Single-Server Private Information Retrieval Schemes are Equivalent to Locally Recoverable Coding Schemes

Swanand Kadhe, *Member, IEEE,* Anoosheh Heidarzadeh, *Member, IEEE,* Alex Sprintson, *Senior Member, IEEE,* and O. Ozan Koyluoglu, *Member, IEEE,*

***Abstract*—The Private Information Retrieval (PIR) problem has recently attracted a significant interest in the information-theory community. In this problem, a client wants to download one or more messages belonging to a database while protecting the identity of the downloaded message(s). In this paper, we focus on the scenarios in which (i) the entire database is stored on a single server and (ii) the client has prior side information, namely a subset of messages unknown to the server. Such prior side information is necessary to enable efficient private information retrieval in the single server scenario.**

In the last decade, there has also been a significant interest in Locally Recoverable Codes (LRCs), a class of storage codes in which each symbol can be recovered from a limited number of other symbols. More recently, there is an interest in *cooperative* locally recoverable codes, i.e., codes in which multiple symbols can be recovered from a small set of other code symbols. The central problem in this context is given a set of code parameters to design an LRC scheme that includes a locally recoverable code along with encoding, decoding, and repair functions.

The paper establishes an equivalence between the single-server PIR schemes and LRC schemes. In particular, we present explicit algorithms that transform a given PIR scheme into an LRC scheme and vice versa. We show that (i) PIR schemes for retrieving a single message are equivalent to classical LRC schemes; and (ii) PIR schemes for retrieving multiple messages are equivalent to cooperative LRC schemes. These equivalence results allow us to recover upper bounds on the download rate for PIR schemes, and to obtain a novel rate upper bound on cooperative LRC schemes. Our results cover schemes based on both linear and non-linear codes.

I. Introduction

The Private Information Retrieval (PIR) problem is one of the important problems in theoretical computer science [2]. The setting of the problem includes a client that needs to retrieve a message belonging to a database with copies stored on a single or multiple remote servers. The message needs to be retrieved in a way that satisfies the privacy condition, i.e., a PIR scheme must prevent the server from identifying the index of the retrieved message. The theoretical computer science community has primarily focused on the settings with small message sizes with the objective to minimize the total number of bits uploaded to and downloaded from the server [3].

Unfortunately, the classical models for private information retrieval were seen limited adoption in practical settings, see, e.g., [4]. One of the reasons is that they do not capture well the requirements and constraints imposed by real-world storage systems. In particular, classical cryptographic PIR schemes are designed to optimize the total amount of communication between the user and the databases, i.e., the sum of lengths of each query (upload) and each answer (download). While this model is adequate for settings with small messages, these techniques do not scale well for large messages.

Starting with the seminal work of Sun and Jafar [5], the multiple-server PIR problem has received a significant attention from the information and coding theory community with breakthrough results in the past few years (see, e.g., [6]–[9], and references therein). The information-theoretic approach has focused on a practical setting with large message sizes with the goal to minimize the ratio of the total number of downloaded bits to the message size.

Recently, the single-server PIR with Side Information (PIR-SI) problem was considered in [10], [11], wherein the client (also called the user) knows a random subset of messages that is unknown to the server. It was shown that the side information enables the user to substantially reduce the download cost and still achieve information-theoretic privacy for the requested message. The multi-message extension of PIR-SI, which enables a user to privately download multiple messages from the server, was considered by Heidarzadeh *et al.* [12] as well as Li and Gastpar [13].

It is well known in theoretical computer science community that there is an *equivalence* between classical multi-server PIR schemes and a class of error-correcting codes called *locally decodable codes* (LDCs) (see, e.g., the surveys [3], [14]). LDCs allow one to *locally* decode an arbitrary message symbol from only a small subset of randomly chosen codeword symbols, even after a fraction of codeword symbols are corrupted by an adversary. The equivalence between PIR and LDC schemes is given in terms of reductions of the following form: the existence of a multi-server PIR scheme guarantees the existence of an LDC (with appropriate parameters), and vice-versa (see, e.g., [3], [14]).

This work was presented in part at the 2019 IEEE Information Theory Workshop (ITW), Aug 2019 [1].

Swanand Kadhe and O. Ozan Koyluoglu are with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mails: {swanand.kadhe, ozan.koyluoglu}@berkeley.edu).

Anoosheh Heidarzadeh and Alex Sprintson are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mails: {anoosheh, spalex}@tamu.edu).

This work was supported in part by NSF grants CCF-1703678 and CNS-1748692.

The work of A. Heidarzadeh and A. Sprintson was supported in part by NSF Grant CCF-1718658.

Continuing with this theme, in this paper, we show that single-server PIR-SI schemes are equivalent to another class of codes with locality called *locally recoverable codes* (LRCs) [15]. To the best of our knowledge, this paper is the first to show an equivalence between a PIR problem and LRCs. LRCs are a class of erasure codes that enable one to recover an erased codeword symbol from only a small subset of other codeword symbols. In particular, in an LRC with block-length n and locality r, every codeword symbol can be recovered (also called *repaired* in the storage context) from at most r other codeword symbols [15]. Rawat *et al.* [16], [17] extended the notion of local recovery to *cooperative local recovery*. Specifically, in a cooperative LRC with block-length n and (r, ℓ) locality, every subset of ℓ codeword symbols can be recovered (or repaired) from at most r other codeword symbols. The central problem in this context is given a set of code parameters to design an LRC scheme that includes a locally recoverable code along with encoding, decoding, and repair functions.

In this paper, we focus on equivalence between PIR-SI and LRC schemes. We say that a PIR-SI scheme and an LRC scheme are *equivalent* if there exists an explicit construction that transforms (reduces) a given PIR-SI scheme into an LRC scheme (with appropriate parameters), and vice versa. In particular, we show that single-message PIR-SI schemes are equivalent to LRC schemes, and multi-message PIR-SI schemes are equivalent to cooperative LRC schemes. We present explicit algorithms that transform a given PIR-SI scheme into an LRC scheme with appropriate parameters, and vice-versa. Our detailed contributions are as follows.

Our Contributions: We focus our attention on the single-server PIR-SI problem in which a user wishes to download D messages from a database of K messages (over a finite field $\mathbb{F}_q$), stored on a single remote server. The user has a random subset of M messages, referred to as *side information*, whose identities are unknown to the server.

First, we focus on the scalar-linear case wherein the answer from the server is of the form $EX_{[K]}$, where E is a $T \times K$ matrix with entries over $\mathbb{F}_q$, and $X_{[K]} = [X_1 \cdots X_K]^T \in \mathbb{F}_q^K$ denotes the set of messages. When the user wishes to protect only the identities of the requested messages, we show the following results:

- Equivalence between single-message ($D = 1$) PIR with Side Information schemes and scalar-linear LRC schemes (Theorem 1). In particular, we show the following: Any solution $E \in \mathbb{F}_q^{T \times K}$ to a single-message PIR-SI problem is a parity-check matrix of an LRC with block-length K, dimension $K - T$, and locality M. Moreover, given a parity check matrix H of an LRC with block-length K, dimension k, and locality M, it is possible to construct a single-message PIR-SI scheme with download rate $1/(K - k)$, where E is a column-permutation of H.
- Equivalence between multi-message ($D \geq 2$) PIR with Side Information schemes and cooperative scalar-linear LRC schemes (Theorem 2).
- As corollaries to Theorems 1 and 2, we derive upper bounds on the download rates for single-message PIR-SI problem (Corollary 1) and multi-message PIR-SI problem (Corollary 3), respectively. In addition, we derive a tight upper bound on the rate of a cooperative LRC for the regime $\ell > r$ (see Corollary 4 and Remark 3). We note that the bound in Corollary 1 coincides with the one derived in [11]. On the other hands, the bounds in corollaries 3 and 4 are new. We compare these bounds with the known bounds in Remarks 2 and 3.

Next, we consider the case when the user wants to protect both the identities of the requested messages and that of the side-information, referred to as (W, S)-PIR-SI.[1] We show an equivalence between capacity-achieving (W, S)-PIR-SI schemes and maximum distance separable (MDS) coding schemes[2] (Theorem 3). Finally, we lift the restriction of scalar-linear solutions, and consider generic PIR-SI and LRC schemes encompassing scalar-linear, vector-linear, and non-linear coding schemes (Theorems 4 and 5).

To show that a PIR-SI scheme can be transformed into an LRC, we use the following key observation. In order to guarantee privacy, it is necessary that the answer from the server in any PIR-SI scheme should satisfy the following condition: any D messages can be recovered from the answer and some set of M other messages.[3] Similarly, an LRC code can be used to construct an answer that satisfies this property. However, this property is not sufficient to ensure privacy. The main challenge in transforming an LRC into a PIR-SI scheme is that, to ensure privacy, one needs to introduce randomization while generating user queries. We achieve this by randomly permuting the messages in a judicious manner such that recovery and privacy are guaranteed.

Related Work: LRCs were introduced in [18]–[20], and the study of the locality property was galvanized with the pioneering work of Gopalan et al. [15], which established a trade-off between the minimum distance of a code and its *locality* analogous to the classical Singleton bound. Since then, a series of results have extended the code distance bound for a given locality for various types of codes along with corresponding *optimal* code constructions achieving the distance bound (see, e.g., [21]–[27], and references therein). The notion of cooperative LRCs was considered in [16], [17].

Both LRCs and PIR-SI schemes have been shown to be related to index coding [28], [29], (see [30] for an excellent recent survey). In [10], [11], it was shown that any solution to the single-message PIR-SI problem must be a solution for a specific class of index coding problems. References [31], [32] showed a duality between index coding and LRCs. We note that it is possible to show that a linear single-message PIR-SI scheme is related to an LRC (with appropriate parameters) by using the results in [11], [31], [32]. However, the results in [11], [31], [32] are not sufficient for proving the reverse direction (i.e., given an LRC scheme, it is possible to design a PIR-SI scheme), and the proof requires new techniques (see

[1]Here, W denotes the demand index set and S denotes the side information index set. We use the term (W, S)-PIR-SI to reflect the fact that the user wants to protect (W, S) jointly.

[2]It is worth noting that an MDS code with dimension k can be considered as an LRC with locality $r = k$.

