, F_i is a function of U, M and $\{F_j,\ j < i\}$ if i is odd, and of V, N and $\{F_j,\ j < i\}$ is i is even (here the messages F_j with j of the same parity as i are redundant). At the end of the protocol, Alice's view will be $(U, M, \mathbf{F})$ and Bob's $(V, N, \mathbf{F})$, where $\mathbf{F} = F_1 \ldots, F_{2t}$.

A *noisy protocol* with n accesses to the DMC $\{W\}$ is described as follows. Alice and Bob, whose initial views are represented by RVs U and V, start the protocol by generating RVs M, N as above. Then Alice selects the DMC input X_1 as a function of U and M, and Bob observes the corresponding output Y_1. After this, in a first session of public communication, they may exchange messages according to a noiseless protocol in which the role of their initial views is played by (U, M) and (V, N, Y_1), respectively; X_1 need not be indicated as part of Alice's view for it is a function of (U, M). In this public communication session, and in subsequent ones, Alice and Bob need not generate new RVs for randomization, the original M and N may be assumed to contain all randomness needed for that purpose.

Next, DMC accesses and public communication sessions alternate. Denote the total public communication in the first i sessions by F^i. Before the i'th access to the DMC, Alice's view is (U, M, F^{i-1}). She selects the DMC input X_i as a function of that view, and Bob observes the corresponding output Y_i. Formally, on the condition that $X_i = x$, the RV Y_i is conditionally independent of $U, V, M, N, Y^{i-1}, F^{i-1}$, and its conditional distribution is $W(\cdot|x)$. Then, in the i'th session of public communication, Alice and Bob perform a noiseless protocol in which their original views are (U, M, F^{i-1}) resp. (V, N, Y^i, F^{i-1}). The protocol ends with the n'th public session, and Alice's and Bob's final views are $(U, M, \mathbf{F})$ and $(V, N, Y^n, \mathbf{F})$ where $\mathbf{F} = F^n$. Alice's knowledge of $X^n = X_1, \ldots, X_n$ need not be indicated for X^n is a function of $(U, M, \mathbf{F})$.

Using the above general concepts, admissible protocols for cost-n oblivious transfer of length-k messages, or briefly (n, k) protocols for OT, are described as follows. Below, $X^n = (X_1, \ldots, X_n)$ and $Y^n = (Y_1, \ldots . Y_n)$ denote, in case of source models, the source output sequences observed by Alice and Bob, and in case of channel models, the sequences of DMC inputs and outputs selected by Alice resp. observed by Bob.

In case of a source model, Alice and Bob may perform any noiseless protocol in which their initial views are $U = (K_0, K_1, X^n)$ and $V = (Z, Y^n)$. Here K_0 and K_1, representing the two binary strings given to Alice, are uniformly distributed on $\{0,1\}^k$, the RV Z, representing the bit given to Bob, is uniformly distributed on $\{0,1\}$, and $K_0, K_1, Z, (X^n, Y^n)$ are mutually independent. In case of a channel model, Alice and Bob may perform any noisy protocol with n accesses to the DMC, in which their initial views are $U = (K_0, K_1)$ and $V = Z$ with K_0, K_1, Z independent and uniformly distributed on $\{0,1\}^k$ resp. $\{0,1\}$. In both cases, upon completing the protocol, Bob produces an estimate $\hat{K}_Z$ of K_Z as a function of his view $(Z, N, Y^n, \mathbf{F})$.

Of course, such an (n,k) protocol is suitable for OT only if it meets the goals stated in the Introduction. These are formalized, in the limit $n \to \infty$, by conditions (1)-(3) below in which the dependence on n of the RVs involved is suppressed to keep the notation transparent. Condition (1) means that Bob learns K_Z with negligible probability of error. Conditions (2) and (3) mean that Alice remains ignorant of Z and Bob of $K_{\overline{Z}}$, in the sense of obtaining negligible amount of information about Z resp. $K_{\overline{Z}}$. In exceptional cases when these conditions hold with equality rather than merely convergence to 0, one speaks of perfect OT.

Definition 1. *A positive number R is an achievable OT rate for a given DMMS or DMC if for $n \to \infty$ there exist (n,k) protocols with $\frac{k}{n} \to R$ such that*

$$\Pr\{\hat{K}_Z \neq K_Z\} \to 0 \tag{1}$$

$$I(K_0 K_1 M X^n \mathbf{F} \wedge Z) \to 0 \tag{2}$$

$$I(Z N Y^n \mathbf{F} \wedge K_{\overline{Z}}) \to 0. \tag{3}$$

The OT capacity of a DMMS or DMC is the supremum of achievable OT rates, or 0 if no $R > 0$ is achievable.

Note that since $I(Z \wedge K_{\overline{Z}}) = 0$, condition (3) is equivalent to

$$I(N Y^n \mathbf{F} \wedge K_1 | Z = 0) \to 0; \quad I(N Y^n \mathbf{F} \wedge K_0 | Z = 1) \to 0. \tag{4}$$

Remark 1. *An alternative definition of achievable OT rates reqiures exponentially fast convergence to 0 in (1)-(3) as $n \to \infty$. Another alternative relaxes (3) to $\frac{1}{n} I(Z N Y^n \mathbf{F} \wedge K_{\overline{Z}}) \to 0$. The results in this paper hold under either definition. Note that Definition 1 admits arbitrarily complex protocols. This is necessary for the generality of our upper bound to OT capacity (Theorem 1). On the other hand, for our achievability results (lower bounds to OT capacity) rather simple protocols will suffice. See also Remark 2.*

Given any DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ and distribution P on $\mathcal{X}$ (referred to as an input distribution), consider a DMMS with generic RVs X, Y whose joint distribution is given by $P(x)W(y|x)$. The OT capacity of this DMMS will be denoted by $C_{\mathrm{OT}}(P, W)$, while the OT capacity of the DMC $\{W\}$ is denoted by $C_{\mathrm{OT}}(W)$.

Lemma 1. *For each DMC $\{W\}$ and input distribution P*

$$C_{\mathrm{OT}}(W) \geq C_{\mathrm{OT}}(P, W).$$

Proof. Let R be an achievable OT rate for the source model given by the DMMS with generic RVs X, Y as above. Then (n, k) protocols achieving OT rate R for the source model give rise to OT protocols for the channel model achieving the same OT rate, simply as follows. In the first stage Alice selects i.i.d. repetitions of X as DMC inputs $X_1, \ldots, X_n$, and Bob observes the corresponding outputs $Y_1, \ldots, Y_n$; in this stage the public channel is not used, thus the first $n-1$ public sessions are empty. Upon completing this stage, Alice and Bob have views as their initial views would be in the source model. Then they perform the given source model protocol.

Remark 2. *Lemma 1 may be applied to the DMC* $\{W^l \colon \mathcal{X}^l \to \mathcal{Y}^l\}$ *defined by*

$$W^l(y_1, \ldots, y_l | x_1, \ldots, x_l) = \prod_{i=1}^{l} W(y_i|x_i),$$

whose OT capacity clearly equals $lC_{\mathrm{OT}}(W)$. *This gives*

$$C_{\mathrm{OT}}(W) \geq \frac{1}{l} C_{\mathrm{OT}}(P^{(l)}, W^l), \quad \textit{for every distribution } P^{(l)} \textit{ on } \mathcal{X}^l.$$

In this paper, for channel models only protocols as in the proof of Lemma 1 will be used, in effect employing the DMC merely to emulate a DMMS (with alphabets $\mathcal{X}$, $\mathcal{Y}$ *or* $\mathcal{X}^2$, $\mathcal{Y}^2$*; we will not use* $l > 2$*). For DMCs with the property that in Lemma 1 some input distribution* P *attains the equality, or at least that* $\frac{1}{l} C_{\mathrm{OT}}(P^{(l)}, W^l) \to C_{\mathrm{OT}}(W)$ *for suitable distributions* $P^{(l)}$ *on* $\mathcal{X}^l$*, the OT capacity can be attained via source model emulating protocols. It remains open whether every DMC has that property.*

Let us briefly mention also a more general concept of OT, where Alice is initially given m strings $K_0, \ldots, K_{m-1}$, and Bob may be interested in any subset $\{K_j, j \in J\}$ of those, with index set J in a specified family $\mathcal{J}$ of subsets of $\{0, \ldots, m-1\}$. Formally, Bob is given a RV Z with $|\mathcal{J}|$ possible values, and an OT protocol is supposed to let him learn all K_j with index j in the set $J \in \mathcal{J}$ specified by the value of Z, while keeping him ignorant of the remaining strings. At the same time, Alice has to remain ignorant of Z, i.e., of which strings of her has Bob chosen to learn. This general OT concept will not be addressed but its simplest special case $m = 1$, $\mathcal{J} = \{\{0\}, \varnothing\}$ will. In that case, referred to below as *one-string OT*, Alice is given only one string K_0, and Bob one bit Z. He is supposed to learn K_0 if $Z = 0$ and remain ignorant of K_0 if $Z = 1$, while Alice should remain ignorant of Z.

