Function-Correcting Codes

Andreas Lenz, Rawad Bitar, Antonia Wachter-Zeh, and Eitan Yaakobi

Abstract—Motivated by applications in machine learning and archival data storage, we introduce function-correcting codes, a new class of codes designed to protect a function evaluation on the data against errors. We show that function-correcting codes are equivalent to irregular-distance codes, i.e., codes that obey some given distance requirement between each pair of codewords. Using these connections, we study irregular-distance codes and derive general upper and lower bounds on their optimal redundancy. Since these bounds heavily depend on the specific function, we provide simplified, suboptimal bounds that are easier to evaluate. We further employ our general results to specific functions of interest and we show that function-correcting codes can achieve significantly less redundancy than standard error-correcting codes which protect the whole data.

I. INTRODUCTION

In standard communication systems a sender transmits a digital message to a receiver via a channel. To protect this message from errors, it is first encoded using an error-correcting code and then transmitted over the channel to the receiver, which decodes the received word to obtain the original message. Within this setup, the goal is to recover the message correctly.

Consider in contrast the setup where instead of the complete message, only a certain function of this message shall be conveyed. Such a setting arises, for example, in machine learning applications [1]–[3] or archival data storage [4], where a large message is available, however only a specific attribute, respectively function, of this message is of interest. We therefore consider a communication scenario, where the receiver’s task is to recover only a certain function of the message, which is illustrated in Fig. 1. This paradigm gives rise to a new class of codes, called function-correcting codes (FCCs), which encode the message to allow a successful recovery of the function value after transmitting the codeword over a channel. In this work, we consider the setup where the message itself is transmitted over the channel, but it is also possible to define the problem for non-systematic encoding. However, in many scenarios it is desired to leave the data in its original form, which makes a systematic encoding necessary. Such scenarios include distributed computing and archival storage, in which the sender and the receiver have access to the message vectors. Interestingly, in contrast to the standard error-correcting problem, allowing for non-systematic encoding in the function-correcting set-up significantly changes the codes and also achievable code redundancies. This is because non-systematic function-correction can be achieved by employing an error-correcting code over the possible outputs of the function.

In general, the key advantage of FCCs over standard error-correcting codes is a reduced redundancy.

Similarly, application-specific error-correcting codes reduce the redundancy, e.g., in order to cope with computation errors in matrix-vector multiplications [5]–[7]; to construct energy-adaptive codes [8]; to optimize the output of a given machine learning algorithm [1]–[3]; to ensure reliable distributed encoding [9]; or to optimize classification [2]. In [3], error-correcting codes are applied to the weights of the neurons in a neural network. The goal is not to protect the stored weights, but to optimize the output model of the neural network.

In contrast to these application-specific works, this paper builds a general understanding for function correction over adversarial channels and derives results for arbitrary functions. Finally, these results are applied to specific functions such as locally binary-valued functions (defined below), the Hamming weight, the Hamming weight distribution, and the min-max function. We restrict our attention to binary channels in this work, however most results can be generalized straightforwardly to larger alphabets.

II. PRELIMINARIES

Let $u \in \mathbb{Z}_2^k$ be the binary message and let $f : \mathbb{Z}_2^k \to \text{Im}(f)$ be a function computed on $u$. The data is encoded via the systematic encoding function $\text{Enc} : \mathbb{Z}_2^k \to \mathbb{Z}_2^{k+r}$, $\text{Enc}(u) = (u, p(u))$, where $p(u) \in \mathbb{Z}_2^r$ is the redundancy vector and $r$ is the redundancy. The resulting codeword $\text{Enc}(u)$ is transmitted over a substitution channel, resulting in $y \in \mathbb{Z}_2^{k+r}$ with $d(y, \text{Enc}(u)) \leq t$, where $d(x, y)$ is the Hamming distance between $x$ and $y$. We call $E \triangleq \frac{\text{Im}(f)}{2^k}$ the expressiveness of $f$ and define FCCs as follows.