[3]This is formalized in lemmas 1, 4, and 10 for the single-message scalar-linear case, multi-message scalar-linear case, and single-message non-linear case, respectively.

Sections IV-A and IV-B and Remark 4 for details). In addition, we give a direct and simple proof for both directions for linear as well as non-linear settings without constructing index codes as an intermediate step.

II. Preliminaries

Notation: For integers a and b, $a \mid b$ denotes that a divides b. For a positive integer K, denote $\{1,\dots,K\}$ by $[K]$. Let $\mathbb{F}_q$ denote the finite field of order q, where q is a power of a prime. For a set $\{X_1,\dots,X_K\}$ and a subset $S \subset [K]$, let $X_S = \{X_j : j \in S\}$. For a positive integer P, let $\mathbf{1}_P$ and $\mathbf{0}_P$, respectively, denote the all-one and all-zero row vectors of length P. Let e_j be a unit (row) vector of length K such that its j-th entry is 1 and the other entries are 0. For a set $W = \{W_1, W_2, \dots, W_D\} \subseteq [K]$, let I_W be a $D \times K$ matrix whose i-th row is e_{W_i}. For a $T \times K$ matrix $E \in \mathbb{F}_q^{T\times K}$, let $\langle E\rangle$ denote the row-space of E. For a subset $S \subset [K]$ and a matrix (resp. vector) E, let E_S denote the $T \times |S|$ submatrix (resp. subvector) consisting of the columns (resp. coordinates) of E indexed by S. For a vector v, let $\mathrm{Supp}\,(v)$ denote the support of $\mathbf{v}$. For a subspace $\mathcal{C} \subset \mathbb{F}_q^K$, let $\mathcal{C}^\perp$ be its dual subspace, i.e., $\mathcal{C}^\perp = \{v \in \mathbb{F}_q^K : v^T c = 0, \forall c \in \mathcal{C}\}$.

A. Single-Server PIR with Side Information

We briefly overview the single-server PIR with side information problem [11], [12]. Consider a server containing a database that consists of a set of K messages $\boldsymbol{X}_{[K]} = [\boldsymbol{X}_1 \cdots \boldsymbol{X}_K]^T$, with each message being independently and uniformly distributed over $\mathbb{F}_q$. A user has the knowledge of a subset $\boldsymbol{X}_S$ of messages for some $S \subset [K]$, $|S| = M$. The user is interested in *privately* downloading D ($1 \le D \le K - M$) messages $\boldsymbol{X}_W$ from the server for some $W \subseteq [K] \setminus S$, $|W| = D$. We refer to W as the *demand index set* and $\boldsymbol{X}_W$ as the *demand*. We refer to S as the *side information index set* and $\boldsymbol{X}_S$ as the *side information*.

Let $\boldsymbol{W}$ and $\boldsymbol{S}$ denote the random variables corresponding to the demand and side information index sets, respectively. We assume that the side information index set $\boldsymbol{S}$ is distributed uniformly over over all subsets of $[K]$ of size M, i.e.,

$$p_{\boldsymbol{S}}(S) = \begin{cases} \frac{1}{\binom{K}{M}}, & S \subset [K], |S| = M, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Further, we assume that the demand index set $\boldsymbol{W}$ has the following conditional distribution given S:

$$p_{\boldsymbol{W}|\boldsymbol{S}}(W \mid S) = \begin{cases} \frac{1}{\binom{K-M}{D}}, & W \subseteq [K] \setminus S, |W| = D, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

We assume that the server does not know the realization of the user's side information index set S and demand index set W and only knows the *a priori* distributions $p_{\boldsymbol{S}}(S)$ and $p_{\boldsymbol{W}|\boldsymbol{S}}(W|S)$.

To download the set of messages $\boldsymbol{X}_W$ given the side information $\boldsymbol{X}_S$, the user sends a query $Q^{[W,S]}$ to the server. We assume that the query is a stochastic function of (W, S) and is independent of the messages in the side information.[4] The server responds to the query it receives with an answer $A^{[W,S]}$ over $\mathbb{F}_q^T$. We assume that the answer is a deterministic function of the query and the messages. Let $\boldsymbol{Q}^{[W,S]}$ and $\boldsymbol{A}^{[W,S]}$ be the corresponding random variables.

Definition 1. *Any scheme consisting of a query and an answer is referred to as the PIR with side information (PIR-SI) scheme if the query and answer satisfy the following two conditions.*

1. ***Recoverability:*** *From the answer $A^{[W,S]}$ and the side information X_S, the user should be able to decode the desired set of messages X_W for any (W, S), i.e.,*
$$H\left(\boldsymbol{X}_{\boldsymbol{W}} \mid \boldsymbol{A}^{[\boldsymbol{W},\boldsymbol{S}]}, \boldsymbol{Q}^{[\boldsymbol{W},\boldsymbol{S}]}, \boldsymbol{X}_{\boldsymbol{S}}, \boldsymbol{W}, \boldsymbol{S}\right) = 0. \tag{3}$$
2. ***Privacy****: We consider two notions of privacy as follows.*
   (i) *W-**privacy**: The server cannot infer any information about the demand index set from the query, answer, and messages, i.e.,*
   $$I\left(\boldsymbol{W}; \boldsymbol{Q}^{[\boldsymbol{W},\boldsymbol{S}]}, \boldsymbol{A}^{[\boldsymbol{W},\boldsymbol{S}]}, \boldsymbol{X}_{[K]}\right) = 0. \tag{4}$$
   (ii) *(W, S)-**privacy**: The server cannot infer any information about the demand index set as well as the side information index set from the query answer, and messages, i.e.,*
   $$I\left(\boldsymbol{W}, \boldsymbol{S}; \boldsymbol{Q}^{[\boldsymbol{W},\boldsymbol{S}]}, \boldsymbol{A}^{[\boldsymbol{W},\boldsymbol{S}]}, \boldsymbol{X}_{[K]}\right) = 0. \tag{5}$$

We refer to the case of $D = 1$ as single-message PIR-SI, while the case of $D \ge 2$ as multi-message PIR-SI.

The *rate* of a PIR-SI scheme is defined as the ratio of the message length ($\log q$ bits) to the total length of the answers (in bits) as follows:[5]

$$R = \frac{D \log_2 q}{H\left(\boldsymbol{A}^{[\boldsymbol{W},\boldsymbol{S}]}\right)}, \tag{6}$$

where $H(\cdot)$ denotes the Shannon entropy measured in bits. The *capacity* of W-PIR-SI (resp. (W, S)-PIR-SI) is defined as the supremum of rates over all W-PIR-SI (resp. (W, S)-PIR-SI) schemes for a given K, M, and D.

B. Locally Recoverable Codes

Let $\mathcal{C}$ denote a linear $[n, k]$ code over $\mathbb{F}_q$ with block-length n and dimension k. We say that the i-th coordinate of a code $\mathcal{C}$ has locality r if, for every codeword $\mathbf{c} \in \mathcal{C}$, the value of $\mathbf{c}_i$ can be recovered from some other r symbols of $\mathbf{c}$. The formal definition of locality is as follows (see [15]).

Definition 2. *An $[n, k]$ code $\mathcal{C}$ has (all-symbol)* locality *r, if for every coordinate $i \in [n]$, there exists a set $R(i) \subset [n] \setminus \{i\}$, $|R(i)| \le r$, called a repair group, such that, for every codeword $\mathbf{c} \in \mathcal{C}$, the symbol $\mathbf{c}_i$ is a linear function of the*

[4] Note that, in general, the query may depend on the content of the side information. However, a query that does not depend on the content of the side information is *universal* in the sense that it achieves privacy for all realizations of the messages. Therefore, we restrict our attention to such universal queries in this work, similar to [10]–[12].

[5] We focus our attention to the download rate similar to [5]. This is because the download rate dominates the total communication rate when the message size is sufficiently large as compared to the size of a query.

symbols $\mathbf{c}_{R(i)}$*. In other words, for every* $\mathbf{c} \in \mathcal{C}$*, it holds that* $\mathbf{c}_i = \sum_{l \in R(i)} \lambda_l \mathbf{c}_l$*, for some* $\lambda_l \in \mathbb{F}_q$*,* $\forall\, l \in R(i)$*. An LRC with these parameters is referred to as an* $[n, k, r]$ *LRC.*

We refer to the linear function used to recover the i-th coordinate as its *repair function*. It is straightforward to see that the i-th coordinate of $\mathcal{C}$ has locality r, if and only if the dual code $\mathcal{C}^{\perp}$ contains a codeword $\mathbf{c}'$ of Hamming weight at most $r+1$ such that the $i \in \text{Supp}\,(\mathbf{c}')$. To see this, note from Definition 2 that we have $\mathbf{c}_i = \sum_{l \in R(i)} \lambda_l \mathbf{c}_l$, and thus, there exists $\mathbf{c}' \in \mathcal{C}^{\perp}$ such that $\text{Supp}\,(\mathbf{c}') = \{i\} \cup R(i)$, $\mathbf{c}'_l = \lambda_l$ for each $l \in R(i)$ and $\mathbf{c}'_i = 1$. Therefore, the repair functions for all coordinates of $\mathcal{C}$ can be concisely represented using a parity-check matrix of $\mathcal{C}$.

Example 1. *Let us consider a* $[7, 3]$ *Simplex code* $\mathcal{C}$*, which is a dual of a* $[7, 4]$ *Hamming code. In particular,* $\mathcal{C}$ *encodes three information symbols* $\{a, b, c\}$ *into seven symbols as* $\{a, b, c, a+b, a+c, b+c, a+b+c\}$*. It is easy to see that any symbol can be recovered from two other symbols. For instance,* a *can be recovered from* $b+c$ *and* $a+b+c$*.*[6]

In [15], it is shown that the minimum distance $d_{min}\,(\mathcal{C})$ of an $[n, k, r]$ LRC $\mathcal{C}$ is upper bounded as

$$d_{min}\,(\mathcal{C}) \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2. \tag{7}$$

Further, it is shown that any systematic code with locality for information symbols that achieves equality in (7) must follow a specific *structure* [15]. We state below the structure theorem [15, Theorem 9], adapted to the form useful for our setup.