The concepts of (n, k) protocol and OT capacity immediately extend to the above general version of OT, and in particular to one-string OT. For the latter case, the analogues of the conditions (1)-(3) in Definition 1 are

$$\Pr\{\hat{K}_0 \neq K_0 | Z = 0\} \to 0 \tag{5}$$

$$I(K_0 M X^n \mathbf{F} \wedge Z) \to 0 \tag{6}$$

$$I(N Y^n \mathbf{F} \wedge K_0 | Z = 1) \to 0. \tag{7}$$

3 Statement of Results

Theorem 1. *The OT capacity of a DMMS with generic RVs X, Y or of a DMC $\{W\}$ is bounded above by*

$$\min\left[I(X \wedge Y), H(X|Y)\right],$$

respectively by the maximum of this expression for RVs X, Y connected by the channel, i.e., satisfying $P_{Y|X} = W$. The same upper bounds hold for one-string OT, as well.

A first example that the upper bound in Theorem 1 may be achievable is provided by the *binary erasure channel* (BEC). A BEC with erasure probability $0 < p < 1$ is a DMC with input alphabet $\{0, 1\}$, output alphabet $\{0, 1, 2\}$, and $W(0|0) = W(1|1) = 1 - p$, $W(2|0) = W(2|1) = p$. It has been shown in [7] that a BEC with erasure probability 1/2 has OT capacity 1/2.

Theorem 2. *If $\{W\}$ is a BEC with erasure probability p, and P is any distribution on $\{0, 1\}$, then*

$$C_{\mathrm{OT}}(W) = \min(p, 1-p), \quad C_{\mathrm{OT}}(P, W) = H(P)\min(p, 1-p).$$

The next theorem addresses a larger class of channels than BECs.

Definition 2. *A generalized erasure channel (GEC) is a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ such that for some nonempty $\mathcal{Y}_1 \subset \mathcal{Y}$ the probabilities $W(y|x), y \in \mathcal{Y}_1$ do not depend on $x \in \mathcal{X}$.*

As outputs $y \in \mathcal{Y}_1$ carry no information about the input, they are interpreted as erasures. The BEC is a special case with $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, 2\}$, $\mathcal{Y}_1 = \{2\}$. The *erasure probabability* of a GEC is $p = \sum_{y \in \mathcal{Y}_1} W(y|x)$ which does not depend on $x \in \mathcal{X}$.

Theorem 3. *If $\{W : \mathcal{X} \to \mathcal{Y}\}$ is a GEC with erasure probability p, and P is any distribution on $\mathcal{X}$, then*

$$\begin{aligned} C_{\mathrm{OT}}(W) &= C(W), & C_{\mathrm{OT}}(P, W) &= I(P, W) & &\text{if } p \geq 1/2 \\ C_{\mathrm{OT}}(W) &\geq \frac{p}{1-p} C(W), & C_{\mathrm{OT}}(P, W) &\geq \frac{p}{1-p} I(P, W) & &\text{if } p < 1/2. \end{aligned}$$

Here $C(W) = \max_P I(P, W)$ is the *Shannon capacity* of the DMC $\{W\}$, and $I(P, W)$ denotes the mutual information of RVs X, Y with joint distribution given by $P(x)W(y|x)$.

The proof technique of the lower bounds in Theorem 3 works beyond the class of GECs. It provides lower bounds to OT capacity for the larger class of DMCs that can be represented as a mixture of two channels with identical input alphabet $\mathcal{X}$ and disjoint output alphabets $\mathcal{Y}_0$ and $\mathcal{Y}_1$, namely as

$$W(y|x) = \begin{cases} (1-p)W_0(y|x), & x \in \mathcal{X}, y \in \mathcal{Y}_0 \\ pW_1(y|x), & x \in \mathcal{X}, y \in \mathcal{Y}_1. \end{cases} \tag{8}$$

Note that if the matrix W_1 has identical rows then (8) gives a GEC.

The following result is not contained in [2]. The auxiliary RV U in its second assertion is unrelated to U appearing in the Preliminaries.

Theorem 4. *For a DMC $\{W\}$ of form (8) and any distribution P on $\mathcal{X}$*

$$C_{\mathrm{OT}}(P,W) \geq [I(P,W_0) - I(P,W_1)] \min(p, 1-p).$$

A possibly better bound is

$$C_{\mathrm{OT}}(P,W) \geq \left[I(U \wedge Y^{(0)}) - I(U \wedge Y^{(1)})\right] \min(p, 1-p),$$

where U is any RV and $X, Y^{(0)}, Y^{(1)}$ are RVs with $P_{XY^{(j)}}(x,y) = P(x)W_j(y|x)$, $j = 0, 1$, such that

$$U \to X \to (Y^{(0)}, Y^{(1)}) \text{ is a Markov chain.} \tag{9}$$

Consequently, $C_{\mathrm{OT}}(W)$ is bounded below by $\min(p, 1-p)$ times the secrecy capacity of the wiretap channel with component channels W_0, W_1.

The model called wiretap channel with component channels W_0, W_1 has been introduced by Wyner [17] assuming a special relationship between W_0, W_1 and by Csiszár and Körner [5] for any W_0, W_1 with the same input alphabet. In this model, Alice selects the inputs, Bob observes the W_0-outputs and an eavesdropper Eve the W_1-outputs. The *secrecy capacity* is the supremum of rates at which Alice can reliably send Bob messages in such a way that Eve remains ignorant about them. According to [5], it equals the maximum of $I(U \wedge Y^{(0)}) - I(U \wedge Y^{(1)})$ for RV's satisfying (9), with X and $Y^{(j)}$ connected by the channel W_j, $j = 0, 1$. Hence the second assertion of Theorem 4 implies the last one by Lemma 1.

Remark 3. *In (8) the indices 0 and 1 can be exchanged if simultaneously p and $1-p$ are exchanged. Hence the bounds in Theorem 4 hold also with the reversed order of W_0 and W_1.*

Theorems 1 and 3 admit to give a necessary and sufficient condition for the positivity of OT capacity.

Theorem 5. *A DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ has positive OT capacity iff there exist x', x'' in $\mathcal{X}$ such that the corresponding rows of the matrix W are not identical, and $W(y|x')W(y|x'') > 0$ for some $y \in \mathcal{Y}$. Further, $C_{\mathrm{OT}}(P,W) > 0$ for an input distribution P iff x', x'' as above exist with $P(x')P(x'') > 0$.*

Remark 4. *A similar result appears in [11,12], but there a stronger condition is claimed necessary and sufficient for $C_{\mathrm{OT}}(W) > 0$; it can be equivalently stated by adding to the requirements on x' and x'' in Theorem 5 that neither of the corresponding rows of W is a convex combination of other rows. That additional requirement, however, is not necessary in the "honest but curious" framework, see Example 3 for a counterexample and additional discussion. Nevertheless, the proof of Theorem 5 uses an idea as [11,12], simplified by the availability of Theorem 3.*

4 Proofs

Proof of Theorem 1. It suffices to prove the claimed bounds for one-string OT capacity. Indeed, (n,k) protocols satisfying (1)-(3) trivially give rise to (n,k) protocols for one-string OT satisfying (5)-(7), just letting the pair of RVs K_1, M in the former protocols play the role of M in the latter. Below, attention is restricted to channel models since the proof for source models is similar but simpler. In the proof, instead of condition (7) only its relaxation

$$I(NY^n\mathbf{F} \wedge K_0|Z=1) = o(n) \tag{10}$$

will be used, see Remark 1 after Definition 1.

