The Generalized Covering Radii of Linear Codes

Dor Elimelech*, Marcelo Firer†, and Moshe Schwartz‡
*Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel, doreli@post.bgu.ac.il
†Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Brazil, mfirer@ime.unicamp.br
‡Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel, schwartz@ee.bgu.ac.il

***Abstract*—Motivated by an application to database linear querying, such as private information-retrieval protocols, we suggest a fundamental property of linear codes – the generalized covering radius. The generalized covering-radius hierarchy of a linear code characterizes the trade-off between storage amount, latency, and access complexity, in such database systems. Several equivalent definitions are provided, showing this as a combinatorial, geometric, and algebraic notion. We derive bounds on the code parameters in relation with the generalized covering radii, study the effect of simple code operations, and describe a connection with generalized Hamming weights.**

I. Introduction

A common query type in database systems involves a linear combination of the database items with coefficients supplied by the user. As examples we mention partial-sum queries [2], and private information retrieval (PIR) protocols [3]. In essence, one can think of the database server as storing m items, $x_1, \ldots, x_m \in \mathbb{F}_{q^\ell}$. A user may query the contents of the database by providing $s_1, \ldots, s_m \in \mathbb{F}_q$, and getting in response the linear combination $\sum_{i=1}^{m} s_i x_i$.

Various aspects of these systems are of interest and in need of optimization, such as the amount of storage at the server, and the required bandwidth for the querying protocol. One important aspect is that of *access complexity*, paralleling a similar concern studied in distributed storage systems [9], [12]. In a straightforward implementation, the time required to access the elements of the database needed to compute the answer to a user query is directly proportional to the number of non-zero coefficients among $s_1, \ldots, s_m$. This may prove to be a bottleneck, in particular since in schemes like PIR, the coefficients are random, and therefore a typical query would require the database server to access a fraction of $1 - \frac{1}{q}$ of the items.

A trade-off between access complexity and storage amount was suggested for PIR in [14], echoing a similar suggestion for databases made in [10]. This suggestion calls for a carefully designed set of linear combinations to be pre-computed and stored by the server. Instead of storing $\overline{x} = (x_1, \ldots, x_m)$ as is, the server stores $\overline{h}_1 \cdot \overline{x}, \ldots, \overline{h}_n \cdot \overline{x}$, where each $\overline{h}_i \in \mathbb{F}_q^m$ describes a linear combination. Assume that the matrix H, whose columns are $\overline{h}_1, \ldots, \overline{h}_n$, is a parity-check matrix for a code with covering radius r. Thus, when the user queries the database using $\overline{s} = (s_1, \ldots, s_m)$, by the properties of the covering code, $\overline{s}$ may be computed using a linear combination of at most r columns of H. Hence, at most r pre-computed combinations that are stored in the database need to be accessed in order to provide the user with the requested linear combination. The trade-off between access complexity and storage amount follows, since instead of storing m elements, the server now stores $n \geqslant m$ linear combinations, and so n is lower bounded by the smallest possible length for a code with covering radius r and redundancy m over $\mathbb{F}_q$. These code parameters have been thoroughly studied and are well understood [4].

We now take access-complexity optimization one step further. The database server naturally receives a stream of queries, say $\overline{s}_1, \overline{s}_2, \ldots$. Those may arrive from the same user, or from multiple distinct users. Instead of handling each of the queries separately, accessing r pre-computed linear combinations for each query, the server may group together t queries, $\overline{s}_1, \ldots, \overline{s}_t$ and, hopefully, access fewer than $r \cdot t$ pre-computed linear combinations as it would in a naive implementation. Thus, both storage amount and latency are traded-off for a reduced access complexity.

The motivation mentioned above leads us to the following combinatorial problem: Design a set of vectors, $\overline{h}_1, \ldots, \overline{h}_n \in \mathbb{F}_q^m$ (describing linear combinations to pre-compute), such that every t vectors, $\overline{s}_1, \ldots, \overline{s}_t \in \mathbb{F}_q^m$ (describing user queries), may be obtained by accessing at most r of the elements $\overline{h}_1, \ldots, \overline{h}_n$. When viewed as columns of a parity-check matrix for a code, this becomes a *generalized covering radius* definition. It bears a resemblance to the generalized Hamming weight of codes, introduced by Wei [13] to characterize the performance of linear codes over a wire-tap channel.

