Tight Asymptotic Bounds for the Deletion Channel with Small Deletion Probabilities

Adam Kalai
Microsoft Research New England
Email: adum@microsoft.com

Michael Mitzenmacher
Harvard University
School of Engineering and Applied Sciences
Email: michaelm@eecs.harvard.edu

Madhu Sudan
Microsoft Research New England
Email: madhu@microsoft.com

***Abstract*—In this paper, we consider the capacity C of the binary deletion channel for the limiting case where the deletion probability p goes to 0. It is known that for any $p < 1/2$, the capacity satisfies $C \geq 1 - H(p)$, where H is the standard binary entropy. We show that this lower bound is essentially tight in the limit, by providing an upper bound $C \leq 1-(1-o(1))H(p)$, where the $o(1)$ term is understood to be vanishing as p goes to 0. Our proof utilizes a natural counting argument that should prove helpful in analyzing related channels.**

I. Introduction

The binary deletion channel is modeled as follows: the sender has an input of n bits, each of which is independently deleted by the channel with a fixed probability p; the receiver obtains $\ell \leq n$ bits, without error and in the order in which they were sent.[1] For example, if 10101010 was sent, the receiver wold obtain 10011 if the third, sixth, and eight bits were deleted. The deletion channel, while simple to describe, has proven remarkably challenging to analyze. Indeed, unlike the standard binary erasure and error channels, there is as yet no known closed form for the capacity of the binary deletion channel as a function of p, or even a computationally efficient method for numerically calculating the capacity to a given precision. See the survey [7] for more background.

In this paper we consider bounds on the capacity of the deletion channel in the regime where $p \to 0$. There has long been known a lower bound on the capacity of $1 - H(p)$ for $p < 1/2$, where H is the standard binary entropy [1], [5], [10]. In this paper, we show that this lower bound is essentially tight in the limit, by providing an upper bound $C \leq 1-(1-o(1))H(p)$, where the $o(1)$ term is understood to be vanishing as p goes to 0. This result helps characterize the interesting behavior of the deletion channel. Recent work has shown that in the regime where $p \to 1$, the deletion channel is "like" an erasure channel, in that the capacity can be bounded between $c_1(1-p)$ and $c_2(1-p)$ for appropriate constants $c_1, c_2 < 1$ [2], [4], [8]. Here, we show that as $p \to 0$ the deletion channel is like a binary symmetric error channel, in terms of its capacity, in a much stronger sense.

Upper bounds for the binary deletion channel have only recently become the subject of study. The first upper bounds specifically for this channel were considered in [2], which also considered the asymptotic regime as $p \to 1$. Further techniques introduced in [4] also allowed analysis of the asymptotic as $p \to 0$; this work gave the best previous bound of $C \leq 1 - 4.19p$ as $p \to 0$. Our work, based on a different technique, offers an essentially tight bound in this regime.

After this paper had originally been submitted, a preprint by Kanoria and Montanari was posted online [6], proving a result analogous to ours, in that they also show that the capacity for the deletion channel has an upper bound of $(1-o(1))H(p)$ in the regime as p goes to 0. In private communication, the authors also have explained that a weaker upper bound of $1 + (3/4)p\log_2 p + o(p\log_2 p)$ can be derived from the framework provided by [2]. The proof of [6] makes use of an entirely different approach to ours, and both should provide new insights in how to push forward on studying deletion channels and related synchronization channels.

II. Proof of the Upper Bound

A. Problem Statement and Notation

The capacity C of the deletion channel, where each bit is deleted with some fixed probability $p < 1/2$, satisfies $C \geq 1 - H(p)$. Our goal is to show that this lower bound is essentially tight in the limit where $p \to 0$. Specifically, we wish to show $C \leq 1-(1-o(1))H(p)$, where the $o(1)$ is understood to be a term that is vanishing as p goes to 0.