Now, given a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$, consider (n,k) protocols for one-string OT that satisfy (5), (6) and (10). By Lemma 3 in Appendix A, the condition (6) implies

$$H(K_0|X^n\mathbf{F},\, Z=0) - H(K_0|X^n\mathbf{F},\, Z=1) = o(n) \tag{11}$$

as well as

$$H(K_0|\mathbf{F},\, Z=0) - H(K_0|\mathbf{F},\, Z=1) = o(n). \tag{12}$$

Since $H(K_0|Z=0) = H(K_0|Z=1) = k$, equation (12) is equivalent to

$$I(K_0 \wedge \mathbf{F}|Z=0) = I(K_0 \wedge \mathbf{F}|Z=1) + o(n)$$

and hence (10) implies

$$I(K_0 \wedge \mathbf{F}|Z=0) = o(n). \tag{13}$$

The conditions (5),(13) are similar to those defining a secret key for Alice and Bob, with (weak sense) security from an eavesdropper who observes their public communication $\mathbf{F}$. If (5),(13) held without the conditioning on $Z=0$ then K_0 would be, by definition, such a secret key, see [10],[1]. Then by these references

$$k = H(K_0) \le \sum_{t=1}^{n} I(X_t \wedge Y_t) + o(n) \tag{14}$$

would hold. Actually, (14) holds also in the present case. Indeed, the conditioning on $Z=0$ affects the mentioned result only by changing the terms $I(X_t \wedge Y_t)$ to $I(X_t \wedge Y_t|Z=0)$. This has a negligible effect if n is large, because (6) implies that $\max_t I(X_t \wedge Z) \to 0$, and hence the conditional distribution of X_t on the condition $Z=0$ differs negligibly from the unconditional one, uniformly in t.

To derive another bound on k, we use that K_0 and NY^nZ are conditionally independent given $X^n\mathbf{F}$. For a formal proof of this, see Lemma 6 in Appendix B. It follows using (5) and Fano's inequality that

$$H(K_0|X^n\mathbf{F},\, Z=0) \le H(K_0|NY^n\mathbf{F},\, Z=0) \le H(K_0|\hat{K}_0,\, Z=0) + o(n),$$

whence by (11) also

$$H(K_0|X^n\mathbf{F},\, Z=1) = o(n). \tag{15}$$

Using (10) and (15) we obtain

$$\begin{aligned} k &= H(K_0|Z=1) = H(K_0|NY^n\mathbf{F},\, Z=1) + o(n) \\ &\leq H(K_0X^n|NY^n\mathbf{F},\, Z=1) + o(n) = H(X^n|NY^n\mathbf{F},\, Z=1) + o(n) \\ &\leq H(X^n|Y^n,\, Z=1) + o(n) \leq \sum_{t=1}^{n} H(X_t|Y_t,\, Z=1) + o(n). \end{aligned}$$

In the last sum, the conditioning on $Z=1$ may be omitted with negligible effect as before. Thus we have shown that

$$k \leq \sum_{t=1}^{n} H(X_t|Y_t) + o(n). \tag{16}$$

Finally, the sums in (14) and (16) may be written as $nI(X_T \wedge Y_T|T)$ and $nH(X_T|Y_T,\, T)$, respectively, where T is a RV uniformly distributed on $\{1,\ldots,n\}$ and independent of (X^n, Y^n). The RVs X_T and Y_T are connected by the channel W and satisfy

$$I(X_T \wedge Y_T|T) \leq I(X_T \wedge Y_T), \quad H(X_T|Y_T,T) \leq H(X_T|Y_T).$$

The proof of Theorem 1 is complete.

Proof of Theorem 2. If X and Y are RVs connected by a BEC with erasure probability p then

$$H(X|Y=0) = H(X|Y=1) = 0, \quad H(X|Y=2) = H(X),$$

hence

$$H(X|Y) = pH(X), \quad I(X \wedge Y) = H(X) - H(X|Y) = (1-p)H(X).$$

It follows by Theorem 1 that

$$C_{\mathrm{OT}}(P,W) \leq H(P)\min(p,1-p), \quad C_{\mathrm{OT}}(W) \leq \min(p,1-p).$$

It remains to show that these upper bounds are achievable.

By Lemma 1, it suffices to show that each $R < H(X)\min(p,1-p)$ is an achievable OT rate for the source model defined by a DMMS with generic RVs X, Y as above. To this end, an OT protocol will be described for this source model. It will involve only two messages sent over the public noiseless channel, the first by Bob and the second by Alice; formally, Alice's message F_1 and Bob's message F_4 will be empty.

Upon observing $Y^n = (Y_1,\ldots,Y_n)$, Bob first determines two subsets G and B of $\{1,\ldots,n\}$, called the good and bad sets, both of size about $n\min(p,1-p)$. If $p \geq 1/2$ then Bob takes for G the set of all indices i with $Y_i \neq 2$, and he assigns the indices i with $Y_i = 2$ to B with probability $(1-p)/p$, independently of each other. If $p < 1/2$ then Bob takes for B the set of all indices with $Y_i = 2$, and he assigns the indices with $Y_i \neq 2$ to G with probability $p/(1-p)$, independently of each other. Formally, in order to comply with the description of protocols in Section 2, Bob may be assumed to use a RV N generated at the outset, when he has to assign indices i to B or to G in a randomized manner. E.g., when $p > 1/2$, this N may consist of n independent bits, each equal to 0 with probability $(1-p)/p$, and an index i with $Y_i = 2$ is assigned to B if the i'th bit of N is 0.

Bob's next action is to send Alice a message telling her the sets G and B but not which is which: he lets her learn two sets S_0, S_1 where $S_0 = G, S_1 = B$ if $Z = 0$, and $S_0 = B, S_1 = G$ if $Z = 1$. Note that the pair of random sets G, B is independent of X^n, the events $\{i \in G\}$, $i = 1, \ldots, n$ are independent and have probability $\min(p, 1-p)$, and the same holds for the events $\{i \in B\}$. This implies, in particular, that Bob's message gives Alice no information about Z.

Consider first the case when X is uniformly distributed on $\{0,1\}$. Suppose Alice's strings K_0, K_1 are of length[1] $k = nr$ where $r < \min(p, 1-p)$ is arbitrarily fixed. If $|G| \geq nr$ and $|B| \geq nr$, which holds with probability going to 1 exponentially fast as $n \to \infty$, let S'_0 and S'_1 denote the subsets of S_0 resp. S_1 consisting of their first nr elements. Then Alice encrypts K_0 and K_1 with the "keys" $\{X_i, i \in S'_0\}$ resp. $\{X_i, i \in S'_1\}$, and sends Bob the "cryptograms" $K_j + \{X_i, i \in S'_j\}$, $j = 0, 1$, where $+$ means componentwise addition mod 2. If $|G| < nr$ or $|B| < nr$ then she sends nothing. Except for the latter case of negligible probability, Bob can decrypt K_Z since $S_Z = G$ implies that he knows $\{X_i, i \in S'_Z\} = \{Y_i, i \in S'_Z\}$. On the other hand, Bob remains fully ignorant of $K_{\overline{Z}}$, since the "key" $\{X_i, i \in S'_{\overline{Z}}\}$ is uniformly distributed on $\{0,1\}^{nr}$ and $S_{\overline{Z}} = B$ implies that Bob has 0 information about it. Note that this already suffices for the proof of $C_{\mathrm{OT}}(W) = \min(p, 1-p)$.

If X is not uniformly distributed on $\{0,1\}$, the strings $\{X_i, i \in S'_j\}$, $j = 0, 1$ are not directly suitable as encryption keys, they have to be transformed to binary strings of length $k < rn$ whose distribution is nearly uniform on $\{0,1\}^k$. It is well-known that given any $\delta > 0$, in the case of large n there exists a mapping $\kappa : \{0,1\}^n \to \{0,1\}^k$ with $k = n(H(X) - \delta)$ such that $k - H(\kappa(X^n))$ is exponentially small (in later proofs we will need a stronger result, Proposition 1). Applying this replacing n by rn, there exists a mapping $\kappa : \{0,1\}^{nr} \to \{0,1\}^k$ with $k = nr(H(X) - \delta)$ such that $\kappa_j = \kappa(\{X_i, i \in S'_j\})$, $j = 0, 1$ are nearly uniformly distributed, in the sense that their entropy differs from k only by an exponentially small amount.

To complete the proof, assume Alice's strings K_0, K_1 are of length $k = nr(H(X) - \delta)$. She encrypts them by the keys κ_0, κ_1, and sends Bob the strings $K_j + \kappa_j$, $j = 0, 1$. Again, Bob can decipher K_Z, and he remains ignorant of $K_{\overline{Z}}$ in the sense that he has an exponentially small amount of information about $K_{\overline{Z}}$, see, e.g. [6, Proposition 17.1].