The goal of this paper is to study the generalized covering radius as a fundamental property of linear codes. Our main contributions are the following:

1) We discuss three definitions for the generalized covering radius of a code, highlighting the combinatorial, geometric, and algebraic properties of this concept, and showing them to be equivalent.
2) We derive bounds that tie the various parameters of codes to the generalized covering radii. In particular, we prove an asymptotic upper bound on the minimum rate of binary codes with a prescribed second generalized covering radius, thus showing an improvement over the naive approach. The bound on the minimal rate is attained by almost all codes.
3) We determine the effect simple code operations have on

The work of D. Elimelech was supported in part by an Israel Science Foundation (ISF) Grant under Grant 1052/18. The work of M. Schwartz was supported in part by a German Israeli Project Cooperation (DIP) Grant under Grant PE2398/1-1. The work of M. Firer was supported in part by Fapesp, grant 13/25977-7 and CNPq 304046/2017-5.

the generalized covering radii: code extension, puncturing, the $(u, u+v)$ construction, and direct sum.

4) We discuss a connection between the generalized covering radii and the generalized Hamming weights of codes by showing that the latter is in fact a packing problem with some rank relaxation.

The paper is organized as follows: Preliminaries and notations are presented in Section II. We study various definitions of the generalized covering radius, and show them to be equivalent, in Section III. Section IV is devoted to the derivation of bounds on the generalized covering radii. Basic operations on codes are studied in Section V, and a relation with the generalized Hamming weights in Section VI. We conclude with a discussion of the results and some open questions in Section VII. Due to space limitations, proofs are omitted or sketched. For the full proofs the reader is referred to [6].

II. Preliminaries

For all $n \in \mathbb{N}$, we define $[n] \triangleq \{1, 2, \ldots, n\}$. If A is a finite set and $t \in \mathbb{N}$, we denote by $\binom{A}{t}$ the set of all subsets of A of size exactly t. We use $\mathbb{F}_q$ to denote the finite field of size q, and denote $\mathbb{F}_q^* \triangleq \mathbb{F}_q \setminus \{0\}$. Given a vector space V over $\mathbb{F}_q$, we denote by $\begin{bmatrix} V \\ t \end{bmatrix}$ the set of all vector subspaces of V of dimension $t \in \mathbb{N}$. We use lower-letters, v, to denote scalars, overlined lower-case letters, $\overline{v}$, to denote vectors, and either bold lower-case letters, $\mathbf{v}$, or upper-case letter, V, to denote matrices. Whether vectors are row vectors or column vectors is deduced from context.

If H is a matrix with n columns, we denote by $\overline{h}_i$ its i-th column. For $I = \{i_1, i_2, \ldots, i_t\} \in \binom{[n]}{t}$, we denote by H_I the restriction of H to the columns whose indices are in I, i.e., $H_I \triangleq (\overline{h}_{i_1}, \ldots, \overline{h}_{i_t})$. We shall also use $\langle H_I \rangle$ to denote the linear space spanned by the columns of H_I

Given $\overline{v} = (v_1, \ldots, v_n) \in \mathbb{F}_q^n$, the support of $\overline{v}$ is defined by $\operatorname{supp}(\overline{v}) \triangleq \{i \in [n] \,|\, v_i \neq 0\}$. Whenever required, for a subset $V \subseteq \mathbb{F}_q^n$ we define $\operatorname{supp}(V) \triangleq \bigcup_{\overline{v} \in V} \operatorname{supp}(\overline{v})$. The Hamming weight of $\overline{v}$ is then defined as $\operatorname{wt}(\overline{v}) \triangleq |\operatorname{supp}(\overline{v})|$. If $\overline{v}' \in \mathbb{F}_q^n$, then the Hamming distance between $\overline{v}$ and $\overline{v}'$ is given by $d(\overline{v}, \overline{v}') \triangleq \operatorname{wt}(\overline{v} - \overline{v}')$. We also extend the definition to the distance between a vector and a set, namely, for a set $C \subseteq \mathbb{F}_q^n$,

$$d(\overline{v}, C) \triangleq \min\{d(\overline{v}, \overline{c}) \,|\, \overline{c} \in C\}.$$

Finally, the Hamming ball of radius r centered at $\overline{v} \in \mathbb{F}_q^n$ is defined as

$$B_{r,n,q}(\overline{v}) \triangleq \left\{ \overline{v}' \in \mathbb{F}_q^n \,\middle|\, d(\overline{v}, \overline{v}') \leqslant r \right\}.$$

We shall omit the subscripts n and q whenever they may be inferred from the context.

