An approach from classical information theory to lower bounds for smooth codes

Abstract

Let $\mathbf{C} : \{0,1\}^n \mapsto \{0,1\}^m$ be a code encoding an n-bit string into an m-bit string. Such a code is called a (q, c, ϵ) *smooth code* if there exists a probabilistic decoding algorithm which while decoding any bit of the input, makes at most q probes on the code word, the probability with which it looks at any location is at most c/m and the error made by the decoding algorithm is at most ϵ. Smooth codes were introduced by Katz and Trevisan [KT00] in connection with *locally decodable codes*.

For 2-probe smooth codes Kerenidis and de Wolf [KdW03] have shown that $m \geq 2^{\Omega(n)}$ in case c and ϵ are constants. Although the final result is about classical codes, their proof goes through *quantum* information theoretic arguments. These arguments do not seem to extend to codes with higher number of probes.

Using very different classical information theoretic arguments, we show that for 2-probe codes if $\epsilon \leq \frac{c^2}{8n^2}$, then $m \geq 2^{\frac{n}{320c^2}-1}$. While we do not match the bounds shown by Kerenidis and de Wolf we hope that the techniques used in this paper extend to match the bounds shown using quantum arguments. More so, we hope that they extend to show bounds for codes with greater number of probes for which the quantum arguments of Kerenidis and de Wolf apparently break down.

Keywords: Codes, locally decodable, smooth, information theory, lower bounds, quantum.

1 Introduction

Error correcting codes encoding say an n-bit message x to an m-bit code-word $\mathbf{C}(x)$ are primarily designed so as to recover the original message x even if there is some error in the code-word $\mathbf{C}(x)$ during transmission. In many contexts we need to recover only part of the initial encoded message while accessing only a limited number of bits of the corrupted code-word (e.g. *Probabilistically checkable proofs, self-correcting computations and extractors*). This was one of the motivations with which Katz and Trevisan [KT00] defined *Locally decodable codes* (LDC). These codes are also closely connected to another important and well studied problem: *Private information retrieval* (PIR) [CKGS98, Amb97, BIKR02, BFG06, RS06]. Katz and Trevisan defined *Smooth codes* in the same context and showed that existence of LDCs implies existence of smooth codes. The main difference between the two is that instead of allowing for errors in the codeword $\mathbf{C}(x)$ as in LDCs, in smooth codes a certain smoothness condition is forced on the decoding algorithm. Lower bounds for smooth codes imply lower bounds for LDCs which in turn imply bounds on communication for PIR schemes. LDCs and smooth codes have been extensively studied in the last few years [KT00, DJT$^+$02, GKST02, Oba02, KdW03, WdW05, Yek06].

Let us review the definition of smooth codes as in Katz and Trevisan [KT00].

Definition 1 *Let $c > 1, 0 \leq \epsilon < 1/2$ and q be an integer. We call a code $\mathbf{C} : \{0,1\}^n \mapsto \{0,1\}^m$ to be a (q, c, ϵ) smooth code if there exists a non-adaptive probabilistic decoding algorithm $\mathcal{A}$ such that:*

1. **Local decoding:** *In every invocation $\mathcal{A}$ reads at most q indices of the code non-adaptively.*
2. **Correctness:** *For every $x \in \{0,1\}^n$ and every $i \in [n]$, $\Pr[\mathcal{A}(\mathbf{C}(x), i) = x_i] \geq 1 - \epsilon$.*
3. **Smoothness:** *For every $i \in [n]$ and every $j \in [m]$, $\Pr[\mathcal{A}(., i)$ 'reads index' $j] \leq c/m$.*

Remark: Here we consider only non-adaptive codes. For codes with constant number of probes, bounds for non-adaptive codes also imply bounds for adaptive codes [KT00].