Remark 5. *The protocol in the above proof achieves more than required in Definition 1: Eve's amount of information about Z is not only asymptotically but*

[1] Here and later on, if a specified length of sequences is not an integer, the next integer is meant.

exactly 0*, and in the case when* X *is uniformly distributed on* $\{0,1\}$*, Bob's information about* $K_{\overline{Z}}$ *is also* 0*. The latter need not hold for the described protocol when* X *is not uniformly distributed, but can be achieved also in that case by a slightly modified protocol. As* $k-H(\kappa_j)$ *equals the* I*-divergence of the distribution of* κ_j *from the uniform distribution on* $\{0,1\}^k$*, its exponential smallness implies that of the variation distance of these distributions. Hence Alice can generate RVs* $\overline{\kappa}_j$ *uniformly distributed on* $\{0,1\}^k$ *with* $\Pr\{\kappa_j \neq \overline{\kappa}_j\}$ *exponentially small,* $j=0,1$*, and send Bob* $K_j+\overline{\kappa}_j$ *rather than* $K_j+\kappa_j$*,* $j=0,1$*. Then Bob can still reconstruct* K_Z *with exponentially small probability of error (an error occurring when* $\kappa_Z \neq \overline{\kappa}_Z$*), and he has* 0 *information about* $K_{\overline{Z}}$*.*

Proof of Theorem 3. Let $\{W\}$ be a GEC. Then (8) holds with $\mathcal{Y}_0 = \mathcal{Y} \setminus \mathcal{Y}_1$, $W_0(y|x) = \frac{1}{1-p}W(y|x)$ $(y \in \mathcal{Y}_0)$ and with $W_1(y|x)$ $(y \in \mathcal{Y}_1)$ not depending on $x \in \mathcal{X}$. Hence by Lemma 7 in Appendix B,

$$I(P,W) = (1-p)I(P,W_0) \ . \tag{17}$$

On account of Theorem 1, Lemma 1 and (17), it suffices to prove that if $\{W\}$ is a GEC then $C_{\mathrm{OT}}(P,W) \geq I(P,W_0)\min(p,1-p)$. This is a special case of the first assertion of Theorem 4, and the proof of that more general result is not really more difficult. Below we proceed directly with the latter.

The following basic proposition about generating a secret key will be used.

Proposition 1. *([10,1]) Let* (X_i,Y_i) $i=1,\ldots,n$ *and* $(\tilde{X}_i,T_i)$ $i=1,\ldots,n$ *be i.i.d. repetitions of pairs of RVs* (X,Y) *resp.* (X,T)*. For any* $\delta>0$ *and* $n\to\infty$ *there exist functions* κ *and* f *on* $\mathcal{X}^n$*, where the range of* κ *is* $\{0,1\}^k$ *with*

$$k = n(I(X \wedge Y) - I(X \wedge T) - \delta) \tag{18}$$

such that $\kappa(X^n)$ *is recoverable from* $f(X^n)$ *and* Y^n *with exponentially small probability of error, and*

$$k - H(\kappa(\tilde{X}^n|f(\tilde{X}^n),T^n) \to 0 \quad \textit{exponentially fast.} \tag{19}$$

Such functions κ *and* f *also exist with*

$$k = n(I(U \wedge Y) - I(U \wedge T) - \delta) \ , \tag{20}$$

for any RV U *satisfying the Markov condition* $U \to X \to (Y,T)$*.*

Remark 6. *In the usual setting, Alice and Bob have to generate a secret key assuming Alice observes* X^n*, Bob observes* Y^n*, only Alice is permitted to send Bob a public message, and the key has to be concealed from Eve who observes Alice's message and has side information* T^n*. This setting is formally less general than that in Proposition 1, for it regards the sequences* X^n *and* $\tilde{X}^n$ *identical rather than only identically distributed. Mathematically, however, this makes no difference, and the stated form of Proposition 1 is more convenient for the* *purpose of this paper. Note that originally weak secrecy had been addressed, i.e., the difference in (19) was shown to be $o(n)$ rather than to approach* 0 *(in [10] for (18) and in [1] also for (20); in [1] the largest key rate k/n asymptotically achievable with unidirectional public communication is also determined). Still, the "strong" version with (19) is also well-known, see, e.g. [6, Theorem 17.21].*

Proof of Theorem 4. Let $\{W : \mathcal{X} \to \mathcal{Y}\}$ with $\mathcal{Y} = \mathcal{Y}_0 \cup \mathcal{Y}_1$ be a DMC of form (8), and consider a DMMS with generic RVs X, Y whose joint distribution is given by $P(x)W(y|x)$, $x \in \mathcal{X}, y \in \mathcal{Y}$. To prove the claimed bounds on $C_{\mathrm{OT}}(P, W)$, protocols for the corresponding source model similar to those in the proof of Theorem 2 will be used.

Upon observing $Y^n = (Y_1, \ldots, Y_n)$, Bob first determines a "good set" G and a "bad set" B as in the proof of Theorem 2, with the only modification that the criteria $Y_i \neq 2$ resp. $Y_i = 2$ are replaced by $Y_i \in \mathcal{Y}_0$ resp. $Y_i \in \mathcal{Y}_1$. As there, the pair of random sets G, B is independent of $X^n = (X_1, \ldots, X_n)$, the events $\{i \in G\}$, $i = 1, \ldots, n$ have probability $\min(p, 1-p)$ and are independent of each other and X^n, and the same holds also for the events $\{i \in B\}$. Then Bob sends Alice a message telling her two sets S_0, S_1 where $S_0 = G, S_1 = B$ if $Z = 0$, and $S_0 = B, S_1 = G$ if $Z = 1$. Thereby Alice receives 0 information about Z.

The i.i.d. pairs (X_i, Y_i) are conditionally independent conditioned on the value of Z and the sets S_0, S_1, moreover, those with $i \in S_0$ as well as those with $i \in S_1$ are conditionally i.i.d. If $i \in S_0$ resp. $i \in S_1$, the conditional distribution of (X_i, Y_i) is given by $P(x)W_0(y|x)$ resp. $P(x)W_1(y|x)$ if $Z = 0$, and by $P(x)W_1(y|x)$ resp. $P(x)W_0(y|x)$ if $Z = 1$. To verify this, suppose first that $Z = 0$. Then $i \in S_0$ means $i \in G$, which implies $Y_i \in \mathcal{Y}_0$, and for $x \in \mathcal{X}, y \in \mathcal{Y}_0$ the conditional probability $\Pr\{X_i = x, Y_i = y | S_0, S_1, Z = 0\} = \Pr\{X_i = x, Y_i = y | G, B\}$ is equal to

$$\Pr\{X_i = x, Y_i = y | i \in G\} = \frac{\Pr\{X_i = x, Y_i = y, i \in G\}}{\Pr\{i \in G\}} = P(x)W_0(y|x) \ ;$$

here the second equality holds because, by the construction of G, the probability in the numerator is equal to $P(x)W(y|x)$ if $p \geq 1/2$ and to $P(x)W(y|x)\frac{p}{1-p}$ if $p < 1/2$, where $W(y|x) = (1-p)W_0(y|x)$ by (8), while the probability in the denominator equals $\min(p, 1-p)$. For $i \in S_1$ the calculation is similar. In the case $Z = 1$ the roles of S_0 and S_1 are simply reversed.

The proof of the first assertion of Theorem 4 will be completed by showing that, for any $r < \min(p, 1-p)$, if Alice's strings K_0, K_1 have length

$$k = rn(I(P, W_0) - I(P, W_1) - \delta)$$

then she, knowing S_0, S_1, can send Bob a message that enables him to recover K_Z while keeping him ignorant of $K_{\overline{Z}}$.

Apply the first assertion of Proposition 1 with rn in the role of n, taking $\{P(x)W_0(y|x), x \in \mathcal{X}, y \in \mathcal{Y}_0\}$ resp. $\{P(x)W_1(y|x), x \in \mathcal{X}, y \in \mathcal{Y}_1\}$ for the joint distribution of X, Y resp. X, T. Let f and κ denote the corresponding functions on $\mathcal{X}^{rn}$ where the range of κ is $\{0,1\}^k$ with the above k, see (18). Supposing

$$|S_0| \geq rn, \quad |S_1| \geq rn \ , \tag{21}$$

denote by S_0' and S_1' the sets of the first rn elements of S_0 resp. S_1. Let Alice compute $f_j = f(\{X_i, i \in S_j'\})$ and $\kappa_j = \kappa(\{X_i, i \in S_j'\})$, $j = 0, 1$, and send Bob a message consisting of f_0, f_1 and the "cryptograms" $K_0 + \kappa_0, K_1 + \kappa_1$; if (21) does not hold then she sends nothing.