III. The Generalized Covering Radii

We would now like to introduce the concept of generalized covering radius. We present several definitions, with varying approaches, be they combinatorial, algebraic, or geometric. We then show that all of the definitions are in fact equivalent (at least, when linear codes are concerned).

Our first definition stems directly from the application outlined in the introduction – database queries.

Definition 1 *Let C be an $[n, k]$ linear code over $\mathbb{F}_q$, given by an $(n-k) \times n$ parity-check matrix $H \in \mathbb{F}_q^{(n-k)\times n}$. For every $t \in \mathbb{N}$ we define the t-*th generalized covering radius*, $R_t(C)$, to be the minimal integer $r \in \mathbb{N}$ such that for every set $S \in \binom{\mathbb{F}_q^{n-k}}{t}$ there exists $I \in \binom{[n]}{r}$ such that $S \subseteq \langle H_I \rangle$, i.e.,*

$$R_t(C) \triangleq \max_{\substack{S \subseteq \mathbb{F}_q^{n-k} \\ |S| = t}} \min_{\substack{I \subseteq [n] \\ S \subseteq \langle H_I \rangle}} |I|.$$

While $R_t(C)$ certainly depends on the code C, for the sake of brevity we sometimes write R_t when we can infer C from the context. At first glance it seems as if R_t does not only depend on C, but also on the choice of parity-check matrix H. However, the following lemma shows this is not the case.

Lemma 2 *Let R_t be given by a full-rank matrix $H \in \mathbb{F}_q^{(n-k)\times n}$ as in Definition* 1*. For any $A \in \mathrm{GL}(n-k, q)$ (the group of $(n-k) \times (n-k)$ invertible matrices with coefficients in $\mathbb{F}_q$), let R'_t be the generalized covering radius, as in Definition* 1*, but using the matrix AH. Then $R_t = R'_t$.*

We observe, in Definition 1, that requiring $S \subseteq \langle H_I \rangle$ also ensures $\langle S \rangle \subseteq \langle H_I \rangle$. We therefore must have for all $t \in [n-k]$,

$$R_t \geqslant t. \tag{1}$$

We also observe that R_1 is in fact the covering radius of the code C, and that the generalized covering radii are naturally monotone increasing, i.e.,

$$R_1 \leqslant R_2 \leqslant \ldots \leqslant R_{n-k} = n - k, \tag{2}$$

as well as $R_t = n - k$ for all $t \geqslant n - k$. Thus, the values $R_1, \ldots, R_{n-k}$ are called the *generalized covering-radius hierarchy*. While being monotone increasing, we do note however, that the generalized covering radius R_t is not necessarily strictly increasing in t (e.g., the shortened binary Hamming code, see [6]).

Aiming for a geometric interpretation of the generalized covering radii, we provide two more equivalent definitions that are increasingly geometric in nature.

Definition 3 *Let C be an $[n, k]$ linear code over $\mathbb{F}_q$. Then for every $t \in \mathbb{N}$ we define the t-th generalized covering radius, $R_t(C)$, to be the minimal integer $r \in \mathbb{N}$ such that for every $\overline{v}_1, \ldots, \overline{v}_t \in \mathbb{F}_q^n$, there exist codewords $\overline{c}_1, \ldots, \overline{c}_t \in C$ and $I \in \binom{[n]}{r}$, such that $\operatorname{supp}(\overline{v}_i - \overline{c}_i) \subseteq I$ for all $i \in [t]$.*

We now move to a "classical" covering in the geometric sense. It involves a covering of a space with certain shapes. We shall require an extension the Hamming metric to a t-Hamming metric. The space we operate in is $\mathbb{F}_q^{t\times n}$. The *t-weight* of a matrix $\mathbf{v} \in \mathbb{F}_q^{t\times n}$, with row vectors denoted $\overline{v}_i$, is defined as

$$\operatorname{wt}^{(t)}(\mathbf{v}) \triangleq \left| \bigcup_{i \in [t]} \operatorname{supp}(\overline{v}_i) \right|.$$

We now define the t-distance between $\mathbf{v}, \mathbf{v}' \in \mathbb{F}_q^{t\times n}$ as

$$d^{(t)}(\mathbf{v}, \mathbf{v}') \triangleq \mathrm{wt}^{(t)}(\mathbf{v} - \mathbf{v}').$$

The t-ball centred in a matrix $v \in \mathbb{F}_q^{t\times n}$ is defined by

$$B_r^{(t)}(\mathbf{v}) = \left\{ \mathbf{v}' \in \mathbb{F}_q^{t\times n} \;\middle|\; d^{(t)}(\mathbf{v}, \mathbf{v}') \leqslant r \right\}.$$

We also note that $d^{(1)}$ is simply the Hamming distance function, hence our previous observation of a 1-ball being a ball in the Hamming metric. A superscript of $^{(1)}$ will generally be omitted unless a special need for emphasis arises.