We will consider codebooks $\mathcal{C} \subseteq \{0,1\}^n$ consisting of messages of n bits and of size $N = |\mathcal{C}|$. We may think of a deletion pattern A as an increasing subsequence of $[n] = \{1, 2, \ldots, n\}$, representing which bits are *not* deleted. We denote a *deletion pattern* by a finite increasing sequence of positive integers, $A = a_1, a_2, \ldots, a_\ell$. The length of sequence is $\mathrm{len}(A) = \ell$, and the number of deletions is $q = n - \ell$. The set of deletion patterns of length ℓ is denoted by

$$P_{\ell,n} = \{a_1, a_2, \ldots, a_\ell \in [n] \mid a_1 < a_2 < \ldots < a_\ell\}.$$

The set of all patterns is $P_n = \bigcup_{\ell=0}^{n} P_{\ell,n}$. For $p \in (0,1)$, the deletion channel can be thought of as choosing a pattern from P_n according to a distribution $\mu_{p,n}$, where each pattern A is chosen with probability $(1-p)^{\mathrm{len}(A)} p^{n-\mathrm{len}(A)}$.

For a string $X \in \{0,1\}^n$, X_A represents the transmission of X through a deletion channel with deletion pattern A in the obvious way: the ith bit of transmission is X_{a_i}. Two transmissions X_A and Y_B are identical if and only if $X_{a_i} = Y_{b_i}$ for all $i \leq \mathrm{len}(A) = \mathrm{len}(B)$. The model of

[1] Several works use d in place of p for the deletion probability.

transmission is that a codeword $Z \in \mathcal{C}$ is chosen uniformly at random, a pattern $A \in P_n$ is chosen according to $\mu_{p,n}$, and then Z_A is received. The decoding algorithm attempts to recover Z. Without loss of generality, we may assume that it is deterministic, and say it computes a function R from the set of received word to codewords. The *success probability* is $\mathbf{Pr}_{Z,A}[R(Z_A) = Z]$.

Let $\lg(x)$ denote the logarithm of x base 2, and H denote the standard entropy function, $H(x) = -x\lg x - (1-x)\lg(1-x)$. We write $\mathbf{Pr}_{x \in_U S}[T(x)]$ to denote the probability of predicate T holding, over x chosen *uniformly* at random from set S.

We make use of the fact that the information capacity and transmission capacity of the deletion channel are the same [3]. Hence, to prove an upper bound on the capacity, we can simply show that a code of sufficiently high rate does not exist. The upper bound on the capacity therefore follows easily from the following theorem, which implies that no code of rate greater than $1 - (1 - o(1))H(p)$ can exist.

Theorem II.1. *Suppose in the setting above there exists a decoding algorithm that succeeds with probability at least δ for a deletion channel with deletion probability p and codeword length $n \geq 12\lg(4/\delta)/p$. Let $q' = (1+\gamma)np$ where $\gamma = 3\lg(4/\delta)/(np)$. Then the number of codewords N satisfies*

$$\lg N \leq n - np(1-\gamma) - \lg\binom{n}{np(1-\gamma)} + \lg\frac{4}{\delta} + \lg\beta$$

where β is given by

$$\beta = \lceil 3q' \lg\frac{ne}{q'} + \lg\frac{4}{\delta}\rceil \left(\frac{6\lceil 3q'\lg\frac{ne}{q'} + \lg\frac{4}{\delta}\rceil}{q'}\right)^{3q'+1}.$$

In particular, $\frac{\lg N}{n} \leq 1-(1-o(1))H(p)$, where the $o(1)$ term is understood as going to 0 in the limit as first n goes to infinity and then as p goes to 0.

We point out no effort has been made to optimize the constants above.

B. Fixed-length deletion channel

For ease of analysis, we first consider the case where the number of received bits is fixed in advance. We then relate this result to the channel with i.i.d. deletions.