Consider first the case $Z = 0$. Then, conditioned on Z and S_0, S_1 satisfying (21), the pairs (X_i, Y_i), $i \in S_0'$ are conditionally i.i.d. with distribution $P(x)W_0(y|x)$. Hence, due to the choice of the mappings f and κ, Bob can recover κ_0 from f_0 and $\{X_i, i \in S_0'\}$ with exponentially small (conditional) probability of error, enabling him to recover K_0. As this always holds when (21) does, the probability of error in recovering K_0 conditioned only on $Z = 0$ is also exponentially small. Further, the pairs (X_i, Y_i), $i \in S_1'$ are conditionally i.i.d. with distribution $P(x)W_1(y|x)$. Hence the choice of κ and f implies that f_1 and $\{Y_i, i \in S_1'\}$ give a negligible amount of information about κ_1; in turn, since κ_1 is nearly uniformly distributed, Bob's amount of information about K_1 provided by $f_1, \{Y_i, i \in S_1'\}$ and $K_1 + \kappa_1$ is also negligible: $I(K_1 \wedge f_1, K_1 + \kappa_1, \{Y_i, i \in S_1'\}|S_0, S_1, Z = 0)$ is exponentially small. To formally verify that the last conditional mutual information coincides with that in the first condition in (4), assuming the RV N has been generated and used by Bob as in the proof of Theorem 2, note that the total communication is now $\mathbf{F} = (S_0, S_1, f_0, K_0 + \kappa_0, f_1, K_1 + \kappa_1)$, and K_1 is independent of (N, S_0, S_1, Z). Hence

$$I(NY^n\mathbf{F} \wedge K_1|Z = 0) = I(Y^n, f_0, K_0 + \kappa_0, f_1, K_1 + \kappa_1 \wedge K_1|N, S_0, S_1, Z = 0)\ .$$

Here, N in the condition may be omitted. It remains to show that

$$I(\{Y_i, i \notin S_0'\}, f_0, K_0 + \kappa_0 \wedge K_1|S_0, S_1, f_1, K_1 + \kappa_1, Z = 0) = 0\ .$$

This follows because $(X_i, Y_i), i = 1, \ldots, n$ are conditionally independent given $S_0, S_1, Z = 0$, and f_j and κ_j are functions of K_j and $\{(X_i, Y_i), i \in S_j'\}$, $j = 0, 1$.

In the case $Z = 1$ it follows similarly that Bob can recover K_1 and he remains ignorant of K_0. This completes the proof of the first assertion of Theorem 4.

The second assertion follows in the same way, applying this time the second assertion of Proposition 1. The third assertion follows from the second one as noted in the passage following Theorem 4.

Remark 7. *Another suitable protocol is obtained by modifying the choice of the sets G and B as follows. According as $p \geq 1/2$ or $p < 1/2$, let G resp. B contain all indices i with Y_i in $\mathcal{Y}_0$ resp. in $\mathcal{Y}_1$ as before, and let the other indices i be assigned to G or B with probabilities $(\pi, 1 - \pi)$. Here π is chosen to make sure that $\Pr\{i \in G\} = \Pr\{i \in B\} = 1/2$, thus π equals $1 - 1/2p$ if $p \geq 1/2$ and $1/2(1 - p)$ if $p < 1/2$. Consider first the case $p \geq 1/2$. Then, by similar calculation as in the proof of Theorem 4,*

$$\Pr\{X_i = x, Y_i = y|i \in G\} = \begin{cases} 2(1-p)P(x)W_0(y|x), & x \in \mathcal{X}, y \in \mathcal{Y}_0 \\ (2p-1)P(x)W_1(y|x), & x \in \mathcal{X}, y \in \mathcal{Y}_1\ , \end{cases}$$

$$\Pr\{X_i = x, Y_i = y|i \in B\} = P(x)W_1(y|x),\ x \in \mathcal{X}, y \in \mathcal{Y}_1\ .$$

It follows, in turn, that the conditional mutual information $I(X_i \wedge Y_i|G,B)$ is equal to $2(1-p)I(P,W_0)+(2p-1)I(P,W_1)$ if $i \in G$ (using Lemma 7) and to $I(P,W_1)$ if $i \in B$. This implies via Proposition 1, again as in the proof of Theorem 4, that with this modified protocol one can achieve OT rate

$$1/2\,[2(1-p)I(P,W_0)+(2p-1)I(P,W_1)-I(P,W_1)] \;\;,$$

the same as with the original protocol. In the case $p < 1/2$ the situation is similar. It follows similarly that OT rates in the second assertion of Theorem 4 can also be achieved with protocols in which G and B are selected as above.

To the proof of Theorem 5 a simple fact is sent forward.

Lemma 2. *If a DMC $\{W'\}$ is obtained from $\{W : \mathcal{X} \to \mathcal{Y}\}$ by restricting the input alphabet $\mathcal{X}$ to a subset $\mathcal{X}'$ then $C_{\mathrm{OT}}(W') \le C_{\mathrm{OT}}(W)$.*

The proof is obvious but depends on the "honest but curious" assumption. Were Alice allowed to deviate from the agreed-upon protocol, a larger input alphabet would give her more room for deviations undetectable for Bob and letting her gain information about Bob's bit Z; this might decrease OT capacity.

Proof of Theorem 5. (i) Necessity. Given a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$, let $\mathcal{X}'$ be a maximal subset of $\mathcal{X}$ such that the rows of the matrix W corresponding to input symbols $x' \in \mathcal{X}'$ are all distinct; let W' be the matrix that has these distinct rows. Clearly $C_{\mathrm{OT}}(W) = C_{\mathrm{OT}}(W')$. If $C_{\mathrm{OT}}(W) > 0$ then $C_{\mathrm{OT}}(W') > 0$ implies by Theorem 1 that the outputs of W' do not unambiguously determine the inputs. In other words, for some $y \in \mathcal{Y}$ there exist x' and x'' in $\mathcal{X}'$ such that $W(y|x')W(y|x'') > 0$; this proves necessity for channel models. For source models the proof is similar, this time using that $C_{\mathrm{OT}}(P,W) = C_{\mathrm{OT}}(P',W')$ where $P'(x')$, $x' \in \mathcal{X}'$ equals the sum of $P(x)$ for all $x \in \mathcal{X}$ such that the rows of W corresponding to x and x' are equal.

(ii) Sufficiency. Let $\{W\}$ be a DMC satisfying the conditions in Theorem 5. Consider an auxiliary DMC $\{\widetilde{W}\}$, restricting the input alphabet $\mathcal{X} \times \mathcal{X}$ of W^2 (see Remark 2) to the pairs (x',x''), (x'',x'), where x',x'' as in Theorem 5 are fixed. Formally, $\{\widetilde{W} : ((x',x''),(x'',x')) \to \mathcal{Y} \times \mathcal{Y}\}$ is defined by

$$\widetilde{W}(y_1,y_2|x',x'') = W(y_1|x')W(y_2|x''),\ \widetilde{W}(y_1,y_2|x'',x') = W(y_1|x'')W(y_2|x'). \tag{22}$$

This auxiliary DMC is a GEC, the role of $\mathcal{Y}_1$ in Definition 2 being played by the subset $\{(y,y) : y \in \mathcal{Y}\}$ of $\mathcal{Y} \times \mathcal{Y}$; hence Theorem 3 implies $C_{\mathrm{OT}}(\widetilde{W}) > 0$. On account of Lemma 2, this proves the positivity of $C_{\mathrm{OT}}(W) = \frac{1}{2}C_{\mathrm{OT}}(W^2)$.