Definition 4 *Let C be an $[n,k]$ linear code over $\mathbb{F}_q$. Then for every $t \in \mathbb{N}$, we define the t-th generalized covering radius, R_t, to be the minimal integer r such that t-balls centered at*

$$C^t \triangleq \left\{ \begin{bmatrix} \bar{c}_1 \\ \vdots \\ \bar{c}_t \end{bmatrix} \;\middle|\; \forall i \in [t], \bar{c}_i \in C \right\},$$

cover $\mathbb{F}_q^{t\times n}$, i.e., $\bigcup_{\mathbf{c}\in C^t} B_r^{(t)}(\mathbf{c}) = \mathbb{F}_q^{t\times n}$.

We would like to comment that if we denote the *columns* of $\mathbf{v} \in \mathbb{F}_q^{t\times n}$ by $\widehat{v}_1, \ldots, \widehat{v}_n \in \mathbb{F}_q^t$, then

$$\mathrm{wt}^{(t)}(\mathbf{v}) = \left|\left\{ j \in [n] \;\middle|\; \widehat{v}_j \neq \bar{0} \right\}\right|.$$

This metric is known in the literature as the *block metric*, introduced independently by Gabidulin [8] and Feng [11].

For our last approach, we make the obvious next step, resulting in an algebraic definition of the generalized covering radii. Using the well known isomorphism $\mathbb{F}_q^t \cong \mathbb{F}_{q^t}$, we view $C^t \subseteq \mathbb{F}_q^{t\times n}$ as a subset of $\mathbb{F}_{q^t}^n$. Under this isomorphism, the t-covering of a code C becomes the first covering radius of C^t in $\mathbb{F}_{q^t}^n$. Thus, we obtain the following equivalent definition:

Definition 5 *Let C be an $[n,k]$ linear code over $\mathbb{F}_q$. Assume $G \in \mathbb{F}_q^{k\times n}$ is a generator matrix for C, namely, $C = \{\overline{u}G \,|\, \overline{u} \in \mathbb{F}_q^k\}$. Let $t \in \mathbb{N}$, and let C' be the linear code over $\mathbb{F}_{q^t}$ generated by the same matrix G, namely, $C' = \{\overline{u}G \,|\, \overline{u} \in \mathbb{F}_{q^t}^k\}$. Then we define the t-th generalized covering radius R_t of C as the covering radius of C', namely,*

$$R_t(C) \triangleq R_1(C').$$

Lemma 6 *Let C be an $[n,k]$ linear code over $\mathbb{F}_q$. Then the values of R_t from Definitions 1, 3, 4, and 5, are the same.*

As a final comment to this section, our original approach to generalize the covering radii of a code, introduced in Definition 1, arises from the interest in querying databases by linear combinations (as, for example, used in PIR), and it uses the parity-check matrix of a code, hence it makes sense only for linear codes. This is not the case for the approach in Definition 4, where R_t is defined intrinsically as a metric invariant. This means that we can use this definition to generalize the covering radii for general (non-linear) codes.

IV. Bounds

A crucial part in our understanding of any figure of merit, is the limits of values it can take. Thus, we devote this section to the derivation of bounds on the generalized covering radii of codes. We put an emphasis on asymptotic bounds, that, given the normalized t-th covering radius, bound the best possible rate. We present a straightforward ball-covering argument for a lower bound. We then also present a trivial upper bound. Our main result is an asymptotic upper bound that improves upon the trivial one, and thus showing there is merit to the usage of generalized covering radii to improve database querying, as described in Section I. Our upper bound is non-constructive, and uses the probabilistic method. It shall be made constructive (albeit, not useful) in Section V.

As is standard, we will require the size of a t-ball. Since the metrics involved are all translation invariant, the size of the ball does not depend on the choice of center. We therefore use

$$V_{r,n,q}^{(t)} \triangleq \left| B_{r,n,q}^{(t)}(\mathbf{0}) \right|.$$

Let $k_t(n,r,q)$ denote the smallest dimension of a linear code C over $\mathbb{F}_q$ with length n and t-covering radius $R_t(C) \leqslant r$. The following theorem was proved in [5].