Proposition 1. *[15] Let* $\mathcal{C}$ *be an* $[n, k, r]$ *code, where* $r \mid k$*,* $r < k$*, and* $n = k + k/r$*. Define* $\Gamma\,(i) = R(i) \cup \{i\}$*. Then, for any* $i, j \in [n]$*,* $i \neq j$*, we have either* $\Gamma\,(i) = \Gamma\,(j)$ *or* $\Gamma\,(i) \cap \Gamma\,(j) = \emptyset$*.*

LRCs (and in general erasure codes) are typically used to add some redundancy to the data. Therefore, along with an $[n, k]$ code $\mathcal{C}$, one needs to specify (i) an encoding function that maps messages from $\mathbb{F}_q^k$ to the codewords of $\mathcal{C}$, (ii) the corresponding decoding function that maps the codewords back to the messages, and (iii) repair functions for all coordinates. We refer to an LRC $\mathcal{C}$ together with its encoding, decoding, and repair functions as a locally recoverable coding (LRC) scheme. For linear codes, a parity-check matrix suffices to concisely represent encoding and decoding functions (see, e.g., [34]).[7] As discussed before Example 1, repair functions for a linear LRC $\mathcal{C}$ can also be concisely represented using a parity-check matrix of $\mathcal{C}$. Thus, we describe a linear LRC scheme using a parity-check matrix of the underlying LRC. Later, when we consider non-linear schemes in Sec. V, we explicitly consider encoding, decoding, and repair functions.

[6]In fact, every symbol of the $[7, 3]$ simplex code has three disjoint repair groups [33]. Further, note that, even though the $[7, 3]$ simplex code is not optimal with respect to the distance upper bound in (7), it is optimal with respect to a field size dependent rate upper bound established in [33].

[7]In practice, a systematic encoding may be preferred. However, we do not focus on this aspect in this paper.

C. Cooperative Locally Recoverable Codes

Let $\mathcal{C}$ denote a linear $[n, k]$ code over $\mathbb{F}_q$ with block-length n and dimension k. We say that the code has (r, ℓ)-cooperative locality if, for every codeword, it is possible to repair any ℓ symbols from at most r other symbols. The formal definition is as follows (see [16]).

Definition 3. *We say that an* $[n, k]$ *code* $\mathcal{C}$ *has* (r, ℓ)*-cooperative locality, if for any subset of* ℓ *coordinates* $\Delta \subset [n]$*,* $|\Delta| = \ell$*, there exists a set* $R(\Delta) \subset [n] \setminus \Delta$*,* $|R(\Delta)| \leq r$*, such that, for every codeword* $\mathbf{c} \in \mathcal{C}$*, the symbols* $\mathbf{c}_{\Delta}$ *are linear functions of the symbols* $\mathbf{c}_{R(\Delta)}$*. An LRC with these parameters is referred to as an* $[n, k, (r, \ell)]$ *cooperative LRC.*

Note that when $\ell = n - k$ and $r = k$, then the above definition coincides with that of an MDS code. In [17], it is shown that the minimum distance $d_{min}\,(\mathcal{C})$ of an $[n, k, (r, \ell)]$ cooperative LRC $\mathcal{C}$ for $r \geq \ell$ is upper bounded as

$$d_{min}\,(\mathcal{C}) \leq n - k + 1 - \ell \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right). \tag{8}$$

We refer to a cooperative LRC $\mathcal{C}$ together with its encoding, decoding, and repair functions as a cooperative locally recoverable coding (LRC) scheme. For linear codes, a parity-check matrix suffices to concisely represent encoding, decoding, and repair functions (similar to LRCs). Thus, we describe a linear cooperative LRC scheme using a parity-check matrix of the underlying cooperative LRC.

III. Equivalence Results for Scalar-Linear Schemes

In this section, we consider scalar-linear PIR-SI schemes. In particular, for any given query $Q^{[W,S]}$, the answer $\boldsymbol{A}^{[W,S]}$ can be specified as

$$\boldsymbol{A}^{[W,S]} = E \boldsymbol{X}_{[K]}, \tag{9}$$

where the matrix $E \in \mathbb{F}_q^{T \times K}$ is a deterministic function of the query $Q^{[W,S]}$. We refer to E as a *solution* to the PIR-SI problem. Note that E should be constructed in such a way that it respects the recoverability condition (3) and privacy condition (4) or (5).[8] Further, notice that T, the number of rows of E, denotes the number of symbols (each of which is an element from $\mathbb{F}_q$) downloaded from the server.

A. Single-Message PIR-SI Schemes and LRCs

First, we show that a single-message W-PIR-SI solution is a dual of an LRC with appropriate parameters.

Theorem 1. *For the single-message* W*-PIR-SI problem with* K *messages and side-information size* M*, any scalar-linear solution* $E \in \mathbb{F}_q^{T \times K}$ *must be a parity check matrix of an LRC with block length* $n = K$*, dimension* $k = K - T$*, and locality* $r = M$*. Moreover, given a parity check matrix* H *of an LRC with block-length* $n = K$*, dimension* k $(< n)$*, and locality* $r = M$*, it is possible to construct a single-message* W*-PIR-SI*

[8]Specific requirements on any feasible solution E that are posed by the recoverability and privacy conditions are characterized in Lemmas 1, 4, and 7 for various settings.

scheme with download rate $1/(n-k)$, such that the solution E is a column-permutation of H.

Theorem 1 enables us to use (7) to obtain an upper bound on the capacity of a (scalar-linear) single-message PIR-SI scheme. As we show next, the bound coincides with the upper bound derived in [10], [11]. The proof is quite straightforward, but we present it in Appendix A for completeness.

Corollary 1. *The scalar-linear capacity of the single-message PIR-SI problem is upper bounded by $\lceil K/(M+1)\rceil^{-1}$.*

Remark 1. *The above result can be directly proved using an upper bound on the rate of an LRC with locality r given as $r/(r+1)$ (see [25, Theorem 1]). It is interesting to note that [25, Theorem 1] uses an argument based on acyclic induced subgraphs, similar to arguments in [11] used to establish the capacity upper bound for the single-message W-PIR-SI problem.*

Now, consider the Partition-and-Code scheme proposed in [11] for the single-message W-PIR-SI problem. Let $K = \alpha(M+1)+\beta$ for some $\alpha > 0$ and $0 \leq \beta < M+1$. In the Partition-and-Code scheme, the user first randomly partitions the K messages into α parts of size $M+1$ and one part of size β, such that one of the parts is $W \cup S'$ for some $S' \subseteq S$. The user then asks the server to send the sum of messages in each subset. Therefore, the server sends $\alpha+1$ symbols.

More generally, instead of asking the server to send the sum of messages in each subset, the user can ask the server to send an arbitrary linear combination for each subset by specifying the coefficient. Therefore, the Partition-and-Code scheme yields a solution E of size $(\alpha+1)\times K$ with the following form (up to column permutation):

$$E = \begin{bmatrix} \times & \cdots & \times & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \times & \cdots & \times & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \times & \cdots & \times \end{bmatrix}, \quad (10)$$

where $\times$ denotes a non-zero element in $\mathbb{F}_q$, and the number of non-zero entries is $M+1$ in all but the last row and β in the last row.

According to the capacity upper bound of Corollary 1, we say that a scalar-linear solution to the single-message W-PIR-SI problem is an *optimal* solution if $T = \lceil K/(M+1)\rceil$. Then, Proposition 1, which dictates the structure of an LRC achieving the Singleton-like bound in (7) with equality, implies the following structure on any optimal scalar-linear solution to the single-message W-PIR-SI problem.

Corollary 2. *When $(M+1) \mid K$, any optimal scalar-linear solution E to the single-message W-PIR-SI problem can be converted to the following form using elementary row operations and column permutations:*

$$E = \begin{bmatrix} \times & \cdots & \times & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \times & \cdots & \times & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \times & \cdots & \times \end{bmatrix}, \quad (11)$$

where $\times$ can be any non-zero element in $\mathbb{F}_q$, i.e., $\times \in \mathbb{F}_q \setminus \{0\}$, and the number of non-zero entries in each row is $M+1$.

Since the solution obtained using the Partition-and-Code scheme (cf. (10)) has the same form as in (11), this shows the *uniqueness* of the solution obtained by the partition-and-code scheme. In other words, any optimal scalar-linear solution can be obtained from the Partition-and-Code solution using elementary row operations and column permutations.

B. Multi-Message PIR-SI and Cooperative LRCs

In this section, we consider the multi-message W-PIR-SI problem, wherein the user wants to privately retrieve $D > 1$ messages, and show that a multi-message W-PIR-SI solution is a dual of a cooperative LRC with appropriate parameters.

Theorem 2. *For the multi-message W-PIR-SI problem with K messages, side-information size M, and demand size D, any scalar-linear solution $E \in \mathbb{F}_q^{T\times K}$ must be a parity check matrix of an LRC with block-length $n = K$, dimension $k = K-T$, and (M,D)-cooperative locality. Moreover, given a parity check matrix H of an LRC with block-length $n = K$, dimension k $(< n)$, and (M,D)-cooperative locality, it is possible to construct a multi-message W-PIR-SI scheme with download rate $1/(n-k)$, such that the solution E is a column-permutation of H.*

Theorem 2 enables us to use (8) to obtain an upper bound on the capacity of a (scalar-linear) multi-message W-PIR-SI scheme. The proof is straightforward and presented in Appendix B for completeness.

Corollary 3. *For $M \geq D$, the scalar-linear capacity of the multi-message PIR-SI problem is upper bounded by $D/\lceil DK/(M+D)\rceil$.*

Remark 2. *It is worth noting that it is an open problem to determine whether the above upper bound is tight (this is unlike Corollary 1, where the upper bound for the single-message case is tight).*[9] *The achievability schemes presented in [12], [13] have smaller rate than $D/\lceil DK/(M+D)\rceil$, see, e.g., [12, Theorem 2].*

Next, using Theorem 2, we can derive an upper bound on the rate of a linear cooperative LRC for $\ell > r$ as follows. The proof is straightforward and presented in Appendix C for completeness.