Consider next a source model defined by a DMMS with generic RVs X,Y whose joint distribution $P_{XY}(x,y) = P(x)W(y|x)$ satisfies the condition in Theorem 5. Fixing x',x'' as there, for $2n$ i.i.d. repetitions of X, viz. $X^{2n} = (X_1,\dots,X_{2n})$ let J denote the set of indices $i \in \{1,\dots,n\}$ for which (X_{2i-1},X_{2i}) equals either (x',x'') or (x'',x'). The tuples $\{(X_{2i-1},X_{2i}),(Y_{2i-1},Y_{2i}), i \in J\}$ are conditionally i.i.d. given J, their (conditional) distribution is equal to $P_{\tilde{X}\tilde{Y}}$

where $P_{\tilde{X}}$ is the uniform distribution on $\{(x', x''), (x'', x')\}$ and $P_{\tilde{Y}|\tilde{X}}$ equals $\widetilde{W}$ in (22). Consider an auxiliary DMMS with generic RVs $\tilde{X}, \tilde{Y}$ as above. Since $\Pr\{i \in J\} = 2P(x')P(x'')$, the size of J exceeds $\ell = nP(x')P(x'')$ with probability approaching 1 exponentially fast as $n \to \infty$. It follows that each (ℓ, k) protocol for the auxiliary DMMS gives rise to a $(2n, k)$ protocol for the original one: Alice tells Bob the set J in her first message, then Alice and Bob perform the given (ℓ, k) protocol using only the first $\ell = nP(x')P(x'')$ tuples $(X_{2i-1}, X_{2i}), (Y_{2i-1}, Y_{2i})$ with $i \in J$. Since the auxiliary DMMS has positive OT capacity by Theorem 3, this completes the proof of Theorem 5.

5 Examples

Example 1 (Binary symmetric channel). *A DMC $\{W : \{0,1\} \to \{0,1\}\}$ is a binary symmetric channel (BSC) with crossover probability $p \neq 1/2$ if $W(1|0) = W(0|1) = p$. To obtain a lower bound to its OT capacity, consider as in the proof of Theorem 5 an auxiliary channel $\{\widetilde{W} : \{(0,1), (1,0)\} \to \{0,1\}^2\}$, see (22) with $x' = 0, x'' = 1$, i.e.,*

$$\widetilde{W}(0,1|\,0,1) = \widetilde{W}(1,0|1,0) = (1-p)^2, \qquad \widetilde{W}(1,0|0,1) = \widetilde{W}(0,1|\,1,0) = p^2,$$
$$\widetilde{W}(0,0|\,0,1) = \widetilde{W}(1,1|0,1) = \widetilde{W}(0,0|1,0) = \widetilde{W}(1,1|1,0) = p(1-p).$$

This $\{\widetilde{W}\}$ is a GEC with erasure probability $\tilde{p} = 2p(1-p) < 1/2$. The role of the set $\mathcal{Y}_1$ in Definition 2 is played by $\{(0,0), (1,1)\}$, and that of $\{W_0\}$ in (8) is played by a channel $\{\widetilde{W_0}\}$ with input and output alphabets equal to $\{(0,1), (1,0)\}$ which is a BSC with crossover probability $\frac{p^2}{1-\tilde{p}} = \frac{p^2}{p^2+(1-p)^2}$.

By Theorem 3 and (17), $C_{\mathrm{OT}}(\widetilde{W}) \geq \frac{\tilde{p}}{1-\tilde{p}} C(\widetilde{W}) = \tilde{p} C(\widetilde{W_0})$. Finally, since Lemma 2 implies $C_{\mathrm{OT}}(\widetilde{W}) \leq C_{\mathrm{OT}}(W^2) = 2C_{\mathrm{OT}}(W)$, we obtain

$$C_{\mathrm{OT}}(W) \geq \frac{1}{2} C_{\mathrm{OT}}(\widetilde{W}) \geq \frac{1}{2} \tilde{p} C(\widetilde{W_0}) = p(1-p)\left[1 - h\left(\frac{p^2}{p^2+(1-p)^2}\right)\right] \ .$$

Example 2 (Z channel). *A Z channel is a DMC $\{W : \{0,1\} \to \{0,1\}\}$ with $W(0|0) = 1$, $W(0|1) = p$, $W(1|1) = 1-p$. To bound its OT capacity from below, consider an auxiliary channel $\{\widetilde{W} : \{(0,1), (1,0)\} \to \{0,1\}^2\}$ as in Example 1, where this time*

$$\widetilde{W}(0,1|0,1) = \widetilde{W}(1,0|1,0) = 1-p, \quad \widetilde{W}(1,1|0,1) = \widetilde{W}(1,1|1,0) = p,$$

and the other entries of the matrix $\widetilde{W}$ are 0. This auxiliary channel is a BEC with erasure probability p, hence $C_{\mathrm{OT}}(\widetilde{W}) = \min(p, 1-p)$ by Theorem 2. It follows that

$$C_{\mathrm{OT}}(W) \geq \frac{1}{2} C_{\mathrm{OT}}(\widetilde{W}) = \frac{1}{2} \min(p, 1-p) \ .$$

Example 3. *The DMC $\{W\colon \{0,1,2\} \to \{0,1\}\}$ with*

$$W(0|0) = W(1|1) = 1,\ W(0|2) = p,\ W(1|2) = 1 - p$$

is, in a sense, a reversed BEC. By Lemma 2, its OT capacity is not smaller than that of the Z channel in Example 2, hence $C_{\mathrm{OT}}(W) \geq \frac{1}{2}\min(p, 1-p)$. Note that while this channel satisfies the condition for $C_{\mathrm{OT}}(W) > 0$ in Theorem 5, it fails to satisfy the stronger condition mentioned in Remark 4. Recall that in the proof of Theorem 5 we have used the fact that the OT capacity of a DMC $\{W\}$ is not changed by a reduction of the input alphabet that keeps only the distinct rows of W. In [11,12] the same is claimed for a further reduction that removes also those rows of W which are convex combinations of others, but that claim is valid only in a "malicious" setting. In the "honest but curious" setting the above DMC is a counterexample, it has positive OT capacity but if the input symbol 2 were removed, the OT capacity would become 0*.*

The lower bounds to OT capacity in the above examples are smaller than the upper bound in Theorem 1, and the exact value of OT capacity remains an open problem. The next example shows that the upper bound in Theorem 1 may be tight even if the channel is not a GEC. The authors have found this example unaware of the work of Wolf and Wullschleger [16] in which the channel below plays a key role and, in particular, another simple $(1,1)$ protocol for perfect OT of 1 bit is given.

Example 4. *For $\mathcal{X} = \mathcal{Y} = \{0,1,2,3\}$, let $\{W : \mathcal{X} \to \mathcal{Y}\}$ be a channel with additive noise such that the RVs X, Y are connected by it if $Y = X + N \pmod 4$ for a RV N uniformly distributed on $\{0,1\}$, independent of X. Theorem 1 gives $C_{\mathrm{OT}}(W) \leq 1$, and $C_{\mathrm{OT}}(P, W) \leq 1$ if P is the uniform distribution on $\mathcal{X}$. These upper bounds are tight; indeed, the next $(1,1)$ protocol achieves perfect OT for the source model with generic RVs X, Y as above and X uniformly distributed on $\mathcal{X}$. Now, Alice has two bits K_0, K_1, Bob one bit Z, independent of each other and (X, Y), and uniformly distributed; Alice observes X and Bob Y. First, let Bob tell Alice the parity of $Y + Z$, sending her $\phi = 0$ or $\phi = 1$ according as $Y + Z$ is even or odd; this gives Alice no information about Z. Then Alice reports Bob the* $\bmod 2$ *sums $K_0 + i_\phi(X)$ and $K_1 + i_{1-\phi}(X)$ where i_0 and i_1 are the indicator functions of the sets $\{1,2\}$ resp. $\{2,3\}$. Note that Bob knowing Y also knows either the bit $i_0(X)$ (if Y is even) or $i_1(X)$ (if Y is odd), but he is fully ignorant of the other bit, in both cases. It follows that Bob can unambiguously determine K_Z but remains fully ignorant of $K_{\overline{Z}}$.*

6 Discussion

Oblivious transfer has been approached from an information theoretic point of view, addressing OT capacity for (discrete memoryless) source and channel models, concentrating on 1 of 2 strings OT.

A general upper bound to OT capacity has been derived, with essential use of inequalities for information measures, see Appendix A. Let us call attention to an improved bound on the difference of conditional entropies via variation distance (Lemma 5), included for its own sake, though a weaker previous bound would also suffice. A remarkable feature of our upper bound to OT capacity is its validity for one-string OT, as well. It remains open whether this is a coincidence caused by the weakness of our method, or perhaps the rate of one-string OT can never exceed the optimal rate of 1 of 2 strings OT.