It has been quite hard to obtain matching upper and lower bounds for general LDCs and smooth codes. Several special cases of interest have been studied and one such important special case has been the 2-probe codes ($q = 2$). For 2-probe LDCs and smooth codes, Kerenidis and de Wolf [KdW03], using *quantum* information theoretic arguments, showed a lower bound on m that is exponential in n, where c, ϵ are constants. This is one of few examples where quantum arguments lead to classical results. A matching upper bound is provided by the *Hadamard code* [KT00]. However for codes with a higher number of probes ($q \geq 3$), there is a large gap in the existing upper and lower bounds [DJT$^+$02, BIKR02, Woo06, WY05, Yek06]. The quantum arguments of [KdW03] do not seem to imply lower bounds that come close to the existing upper bounds for $q \geq 3$.

In the wake of the above facts it is desirable to get different arguments for showing lower bounds for 2-probe codes, which could potentially be extended to show stronger

bounds for codes with $q \geq 3$. Prior to the work of [KdW03], Goldreich, Karloff, Schulman and Trevisan [GKST02] studied the special case of 2-probe *linear* LDCs; codes in which the encoding function is linear and showed a lower bound on m that is exponential in n. More recently Beigel, Fortnow and Gasarch [BFG06] have shown tighter exponential lower bounds on communication for 2-prover zero-error PIR schemes implying similar bounds for 2-probe zero-error LDCs (i.e. $\epsilon = 0$).

We attempt here one such approach for obtaining alternative arguments for the result of [KdW03] for general 2-probe smooth codes. We show the following:

Theorem 1.1 *Let* $\mathbf{C} : \{0,1\}^n \mapsto \{0,1\}^m$ *be a* $(2, c, \epsilon)$ *smooth code with* $\epsilon \leq \frac{c^2}{8n^2}$. *Then,* $m \geq 2^{\frac{n}{320c^2}-1}$.

Note: Usually bounds for smooth codes imply corresponding bounds for LDCs. However such is not the case with our result since the error parameter that we are considering is much smaller than the smoothness parameter which we are letting to be a constant. Lower bounds for smooth codes when the error is allowed to be a constant do imply bounds for corresponding LDCs.

Our techniques: Our techniques rely on classical information theory as opposed to the quantum information theory arguments of [KdW03]. We avoid the use of the essentially quantum concept of *superpositions* which was critically used by [KdW03] in their reduction of classical 2-probe codes to quantum 1-probe codes. This 2 to 1 quantum reduction, we believe is a critical bottleneck in the arguments of [KdW03] which prevents them from extending to 3-probe codes and beyond. We instead give direct arguments for classical 2-probe codes.

We think of the n-bit message $X \stackrel{\Delta}{=} X_1, \ldots, X_n$ (that is encoded) and the m-bit codeword $Y \stackrel{\Delta}{=} Y_1, \ldots, Y_m$ as correlated random variables (where X_i represents the i-th bit of X and similarly for Y). For each index $i \in \{1, \ldots, n\}$, let us now consider a graph G_i on m vertices with edges being the pair of vertices that are possibly probed by the decoding algorithm $\mathcal{A}$ on input i. Let the error parameter of $\mathcal{A}$ be ϵ (we let ϵ to be $o(1/n^2)$, the reason for which we briefly mention a little later). It was noted by [KT00] that in each G_i there exists a matching M_i of size $\Omega(m)$, of *good* edges such that for each edge in M_i, the decoding algorithm on probing that edge, determines answer correctly with good probability, say $1 - 2\epsilon$. We first note that for a typical good edge say $(j, k) \in M_i$, Y_j and Y_k get *highly* correlated after conditioning on X_i. We then make the central observation that if Y_j, Y_k get highly correlated by conditioning on X_{i_1} and later Y_j, Y_l get highly correlated by conditioning on X_{i_2}, then it implies that Y_k, Y_l get highly correlated on conditioning on both X_{i_1} and X_{i_2}. We then finally observe that this way a lot of random variables in the set $\{Y_1, \ldots, Y_m\}$ start to get correlated with each other fast and this process cannot continue for long.