In this subsection, we assume n, ℓ and $q = n - \ell$ are known and fixed. We define the (q, n) deletion channel in the natural way: codeword $Z \in \mathcal{C}$ and pattern $A \in P_{\ell,n}$ are chosen uniformly at random, and Z_A is received. A decoding algorithm is successful when, on input Z_A, it outputs Z. We now prove the following:

Theorem II.2. *Let $q \leq n$ and suppose there exists a decoding algorithm that succeeds on the (q, n) deletion channel with probability at least $\delta > 0$, where $n \geq 12\lg(2/\delta)/p$. Then the size of the codebook $N = |\mathcal{C}|$ satisfies*

$$\lg N \leq n - q - \lg\binom{n}{q} + \lg\frac{2}{\delta} + \lg\alpha,$$

where α is given by

$$\alpha = \lceil 3q\lg\frac{ne}{q} + \lg\frac{2}{\delta}\rceil\left(\frac{6\lceil 3q\lg\frac{ne}{q} + \lg\frac{2}{\delta}\rceil}{q}\right)^{3q+1}.$$

While the $\lg\alpha$ term in Theorem II.2 is somewhat difficult, some manipulation gives that when $q = pn$, the result yields $\lg N \leq n(1-(1-o(1))H(p))$ as desired. To see this, note that the $\lg\binom{n}{q}$ term is $n(1-o(1))H(p)$ using standard results (see, e.g., [9][Lemma 9.2]). The $\lg\alpha$ term is dominated by $O(q \cdot \lg\lg(n/q)) = O(np\lg\lg(1/p))$, so the entire expression $n(1-(1-o(1))H(p))$.

We provide some high-level intuition behind the analysis. It is worth first expressing the intuition in terms of the standard binary symmetric error channel. The argument is based on a reduction. Suppose one had a codebook for this channel with N codewords of n bits and a decoding algorithm that could correct for any collection of pn errors perfectly. We could use this codebook and decoding algorithm as a means to represent information as follows. Since there are $\binom{n}{pn} \approx 2^{H(p)n}$ possible error sequences, one could encode (approximately, up to lower order terms) $\lg(N2^{H(p)n})$ bits of information into n bits by taking a codeword, purposely introducing a collection of pn errors, and using the resulting string to represent the information; one could recover the original information by running the decoding algorithm to determine the original codeword and the locations of the errors introduced. Hence, we must have that $\lg(N2^{H(p)n}) \leq n$, or $\frac{\lg N}{n} \leq 1 - H(p)$. This argument, when made suitably rigorous and taking into account the possibility of decoding errors, is a slightly atypical but perfectly reasonable way of viewing the standard Shannon bound.

We utilize the same type of argument here. We show for the deletion channel that if we had a codebook of N codewords with a corresponding decoding algorithm, then when the deletion probability p is suitably small, from the received string and decoding algorithm we can also recover the deletion pattern A itself with nonnegligible probability. Intuitively, this means that if one had a codebook of size N, one could use it to represent information in the same manner as above, so the capacity, given by $\frac{\lg N}{n}$, is also bounded by (approximately) $1 - H(p)$. This argument has a few more complexities in the setting of the deletion channel. For example, if one of the codewords is the all 0's string, we learn nothing about the deletion pattern from the received string. Hence, part of our argument is that there are not so many such "bad" strings where we cannot recover A.

To begin we introduce the *distance* between two deletion patterns of equal length, A and B, denoted by $\Delta(A, B)$, by defining it to be the number of disagreements between a_i and b_i:

$$\Delta(A, B) = \left|\{i \mid a_i \neq b_i\}\right|.$$

We do not define $\Delta(A, B)$ for patterns of unequal length. This definition has the following property.

Lemma II.1. *Take any two length-ℓ deletion patterns, A and B. For uniformly random $X \in \{0,1\}^n$,*

$$\mathbf{Pr}_{X \in_U \{0,1\}^n}[X_A = X_B] \le 2^{-\Delta(A,B)}.$$

Proof: Consider picking the random bits of X in order, one at a time. We call each value i with $a_i \neq b_i$ a discrepancy. Each discrepancy imposes the constraint for $k = \max(a_i, b_i)$ and $j = \min(a_i, b_i)$ that when bit X_k is chosen, it must be equal to bit X_j. This happens with probability exactly $1/2$, independent of which previous constraints have or haven't been satisfied. Moreover, each discrepancy imposes a constraint on a different bit, because each bit is constrained to be equal to at most one of the previous bits; if $i < j$ then $\max(a_i, b_i) < \max(a_j, b_j)$. By independence, the probability that all constraints are satisfied is $2^{-\Delta(A,B)}$. ■

A key technical step is bounding the number of patterns "close" to a given pattern A.