Corollary 4. *For $\ell > r$, the rate of a linear $[n,k,(r,\ell)]$ cooperative LRC is upper bounded by r/n.*

Remark 3. *Corollary 4 yields a better bound on the rate of a cooperative LRC for $\ell > r$ than [17, Corollary 1], which gives the rate upper bound of $r/(r+\ell)+\ell^2/(nr)$. Note that $r/n < r/(r+\ell)+\ell^2/(nr)$, since $n \geq (r+\ell)$. Moreover, the rate bound of r/n is tight for $n > 2r$. This is because an $[n,r]$ MDS code trivially has (r,ℓ)-cooperative locality for*

[9] We note that in the proof of Corollary 3 (in Appendix B), we use the rate upper bound on $[n,k,(r,\ell)]$ cooperative LRC of $r/(r+\ell)$ [17, Corollary 1]. It is an open question whether this rate bound is tight for cooperative LRCs.

any ℓ. Specifically, in an $[n, r]$ MDS code, any r symbols can recover all the remaining $n - r$ symbols.

Interestingly, Theorem 2 also enables us to obtain *computationally efficient* solutions for the multi-message W-PIR-SI problem. In particular, for $D \leq M$, the schemes in [35] (see also [36]) rely on generalized Reed-Solomon codes, and thus, require a finite field size at least $D + \lfloor M/D \rfloor$. On the other hand, it is possible to use constructions of cooperative LRCs to obtain PIR-SI schemes over smaller field size.[10] As an example, an $[n = 2^k - 1, k]$ simplex code has $(\ell + 1, \ell)$-cooperative locality for any $1 \leq \ell \leq (n-1)/2$ (see [17]). Thus, it is possible to obtain multi-message W-PIR-SI solutions over the binary field when $K = 2^t - 1$ for a positive integer t, $1 \leq D \leq (K-1)/2$, and $M = D + 1$.

C. (W, S)-Private PIR-SI Schemes and MDS Codes

Finally, we show an equivalence between a solution to the (W, S)-PIR-SI problem and a maximum distance separable (MDS) code. The result holds for any $1 \leq D \leq K - M$.

Theorem 3. *Any scalar-linear solution E to the (W, S)-PIR-SI problem with $T = K - M$ must be a parity-check matrix of a $[K, M]$ MDS code. Moreover, if H is a parity check matrix of a $[K, M]$-MDS code, then $E = H$ is a solution to the (W, S)-PIR-SI problem.*

It is worth noting that the achievability schemes in [11], [12] for (W, S)-privacy are based on MDS codes. Here, we are showing a stronger result that *any* scalar-linear solution E to the (W, S)-PIR-SI problem (with $T = K - M$) must be a parity-check matrix of an MDS code. Moreover, it is proved in [11], [12] that for any (W, S)-PIR-SI scheme, the server should send at least $K - M$ symbols. Therefore, Theorem 3 essentially considers *capacity-achieving* (W, S)-PIR-SI schemes.

IV. Proofs of Main Results

A. Proof of Theorem 1

Single-Message W-PIR-SI $\implies$ LRC: First, we note that the following necessary condition is imposed by the privacy and recoverability conditions.

Lemma 1. *For any query $Q^{[W,S]}$, the solution E to the single-message W-PIR-SI problem must satisfy the following necessary condition: for any candidate demand index $W' \in [K]$, there must exist a potential side information index set $S' \subseteq [K] \backslash W'$, $|S'| \leq M$ such that it is possible to recover W' from $EX_{[K]}$ and $X_{S'}$. In other words, the following condition must hold:*

$$e_{W'} \in \left\langle \begin{bmatrix} E \\ I_{S'} \end{bmatrix} \right\rangle. \tag{12}$$

Proof: If the aforementioned necessary condition does not hold, then the server will learn from E that W' is not the user's demand index. Indeed, since E is the solution corresponding to the query $Q^{[W,S]}$, we have

$$\mathbb{P}\left(\boldsymbol{W} = W' \mid \boldsymbol{Q}^{[\boldsymbol{W},\boldsymbol{S}]} = Q^{[W,S]}, \boldsymbol{A}^{[\boldsymbol{W},\boldsymbol{S}]} = A^{[W,S]}, \boldsymbol{X}_{[K]} = X_{[K]}\right) = \tag{13}$$

which, in turn, implies that $I\left(\boldsymbol{W}; \boldsymbol{Q}^{[\boldsymbol{W},\boldsymbol{S}]}, \boldsymbol{A}^{[\boldsymbol{W},\boldsymbol{S}]}, \boldsymbol{X}_{[K]}\right) > 0$. This violates the W-privacy condition (4). ∎

Next, we show that a solution to the SM-PIR-SI problem is a parity-check matrix of an LRC with appropriate parameters.

Lemma 2. *Any scalar-linear solution E to the single-message W-PIR-SI problem must be a parity-check matrix of a $[K, K - T, M]$ LRC.*

Proof: The necessary condition (12) in Lemma 1 implies that for every $W' \in [K]$, $\langle E \rangle$ must contain a vector $\mathbf{v}$ of Hamming weight at most $M + 1$ such that $W' \in \text{Supp}(\mathbf{v})$. According to Definition 2, $\langle E \rangle^{\perp}$ is an LRC with block-length K and all-symbol locality M. ∎

Example 2. *Consider a single-message PIR-SI problem with $K = 9$ messages $[X_1\ X_2\ \cdots\ X_9]$. Suppose the user has $\{X_4, X_9\}$ and wants X_1. Consider the solution generated by the Partition-and-Code scheme in [11]. For the case of $(M+1) \mid K$, the Partition-and-Code scheme first partitions the messages into $K/(M+1)$ sets such that one of the sets consists of the demand and the side-information and the remaining sets are formed by randomly partitioning the remaining messages. Suppose these sets are $P_1 = \{X_1, X_4, X_9\}$, $P_2 = \{X_5, X_7, X_8\}$, and $P_3 = \{X_2, X_3, X_6\}$. The server sends the sum of messages in each set P_i. Thus, the answer from the server is $A = \{X_1 + X_4 + X_9, X_5 + X_7 + X_8, X_2 + X_3 + X_4\}$. The solution matrix E for this scheme is as follows:*

$$E = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{14}$$

It is easy to see that E above is a parity-check matrix of a $[n = 9, k = 6, r = 2]$ LRC. To see that the locality is $r = 2$, observe that any symbol can be recovered from the two other symbols due to the parity-check equations.

On exploiting the connection with index coding: At a high level, *index coding* [28], [29] consists of a server with K messages, and a number of clients, each of which is interested in one message and knows some subset of the other messages as side information. The set of all clients' demand and side information is referred to as an *instance* of the index coding problem. The server broadcasts some coded symbols, each of which is some function of the messages (e.g., a linear combination of the messages, in the case of scalar-linear schemes), and the goal is to minimize the number of coded symbols broadcast by the server. When the number of clients is K with every client requiring a distinct message, the instance can be concisely represented by using a directed graph, called the *side information graph*. The side information graph consists of K vertices, one for each message, and there is an arc (i, j) if the client requiring message i knows message j.

[10] Note that small field size schemes obtained from cooperative LRCs may have smaller download rate than those in [35], [36].

It was shown in [11, Lemma 1] that, for the single-message W-PIR-SI problem, the answer from the server must be a solution to an index coding problem in which the out-degree of each vertex in the side information graph is at most M. In [31], the authors consider *generalized* locally repairable codes, in which, every coordinate is decodable from a specific recoverability set of other coordinates. These requirements can be represented using a directed graph, called the *recoverability graph*. Theorem 1 in [31] proves that the dual linear subspace of an LRC is a solution to an index coding instance where the side information graph is the recoverability graph of the LRC.

Using these two results, it is straightforward to show that the dual linear subspace of a single-message W-PIR-SI solution for K messages and side information size M must be an LRC with length K and locality M. Using Lemmas 1 and 2, we presented a direct and simple proof for this result, without using the connection to index coding. In the reverse direction, the dual linear subspace of an LRC with locality M is a solution to some index coding instance in which the out-degree of each vertex is at most M by [31, Theorem 1]. However, any index coding solution does not guarantee privacy and recoverability of the desired demand index with a given side information index set. Therefore, the results of [11], [31], [32] are not sufficient to show the reverse direction.

LRC $\implies$ single-message W-PIR-SI: Here, we show that one can construct a solution to the single-message W-PIR-SI problem given a parity-check matrix of an LRC with appropriate parameters.[11]

Lemma 3. *Given a parity-check matrix of a $[K, K-T, M]$ LRC, it is possible to construct a single-message W-PIR-SI scheme such that the solution E is a column-permutation of H.*

Proof: We present a constructive proof. In the rest of the proof, we consider all sets as ordered sets (with a natural ascending order). For a given W and S, the user first finds a permutation π on $[K]$ as follows. Choose an index W' uniformly at random from $[K]$, independent of W and S. Let $R(W')$ be a repair group of W'. If a coordinate has multiple repair groups, arbitrarily choose one repair group.[12] By the definition of locality, we have $|R(W')| \leq M$. For simplicity, we assume that every repair group of any symbol is of size M. (The arguments can be easily generalized to the case when some repair groups are smaller than M.) Let $R'(W')$ be a random permutation of $R(W')$. Let $P = [K] \setminus \{W \cup S\}$, and P' be a random permutation of $[K] \setminus \{W' \cup R(W')\}$. Let π be the permutation that maps W to W', S to $R'(W')$, and P to P'. The user sends π as its query $Q^{[W,S]}$. The server then applies π to the columns of H to obtain E, i.e., $E_i = H_{\pi(i)}$ for each $i \in [K]$, where the subscript denotes the column of the matrix. Then, the server computes the answer as EX.

[11]Similar to the forward direction, we do not use any connection to index coding, and present a direct proof. However, it is worth noting that the proof technique can be easily adapted to show that it is possible to construct a single-message W-PIR-SI scheme using a solution to an index coding instance in which out-degree of each vertex in the side information graph is at most M.

[12]This arbitrary choice of a repair group for each coordinate is made *a priori*, and is known to the server as part of the scheme.

Next, we show that the above scheme satisfies the recoverablity and W-privacy conditions. Indeed, by the definition of locality for W', $\langle H \rangle$ contains a vector whose support is $W' \cup R(W')$. Therefore, by the construction of E, $\langle E \rangle$ contains a vector whose support is $W \cup S$. Hence, the recoverability condition in (3) is satisfied.