To make the calculations, we identify a subset GOOD of $\{1, \ldots, n\}$ such that $|\mathsf{GOOD}| = \Omega(n)$ and then successively condition on $X_i, i \in \mathsf{GOOD}$ one by one. For every $j \in \{1, \ldots, m\}$ and $i \in \mathsf{GOOD}$ we maintain a set S_j^i of the random variables Y_k, such that Y_j and Y_k are

highly correlated conditioned on $X_1 \ldots X_i$. We then argue that if $\epsilon = o(1/n^2)$ then for each $i \in \mathsf{GOOD}$, there exists $\Omega(m)$ values of j such that $|S_j^i| \geq 2|S_j^{i-1}|$. This combined with the fact that $|\mathsf{GOOD}| = \Omega(n)$ and $|S_j^i| \leq m$, then finally implies $m \geq 2^{\Omega(n)}$.

Here the reason for $\epsilon = o(1/n^2)$ is roughly that as we increase i, the correlation of random variables $Y_k \in S_j^i$ with Y_j gets weaker with each successive conditioning of X_i's. The reason for this is roughly that as in the example above the *derived* correlation between Y_k, Y_l could be slightly weaker than the correlations between Y_k, Y_j and Y_j, Y_l. Hence, in order to progress till the last element in GOOD with the desired condition on S_j^i's for every i, we get such a requirement on ϵ.

We would like to point out that in the special case when $\epsilon = 0$, our arguments might appear to bear similarities with the arguments of [BFG06]. In this special case [BFG06] could use counting and combinatorial arguments. However we have to resort to information theory to handle the case of $\epsilon > 0$.

Discussion and further work: Although our result here falls short of the optimum bounds shown by [KdW03], we believe that it is a step in the direction of de-quantifying their proof. We hope that the different approach that we take, might possibly be extended to match their bounds. More importantly we hope that these arguments might extend to derive stronger bounds for codes with 3-probes and beyond.

As of now, we do not present here a clean conjecture whose resolution would lead to the extension of the bounds to constant error using classical arguments. However we believe that it should be possible to extent the main intuition in this paper that each good edge (or hyperedge in case of higher number of queries) leads to correlations in corresponding random variables in the code word Y. Since (because of the smoothness of the code), there are several disjoint good (hyper)edges for each $i \in [n]$, several disjoint set of random variables in the set $\{Y_1, \ldots, Y_m\}$ get affected due to conditioning on each X_i. This in general collapses the entropy of the combined random variable Y fast and this process can continue only till the entropy in Y lasts, at the end of which Y can no longer absorb conditioning of further random variables X_i's.

1.1 Organization

In the next section we present some classical information theoretic preliminaries and definitions which we will use in our proof of Theorem 1.1 which we present in the subsequent Section 3.

2 Preliminaries

In this section we briefly review some of the information theory facts that will be useful for us in our proofs in the next section. For an excellent introduction to classical information theory, please refer to the book by Cover and Thomas [CT91]. Several of the facts mentioned below which we state without a proof are taken from this book.

For an integer $n \geq 1$, we let $[n]$ represent the set $\{1, \ldots, n\}$. We let our random variables to be finite valued. Let X, Y be random variables. We will let $H(X), H(X|Y)$ represent the *entropy* of X and the *conditional entropy* of X given Y. We let $I(X:Y) \stackrel{\Delta}{=} H(X) + H(Y) - H(XY) = H(X) - H(X|Y)$ represent the *mutual information* between X and Y. We will use the fact $I(X:Y) \geq 0$, equivalently $H(X) + H(Y) \geq H(XY)$ and $H(X) \geq H(X|Y)$, several times without explicitly mentioning it. We will also use the *monotonicity of entropy* i.e. $H(XY) \geq H(X)$, equivalently $H(Y) \geq I(X:Y)$ several times without explicitly mentioning it. Let X be an m valued random variable, then it follows easily that $H(X) \leq \log_2 m$ (below we always take logarithm to the base 2).