Lemma II.2. *For any pattern $A \in P_{\ell,n}$ and integer $t \ge 1$, the number of patterns $B \in P_{\ell,n}$ such that $\Delta(A,B) \le t$ is at most*

$$(t+1)\binom{2q+t+1}{2q+1}\binom{q+t}{q}.$$

Proof: Fix A. Let $q = n - \ell$ be the number of deletions. Call a bit $i \in [n]$ *clean* with respect to A and B if there is some $j \in [\ell]$ such that $a_j = b_j = i$, i.e., the bit is transmitted in both patterns, in the same position. Call a bit *dirty* otherwise. Let $D(A,B)$ denote the set of dirty bits with respect to patterns A and B. All deletions occur in the dirty bits. The idea is to upper bound the number of *sets* of dirty bits and then upper bound the number of deletion patterns within them.

Intuitively, the idea is that the there are not too many dirty bits and they all must lie "near" to the deletions in A, since a great many bits are clean. There is a simple upper bound on the number of dirty bits:

$$\Delta(A,B) \le t \ \Rightarrow\ |D(A,B)| \le q + t. \qquad (1)$$

This is because, if there are u discrepancies, then there are $q - u$ bits that are dirty because they are deleted in both patterns, and at most $2u$ dirty bits corresponding to the discrepancies where $a_i \neq b_i$ (namely a_i and b_i). Hence there are at most $q + u \le q + t$ dirty bits.

Next, we upper bound the number of possibilities for dirty sets $D(A,B)$. In particular, we will show, that for any fixed A,

$$\left|\left\{D(A,B) \mid B \in P_{\ell,n} \wedge \Delta(A,B) \le t\right\}\right| \le (t+1)\binom{2q+t+1}{2q+1}. \qquad (2)$$

Together with (1), this implies that the number of possible patterns B within t of A is at most $(t+1)\binom{2q+t+1}{2q+1}\binom{q+t}{q}$, because all q deleted bits occur within the set of dirty bits, there are at most $(t+1)\binom{2q+t+1}{2q+1}$ such sets, and each set is of size at most $q+t$.

It remains to show equation (2). For integers $i \le j$, denote by $[i,j]$ the *discrete block* $\{i, i+1, \ldots, j\}$. For $i > j$, let $[i,j] = \emptyset$. The set of bits $[n]$ can be partitioned into alternating discrete blocks of all clean and all dirty bits (it may start with a clean or dirty block). Let $Q = \{d_1, d_2, \ldots, d_q\}$ denote the set of q bits deleted by A. Clearly these are all dirty bits. Moreover, between any two clean blocks, there must be a bit of Q. To see this, consider bits $i < j < k$ such that i and k are clean and j is dirty. Now $a_i = b_i$ and $a_k = b_k$, and some bit between i and k must have been deleted from one of the patterns or else $a_j = b_j$ and j would be clean. If a bit was deleted from pattern B, then a bit from pattern A must also have been deleted in order for the patterns to align at both i and k. Thus, between each two clean blocks, there must be an element of Q.