For the W-privacy, it suffices to show that, for any $W \in [K]$ and any permutation π,

$$\mathbb{P}\left(\boldsymbol{Q}^{[\boldsymbol{W},\boldsymbol{S}]} = \pi \mid \boldsymbol{W} = W\right) = \frac{1}{K!}. \tag{15}$$

This is because, using (15), it is straightforward to show that

$$\mathbb{P}\left(\boldsymbol{W} = W \mid \boldsymbol{Q}^{[\boldsymbol{W},\boldsymbol{S}]} = \pi, \boldsymbol{X}_{[K]}, \boldsymbol{A}^{[\boldsymbol{W},\boldsymbol{S}]}\right) = \mathbb{P}\left(\boldsymbol{W} = W\right),$$

from which the privacy condition (4) follows.

Now, we give a proof of (15). Observe that the query generation process first maps the demand index to a random index in $[K]$. Let $\boldsymbol{W'}$ denote that random index. Let $\boldsymbol{R'}(\boldsymbol{W'})$ and $\boldsymbol{P'}$ be random variables corresponding to (independent) uniform random permutations of $R(\boldsymbol{W'})$ and $[K] \setminus \{\boldsymbol{W'} \cup R(\boldsymbol{W'})\}$, respectively. Now, given a permutation π on $[K]$ as a query, define the following events:

$$\begin{aligned}
\mathcal{E}_1 &= \left\{\boldsymbol{W'} = \pi\left(\boldsymbol{W}\right)\right\}, \\
\mathcal{E}_2 &= \left\{\pi(\boldsymbol{S}) = (\boldsymbol{R'}(\boldsymbol{W'}))\right\}, \\
\mathcal{E}_3 &= \left\{\pi\left([K] \setminus \{\boldsymbol{W} \cup \boldsymbol{S}\}\right) = \boldsymbol{P'}\right\}.
\end{aligned}$$

Then, for any $W \in [K]$ and a permutation π on $[K]$, the probability of choosing π as a query can be written as

$$\begin{aligned}
&\mathbb{P}\left(\boldsymbol{Q}^{[\boldsymbol{W},\boldsymbol{S}]} = \pi \mid \boldsymbol{W} = W\right) \\
&\quad \overset{(a)}{=} \mathbb{P}\left(\mathcal{E}_1 | \boldsymbol{W} = W\right) \times \mathbb{P}\left(\mathcal{E}_2 \mid \mathcal{E}_1, \boldsymbol{W} = W\right) \\
&\qquad \times \mathbb{P}\left(\mathcal{E}_3 \mid \mathcal{E}_2, \mathcal{E}_1, \boldsymbol{W} = W\right), \\
&\quad \overset{(b)}{=} \frac{1}{K} \times \frac{1}{M!\binom{K-1}{M}} \times \frac{1}{(K-1-M)!}, \\
&\quad = \frac{1}{K!},
\end{aligned}$$

where (a) follows from the query generation procedure, and (b) uses (1) and (2) to compute $\mathbb{P}\left(\mathcal{E}_2 \mid \mathcal{E}_1, \boldsymbol{W} = W\right)$. This completes the proof of (15), and concludes the proof. ∎

Theorem 1 immediately follows from Lemmas 2 and 3.

Example 3. *Consider the $[7, 3]$ Simplex code with the parity-check matrix given below:*

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{16}$$

It is well-known that the Simplex code has locality $r = 2$, (see, e.g., [37]). Let us construct a PIR-SI scheme using H for the case when there are seven messages $[X_1\, X_2\, \cdots X_7]$, the user wants X_1, and has $\{X_2, X_3\}$ as a side-information. First, we choose an index W' uniformly at random from $[7]$, say $W' = 7$. Then, we choose an arbitrary repair group of the seventh coordinate, say $R(W') = \{4, 6\}$ (corresponding to the last row of H). Consider a random permutation of $R(W')$ as $\{6, 4\}$. Let $P = [K] \setminus \{W \cup S\} = \{4, 5, 6, 7\}$, and $P' =$

$\{1,3,5,2\}$ *be a random permutation of* $[K]\setminus\{W'\cup R(W')\}$. *Denote* π *to be the permutation that maps* W *to* W', S *to* $R'(W')$, *and* P *to* P'. *In particular,* $\pi(1)=7$, $\pi(2)=6$, $\pi(3)=4$, $\pi(4)=1$, $\pi(5)=3$, $\pi(6)=5$, $\pi(7)=2$. *The user sends* π *as its query. The server applies* π *to the columns of* H *to obtain* E, *i.e.,* $E_i=H_{\pi(i)}$, $i\in[K]$, *and sends* EX. *In other words, we have*

$$E=\begin{bmatrix}1&1&0&1&0&1&0\\1&0&0&0&0&1&1\\0&1&0&0&1&1&0\\1&1&1&0&0&0&0\end{bmatrix}, \tag{17}$$

and the answer is $A=\{X_1+X_2+X_4+X_6, X_1+X_6+X_7, X_2+X_5+X_6, X_1+X_2+X_3\}$. *Observe that the user can decode* X_1 *from* $X_1+X_2+X_3$, *since it knows* X_2 *and* X_3. *As proved in the lemma above, the permutation chosen by the user is uniform random, which ensures privacy.*

B. Proof of Theorem 2

The proof is similar to that of Theorem 2.

multi-message W-PIR-SI $\implies$ Cooperative LRC: First, we note that the following necessary condition is imposed by the privacy and recoverability conditions. The proof follows the same steps as in the case of Lemma 1.

Lemma 4. *For any query* $Q^{[W,S]}$, *the solution* E *to the multi-message* W*-PIR-SI problem must satisfy the following necessary condition: for every candidate demand index set* $W'\subset[K]$, $|W'|=D$, *there must exist a potential side information index set* $S'\subseteq[K]\setminus W'$, $|S'|\leq M$ *such that it is possible to recover* $X_{W'}$ *from* $EX_{[K]}$ *and* $X_{S'}$. *In other words, for every* $W'\subset[K]$, $|W'|=D$, *the following condition must hold:*

$$e_{i_j}\in\left\langle\begin{bmatrix}E\\I_{S'}\end{bmatrix}\right\rangle,\quad\forall\, i_j\in W'. \tag{18}$$

Next, using the above lemma, we show that any solution E to the multi-message W-PIR-SI problem is a parity-check matrix of a cooperative LRC with appropriate parameters.

Lemma 5. *Any scalar-linear solution* E *to the multi-message* W*-PIR-SI problem must be a parity-check matrix of a* $[K,K-T,(M,D)]$ *cooperative LRC.*

Proof: The necessary condition (18) in Lemma 4 implies that for every subset $W'=\{i_1,i_2,\ldots,i_D\}\subset[K]$ of size D, $\langle E\rangle$ must contain D vectors v_1,v_2,...,v_D with the following two properties: (i) $|\cup_{j=1}^{D}\text{Supp}\,(v_j)|\leq D+M$, and (ii) for each $1\leq j\leq D$, $\text{Supp}\,(v_j)\cap W'=\{i_j\}$. Then, it is easy to verify from Definition 3 that $\langle E\rangle^{\perp}$ is an (M,D) cooperative LRC with block-length K. ∎

On exploiting the connection with index coding: The multi-message PIR-SI problem can be shown to be related to multiple-groupcast index coding problem [38] by extending the arguments in [11, Lemma 1]. Also, it is possible to extend the result in [31, Theorem 1] to show a duality between multiple-groupcast index coding and cooperative LRCs. Using these two results, it is straightforward to prove the forward direction (i.e, multi-message W-PIR-SI solution is a parity-check matrix of a cooperative LRC). In Lemmas 4 and 5, we gave a direct and simple proof without using the connection to multiple-groupcast index coding. Similar to the single-message case, the results in [11], [31], [32] are also not sufficient to show the reverse direction.

Cooperative LRC $\implies$ multi-message W-PIR-SI: Here, we show that one can construct a multi-message W-PIR-SI solution using a parity-check matrix of a cooperative LRC.

Lemma 6. *Given a parity-check matrix of a* $[K,K-T,(M,D)]$ *cooperative LRC, it is possible to construct a multi-message* W*-PIR-SI scheme such that the solution* E *is a column-permutation of* H.

Proof: The proof is similar to that of Lemma 3. We outline only the steps that are pertinent to the multi-message setting. In the rest of the proof, we consider all sets as ordered sets (with a natural ascending order).

For a given W and S, the user first finds a permutation π on $[K]$ as follows. Choose an index set W' of size D uniformly at random from $[K]$, independent of W and S. Let $R(W')$ be a repair group of W'. If a D-set of coordinates has multiple repair groups, arbitrarily choose one repair group. This arbitrary choice of a repair group for each coordinate is made *a priori*, and are known to the server as a part of the scheme. By the definition of cooperative locality, we have $|R(W')|\leq M$. For simplicity, we assume that every repair group of any D-set is of size M. The arguments can be easily generalized to the case when some repair groups are smaller than M. Let $R'(W')$ be a random permutation of $R(W')$.

Let $P=[K]\setminus\{W\cup S\}$, and P' be a random permutation of $[K]\setminus\{W'\cup R(W')\}$. Let π be the permutation that maps W to W', S to $R'(W')$, and P to P'. The user sends π as its query $Q^{[W,S]}$. The server then applies π to the columns of H to obtain E, i.e., $E_i=H_{\pi(i)}$ for each $i\in[K]$, where the subscript denotes the column of the matrix. Then, the server computes the answer as EX. The proof that the above scheme satisfies the recoverablity and W-privacy conditions follows the same steps as that of Lemma 3, and hence omitted. ∎

Theorem 2 immediately follows from Lemmas 5 and 6.

C. Proof of Theorem 3

(W,S)-PIR-SI $\implies$ MDS code: First, we note that the (W,S)-privacy condition implies the following necessary condition irrespective of the value of D. The proof is similar to Lemma 1.

Lemma 7. *Any solution* E *to the* (W,S)*-PIR-SI problem must satisfy the following: for each message* X_i *and every set* $S_i\subseteq[K]\setminus\{i\}$ *of size* M, *it is possible to recover* X_i *from* $EX_{[K]}$ *and* X_{S_i}.

Using Lemma 7, the following result is immediate.