Definition 1. An encoding function $\text{Enc} : \mathbb{Z}_2^k \to \mathbb{Z}_2^{k+r}$ defines a function-correcting code for the function $f : \mathbb{Z}_2^k \to \text{Im}(f)$ if for all $u_1, u_2 \in \mathbb{Z}_2^k$ with $f(u_1) \neq f(u_2)$,

$$d(\text{Enc}(u_1), \text{Enc}(u_2)) \geq 2t + 1.$$ 

By this definition, given any $y$, which is obtained by at most $t$ errors from $\text{Enc}(u)$, the receiver can uniquely recover $f(u)$, if it has knowledge about $f(\bullet)$ and $\text{Enc}(\bullet)$. Noteworthy, only codewords that originate from information vectors that evaluate to different function values need to have distance at least $2t + 1$. Throughout the paper, a standard error-correcting code is an FCC for $f(u) = u$, i.e., a code that allows to reconstruct the whole message $u$. We summarize some basic properties of FCCs in the following.
1) For any bijective function $f$, any FCC is a standard error-correcting code.
2) For any constant function $f$, the encoder $\text{Enc}(u) = u$ is an FCC with redundancy 0.
3) For any function $f$, if the encoder has no knowledge about the function $f$, function-correction is only possible using standard error-correcting codes.

The main quantity of interest in this paper is the optimal redundancy of an FCC that is designed for a function $f$.

**Definition 2.** The optimal redundancy $r_f(k,t)$ is defined as the smallest $r$ such that there exists an FCC with encoding function $\text{Enc} : \mathbb{Z}_2^k \to \mathbb{Z}_2^{k+r}$ for the function $f$.

We define $V(n,d) = \sum_{i=0}^{d} \binom{n}{i}$ as the volume of the Hamming ball of radius $d$. For any integer $M$, we write $[M] = \max\{M, 0\}$ and we let $[M] \triangleq \{1, \ldots, M\}$. Note that while our quantitative results in this paper are for substitution channels, the concepts can be generalized to other channels.

We will show that FCCs are closely related to so-called irregular-distance codes, which will be introduced and discussed in Section III. Irregular-distance codes are codes that have a specific distance between each pair of codewords. In particular, we will show in Section IV that the optimal redundancy of an FCC is given by the smallest length $r$ for which an irregular-distance code exists. Based on this, we derive generic results about FCCs and apply these to specific functions in Section V.

### III. Irregular Distance Codes

Let $P = \{p_1, p_2, \ldots, p_M\} \subseteq \mathbb{Z}_2^k$ be a code of length $r$ and cardinality $M$. Note that we choose $r$ as the code blocklength, as we will relate the code length $r$ to the redundancy of FCCs later.

We call a symmetric matrix $D \in \mathbb{N}_0^{M \times M}$ a distance matrix if $D_{ij} \geq 0$, $i, j \in [M]$ and $D_{ii} = 0$.

**Definition 3.** We call $P = \{p_1, p_2, \ldots, p_M\}$ an $[M,D]$ code, if $d(p_i, p_j) \geq D_{ij}$ for all $i, j \in [M]$.

Note that since the codewords $p_i$ are not ordered, the association of codewords with columns and rows of the matrix $D$ is not fixed and we therefore assume that they are arranged such that the distance requirements are fulfilled.

**Definition 4.** Let $M \in \mathbb{N}$ and the distance matrix $D \in \mathbb{N}_0^{M \times M}$ be given. We define $N(M,D)$ to be the smallest integer $r$ such that there exists an $[M,D]$ code of length $r$. For the case, where $D_{ij} = D$ for all $i \neq j$ we write $N(M,D)$.

We summarize some results about $N(M,D)$ here, which allow us to obtain results on the redundancy of FCCs using Theorems 1 and 2. We start by a generalization of the Plotkin bound [10] on codes with irregular distance requirements.

**Lemma 1.** For any $M \in \mathbb{N}$ and distance matrix $D$,

$$N(M,D) \geq \left\{ \begin{array}{ll} \frac{1}{M^2} \sum_{i<j}^M |D_{ij}|, & \text{if } M \text{ is even,} \\ \frac{1}{M^2} \sum_{i<j}^M |D_{ij}| - 1, & \text{if } M \text{ is odd.} \end{array} \right.$$ 

The proof is obtained analogously to the proof of the standard Plotkin bound [10]. For the case of regular distance codes, we can deduce the following form of Plotkin’s bound.