For random variables $X_1, \ldots, X_n$, we have the following *chain rule of entropy*:

$$H(X_1, \ldots X_n) = \sum_{i=1}^{n} H(X_i | X_1 \ldots X_{i-1}) \tag{1}$$

Similarly for random variables $X_1, \ldots, X_n, Y$, we have the following *chain rule of mutual information*:

$$I(X_1 \ldots X_n : Y) = \sum_{i=1}^{n} I(X_i : Y | X_1 \ldots X_{i-1}) \tag{2}$$

Let X, Y, Z be random variables. Then we have the following important *monotonicity relation* of mutual information:

$$I(XY : Z) \geq I(X : Z) \tag{3}$$

All the above mentioned relations also hold for conditional random variables for example, for random variables $X, Y, Z, I(X:Y|Z) \geq 0, H(XY|Z) \geq H(X|Z)$ and so on. Again we may be using the conditional versions of the above relations several times without explicitly mentioning it.

For correlated random variables X, Y, we have the following Fano's inequality. Let $\epsilon \stackrel{\Delta}{=} \Pr[X \neq Y]$ and let $|X|$ represent the size of the range of X. Then

$$H(\epsilon) + \epsilon \log(|X| - 1) \geq H(X|Y) \tag{4}$$

For $0 \leq p \leq 1/2$, we have the bound $H(p) \leq 2\sqrt{p}$.

3 Proof of Theorem 1.1

Let $X \stackrel{\Delta}{=} X_1 \ldots X_n$ be a random variable uniformly distributed in $\{0,1\}^n$ (corresponding to the input being encoded) and X_i correspond to the i-th bit of X. This implies that X_i's are distributed independently and uniformly in $\{0,1\}$. Let $Y \stackrel{\Delta}{=} Y_1 \ldots Y_m$ be a random variable (correlated with X) corresponding to the code, i.e $Y = \mathbf{C}(X)$. Here $Y_j, j \in [m]$ corresponds to the j-th bit of the code.

Let $\mathcal{A}$ be as in Definition 1. Let $0 \leq \epsilon < 1/2$, for $i \in [n]$ let E_i^{ϵ} be the graph on $[m]$ consisting of edges (j,k) such that,

$$\Pr[\mathcal{A}(\mathbf{C}(X), i) = X_i | \mathcal{A} \text{ reads } (Y_j, Y_k)] \geq 1 - \epsilon \tag{5}$$

Following interesting fact can be shown using arguments of Katz and Trevisan [KT00]:

Lemma 3.1 *Let $\mathbf{C} : \{0,1\}^n \mapsto \{0,1\}^m$ be a $(2, c, \epsilon)$ smooth code. Let $E_i^{2\epsilon}$ be as described above. Then for each $i \in [n]$, $E_i^{2\epsilon}$ has a matching M_i of size at least $\frac{m}{4c}$.*

Proof: Using the definition of smooth code we have,

$$\begin{aligned}
1 - \epsilon &\leq \Pr[\mathcal{A}(\mathbf{C}(X), i) = X_i | \mathcal{A}(\mathbf{C}(X), i) \text{ reads } E_i^{2\epsilon}] \Pr[\mathcal{A}(\mathbf{C}(X), i) \text{ reads } E_i^{2\epsilon}] \\
&+ \Pr[\mathcal{A}(\mathbf{C}(X), i) = X_i | \mathcal{A}(\mathbf{C}(X), i) \text{ reads complement of } E_i^{2\epsilon}] \\
&\quad \cdot \Pr[\mathcal{A}(\mathbf{C}(X), i) \text{ reads complement of } E_i^{2\epsilon}] \\
&\leq \Pr[\mathcal{A}(\mathbf{C}(X), i) \text{ reads } E_i^{2\epsilon}] + (1 - 2\epsilon)(1 - \Pr[\mathcal{A}(\mathbf{C}(X), i) \text{ reads } E_i^{2\epsilon}])
\end{aligned}$$