Lemma 8. *Any solution* E *to the* (W,S)*-PIR-SI problem with* $T=K-M$ *is a parity-check matrix of a* $[K,M]$ *MDS code.*

Proof: The aforementioned necessary condition in Lemma 8 implies that, for any set $S\subset[K]$ of size M and for every $i\in[K]\setminus S$, we must have

$$e_i\in\left\langle\begin{bmatrix}E\\I_S\end{bmatrix}\right\rangle. \tag{19}$$

Equation (19), in turn, implies that the $T\times(K-M)$ submatrix of E formed by the columns indexed by $[K]\setminus S$ must have the rank at least $K-M$. As $T=M$, this guarantees that the columns of E indexed by $[K]\setminus S$ must be linearly independent. Since this should hold for each subset $S\subset[K]$ of size M, we have that every subset of columns of E of size $K-M$ are linearly independent. Thus, E must be a parity check matrix of a $[K,M]$ MDS code. ∎

MDS code $\implies$ (W,S)-PIR-SI: Next, we establish a relation from a parity check matrix of an MDS code to a solution of the (W,S)-PIR-SI problem.

Lemma 9. *Let H be a parity check matrix of a (K,M)-MDS code. Then, $E=H$ is a solution to the (W,S)-PIR-SI problem.*

Proof: First, note that the scheme with $E=H$ is private, since the solution is independent of the particular realization of $\boldsymbol{W}$ and $\boldsymbol{S}$. As the server already knows the size of the side information index set, it does not obtain any other information about $\boldsymbol{W}$ and $\boldsymbol{S}$ from E.

To see the recoverability, note that any $K-M$ columns of H are linearly independent. Thus, given the side information X_S for any $S\subset[K]$ of size M, the user can recover all the messages X_i, $i\in[K]\setminus S$, including the demand message(s) X_W. ∎

Theorem 3 immediately follows from Lemmas 8 and 9.

V. Equivalence Results for Non-Linear Schemes

In this section, we consider generic PIR-SI and LRC schemes, which encompass scalar-linear, vector-linear, and non-linear schemes. We present the results for multi-message W-PIR-SI and cooperative LRCs, which reduce to the case of single-message W-PIR-SI and LRCs when $D=1$. Our proof techniques build up on those used in [32].

We begin with the definition of a generic cooperative LRC scheme. Unlike the linear case, here we explicitly consider encoding, decoding, and repair functions along with the code. In order to encompass vector code designs that allow *sub-packetization* (i.e., splitting symbols into multiple sub-symbols and designing encoding, decoding, and repair functions that operate over sub-symbols) and to acoommodate different sizes for input and output symbols, we assume that input symbols lie in $\mathbb{F}_q$ while codeword symbols lie in an extension field $\mathbb{F}_{q^m}$. More generally, one can consider arbitrary (finite) alphabets for the input and output symbols.

Definition 4. *An (n,k,r,ℓ) cooperative LRC scheme consists of*

1) *a code $\mathcal{C}\subseteq\mathbb{F}_{q^m}^n$ (with $k=\log_q|\mathcal{C}|$) containing a set of vectors in $\mathbb{F}_{q^m}^n$, referred to as* codewords*;*
2) *a bijection $f:\mathbb{F}_q^k\to\mathcal{C}$ referred to as the encoding function, and the inverse f^{-1} referred to as the decoding function; and*
3) *a set of $\binom{n}{\ell}$ deterministic functions $\{g_\Delta:\Delta\subset[n],|\Delta|=\ell\}$, $g_\Delta:\mathbb{F}_{q^m}^r\to\mathbb{F}_{q^m}^\ell$, referred to as repair functions, such that, for every subset of ℓ coordinates $\Delta\subset[n]$, there exists a set of coordinates $R(\Delta)\subset[n]\setminus\Delta$, $|R(\Delta)|=r$, satisfying $g_\Delta(\mathbf{c}_{R(\Delta)})=\mathbf{c}_\Delta$ for every codeword $\mathbf{c}\in\mathcal{C}$. We say that $R(\Delta)$ is a repair group of the set of coordinates Δ.*

Next, for the multi-message W-PIR-SI problem, we define a PIR-SI scheme. Here, we rigorously define the query and answer functions along with recovery functions used to decode the demand. Towards this end, we introduce the following notation:

$$\mathcal{W}=\{(W,S)\mid W\subset[K],|W|=D,S\subseteq[K]\setminus\{W\},|S|=M\}. \tag{20}$$

That is, $\mathcal{W}$ is the set of all possible combinations of the demand index set and the side information index set. In order to encompass vector code designs that allow sub-packetization, we assume that message symbols lie in an extension field $\mathbb{F}_{q^m}$, and the answer from the server consists of symbols from $\mathbb{F}_q$.

Definition 5. *A (K,T,M,D) PIR-SI scheme consists of*

1) *a set of vectors in $\mathbb{F}_q^T$, referred to as* codewords*,*
2) *a class of deterministic answer functions $\mathcal{A}$, where each function $A\in\mathcal{A}$ maps a message vector from $\mathbb{F}_{q^m}^K$ to a codeword, i.e., $A:\mathbb{F}_{q^m}^K\to\mathbb{F}_q^T$,*
3) *a class of deterministic recovery functions $\mathcal{D}$, where each function $D\in\mathcal{D}$ is from $\mathbb{F}_q^T\times\mathbb{F}_{q^m}^M$ to $\mathbb{F}_{q^m}^D$, and*
4) *a stochastic query function $Q:\mathcal{W}\to\mathcal{A}$ that maps (W,S) to an answer function $A\in\mathcal{A}$ (independently of the value of X_S) such that:*
   (i) *for every $W',W\subset[K]$, $|W|=|W'|=D$, $W\neq W'$, $S\subseteq[K]\setminus\{W\}$, $|S|=M$, and for each $A\in\mathcal{A}$,*
   $$\mathbb{P}\left(\boldsymbol{W}=W'\mid Q(W,S)=A\right)=\mathbb{P}\left(\boldsymbol{W}=W'\right), \tag{21}$$
   and
   (ii) *there exists a recovery function $D\in\mathcal{D}$, which depends on A, satisfying*
   $$D\left(A(X_1,\cdots,X_K),X_S\right)=X_W. \tag{22}$$

It is straightforward to show that the W-privacy condition (21) and recovery condition (22) implies the following necessary condition on a PIR-SI code. The proof is similar to that of Lemma 1, and is omitted.

Lemma 10. *In a (K,T,M,D) PIR-SI coding scheme, for any $A\in\mathcal{A}$, for every $W'\in[K]$, $|W'|=D$, there must exist a set $S_{W'}\subseteq[K]\setminus\{W'\}$, $|S_{W'}|=M$, and a recovery function $D_{W'}$ such that $D_{W'}\left(A(X_1,\cdots,X_K),X_{S_{W'}}\right)=X_{W'}$.*

Now, we show a relation from a PIR-SI code to a cooperative LRC.

Theorem 4. *Given a (K,T,M,D) PIR-SI scheme, it is possible to construct a (K,k,M,D) cooperative LRC scheme with $k\geq mK-T$.*

Proof: First, note that, for any $A\in\mathcal{A}$, there must exist a vector $\mathbf{a}\in\mathbb{F}_q^T$ such that $\left|\left\{X\in\mathbb{F}_q^K\mid A(X)=\mathbf{a}\right\}\right|\geq q^{mK-T}$. This is because every $A\in\mathcal{A}$ maps $\mathbb{F}_{q^m}^K$ to $\mathbb{F}_q^T$. Next, for an arbitrary $A\in\mathcal{A}$ and the corresponding $\mathbf{a}$, let us define $\mathcal{C}_{\mathbf{a}}=\left\{X\in\mathbb{F}_{q^m}^K\mid A(X)=\mathbf{a}\right\}$. Now, from Lemma 10, for every $W'\subset[K]$ of size D, there must exist a deterministic decoding function $D_{W'}$ and a set $S_{W'}\subseteq[K]\setminus\{i\}$ of size M,

such that $D_{W'}\left(\mathbf{a}, X_{S_{W'}}\right) = X_{W'}$. Using this, define, for every $W' \subset [K]$ of size D, $R(W') = S_{W'}$, and $g_{W'}\left(\mathbf{c}_{R(W')}\right) = D_{W'}\left(\mathbf{a}, X_{S_{W'}}\right)$. It is easy to verify that the code $\mathcal{C}_{\mathbf{a}}$ along with an arbitrary bijection $f : \mathbb{F}_q^{\lfloor \log_q |\mathcal{C}_{\mathbf{a}}| \rfloor} \to \mathcal{C}_{\mathbf{a}}$ and repair functions $\{g_{W'} : W' \subset [K], |W'| = D\}$ is a (K, k, M, D) cooperative LRC with $k \geq mK - T$. ∎

To complete the equivalence, we establish a relation from cooperative LRC schemes to PIR-SI schemes. In particular, we consider a class of (K, k, M, D) cooperative LRC schemes for which the code $\mathcal{C}$ contains k coordinates such that values on these coordinates determine the values on the remaining coordinates.[13]

Let us consider codewords of an LRC as length-(mK) vectors with each coordinate taking a value from $\mathbb{F}_q$. Notice that this is always possible since there is a bijection between $\mathbb{F}_{q^m}$ and $\mathbb{F}_q^m$. For a code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^K$ and a set $P \subset [mK]$, let $\mathcal{C}_P$ denote the code obtained by puncturing $\mathcal{C}$ on the coordinates outside of P. Note that, for a code $\mathcal{C}$ of size q^k, if there exists a set I of k coordinates such that $|\mathcal{C}_I| = q^k$, then the coordinates in I determine the remaining coordinates. We refer to such an LRC as a *good* LRC. As an example of a good LRC, we show in Appendix D that a *rate-optimal* LRC is a good LRC.

Next, we show that, given a good (K, k, M, D) cooperative LRC, it is possible to construct a $(K, mK - k, M, D)$ PIR-SI scheme.