Theorem 7 ([5]) *For all $n, r \in \mathbb{N}$, and a prime power q,*

$$\begin{aligned} n - \log_q V_{r,n,q}^{(1)} &\leqslant k_1(n,r,q) \\ &\leqslant n - \log_q V_{r,n,q}^{(1)} + 2\log_2 n - \log_q n + O(1). \end{aligned}$$

It is convenient to study normalized parameters with respect to the length of the code. If C is an $[n,k]$ linear code, we define its normalized parameters, $\kappa \triangleq \frac{k}{n}$, and $\rho_t \triangleq \frac{R_t}{n}$. Note that we use κ for the rate of the code, and not R, to avoid confusion with the covering radius. For $t \in \mathbb{N}$ and a normalized covering radius $0 \leqslant \rho \leqslant 1$, the minimal rate achieving ρ is defined as

$$\kappa_t(\rho, q) \triangleq \liminf_{n\to\infty} \frac{k_t(n, \rho n, q)}{n}.$$

In this notation, Theorem 7 gives an asymptotically tight expression,

$$\kappa_1(n,\rho) = \begin{cases} 1 - H_q(\rho) & 0 \leqslant \rho < 1 - \frac{1}{q}, \\ 0 & 1 - \frac{1}{q} \leqslant \rho \leqslant 1, \end{cases} \tag{3}$$

where $H_q(x)$ is the q-ary entropy function given by

$$H_q(x) = x\log_q(q-1) - x\log_q(x) - (1-x)\log_q(1-x).$$

A. General Bounds

For a simple lower bound we use ball-covering.

Proposition 8 *For any $n, t \in \mathbb{N}$, prime power q, and $0 \leqslant \rho \leqslant 1 - \frac{1}{q^t}$, we have $\kappa_t(\rho, q) \geqslant 1 - H_{q^t}(\rho)$.*

For an upper bound, we first observe the following:

Proposition 9 *Let C be an $[n,k]$ code over $\mathbb{F}_q$. Then for all $t \in \mathbb{N}$, we have $R_t \leqslant t \cdot R_1$.*

We can now give the following naive upper bound.

Proposition 10 *For any $n, t \in \mathbb{N}$, $t \geqslant 2$, prime power q, and $0 \leqslant \rho \leqslant 1$, we have $\kappa_t(\rho, q) \leqslant 1 - H_q\left(\frac{\rho}{t}\right)$.*

Proposition 9 is in fact a consequence of the following, more general, upper bound. This upper bound shows that the generalized covering radii are sub-additive.

Proposition 11 *Let C be an $[n,k]$ code over $\mathbb{F}_q$. Then for all $t_1, t_2 \in \mathbb{N}$, we have $R_{t_1+t_2} \leqslant R_{t_1} + R_{t_2}$.*

B. Upper Bounding the Binary Case with $t = 2$

The upper bound we now present improves upon the trivial one from Proposition 10. Since it is significantly more complex, and has many moving parts, we focus on the binary case with $t = 2$ only. We follow a similar strategy to the one employed by [4, Theorem 12.3.5] for the covering radius, though major adjustments are required due to the more involved nature of this generalized problem. In essence, we show the existence of a covering code using the probabilistic method. The probability is nearly 1, implying almost *all* codes are at least as good as this bound. The main result is Theorem 12.

We outline the proof strategy to facilitate reading this section. We use the probabilistic method by choosing a random generator matrix for a code and bounding the probability that balls centered at the codewords indeed cover the entire space. To do so, we study the random variable that counts how many codewords cover a given point in space. To get a handle on this variable, we bound its expectation and variance.

We consider the function $f : [0,1] \to \mathbb{R}$ defined by

$$f(\rho) = \begin{cases} H_2(s(\rho)) + 2s(\rho) \\ \quad +2(1-s(\rho))H_2\left(\frac{\rho - s(\rho)}{1-s(\rho)}\right) & 0 \leqslant \rho < \frac{3}{4}, \\ 3 & \frac{3}{4} \leqslant \rho \leqslant 1, \end{cases}$$

where

$$s(\rho) \triangleq \frac{1}{10}\left(1 + 8\rho - \sqrt{1 + 16\rho - 16\rho^2}\right).$$

All the required supporting lemmas appear in the full version [6]. The main result is the following:

Theorem 12 *For any $0 < \rho \leqslant 1$,*

$$\kappa_2(\rho, 2) \leqslant \begin{cases} 1 - (4H_4(\rho) - f(\rho)) & 0 \leqslant \rho < \frac{3}{4}, \\ 0 & \frac{3}{4} \leqslant \rho \leqslant 1. \end{cases}$$

A comparison of the various asymptotic bounds is shown in Figure 1. It is interesting to note that the upper bound of Theorem 12 matches the lower ball-covering bound at $\rho = \frac{3}{4}$, particularly so because the function $f(\rho)$ is defined by the binary entropy function, and not the quaternary entropy function. We also note that the naive upper bound of Proposition 10 is better than the upper bound of Theorem 12 for $\rho \lesssim 0.145$.

We would like to remark that the analysis performed in the proof of our upper bound is already involved for the presented case (of $q = 2$ and $t = 2$). A case-by-case analysis is preformed, resulting in a complicated optimization problem. Using current techniques, this approach is unscalable, becoming intractable for larger values of q and t.

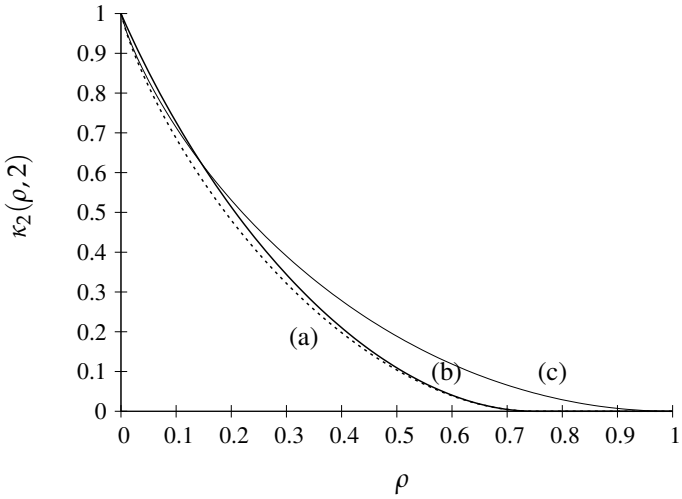

Fig. 1. A comparison of the bounds on $\kappa_2(\rho, 2)$: (a) the ball-covering lower bound, (b) the upper bound of Theorem 12, and (c) the naive upper bound of Proposition 10.

V. Simple Code Operations

Some code operations are very common. Among these we can find code extension, code puncturing, the $(u, u+v)$ construction, and direct sum. In this section we show the effect these operations have on the generalized covering radii mimics their effect on the (regular) covering radius. We use the direct product to turn the non-constructive upper bound of Theorem 12 to an explicit construction, albeit, not a very useful one.

Given a code $C \subseteq \mathbb{F}_q^n$, let

$$C^* \triangleq \{(c_1, \ldots, c_{n-1}) \,|\, (c_1, \ldots, c_{n-1}, c_n) \in C\},$$

be the *punctured code*, and

$$\overline{C} \triangleq \left\{ (c_1, \ldots, c_n, -\sum_{i=1}^{n} c_i) \;\middle|\; (c_1, \ldots, c_n) \in C \right\},$$

be the *extended code*. Even though puncturing is defined as the removal of the last coordinate, the following results apply to the removal of any single coordinate.

By [4, Theorem 3.1.1, p. 62], $R_1(C^*)$ equals $R_1(C)$ or $R_1(C) - 1$ and $R_1(\overline{C})$ equals $R_1(C)$ or $R_1(C) + 1$. The same result holds for the generalized covering radii.

Proposition 13 *Let C be an $[n,k]$ linear code. Then for any $t \in \mathbb{N}$,*

1) *$R_t(C^*)$ equals $R_t(C)$ or $R_t(C) - 1$;*
2) *$R_t(\overline{C})$ equals $R_t(C)$ or $R_t(C) + 1$.*

Assume C_1 and C_2 are $[n, k_1]$ and $[n, k_2]$ codes, respectively. The $(u, u+v)$ construction uses C_1 and C_2 to produce a code

$$C = \{(\overline{u}, \overline{u} + \overline{v}) \,|\, \overline{u} \in C_1, \overline{v} \in C_2\},$$

and by [4, Theorem 3.4.1, p. 66], its covering radius is upper bounded by $R_1(C) \leqslant R_1(C_1) + R_1(C_2)$.