**Corollary 1 (cf. [10]).** For any $M, D \in \mathbb{N}$,

$$N(M,D) \geq 2D \frac{M - 1}{M}.$$ 

Conversely, we can derive an achievable bound, which is a generalization of the well-known Gilbert-Varshamov bound [11], [12] to irregular-distance codes.

**Lemma 2.** For any $M \in \mathbb{N}$, distance matrix $D$, and any permutation $\pi : [M] \to [M]$,

$$N(M,D) \leq \min_{r \in \mathbb{N}} \left\{ r : 2^r > \max_{j \in [M]} \sum_{i=1}^{j-1} V(r, |D_{\pi(i)\pi(j)} - 1|) \right\}.$$ 

The lemma can be proven by a greedy selection of codewords and the full proof can be found in the extended version of this paper [13]. Note that for regular distance codes with $|D_{ij}| = D$, the bound results in the well-known Gilbert-Varshamov bound [11], [12]. Several of our results in the following require codes of small cardinality, i.e., the code size is in the same order of magnitude as the minimum distance. The following result is based on Hadamard codes [14], [15].

**Lemma 3.** (cf. [15, Def. 3.13]) Let $D \in \mathbb{N}$ be such that there exists a Hadamard matrix of order $D$ and $M \leq 4D$. Then,

$$N(M,D) \leq 2D.$$ 

The range of the parameter $D$ is restricted to the lengths for which Hadamard codes are known to exist. Note that there exist other good codes of small size, such as weak flip codes [16], however, they only attain the Plotkin bound for a limited range of parameters. In general it is possible to puncture or juxtapose Hadamard codes (cf. Levenshtein’s theorem [14, Section 2.3]) to obtain codes for a larger range of parameters. However, for our discussion, the application of the Gilbert-Varshamov bound is sufficient and further allows to prove existence of codes whose size is quadratic in their minimum distance as follows.

**Lemma 4.** For any $M, D \in \mathbb{N}$ with $D \geq 10$ and $M \leq D^2$,

$$N(M,D) \leq \frac{2D}{1 - 2\sqrt{\ln(D)/D}}.$$ 

The proof of this Lemma is obtained using Lemma 2 together with [17, Lemma 4.7.2] and is contained in the full version [13]. This result means, that given the size of the code is moderate, i.e., $M \leq D^2$, for large $D$, the optimal length of an error-correcting code approaches $2D$. While Lemma 4 gives a slightly weaker bound than Lemma 3, it holds for any $D$ and for larger code sizes $M$. Note that a similar bound as in Lemma 4 can be derived also for larger $M$, i.e., $M \leq D^m$, $m > 2$, however $m = 2$ is sufficient for the subsequent analysis. These existence bounds allow to narrow down the optimal length of irregular-distance codes as follows.

**Lemma 5.** Let $M \in \mathbb{N}$ and $D \in \mathbb{N}_0^{M \times M}$ and denote $D_{\text{max}} = \max_{i,j} |D_{ij}|$, if $M \leq D_{\text{max}}$. We have

$$D_{\text{max}} \leq N(M,D) \leq \frac{2D_{\text{max}}}{1 - 2\sqrt{\ln(D_{\text{max}})/D_{\text{max}}}.$$ 

### IV. Generic Functions

This section is devoted to establishing general results on FCCs. We start by showing the relationship of FCCs and irregular-distance codes and proceed by establishing several lower and upper bounds on the optimal redundancy of FCCs. We define the distance matrix of the function $f$ as follows.
Definition 5. Let $u_1, \ldots, u_M \in \mathbb{Z}_2^k$. We define the distance requirement matrix $D_f(t, u_1, \ldots, u_M)_{ij}$ of a function $f$ as the $M \times M$ matrix with entries

$$[D_f(t, u_1, \ldots, u_M)]_{ij} = \begin{cases} 2t+1-d(u_i, u_j), & \text{if } f(u_i) \neq f(u_j), \\ 0, & \text{otherwise.} \end{cases}$$

We now develop bounds on the redundancy of FCCs. Based on Definitions 1 and 4, we find the following connection between the redundancy of optimal FCCs and irregular-distance codes.