Theorem 5. *Given a (K, k, M, D) LRC scheme with a good code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^K$, (i.e., there exists a set I of k coordinates satisfying $|\mathcal{C}_I| = q^k$), it is possible to construct a $(K, mK - k, M, D)$ PIR-SI scheme.*

In order to prove Theorem 5, we need another lemma. We first define the following notation. Given a vector $\mathbf{u} \in \mathbb{F}_{q^m}^K$, we define a translation of an LRC $\mathcal{C} \subseteq \mathbb{F}_{q^m}^K$ as

$$\mathcal{C} + \mathbf{u} = \{\mathbf{c} + \mathbf{u} \mid \mathbf{c} \in \mathcal{C}\}. \tag{23}$$

Now, we show that there exist q^{mK-k} translations of a good LRC that partition $\mathbb{F}_{q^m}^K$.

Lemma 11. *Consider a good (K, k, M, D) LRC $\mathcal{C} \subseteq \mathbb{F}_{q^m}^K$ with a set I of k coordinates satisfying $|\mathcal{C}_I| = q^k$. Then, there exist q^{mK-k} distinct vectors $\mathbf{u}_j \in \mathbb{F}_{q^m}^K$, $j = 0, \ldots, q^{mK-k}-1$, such that the translations $\left\{\mathcal{C} + \mathbf{u}_j \mid j = 0, \ldots, q^{mK-k} - 1\right\}$ partition the space $\mathbb{F}_{q^m}^K$. That is,*

$$(\mathcal{C} + \mathbf{u}_i) \cap (\mathcal{C} + \mathbf{u}_j) = \emptyset, \quad \forall\, i \neq j, \tag{24}$$

and

$$\cup_{j=0}^{q^{mK-k}-1} (\mathcal{C} + \mathbf{u}_j) = \mathbb{F}_{q^m}^K. \tag{25}$$

Proof: We give a constructive proof. Let I be the set of coordinates of $\mathcal{C}$ as described in the statement of the lemma. Without loss of generality, let I be the first k coordinates. Let $I' = [mK] \setminus I$. Let $\left\{\mathbf{v}_i \mid 0 \leq i \leq q^{mK-k} - 1\right\}$ denote the set of vectors in $\mathbb{F}_q^{mK-k}$ in a lexicographic order. For each $0 \leq i \leq q^{mK-k} - 1$, define $\mathbf{u}_i = [\mathbf{0}\ \mathbf{v}_i]$, where $\mathbf{0}$ is the all-zero vector of length k. Notice that, since there is a bijection between $\mathbb{F}_{q^m}$ and $\mathbb{F}_q^m$, then $\mathbf{u}_i \in \mathbb{F}_{q^m}^K$ for each $0 \leq i \leq q^{mK-k} - 1$.

Now, note that any translation of $\mathcal{C}$ has the same size as $\mathcal{C}$. Thus, to prove (25), it suffices to show (24). We prove this by the way of contradiction. Suppose, for contradiction, that there exists a pair of codewords $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$ such that $\mathbf{c} + \mathbf{u}_i = \mathbf{c}' + \mathbf{u}_j$. This implies that

$$[\mathbf{c}_I\ \ \mathbf{c}_{I'} + \mathbf{v}_i] = [\mathbf{c}'_I\ \ \mathbf{c}'_{I'} + \mathbf{v}_j]. \tag{26}$$

Therefore, $\mathbf{c}_I = \mathbf{c}'_I$. Further, since the coordinates in I' can be recovered from those in I, we must have $\mathbf{c}_{I'} = \mathbf{c}'_{I'}$. This is because, since coordinates in I determine the rest of the coordinates, if two codewords $\mathbf{c}$ and $\mathbf{c}'$ have the same values on the coordinates in I, then they must have the same values on all other coordinates. However, as $\mathbf{v}_i \neq \mathbf{v}_j$, we have a contradiction to (26). ∎

Remark 4. *In [32, Lemma 4], it is shown that given an LRC of length K and size q^k with a recoverability graph[14] G, one can construct an index code for an instance with side information graph G such that the length of the index code is $K - k + f(K, k, q)$, where q is the field size and $f(K, k, q)$ is a positive number greater than 1 whose expression can be found in [32]. (Note that [32, Lemma 4] assumes $m = 1$.) In other words, unlike linear schemes, there is a 'duality gap' for non-linear LRCs and index codes. The key idea in the proof of [32, Lemma 4] is to show that for any LRC of size q^k, there exist $q^{K-k} f(K, k, q)$ translations of the LRC that cover $\mathbb{F}_q^K$. In Lemma 11, for any good LRC of size q^k, we explicitly construct its q^{K-k} translations that 'optimally' cover $\mathbb{F}_q^K$ (assuming $m = 1$). Using this result, it is easy to show that given a good LRC of length K and size q^k with a recoverability graph G, one can construct an index code for an instance with side information graph G such that the length of the index code is $K - k$. This implies that there is no duality gap for good LRCs.*

Proof of Theorem 5: Lemma 11 enables us to construct a $(K, mK - k, M, D)$ PIR-SI scheme using a good LRC as follows. To simplify the notation, define $T = mK - k$.

Answer functions: We construct a set $\mathcal{A}$ of $K!$ answer functions, and associate every answer function with a permutation on $[K]$. Towards this end, we need the following additional notation. For $0 \leq a \leq q^T - 1$, let $\bar{a}_q$ denote the length-T q-ary expansion of a. In addition, for a permutation π on $[K]$ and a vector $X_{[K]} = [X_1 \cdots X_K] \in \mathbb{F}_{q^m}^K$, let $\pi(X_{[K]}) = X_{\pi([K])}$.

Let $\mathcal{U} = \{\mathbf{u}_j \in \mathbb{F}_{q^m}^K, j = 0, \ldots, q^T - 1\}$ be a set of vectors as described in Lemma 11. For a given $X_{[K]} \in \mathbb{F}_{q^m}^K$ and a permutation π on $[K]$, let $0 \leq a^* \leq q^T - 1$ be the index of the translation that contains $\pi(X_{[K]})$. That is, for a given $X_{[K]} \in \mathbb{F}_{q^m}^K$ and a permutation π on $[K]$, a^* is such that $\pi(X_{[K]}) \in \mathcal{C} + \mathbf{u}_{a^*}$. Note that, by Lemma 11, the translations $\left\{\mathcal{C} + \mathbf{u}_j \mid 0 \leq j \leq q^T - 1\right\}$ partition the space $\mathbb{F}_{q^m}^K$. Hence, there exists a unique such $\mathbf{u}_{a^*} \in \mathcal{U}$ for every $X_{[K]} \in \mathbb{F}_{q^m}^K$

[13] Note that, unlike linear codes, for an arbitrary non-linear code of size q^k, there may not exist any subset of k coordinates that determine the values of the remaining coordinates. As an example, consider the following $(3, 2)$ code: $\{000, 010, 100, 001\}$.

[14] See Sec. IV-A for the definitions of index coding, recoverability graph, and side information graph, and the duality result for linear schemes.

and any permutation π on $[K]$. Define the answer functions for every $X_{[K]} \in \mathbb{F}_{q^m}^K$ and every permutation π on $[K]$ as

$$A_\pi\left(X_{[K]}\right) = \bar{a^*}_q, \tag{27}$$

where $\bar{a^*}_q$ is the length-T q-ary expansion of a^*. Note that the server will transmit the T symbols of $\bar{a^*}_q$ to the user.

Query function: We are given a demand index set $W \subset [K]$ of size D and a side information index set $S \subseteq [K] \setminus \{W\}$ of size M. First, choose a set of D indices $W' \in [K]$ uniformly at random, independent of W and S. Let $R(W')$ be a repair group of W' in $\mathcal{C}$. If a D-set of coordinates has multiple repair groups, arbitrarily choose one repair group. This arbitrary choice of a repair group for each coordinate is made *a priori*, and is known to the server as part of the scheme.

Let $P = [K] \setminus (W \cup S)$. Let $R'(W')$ and P' be random permutations of sets $R(W')$ and $[K] \setminus (W' \cup R(W'))$, respectively. Let π be a permutation on the set $[K]$ that maps W to W', S to $R'(W')$, and P to P'. Then, the query function Q maps (W, S) to A_π in $\mathcal{A}$. Note that it suffices for the user to send π as their query.

Recovery functions: For a set $P \subset [K]$, let $\mathbf{u}_{a^*}|_P$ denote the length-$|P|$ vector obtained by deleting the coordinates of $\mathbf{u}_{a^*}$ outside P. Now, given π and A_π, define the recovery function as

$$D\left(A_\pi(X), X_S\right) = g_{W'}\left(X_{R'(W')} - \mathbf{u}_{a^*}|_{R'(W')}\right) + \mathbf{u}_{a^*}|_{W'}, \tag{28}$$

where $g_{W'}(\cdot)$ is the repair function of $\mathcal{C}$ for coordinates $\mathbf{c}_{W'}$ (see Definition 4).

Recoverability and Privacy: It is straightforward to verify that $D\left(A_\pi(X), X_S\right) = X_W$ (cf. (28)). The W-privacy condition (21) can be proven in the same way as in the proof of Lemma 3, and thus, the proof is omitted.

VI. Conclusion

The theoretical computer science community has established a strong relationship between multi-server private information retrieval (PIR) schemes and schemes for locally decodable codes (LDCs). This paper extends this theme by establishing an equivalence between single-server PIR schemes for a recently proposed PIR with side information problem and schemes for locally recoverable codes (LRCs). In particular, we present explicit algorithms that transform a given PIR scheme into an LRC scheme and vice versa. As corollaries to these equivalence results, we obtain upper bounds on the download rate for PIR-SI schemes, and a novel rate upper bound on cooperative LRCs.

Acknowledgement

S. Kadhe would like to thank Kannan Ramchandran for helpful discussions.

References

[1] S. Kadhe, A. Heidarzadeh, A. Sprintson, and O. O. Koyluoglu, "On an equivalence between single-server pir with side information and locally recoverable codes," in *2019 IEEE Information Theory Workshop (ITW)*, 2019, pp. 1–5.

[2] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *Journal of the ACM*, vol. 45, no. 6, pp. 965–981, 1998.

[3] S. Yekhanin, "Private information retrieval," *Communications of the ACM*, vol. 53, no. 4, pp. 68–73, 2010.

[4] T. Mayberry, E.-O. Blass, and A. H. Chan, "Pirmap: Efficient private information retrieval for mapreduce," in *Financial Cryptography and Data Security*, A.-R. Sadeghi, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 371–385.

[5] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, 2017.