Hence, each dirty block contains at least one bit from Q, with the possible exception of a dirty block containing 1 and a dirty block containing n. The set $D(A,B)$ can then be described by $2(q+1)$ nonnegative integers, say $r_0, r_1, \ldots, r_q$ and $l_1, l_2, \ldots, l_{q+1}$, where the dirty bits are

$$\begin{aligned} D(A,B) \quad = \quad & Q \cup [1, r_0] \cup [n - l_{q+1} + 1, n] \cup \\ & \bigcup_{i=1}^{q} \big([d_i + 1, d_i + r_i] \cup [d_i - l_i, d_i - 1]\big). \end{aligned}$$

Such a description is not unique (e.g., the above intervals may overlap), but there is always at least one such description that marks each dirty bit exactly once. That it, the description is *frugal*, meaning $r_0 + l_{q+1} + \sum_1^q (r_i + l_i) = |D(A,B)| - q$. A well-known combinatorial fact is that the number of r-tuples of nonnegative integers that sum to s is $\binom{r+s-1}{r-1}$. Hence, if we fix the number of dirty bits to be $d = |D(A,B)|$, then the number of frugal descriptions is at most the number of $(2q+2)$-tuples that sum to $d - q$, or

$$\binom{2q+1+d-q}{2q+1}.$$

As $d \le q + t$ from equation (1), the number of frugal descriptions is at most

$$\binom{2q+t+1}{2q+1}.$$

The number of possible sets of dirty bits is then at most the number of possibilities for $d \in [q, q+t]$, which is $t+1$, times the number of frugal descriptions, which is at most $\binom{2q+t+1}{2q+1}$. This gives equation (2). ■

Again, our high-level goal is to show that if one can decode one can also, with non-negligible probability, recover the deletion pattern A itself. So far we have shown that there are not too many deletion patterns close to any deletion pattern. Now we will use this to show that, for most codewords, we can recover the deletion pattern based on the received sequence. Naturally, this will lead us to an upper bound on the number of possible codewords.

Of course, there are certainly bad possible codewords, like the all 0's string, where we cannot recover the deletion pattern based on the received sequence. To begin, we show there are not too many such strings.

Definition II.1. For $t \geq 1$, We say $X \in \{0,1\}^n$ is *t-bad* if there exist two deletion patterns $A, B \in P_{\ell,n}$ such that $\Delta(A,B) \geq t$ and $X_A = X_B$.

For example, the all 0's and all 1's strings are both bad for all $t \leq \ell$.

Lemma II.3. *For any $t \geq 1$, there are at most $\binom{n}{q}^2 2^{n-t}$ different t-bad strings $X \in \{0,1\}^n$.*

Proof: It is equivalent to show that the probability that a random X is t-bad is at most $\binom{n}{q}^2 2^{-t}$. For any fixed length-ℓ patterns A, B of distance $\Delta(A,B) \geq t$, the probability that a random X has $X_A = X_B$ is at most 2^{-t} by Lemma II.1. By the union bound over all pairs of patterns,

$$\mathbf{Pr}_{X \in_U \{0,1\}^n}\left[\exists A, B \in P_{\ell,n}\ X_A = X_B\right] \leq \binom{n}{q}^2 2^{-t},$$

proving the lemma. ■

The following easy lemma proves useful for bounding the probability of both successfully decoding and recovering the deletion pattern.

Lemma II.4. *Let ρ by a joint distribution over $S \times T$, for finite sets S, T, such that the marginal distribution over S is uniform. Let $g : T \to S$ be a function. Then,*

$$\mathbf{Pr}_{(a,b)\sim\rho}[g(b) = a] \leq \frac{|T|}{|S|}.$$

Proof: This follows from the fact that $g(b) = a$ only if a is in the range of g, which has size at most $|T|$, and hence happens with probability at most $|T|/|S|$. ■

We are now prepared to prove Theorem II.2.

Proof of Theorem II.2: We create a hypothetical *guesser* that, given Z_A for $Z \in \mathcal{C}$ and $A \in P_{\ell,n}$ chosen uniformly at random, will be able to guess both Z and A with nonnegligible probability. Let $q = n - \ell$ be the number of deleted bits. Take $t = \lceil 3q \lg \frac{ne}{q} + \lg \frac{2}{\delta} \rceil$.