Proposition 14 *Let C_i be an $[n,k_i]$ code over $\mathbb{F}_q$, $i=1,2$, and let C be the code constructed from C_1 and C_2 using the $(u,u+v)$ construction. Then for any $t\in\mathbb{N}$,*

$$R_t(C)\leqslant R_t(C_1)+R_t(C_2).$$

We now look at the direct sum. Given an $[n_1,k_1]$ code C_1, and an $[n_2,k_2]$ code C_2, both over $\mathbb{F}_q$, the direct sum is defined as

$$C_1\oplus C_2\triangleq\{(\bar{c}_1,\bar{c}_2)\,|\,\bar{c}_1\in C_1,\bar{c}_2\in C_2\},$$

which is an $[n_1+n_2,k_1+k_2]$ code over $\mathbb{F}_q$. It is well known [4, Theorem 3.2.1, p. 63] that

$$R_1(C_1\oplus C_2)=R_1(C_1)+R_1(C_2).$$

Proposition 15 *Let C_i be an $[n_i,k_i]$ code over $\mathbb{F}_q$, for $i=1,2$. Then for any $t\in\mathbb{N}$,*

$$R_t(C_1\oplus C_2)=R_t(C_1)+R_t(C_2).$$

Remark 16 *The upper bound presented in Theorem 12 is proved by showing the existence of a sequence of codes in a non-constructive way. Using Proposition 15 and a probabilistic argument employed in the proof of Theorem 12, we find an explicit construction for a code attaining the bound of Theorem 12. We construct our codes by a standard procedure of taking the direct sums of codes. The construction is explained in detail in [6].*

The chief disadvantage of our construction is the enormous block length of the resulting code. In our construction, in order to ensure a normalized covering radius at most $\rho+\varepsilon$ the required block length is $\Omega(2^{2^{1/\varepsilon}+1/\sqrt{\varepsilon}})$.

VI. The Generalized Packing Radii

Given an $[n,k]$ linear code C over $\mathbb{F}_q$, the generalized Hamming weight of the code, d_t, $t\in\mathbb{N}$, is defined as the minimal support size containing a linear subcode of C of dimension t, i.e.,

$$d_t\triangleq\min_{C'\in\begin{bmatrix}C\\t\end{bmatrix}}\left|\mathrm{supp}(C')\right|.$$

In particular, d_1 is the usual minimum distance of C.

Generalized Hamming weights were introduced by Wei in 1991 [13], as a figure of merit to analyze the security performance of a code on a wire-tap channel. Wei proved that the weight hierarchy is strictly increasing and proved the duality theorem, relating the weight hierarchy of a code and its dual.

In the following we shall require the size $\lfloor(d_t-1)/2\rfloor$. To simplify the presentation we define for all $t\in\mathbb{N}$, $\delta_t\triangleq\left\lfloor\frac{d_t-1}{2}\right\rfloor$. We also define the set

$$\mathcal{L}^{(t)}(\mathbb{F}_q^n)\triangleq\left\{\mathbf{v}\in\mathbb{F}_q^{t\times n}\,\middle|\,\mathrm{rank}(\mathbf{v})=t\right\}.$$

Lemma 17 *Let C be an $[n,k]$ linear code over $\mathbb{F}_q$. Then for every $t\in[k]$, δ_t is the largest integer satisfying that for all $\mathbf{c},\mathbf{c}'\in C^t$ such that $\mathbf{c}-\mathbf{c}'\in\mathcal{L}^{(t)}(\mathbb{F}_q^n)$,*

$$B_{\delta_t}^{(t)}(\mathbf{c})\cap B_{\delta_t}^{(t)}(\mathbf{c}')=\emptyset.$$

We observe that for $t=1$, Lemma 17 becomes the standard packing of Hamming error balls induced by the code C, and δ_1 is the packing radius of the code, and hence, $\delta_1\leqslant R_1$. It is therefore tempting to conjecture that $\delta_t\leqslant R_t$ for all $t\in[\min\{k,n-k\}]$. However, Lemma 17 does not describe a packing of t-balls, when $t\geqslant 2$, since these may intersect if the difference between their centers is not of full rank.

VII. Conclusion

We proposed a fundamental property of linear codes – the generalized covering-radius hierarchy. It characterizes the trade-off between storage amount, latency, and access complexity in databases queried by linear combinations, as is the case, for example, in PIR schemes. We showed three equivalent definitions for these radii, highlighting their combinatorial, geometric, and algebraic aspects. We derived bounds on the code parameters in relation with the generalized covering radii, and studied the effect simple code operations have on them. Finally, we described a connection between the generalized covering-radius hierarchy and the generalized Hamming weight hierarchy.