This implies $\Pr[\mathcal{A}(\mathbf{C}(X), i) \text{ reads } E_i^{2\epsilon}] \geq 1/2$. For an edge $e \in E_i^{2\epsilon}$, let $P_e \stackrel{\Delta}{=} \Pr[\mathcal{A}(\mathbf{C}(X), i) \text{ reads } e]$. This implies $\sum_{e \in E_i^{2\epsilon}} P_e \geq 1/2$. Furthermore since $\mathbf{C}$ is a $(2, c, \epsilon)$ smooth code, for every $j \in [m]$, it implies $\sum_{e \in E_i^{2\epsilon} | j \in e} P_e \leq c/m$. Let V be a vertex cover of $E_i^{2\epsilon}$. Therefore,

$$1/2 \leq \sum_{e \in E_i^{2\epsilon} | e \cap V \neq \emptyset} P_e \leq \sum_{j \in V} \sum_{e \in E_i^{2\epsilon} | j \in e} P_e \leq |V| c/m$$

This implies that minimum vertex cover of $E_i^{2\epsilon}$ has size at least $m/2c$. This now implies that $E_i^{2\epsilon}$ has a matching of size at least $m/4c$. ■

We start with the following claim.

Claim 3.2 *Let $(j, k) \in M_i$ and $\epsilon' \stackrel{\Delta}{=} \sqrt{8\epsilon}$. Then, $I(X_i : Y_j Y_k) \geq 1 - \epsilon'$.*

Proof:

$$\begin{aligned}
I(X_i : Y_j Y_k) &= H(X_i) - H(X_i | Y_j Y_k) \\
&\geq 1 - H(2\epsilon) \text{ (from(4) and (5))} \\
&\geq 1 - \sqrt{8\epsilon} \text{ (from the bound } H(p) \leq 2\sqrt{p})
\end{aligned}$$

■

We make the following claim which roughly states that the information about various X_is do not quite go into the individual bits of Y. For $i \in [n]$ let, $\tilde{X}_i \stackrel{\Delta}{=} X_1 \ldots X_{i-1}$.

Claim 3.3

$$\sum_{i \in [n]} \sum_{(j,k) \in M_i} (I(X_i : Y_j | \tilde{X}_i) + I(X_i : Y_k | \tilde{X}_i)) \leq m$$

Proof:

$$\begin{aligned}
\sum_{i\in[n]}\sum_{(j,k)\in M_i} I(X_i:Y_j|\tilde{X}_i)+I(X_i:Y_k|\tilde{X}_i) &\leq \sum_{i\in[n]}\sum_{j\in[m]} I(X_i:Y_j|\tilde{X}_i) \text{ (since } M_i\text{s are matchings)}\\
&= \sum_{j\in[m]}\sum_{i\in[n]} I(X_i:Y_j|\tilde{X}_i)\\
&= \sum_{j\in[m]} I(X:Y_j) \text{ (from (2))}\\
&\leq m \text{ (since } \forall j\in[m], Y_j \text{ is a binary random variable)}
\end{aligned}$$

■

We now have the following claim which roughly states that for a typical edge $(j,k)\in M_i$ there is a substantial increase in correlation between Y_j and Y_k after conditioning on X_i.