On input X, the guesser can run the decoding algorithm to compute the proposed decoding $R(X)$, and then outputs $g(x) = (R(X), B)$, where B is the lexicographically first pattern that satisfies $R(X)_B = X$ if one exists, or is the pattern $B = 1, 2, \ldots, \ell$, otherwise. The success probability of the guesser may be lower-bounded as follows. Let the uniformly random codeword and deletion pattern be $Z \in \mathcal{C}$ and $A \in P_{\ell,n}$, respectively. The decoding succeeds ($R(Z_A) = Z$) *and* the codeword $Z \in \mathcal{C}$ is not t-bad with probability at least

$$\mathbf{Pr}_{Z \in_U \mathcal{C}}[R(Z_A) = Z \wedge Z \text{ is not } t\text{-bad}] \geq \delta - \binom{n}{q}^2 \frac{2^{n-t}}{N} \geq \frac{\delta}{2}.$$

This holds because the probability of success is δ, the probability of a t-bad $C \in \mathcal{C}$ is at most $\binom{n}{q}^2 2^{n-t}/N$ by Lemma II.3, and by our choice of parameters. To see this, note that $\binom{n}{q}^2 \leq (ne/q)^{2q}$ and if $N \geq 2^{n - q\lg(n/q)}$, then the above inequality above holds. If $N < 2^{n-q\lg(n/q)}$, the inequality may not hold, but in this case the theorem follows trivially. By the definition of t-bad, B must satisfy $\Delta(A,B) \leq t-1$. However, by Lemma II.2 and again the fact that $\binom{n}{k} \leq (ne/k)^k$, the number of such patterns is at most

$$\begin{aligned}
&(t-1)\binom{2q+t}{2q+1}\binom{q+t-1}{q} \\
&\quad\leq\quad t\left(e\frac{2q+t}{2q+1}\right)^{2q+1}\left(e\frac{q+t-1}{q}\right)^{q} \\
&\quad\leq\quad t\left(\frac{6t}{q}\right)^{3q+1}.
\end{aligned}$$

Recall that, as given in the statement of the theorem, $\alpha = t\left(\frac{6t}{q}\right)^{3q+1}$.

Conditioned on the decoding succeeding and the codeword not being t-bad, each deletion pattern is equally likely, and hence the lexicographically first pattern is correct with probability at least α^{-1}. Hence, the total success probability of the guesser is at least

$$\mathbf{Pr}_{Z \in_U \mathcal{C}, A \in_U P_{\ell,n}}[g(Z_A) = (Z, A)] \geq \delta\alpha^{-1}/2.$$

However, using Lemma II.4 with the sets $S = \mathcal{C} \times P_{\ell,n}$ and $T = \{0,1\}^\ell$, we also have that this probability is at most $\frac{2^\ell}{N\binom{n}{q}}$. Rearranging terms, we have

$$\lg N \leq \ell - \lg\binom{n}{q} + \lg\frac{2}{\delta} + \lg\alpha.$$

■

C. Proof of Theorem II.1

Going from the exact case of Theorem II.2 to the case where the number of deletions is itself random as in Theorem II.1 merely involves taking advantage of the concentration of the number of deletions around its mean p.

Proof of Theorem II.1: Suppose we have a decoding algorithm for the deletion channel that succeeds on codebook $\mathcal{C}$ with probability $\delta > 0$. We let $\gamma = \sqrt{3\lg(4/\delta)/(np)}$. Assuming as in the theorem statement that $n \geq 12\lg(4/\delta)/p$, we have $\gamma \leq 1/2$. Standard multiplicative Chernoff bounds (such as [9][Corollary 4.6]) guarantee that, with probability at least $\delta/2$, a random deletion pattern will have $q \in [(1-\gamma)pn, (1+\gamma)pn]$. Hence there must be some q^* in this range such that the success probability of the exact (q^*, n) deletion channel is at least $\delta/2$.