Our achievability results (lower bounds to OT capacity) rely on rather simple protocols, still they shed light on relationships of OT and other problems of information theoretic security, such as secret key agreement using public discussion [10,1] and secure transmission over insecure channels [17,5]. It remains open whether the OT capacity of channel models can always be attained via source model emulating protocols, as in those cases when we were able to determine OT capacity. These cases are the binary erasure channels with any erasure probability p, and generalized erasure channels (introduced here) with $p \geq 1/2$. An additional such channel appears in Example 4; it remains open whether this is exceptional, or perhaps a member of another "good" class.

Throughout this paper, only models with "honest but curious" participants are studied. Still, let us briefly address some issues arising in "malicious" settings. In case of a BEC or GEC, with agreed-upon protocol as in the proofs of Theorems 2 and 3, a malicious Alice has no opportunity to learn about Bobs bit Z if he follows the protocol. In Examples 1-2, however, a malicious Alice can well gain information about Z if she deviates from using DMC input pairs $(0, 1)$ and $(1, 0)$ only. In Example 3, the malicious model admits no OT at all, see [11,12]. Indeed, Eve may send instead of DMC input 2 always 0 or 1, with probabilities $(p, 1 - p)$; this cheating is undetectable to Bob, and reduces any protocol, in effect, to one for a noiseless channel.

Even the BEC and GEC models are vulnerable to cheating by Bob, who may gain illegitimate information by deviating from the agreed-upon protocol, maliciously selecting the set B. Suppose $p \leq 1/2$, when the protocol requires Bob to take for B the set of indices i with $Y_i = 2$ (or $Y_i \in \mathcal{Y}_1$). He may instead chose B as follows, not modifying the choice of G. If $p \leq 1/3$, he may take B to consist only of indices with $Y_i \neq 2$ (or $Y_i \in \mathcal{Y}_0$), assigning each such index with the same probability $p/(1 - p)$ to B as to G. If $1/3 < p < 1/2$, Bob may assign to B all indices with $Y_i \neq 2$ (or $Y_i \in \mathcal{Y}_0$) not assigned to G, and assign to B the remaining indices with probability $(3p - 1)/p$. If Bob uses this fake B in giving Alice the sets S_0, S_1, she has no way to detect cheating; in case $p \leq 1/3$ Bob will learn both of Alice's strings, and also when $1/3 < p < 1/2$, he will get nonzero information about $K_{\overline{Z}}$, in addition to learning K_Z.

Note, however, that if $p = 1/2$ then the sets G and B provided by the agreed-upon protocol are complements of each other, thus no deviation in selecting B is possible without one in selecting G. This amounts to a kind of limited protection against Bob's cheating: while a malicious Bob can still gain information about both of Alice's strings, to do so he has to give up his goal of fully learning K_Z (the situation is similar if $p > 1/2$). Recall that protocols as in the proof of Theorem 4 can always be modified to protocols of equal power that use complementary sets G and B, see Remark 7. It is plausible that for a BEC or GEC, modified protocols of this kind provide limited protection as above against Bob's cheating also when $p < 1/2$.

This issue is not pursued here any further, since by a recent result of Pinto et al. [13] the OT capacity of a GEC, determined in this paper, is actually achievable also in the "malicious" model. Another recent work, Ishai et al. [8], regarded Alice's pair of strings (K_0, K_1) as a sequence of k pairs (K_{0i}, K_{1i}), $i = 1, \ldots, k$. Bob selects one component of each pair he wants to learn, this selection is specified by a k-bit string $Z = Z_1, \ldots, Z_k$. Then an (n, k) protocol is supposed to let Bob learn $K_{Z_1 1}, \ldots, K_{Z_k k}$ and keep him ignorant of $K_{\overline{Z}_1 1}, \ldots, K_{\overline{Z}_k k}$, while Eve remains ignorant of Z. Ishai et al. show that this goal is achievable with k/n bounded away from 0, see [8] for details. Finally, the reader's attention is called to recent works that address more general problems via similar techniques, and also contain results relevant for OT capacity, as pointed out by an anonymous referee. See Prabhakaran and Prabhakaran [14] and references there.

References

1. Ahlswede, R., Csiszár, I.: Common randomness in information theory and cryptography, part I. IEEE Trans. Inf. Theory 39, 1121–1132 (1993)
2. Ahlswede, R., Csiszár, I.: On oblivious transfer capacity. In: Proc. ISIT 2007, Nice, pp. 2061–2064 (2007)
3. Alicki, R., Fannes, M.: Continuity of quantum conditional information. J. Phys. A: Math. Gen. 37, L55–L57 (2004)
4. Audenaert, K.M.R.: A sharp Fannes-type inequality for the von Neumann entropy. J. Phys. A. 40, 8127–8136 (2007)
5. Csiszár, I., Körner, J.: Broadcast channels with confidential messages. IEEE Trans. Inf. Theory 24, 339–348 (1978)
6. Csiszár, I., Körner, J.: Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd edn. Cambridge University Press (2011)
7. Imai, H., Nascimento, A., Morozov, K.: On the oblivious transfer capacity of the erasure channel. In: Proc. ISIT 2006, Seattle, pp. 1428–1431 (2006)
8. Ishai, Y., Kushilevitz, E., Ostrovsky, R., Prabhakaran, M., Sahai, A., Wullschleger, J.: Constant-Rate Oblivious Transfer from Noisy Channels. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 667–684. Springer, Heidelberg (2011)
9. Kilian, J.: Founding cryptography on oblivious transfer. In: Proc. STOC 1988, pp. 20–31 (1988)
10. Maurer, U.: Secret key agreement by public discussion. IEEE Trans. Inf. Theory 39, 733–742 (1993)
11. Nascimento, A., Winter, A.: On the oblivious transfer capacity of noisy correlations. In: Proc. ISIT 2006, Seattle, pp. 1871–1875 (2006)
12. Nascimento, A., Winter, A.: On the oblivious transfer capacity of noisy resources. IEEE Trans. Inf. Theory 54, 2572–2581 (2008)
13. Pinto, A., Dowsley, R., Morozov, K., Nascimento, A.: Achieving oblivious transfer apacity of generalized erasure channels in the malicious model. IEEE Trans. Inf. Theory 57, 5566–5571 (2011)
14. Prabhakaran, V., Prabhakaran, M.: Assisted common information with an application to secure two-party sampling. arXiv:1206.1282v1 [cs.IT] (2012)

15. Winter, A., Nascimento, A.C.A., Imai, H.: Commitment Capacity of Discrete Memoryless Channels. In: Paterson, K.G. (ed.) Cryptography and Coding 2003. LNCS, vol. 2898, pp. 35–51. Springer, Heidelberg (2003)
16. Wolf, S., Wullschleger, J.: Oblivious Transfer Is Symmetric. In: Vaudenay, S. (ed.) EUROCRYPT 2006. LNCS, vol. 4004, pp. 222–232. Springer, Heidelberg (2006)
17. Wyner, A.: The wiretap channel. Bell System Tech. J. 54, 1355–1387 (1975)
18. Zhang, Z.: Estimating mutual information via Kolmogorov distance. IEEE Trans. Inf. Theory 53, 3280–3283 (2007)

Appendix A

Let U, V, Z denote RVs with values in finite sets $\mathcal{U}, \mathcal{V}, \mathcal{Z}$. Suppose $z_1, z_2 \in \mathcal{Z}$ with $\Pr\{Z = z_1\} = p > 0$, $\Pr\{Z = z_2\} = q > 0$.