Claim 3.4 *Let $\epsilon' \leq \frac{c}{n}$. Then,*

$$\mathbb{E}_{i\in_U[n],(j,k)\in_U M_i}[I(Y_j:Y_k|X_i\tilde{X}_i)-I(Y_j:Y_k|\tilde{X}_i)] \geq 1-5c/n$$

Proof: Let $(j,k)\in M_i$. Since X_i and $\tilde{X}_i$ are independent random variables, this implies $I(X_i:\tilde{X}_i)=0$ and we get:

$$\begin{aligned}
I(X_i:Y_jY_k) &\leq I(X_i:\tilde{X}_iY_jY_k) \text{ (from (3))}\\
&= I(X_i:\tilde{X}_i)+I(X_i:Y_jY_k|\tilde{X}_i) \text{ (from (2))}\\
&= I(X_i:Y_jY_k|\tilde{X}_i)\\
&= I(X_i:Y_j|\tilde{X}_i)+I(X_i:Y_k|\tilde{X}_i)+I(Y_j:Y_k|X_i\tilde{X}_i)-I(Y_j:Y_k|\tilde{X}_i) \text{ (from (2))}
\end{aligned}$$

From Claim 3.2 we get,

$$\begin{aligned}
(1-\epsilon')\sum_i |M_i| &\leq \sum_i\sum_{(j,k)\in M_i} I(X_i:Y_jY_k)\\
&\leq \sum_i\sum_{(j,k)\in M_i} I(X_i:Y_j|\tilde{X}_i)+I(X_i:Y_k|\tilde{X}_i)+I(Y_j:Y_k|X_i\tilde{X}_i)-I(Y_j:Y_k|\tilde{X}_i)
\end{aligned}$$

Claim 3.3 now implies:

$$\begin{aligned}
&\sum_i\sum_{(j,k)\in M_i} I(Y_j:Y_k|X_i\tilde{X}_i)-I(Y_j:Y_k|\tilde{X}_i) \geq (1-\epsilon')\sum_i|M_i|-m\\
\geq\ & (\sum_i|M_i|)(1-\epsilon'-\frac{m}{\sum_i|M_i|}) \geq (\sum_i|M_i|)(1-c/n-4c/n) \text{ (from Lemma 3.1)}
\end{aligned}$$

■

Applying Markov's inequality on the above claim we get:

Claim 3.5 *Let $0 < \delta_1, \delta_2 \leq 1$. There exists a set $\mathsf{GOOD} \subseteq [n]$ and sets $\mathsf{GOOD}_i \subseteq M_i$ such that:*

1. $|\mathsf{GOOD}| \geq (1-\delta_1)n$ *and* $i \in \mathsf{GOOD}, \mathbb{E}_{(j,k)\in M_i}[I(Y_j : Y_k|X_i\tilde{X}_i) - I(Y_j : Y_k|\tilde{X}_i)] \geq 1 - \frac{5c}{\delta_1 n}$
2. $\forall i \in \mathsf{GOOD}, |\mathsf{GOOD}_i| \geq (1-\delta_2)|M_i|$ *and for* $(j,k) \in \mathsf{GOOD}_i, I(Y_j : Y_k|X_i\tilde{X}_i) - I(Y_j : Y_k|\tilde{X}_i) \geq 1 - \frac{5c}{\delta_1\delta_2 n}$

Let $\delta_1 = \delta_2 = 1/2$. Let $\tilde{\epsilon} \triangleq \frac{20c}{n}$. Therefore for $i \in \mathsf{GOOD}$ and $(j,k) \in \mathsf{GOOD}_i$ we have from above,

$$I(Y_j : Y_k|X_i\tilde{X}_i) - I(Y_j : Y_k|\tilde{X}_i) \geq 1 - \tilde{\epsilon} \tag{6}$$

We fix GOOD to have exactly $\frac{1}{2\tilde{\epsilon}} - 2$ elements. For $i \in \mathsf{GOOD}$, let a_i be the index of i in GOOD. For $i \notin \mathsf{GOOD}$, let a_i be the index of largest $i' < i$ in GOOD. For $j \in [m], i \in [n]$, let $S_j^i \triangleq \{l \in [m] : H(Y_j|Y_lX_i\tilde{X}_i) \leq a_i\tilde{\epsilon}\}$. Let $S_j^0 \triangleq \{j\}$.

We show the following main lemma.