Let α^* be given by

$$\alpha^* = \lceil 3q^* \lg\frac{ne}{q^*} + \lg\frac{4}{\delta}\rceil\left(\frac{6\lceil 3q^*\lg\frac{ne}{q^*} + \lg\frac{4}{\delta}\rceil}{q^*}\right)^{3q^*+1}.$$

By Theorem II.2

$$\lg N \leq n - q^* - \lg\binom{n}{q^*} + \lg\frac{4}{\delta} + \lg\alpha^*.$$

Noting that α^* is maximized for the largest possible value of q^* in the range and the other terms are maximized for the smallest values of q^* we have

$$\lg N \leq n - \lg\binom{n}{np(1-\gamma)} - np(1-\gamma) + \lg\frac{4}{\delta} + \lg\beta$$

where

$$\beta = \lceil 3q' \lg \frac{ne}{q'} + \lg \frac{4}{\delta} \rceil \left(\frac{6\lceil 3q' \lg \frac{ne}{q'} + \lg \frac{4}{\delta} \rceil}{q'} \right)^{3q'+1}$$

and $q' = (1+\gamma)np$. To conclude, consider any fixed p with n going to infinity. Note that

$$\lg \binom{n}{np(1-\gamma)} = n(H(p) + o(1))$$

as γ is vanishing. Further, similarly to as we have described previously, the first two terms $n - \lg \binom{n}{np(1-\gamma)}$ dominate the right hand side of the equation; the $\lg \beta$ term can be seen to be $O(np \log\log(1/p)) = o(nH(p))$ as p goes to 0. Dividing through by n we obtain $\frac{\lg N}{n} \leq 1 - (1 - o(1))H(p)$, where the $o(1)$ term is understood as going to 0 in the limit as first n goes to infinity and then as p goes to 0. ■

III. Conclusion

We have considered deletion channels in the limit as the deletion probability $p \rightarrow 0$ and shown that its capacity is at most $1 - (1 - o(1))H(p)$. The intuition behind our argument is simple; one could use a code for such a channel to store information in both the message and deletion pattern, which can be recovered with non-trivial probability given a decoding algorithm. This necessarily limits the capacity of the underlying code. In the full version of the paper, we consider the natural generalizations to insertion channels and other related channels.

IV. Acknowledgments

Michael Mitzenmacher was funded in part by NSF grant CCF-0634923. Part of this work was done while Michael Mitzenmacher was visiting Microsoft Research New England.

References

[1] S. Diggavi and M. Grossglauser. On information transmission over a finite buffer channel. *IEEE Transactions on Information Theory*, 52(3):1226–1237, 2006.

[2] S. Diggavi, M. Mitzenmacher, and H. Pfister. Capacity Upper Bounds for Deletion Channels. In *Proceedings of the International Symposium on Information Theory*, pp. 1716-1720, Nice, France, June 2007.

[3] R. L. Dobrushin. Shannon's Theorems for Channels with Synchronization Errors. *Problems of Information Transmission*, 3(4):11-26, 1967. Translated from *Problemy Peredachi Informatsii*, vol. 3, no. 4, pp 18-36, 1967.

[4] D. Fertonani and T. M. Duman. Novel bounds on the capacity of binary channels with deletions and substitutions. In *Proceedings of the 2009 IEEE International Symposium on Information Theory (ISIT)*, 2009.

[5] R. G. Gallager. Sequential decoding for binary channels with noise and synchronization errors. Lincoln Lab. Group Report, October 1961.

[6] Y. Kanoria and A. Montanari. On the deletion channel with small deletion probability. arXiv:0912.5176v1 [cs.IT]. http://arxiv.org/abs/0912.5176v1.

[7] M. Mitzenmacher. A survey of results for deletion channels and related synchronization channels. *Probability Surveys*, vol. 6, pp. 1-33, 2009.

[8] M. Mitzenmacher and E. Drinea. A simple lower bound for the capacity of the deletion channel. *IEEE Transactions on Information Theory*, 52:10, pp. 4657-4660, 2006.

[9] M. Mitzenmacher and E. Upfal. *Probability and Computing : Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[10] K. S. Zigangirov. Sequential decoding for a binary channel with drop-outs and insertions. *Problems of Information Transmission*, vol. 5, no. 2, pp. 17–22, 1969. Translated from *Problemy Peredachi Informatsii*, vol. 5, no. 2, pp 23–30, 1969.