Lemma 3

$$|H(U|V, Z = z_1) - H(U|V, Z = z_2)| \leq 3\sqrt{\frac{(p+q)\ln 2}{2pq} I(UV \wedge Z) \log |\mathcal{U}| + 1} \ .$$

Remark 8. *It will be clear from the proof that the constant term* $+1$ *could be replaced by a term that goes to* 0 *as* $I(UV \wedge Z)$ *does, which may be relevant for some purposes but not here.*

The proof of Lemma 3 will rely on two auxiliary lemmas. The variation distance of probability distributions P and Q on the same finite set, say $\mathcal{S}$, is

$$|P - Q| = \sum_{s \in \mathcal{S}} |P(s) - Q(s)| \ .$$

Lemma 4. *The variation distance of the conditional distributions of* U *on the conditions* $Z = z_1$ *resp.* $Z = z_2$ *is bounded as*

$$\left|P_{U|Z=z_1} - P_{U|Z=z_2}\right| \leq \sqrt{\frac{2(p+q)\ln 2}{pq} I(U \wedge Z)} \ .$$

Proof.

$$\begin{aligned} I(U \wedge Z) &= \sum_{z \in \mathcal{Z}} \Pr\{Z = z\} D(P_{U|Z=z} \| P_U) \\ &\geq p D(P_{U|Z=z_1} \| P_U) + q D(P_{U|Z=z_2} \| P_U) \\ &\geq \frac{p |P_{U|Z=z_1} - P_U|^2}{2\ln 2} + \frac{q |P_{U|Z=z_2} - P_U|^2}{2 \ln 2} \ ; \end{aligned}$$

the last step is by Pinsker inequality. Since

$$|P_{U|Z=z_1} - P_U| + |P_{U|Z=z_2} - P_U| \geq |P_{U|Z=z_1} - P_{U|Z=z_2}| \ ,$$

it follows by the easily checked inequality $pa^2 + qb^2 \geq \frac{pq}{p+q}(a+b)^2$ that $I(U \wedge Z)$ is further bounded below by

$$\frac{pq}{2(p+q)\ln 2} |P_{U|Z=z_1} - P_{U|Z=z_2}|^2 \ .$$

Lemma 5. *For RVs U_1, U_2 with values in $\mathcal{U}$, and V_1, V_2 with values in $\mathcal{V}$,*

$$|H(U_1|V_1) - H(U_2|V_2)| \leq \left[\frac{1}{2}\left|P_{U_1V_1} - P_{U_2V_2}\right| + \left|P_{V_1} - P_{V_2}\right|\right] \log|\mathcal{U}|$$
$$+h\left(\frac{1}{2}\min\left[1, |P_{U_1V_1} - P_{U_2V_2}| + |P_{V_1} - P_{V_2}|\right]\right)$$
$$\leq \frac{3}{2}\left|P_{U_1V_1} - P_{U_2V_2}\right| \log|\mathcal{U}| + h\left(\min\left[\frac{1}{2}, |P_{U_1V_1} - P_{U_2V_2}|\right]\right) \quad ,$$

where $h(t) = -t\log t - (1-t)\log(1-t)$, $0 \leq t \leq 1$.

Remark 9. *The main feature of this lemma, for our purposes, is that it does not involve the cardinality of $\mathcal{V}$, only that of $\mathcal{U}$. A previous bound of this kind to the difference of conditional entropies, due to Alicki and Fannes [3], would also suffice for the proof of Theorem 1. but we preferred to sharpen it to obtain Lemma 3 in the stated form.*

Proof. The following bound for the entropy difference of two distributions on $\mathcal{U}$ will be used:

$$|H(P) - H(Q)| \leq \frac{1}{2}|P - Q| \log|\mathcal{U}| + h\left(\frac{1}{2}|P - Q|\right) \quad . \tag{23}$$

This sharpening of a more familiar weaker bound is rather recent [4,18]. Let us recall its simple proof: Let X and Y be RVs with $P_X = P$, $P_Y = Q$ such that $\Pr\{X \neq Y\}$ is smallest possible subject to these conditions, thus $\Pr\{X \neq Y\} = \frac{1}{2}|P - Q|$. Then, as $H(P) - H(Q) \leq H(X|Y)$ and $H(Q) - H(P) \leq H(Y|X)$, (23) follows from Fano's inequality.

Now,

$$H(U_1|V_1) - H(U_2|V_2) = \sum_{v\in\mathcal{V}} \left[P_{V_1}(v)H(P_{U_1|V_1=v}) - P_{V_2}(v)H(P_{U_2|V_2=v})\right]$$
$$\leq \sum_{v\in\mathcal{V}} P_{V_1}(v)\left[H(P_{U_1|V_1=v}) - H(P_{U_2|V_2=v})\right]$$
$$+ \sum_{v:P_{V_1}(v)>P_{V_2}(v)} \left[P_{V_1}(v) - P_{V_2}(v)\right] H(P_{U_2|V_2=v}) \quad .$$

Bounding the first sum via (23), and the entropies in the second sum by $\log|\mathcal{U}|$, this can be continued as

$$\leq \frac{1}{2}\sum_{v\in\mathcal{V}} P_{V_1}(v)|P_{U_1|V_1=v} - P_{U_2|V_2=v}| \log|\mathcal{U}|$$
$$+ \sum_{v\in\mathcal{V}} P_{V_1}(v) h\left(\frac{1}{2}|P_{U_1|V_1=v} - P_{U_2|V_2=v}|\right) + \frac{1}{2}|P_{V_1} - P_{V_2}| \log|\mathcal{U}| \quad .$$

Let U_3 be an auxiliary RV such that $P_{U_3V_1}(u,v) = P_{V_1}(v)P_{U_2|V_2=v}(u)$. Then

$$\begin{aligned}
&\sum_{v\in\mathcal{V}} P_{V_1}(v)|P_{U_1|V_1=v} - P_{U_2|V_2=v}| = |P_{U_1V_1} - P_{U_3V_1}| \\
&\le |P_{U_1V_1} - P_{U_2V_2}| + |P_{U_3V_1} - P_{U_2V_2}| \\
&= |P_{U_1V_1} - P_{U_2V_2}| + |P_{V_1} - P_{V_2}| \le 2|P_{U_1V_1} - P_{U_2V_2}| \ .
\end{aligned}$$

Using this, and that the concave function $h(t)$ is increasing in $[0, 1/2]$, and noting that the above arguments hold also with the roles of (U_1, V_1) and (U_2, V_2) interchanged, Lemma 5 follows.

Proof of Lemma 3. Apply Lemma 5 to RVs U_1, V_1 with joint distribution $P_{U_1V_1} = P_{UV|Z=z_1}$ and U_2, V_2 with $P_{U_2V_2} = P_{UV|Z=z_2}$, replacing the $h()$ term by its upper bound 1. This gives

$$|H(U|V, Z = z_1) - H(U|V, Z = z_2)| \le \frac{3}{2}|P_{UV|Z=z_1} - P_{UV|Z=z_2}| \log |\mathcal{U}| + 1 \ .$$

Combining this with Lemma 4 completes the proof of Lemma 3.

Appendix B

Lemma 6. *With the notation in the proof of Theorem 1,*

$$I(K_0M \wedge NY^nZ|X^n\mathbf{F}) = 0 \ .$$

Proof. Recall that $\mathbf{F} = F^n$ where F^t denotes the total public communication in the first t sessions. For each $1 \le t \le n$ we have

$$\begin{aligned}
I(K_0M \wedge NY^tZ|X^tF^t) &\le I(K_0M \wedge NY^tZ|X^tF^{t-1}) \\
= I(K_0M \wedge NY^{t-1}Z|X^tF^{t-1}) &\le I(K_0MX_t \wedge NY^{t-1}Z|X^{t-1}F^{t-1}) \\
&= I(K_0M \wedge NY^{t-1}Z|X^{t-1}F^{t-1}) \ .
\end{aligned}$$

Here the first inequality holds by [6, Lemma 17.18] (or previous similar results in [10,1]), the next equality holds because $I(K_0M \wedge Y_t|X^tF^{t-1}NY^{t-1}Z) = 0$ due to the conditional independence of Y_t given X_t from the other RVs, and the last equality holds since X_t is a function of K_0, M and F^{t-1}. The lemma follows since $I(K_0M \wedge NY^{t-1}Z|X^{t-1}F^{t-1}) = 0$ trivially holds for $t = 1$.

Lemma 7. *For $\{W : \mathcal{X} \to \mathcal{Y}_0 \cup \mathcal{Y}_1\}$ as in (8), the identity*

$$I(P, W) = (1 - p)I(p, W_0) + pI(P, W_1)$$

holds for each input distribution P.

Proof. Let X and Y have joint distribution $P(x)W(y|x)$. Define $T=j$ if $Y\in\mathcal{Y}_j$, $j=0,1$, then $P_T=(1-p,p)$ and T is independent of X. The claimed identity follows since

$$I(P,W)=I(X\wedge Y)=I(X\wedge YT)=I(X\wedge Y|T),$$

and for each $x\in\mathcal{X}$ and $y\in\mathcal{Y}_j$, $j=0,1$,

$$\Pr\{X=x,Y=y|T=j\}=\frac{\Pr\{X=x,Y=y\}}{\Pr\{T=j\}}=P(x)W_j(y|x)\ \ .$$