Lemma 3.6 *Let $i \in \mathsf{GOOD}, (j,k) \in \mathsf{GOOD}_i$. Then,*

1. $S_j^{i-1} \cap S_k^{i-1} = \emptyset$
2. $S_j^{i-1} \cup S_k^{i-1} \subseteq S_j^i \cap S_k^i$.

Proof: **Part 1:** Let $l \in S_j^{i-1} \cap S_k^{i-1}$. Using standard information theoretic relations it follows:

$$H(Y_kY_j|Y_l\tilde{X}_i) \leq H(Y_k|Y_l\tilde{X}_i) + H(Y_j|Y_l\tilde{X}_i) \leq 2(a_i - 1)\tilde{\epsilon}$$

Since $(j,k) \in \mathsf{GOOD}_i$ and from(6),

$$H(Y_k|\tilde{X}_i) \geq H(Y_k|X_i\tilde{X}_i) \geq I(Y_k : Y_j|X_i\tilde{X}_i) \geq 1 - \tilde{\epsilon}$$

Similarly $H(Y_j|\tilde{X}_i) \geq 1 - \tilde{\epsilon}$. Therefore again from(6),

$$\begin{aligned} H(Y_jY_k|\tilde{X}_i) &= H(Y_j|\tilde{X}_i) + H(Y_k|\tilde{X}_i) - I(Y_j : Y_k|\tilde{X}_i) \\ &\geq 2 - 2\tilde{\epsilon} - \tilde{\epsilon} = 2 - 3\tilde{\epsilon} \end{aligned}$$

Now,

$$\begin{aligned} I(Y_l : Y_jY_k|\tilde{X}_i) &= H(Y_jY_k|\tilde{X}_i) - H(Y_jY_k|Y_l\tilde{X}_i) \\ &\geq 2 - 3\tilde{\epsilon} - 2(a_i - 1)\tilde{\epsilon} \geq 2 - 2(a_i + 1)\tilde{\epsilon} > 1 \text{ (since } a_i \leq \tfrac{1}{2\tilde{\epsilon}} - 2) \end{aligned}$$

This is a contradiction since Y_l is a binary random variable.

Part 2: We show $S_j^{i-1} \cup S_k^{i-1} \subseteq S_j^i$ and $S_j^{i-1} \cup S_k^{i-1} \subseteq S_k^i$ follows similarly. It is easily seen that $S_j^{i-1} \subseteq S_j^i$. Let $l \in S_k^{i-1}$. Since $(j,k) \in \mathsf{GOOD}_i$, from(6),

$$H(Y_j|Y_kX_i\tilde{X}_i) = H(Y_j|X_i\tilde{X}_i) - I(Y_j : Y_k|X_i\tilde{X}_i) \leq 1 - (1 - \tilde{\epsilon}) = \tilde{\epsilon}$$

Now,

$$\begin{aligned} H(Y_j|Y_l X_i \tilde{X}_i) &\leq H(Y_j Y_k|Y_l X_i \tilde{X}_i) \\ &= H(Y_k|Y_l X_i \tilde{X}_i) + H(Y_j|Y_l Y_k X_i \tilde{X}_i) \text{ (from (1))} \\ &\leq H(Y_k|Y_l \tilde{X}_i) + H(Y_j|Y_k X_i \tilde{X}_i) \\ &\leq (a_i - 1)\tilde{\epsilon} + \tilde{\epsilon} = a_i \tilde{\epsilon} \end{aligned}$$

Hence $l \in S_j^i$ and therefore $S_k^{i-1} \subseteq S_j^i$. ■

Our theorem now finally follows.