[6] H. Sun and S. A. Jafar, "The capacity of robust private information retrieval with colluding databases," *IEEE Trans. on Info. Theory*, vol. 64, no. 4, pp. 2361–2370, April 2018.

[7] R. Tajeddine and S. El Rouayheb, "Robust private information retrieval on coded data," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017.

[8] K. Banawan and S. Ulukus, "Multi-message private information retrieval: Capacity results and near-optimal schemes," *CoRR*, vol. abs/1702.01739, 2017. [Online]. Available: http://arxiv.org/abs/1702.01739

[9] ——, "The capacity of private information retrieval from coded databases," *IEEE Trans. on Info. Theory*, vol. 64, no. 3, pp. 1945–1956, March 2018.

[10] S. Kadhe, B. Garcia, A. Heidarzadeh, S. E. Rouayheb, and A. Sprintson, "Private information retrieval with side information: The single server case," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2017, pp. 1099–1106.

[11] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson, "Private information retrieval with side information," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2032–2043, 2020.

[12] A. Heidarzadeh, B. Garcia, S. Kadhe, S. E. Rouayheb, and A. Sprintson, "On the capacity of single-server multi-message private information retrieval with side information," in *2018 56th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2018.

[13] S. Li and M. Gastpar, "Single-server multi-message private information retrieval with side information," in *2018 56th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2018.

[14] S. Yekhanin, "Locally decodable codes," in *Computer Science–Theory and Applications*. Springer, 2011, pp. 289–290.

[15] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *Information Theory, IEEE Transactions on*, vol. 58, no. 11, pp. 6925–6934, Nov 2012.

[16] A. S. Rawat, A. Mazumdar, and S. Vishwanath, "On cooperative local repair in distributed storage," in *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, March 2014, pp. 1–5.

[17] ——, "Cooperative local repair in distributed storage," *Journal on Advances in Signal Processing*, vol. 2015, no. 1, p. 107, Dec 2015.

[18] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," in *Network Computing and Applications, 2007. NCA 2007. Sixth IEEE International Symposium on*, July 2007, pp. 79–86.

[19] J. Han and L. Lastras-Montao, "Reliable memories with subline accesses," in *2007 IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 2531–2535.

[20] F. Oggier and A. Datta, "Self-repairing homomorphic codes for distributed storage systems," in *INFOCOM, 2011 Proceedings IEEE*, April 2011, pp. 1215–1223.

[21] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 5843–5855, Oct 2014.

[22] A. Rawat, O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal locally repairable and secure codes for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 212–236, Jan 2014.

[23] A. Wang and Z. Zhang, "An integer programming-based bound for locally repairable codes," *IEEE Transactions on Information Theory*, vol. 61, no. 10, pp. 5280–5294, Oct 2015.

[24] N. Silberstein, A. Rawat, O. Koyluoglu, and S. Vishwanath, "Optimal locally repairable codes via rank-metric codes," in *2013 IEEE International Symposium on Information Theory Proceedings (ISIT)*, July 2013, pp. 1819–1823.

[25] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *Information Theory, IEEE Transactions on*, vol. 60, no. 8, pp. 4661–4676, Aug 2014.

[26] P. Huang, E. Yaakobi, H. Uchikawa, and P. H. Siegel, "Linear locally repairable codes with availability," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1871–1875.
[27] ——, "Binary linear locally repairable codes," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6268–6283, Nov 2016.
[28] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1479–1494, March 2011.
[29] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim, "Broadcasting with side information," in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, Oct 2008, pp. 823–832.
[30] F. Arbabjolfaei and Y.-H. Kim, "Fundamentals of index coding," *Foundations and Trends in Communications and Information Theory*, vol. 14, no. 3-4, pp. 163–346, 2018. [Online]. Available: http://dx.doi.org/10.1561/0100000094
[31] K. Shanmugam and A. G. Dimakis, "Bounding multiple unicasts through index coding and locally repairable codes," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 296–300.
[32] A. Mazumdar, "Storage capacity of repairable networks," *IEEE Trans. on Info. Theory*, vol. 61, no. 11, pp. 5810–5821, Nov 2015.
[33] V. R. Cadambe and A. Mazumdar, "Bounds on the size of locally recoverable codes," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 5787–5794, Nov 2015.
[34] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY, USA: Cambridge University Press, 2008.
[35] A. Heidarzadeh, B. Garcia, S. Kadhe, S. E. Rouayheb, and A. Sprintson, "On the capacity of single-server multi-message private information retrieval with side information," *CoRR*, vol. abs/1807.09908, 2018.
[36] S. Li and M. Gastpar, "Single-server multi-message private information retrieval with side information," *CoRR*, vol. abs/1808.05797, 2018. [Online]. Available: http://arxiv.org/abs/1808.05797
[37] V. R. Cadambe and A. Mazumdar, "Bounds on the size of locally recoverable codes," *IEEE transactions on information theory*, vol. 61, no. 11, pp. 5787–5794, 2015.
[38] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Lexicographic products and the power of non-linear network coding," in *Proceedings of the 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, ser. FOCS 11. USA: IEEE Computer Society, 2011, p. 609618. [Online]. Available: https://doi.org/10.1109/FOCS.2011.39

APPENDIX

A. Proof of Corollary 1

Let E be a scalar-linear solution to the single-message W-PIR-SI problem. Let $\mathcal{C} = \langle E \rangle^{\perp}$. Suppose the minimum distance of $\mathcal{C}$ is d. Note that we must have $d \geq 2$. Otherwise, if $d = 1$, then E must contain a column of all zeros. Let W' denote the index of this all-zero column. This, however, implies that $X_{W'}$ cannot be the demand, and this will violate the privacy condition.[15] Now, from Theorem 1, $\langle E \rangle^{\perp}$ is an LRC with block-length $n = K$, dimension $k = K - T$, and locality $r = M$. Therefore, we have from (7) that

$$K \geq K - T + \left\lceil \frac{K-T}{M} \right\rceil - 2 + d.$$

After re-arranging, and noting that $d \geq 2$ and T is an integer, we get

$$T \geq \left\lceil \frac{K}{M+1} \right\rceil.$$

As the messages are independent and uniformly distributed over $\mathbb{F}_q$, we have $H\left(\boldsymbol{A}^{[W,S]}\right) = T \log q$. The result then follows from (6).

[15]Note that here we are using the same argument as in the proof of Theorem 1 (cf. (13)).

B. Proof of Corollary 3

Let $\mathcal{C} = \langle E \rangle^{\perp}$. From Theorem 2, $\mathcal{C}$ must be a code with blocklength K and (M, D)-cooperative locality. Using (8), it is shown in [17, Corollary 1] that the rate of a code with (M, D)-cooperative locality for $M \geq D$ is upper bounded as $M/(M + D)$. Therefore, we have $T/K \geq 1 - M/(M + D)$, which yields $T \geq \lceil DK/(D + M) \rceil$. As the messages are independent and uniformly distributed over $\mathbb{F}_q$, we have $H\left(\boldsymbol{A}^{[W,S]}\right) = T \log q$. The result then follows from (6).

C. Proof of Corollary 4

Let H be a parity check matrix of an $[n, k, (r, \ell)]$ cooperative LRC. From Theorem 2, H is a solution (up to a column-permutation) of a multi-message PIR-SI problem such that $K = n$, $M = r$, and $D = \ell$. Now, in [35, Lemma 1], it is shown that, when $D > M$, the number of transmissions in any multi-message PIR-SI scheme is at least $K - M$. Therefore, we have $n - k \geq n - r$, from which the result follows.

D. Rate-Optimal LRCs are 'Good' LRCs

From [25, Theorem 2.1], we have the following upper bound on the size of an (n, k, r) LRC: for any (n, k, r) LRC $\mathcal{C} \subset \mathbb{F}_q^n$, the size $|\mathcal{C}| \leq q^{n - \lceil n/(r+1) \rceil}$. We refer to an (n, k, r) LRC $\mathcal{C}$ satisfying the equality $|\mathcal{C}| = q^{n - \lceil n/(r+1) \rceil}$ to be a *(rate-) optimal* LRC. To simplify the presentation, we define $T_{OPT} \triangleq \lceil K/(M + 1) \rceil$.

In the following, we show that any optimal LRC must contain $K - T_{OPT}$ coordinates whose values determine the values of the remaining T_{OPT} coordinates.

Lemma 12. *For an optimal $(K, K - T_{OPT}, M)$ LRC $\mathcal{C} \subset \mathbb{F}_q^K$, there exists a partition of K coordinates into sets P_1 and P_2 such that $|P_1| = K - T_{OPT}$, $|P_2| = T_{OPT}$, and for any codeword $\mathbf{c} \in \mathcal{C}$, the symbols $\mathbf{c}_{P_2}$ can be recovered from the symbols $\mathbf{c}_{P_1}$.*

Proof: We iteratively construct P_1 and P_2 as follows.

1. Initialize $P_1 = P_2 = \emptyset$
2. While $|P_1 \cup P_2| < K$:
   2.1 Choose a coordinate $i \notin P_1 \cup P_2$;
   2.2 Set $P_1 \leftarrow P_1 \cup R(i)$, for a repair group $R(i)$ of i;
   2.3 Set $P_2 \leftarrow P_2 \cup \{i\}$.

By this construction, the coordinates in P_2 can be recovered from the coordinates in P_1.

Note that, in each step, P_2 grows by one, and P_1 grows by at most M as the locality of the code is M. In other words, in each step, $P_1 \cup P_2$ grows by at most $M + 1$. Therefore, the number of steps for which the while loop runs is at least $\lceil K/(M + 1) \rceil = T_{OPT}$. This gives $|P_2| \geq T_{OPT}$.

Next, we show that $|P_2| \leq T_{OPT}$. Since there is a bijection between $\mathbb{F}_q^{K - T_{OPT}}$ and $\mathcal{C}$, and since each coordinate in P_2 is a function of the coordinates in P_1, there must be a bijection between $\mathbb{F}_q^{K - T_{OPT}}$ and $\mathcal{C}_{P_1}$. This implies that $|P_1| \geq K - T_{OPT}$, and thus, $|P_2| \leq T_{OPT}$.

We conclude that $|P_2| = T_{OPT}$, which completes the proof. ∎