While the study of the generalized covering-radius hierarchy has its own independent intellectual merit, let us also place the bound of Theorem 12 back in the context of PIR schemes. Consider the binary case, and assume we allow a latency of $t=2$, namely, the server waits until two queries arrive and then handles them both. Further assume, that to handle the two queries we allow the server to access at most $\frac{1}{2}$ of its storage. Stated alternatively, the average access per query is a $\frac{1}{4}$ of the storage. By Theorem 12, since $\kappa_2(\frac{1}{2},2)\approx 0.11$, there exists a code allowing 89% of the server storage for user information and only 11% for overhead. A naive approach, using $\kappa_1(\frac{1}{4},2)\approx 0.19$, implies the storage may contain only 81% user information and 19% overhead.

Many other open problems remain, and we mention but a few. First, extending Theorem 12 to address non-binary generalized covering radii for all t is still an open question, as is closing the gap with the lower bound of Proposition 8.

It would also be interesting to determine the generalized covering-radius hierarchy of known codes. These may be extreme in some cases. We can show that the Hamming code satisfies $R_t=t$, and in particular the covering-radius hierarchy is strictly increasing, that is, $R_t<R_{t+1}$ for all $t\in[n-k-1]$. We can also show the only non-trivial code with this property is the Hamming code. In contrast with the Hamming code, whose generalized covering radii are all distinct, the opposite occurs with MDS codes. As was shown in [1], [7], the (first) covering radius of $[n,k]$ MDS codes is $n-k$, except in rare cases where it is $n-k-1$. Since the upper limit on the generalized covering radius is $n-k$, the entire hierarchy is either constant, or is a step function.

Finally, we have an algorithmic question: Given a parity-check matrix H for an $[n,k]$ code over $\mathbb{F}_q$, and given vectors $\bar{s}_1,\ldots,\bar{s}_t\in\mathbb{F}_q^{n-k}$, how do we efficiently find R_t columns of H that span the t vectors? These questions, and many others, are left for future research.

REFERENCES

[1] D. Bartoli, M. Giulietti, and I. Platoni, "On the covering radius of MDS codes," *IEEE Trans. Inform. Theory*, vol. 61, no. 2, pp. 801–811, Feb. 2015.

[2] B. Chazelle and B. Rosenberg, "Computing partial sums in multidimensional arrays," in *Proceedings of the fifth annual symposium on Computational geometry*, 1989, pp. 131–139.

[3] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *Proceedings of IEEE 36th Annual Foundations of Computer Science*. IEEE, 1995, pp. 41–50.

[4] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering codes*. North-Holland, 1997.

[5] G. Cohen and P. Frankl, "Good coverings of Hamming spaces with spheres," *Discrete Mathematics*, vol. 56, no. 2-3, pp. 125–131, 1985.

[6] D. Elimelech, M. Firer, and M. Schwartz, "The generalized covering radii of linear codes," *arXiv preprint arXiv:2012.06467*, 2020.

[7] E. M. Gabidulin and T. Kløve, "The Newton radius of MDS codes," in *1998 Information Theory Workshop (ITW) Killarney, Ireland*, Jun. 1998, pp. 50–51.

[8] E. Gabidulin, "Combinatorial metrics in coding theory," in *2nd International Symposium on Information Theory, Armenia, USSR*, 1971.

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

[10] C.-T. Ho, J. Bruck, and R. Agrawal, "Partial-sum queries in OLAP data cubes using covering codes," *IEEE Trans. Comput.*, vol. 47, no. 12, pp. 1326–1340, Dec. 1998.

[11] L. X. K. Feng and F. J. Hickernell, "Linear error-block codes," *Finite Fields and Their Applications*, vol. 12, pp. 638–652, 2006.

[12] I. Tamo, Z. Wang, and J. Bruck, "Access versus bandwidth in codes for storage," *IEEE Trans. Inform. Theory*, vol. 60, no. 4, pp. 2028–2037, Apr. 2014.

[13] V. K. Wei, "Generalized Hamming weights for linear codes," *IEEE Trans. Inform. Theory*, vol. 37, no. 5, pp. 1412–1418, Sep. 1991.

[14] Y. Zhang, E. Yaakobi, T. Etzion, and M. Schwartz, "On the access complexity of PIR schemes," in *Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT2019), Paris, France*, Jul. 2019, pp. 2134–2138.