Proof: **[Theorem 1.1]** Let $i \in \mathsf{GOOD}$. Since $\epsilon \leq \frac{c^2}{8n^2}$, Claim 3.4 holds. Lemma 3.6 implies that for $(j,k) \in \mathsf{GOOD}_i$, either $|S_j^i| = 2|S_j^{i-1}|$ or $|S_k^i| = 2|S_k^{i-1}|$. Then,

$$\sum_j \log |S_j^{i-1}| + |\mathsf{GOOD}_i| \leq \sum_j \log |S_j^i| \tag{7}$$

Let $\tilde{i}$ be the largest $i \in \mathsf{GOOD}$. Now,

$$\begin{aligned} (\frac{n}{40c} - 2)\frac{m}{8c} &\leq \sum_{i \in \mathsf{GOOD}} \frac{|M_i|}{2} \text{ (from Lemma 3.1)} \\ &\leq \sum_{i \in \mathsf{GOOD}} |\mathsf{GOOD}_i| \text{ (from Claim(3.5) and } \delta_2 = 1/2) \\ &\leq \sum_j \log |S_j^{\tilde{i}}| \text{ (from(7))} \\ &\leq m \log m \\ \Rightarrow m &\geq 2^{(\frac{n}{40c} - 2)\frac{1}{8c}} \geq 2^{\frac{n}{320c^2} - 1} \end{aligned}$$

■

Acknowledgment: We thank Andrew Mills, Ashwin Nayak, Jaikumar Radhakrishnan, Alex Samorodnitsky, Pranab Sen, Ronald de Wolf and David P. Woodruff for many useful discussions and comments.

References

[Amb97] A. Ambainis. Upper bound on communication complexity of private information retrieval. In *Proceedings of the 24th International Colloquium on Automata, Languages and Programming, (ICALP)*, pages 401–407, 1997.

[BFG06] R. Beigel, L. Fortnow, and W. Gasarch. A tight lower bound for restricted PIR protocols. In *Proceedings of the 21st IEEE Conference on Computational Complexity*, volume 15(1), pages 82–91, 2006.

[BIKR02] A. Beimel, Y. Ishai, E. Kushilevitz, and J. F. Raymond. Breaking the Order($n^{1/(2k-1)}$) barrier for information-theoretic private information retrieval. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, pages 261–270, 2002.

[CKGS98] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. In *Journal of ACM*, volume 45, pages 965–981, 1998.

[CT91] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley and Sons, 1991.

[DJT+02] A. Deshpande, R. Jain, K. Telikepalli, J. Radhakrishnan, and S.V. Lokam. Better lower bounds for locally decodable codes. In *Proceedings of the 17th IEEE Conference on Computational Complexity*, pages 184–193, 2002.

[GKST02] Oded Goldreich, Howard J. Karloff, Leonard J. Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. In *Proceedings of the 17th IEEE Conference on Computational Complexity*, 2002.

[KdW03] I. Kerenidis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes via quantum argument. In *Proceedings of the 35th ACM Symposium on Theory of Computing*, pages 106–115, 2003.

[KT00] J. Katz and L. Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pages 80–86, 2000.

[Oba02] K. Obata. Optimal lower bounds for 2-query locally decodable linear codes. In *Proceedings of Randomization and Approximation Techniques in Computer Science (RANDOM)*, volume 2483, pages 39–50, 2002.

[RS06] A. Razborov and Yekhanin. S. An Omega($n^{1/3}$) lower bound for bilinear group based private information retrieval. In *Proceedings of the 47th IEEE Symposium on Foundations of Computer Science*, 2006.

[Sam06] Alex Samorodnitsky. An attempt to de-quantify the lower bound for 2-query locally decodable code. Unpublished, 2006.

[WdW05] S. Wehner and R. de Wolf. Improved lower bounds for locally decodable codes and private information retrieval. In *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming, (ICALP)*, pages 1424–1436, 2005.

[Woo06] D.P. Woodruff. New lower bounds for general locally decodable codes. Unpublished, 2006.

[WY05] D.P. Woodruff and S. Yekhanin. A geometric approach to information-theoretic private information retrieval. In *Proceedings of the 20th IEEE Conference on Computational Complexity*, pages 275–284, 2005.

[Yek06] S Yekhanin. New locally decodable codes and private information retrieval schemes. Unpublished, 2006.