Theorem 1. For any function $f : \mathbb{Z}_2^k \to \text{Im}(f)$,

$$r_f(k, t) = N(2^k, D_f(t, u_1, \ldots, u_{2^k})).$$

where $\{u_1, \ldots, u_{2^k}\} = \mathbb{Z}_2^k$ are all vectors of length $k$.

Proof. We first give a lower bound on the optimal redundancy, $r_f(k, t) \geq N(2^k, D_f(t, u_1, \ldots, u_{2^k}))$. By Definition 1, any FCC satisfies $d(\text{Enc}(u_i), \text{Enc}(u_j)) \geq 2t + 1$ for any $i \neq j$ with $f(u_i) \neq f(u_j)$. Let $u_1, \ldots, u_{2^k}$ be the information vectors from the statement of the theorem and let $p_1, \ldots, p_{2^k}$ be the corresponding redundancy vectors, i.e., $\text{Enc}(u_i) = (u_i, p_i)$. We prove the lower bound by contradiction. Assume on the contrary that $r_f(k, t) < N(2^k, D_f(t, u_1, \ldots, u_{2^k}))$. This implies that there must exist two redundancy vectors $p_i$ and $p_j$, $i \neq j$ with $d(p_i, p_j) < 2t + 1 - d(u_i, u_j)$ and henceforth $d(\text{Enc}(u_i), \text{Enc}(u_j)) = d(u_i, u_j) + d(p_i, p_j) < 2t + 1$ which is a contradiction.

On the other hand $r_f(k, t) \leq N(2^k, D_f(t, u_1, \ldots, u_{2^k}))$, as using a correctly assigned $[2^k, D_f(t, u_1, \ldots, u_{2^k})]_{ij}$ code for the redundancy vectors gives an FCC.

Theorem 1 is defined over all possible $2^k$ function vectors $u$. However, since both the distance matrix $D_f(t, u_1, \ldots, u_{2^k})$ and the resulting code length $N(2^k, D_f(t, u_1, \ldots, u_{2^k}))$ are often hard to analyze or compute, we now continue by deriving results that act on a smaller set of information vectors. In particular, using an arbitrary subset of information vectors $u_1, \ldots, u_M$ we can obtain a lower bound on the redundancy as follows.

Corollary 2. Let $u_1, \ldots, u_M \in \mathbb{Z}_2^k$ be arbitrary different information vectors. Then, the redundancy is at least

$$r_f(k, t) \geq N(M, D_f(t, u_1, \ldots, u_M)).$$

Since finding $N(2^k, D_f(t, u_1, \ldots, u_{2^k}))$ is in general quite difficult, it can be easier to focus only on a small but representative subset of information vectors. However, the particular subset heavily depends on the function itself and it seems quite challenging to give a generic approach on how a good subset can be found. Loosely speaking, good bounds are obtained for information vectors that have distinct function values and are close in Hamming distance. Throughout this paper, we will provide some insights on good choices of information vectors using illustrative examples. Specifically, we can derive the following corollary for arbitrary functions.

Corollary 3. Let $f : \mathbb{Z}_2^k \to \text{Im}(f)$ be an arbitrary function. Let $e^* \in \mathbb{N}$ denote the smallest integer that there exist information vectors $u_1, \ldots, u_E \in \mathbb{Z}_2^k$ with $\{f(u_1), \ldots, f(u_E)\} = \text{Im}(f)$ and $d(u_i, u_j) \leq e^*$. Then,

$$r_f(k, t) \geq N(E, 2t + 1 - e^*).$$

Corollary 3 will be interesting when comparing with explicit code constructions provided in subsequent sections. The following universal bound directly follows from Corollary 2.

Corollary 4. For any function $f$ with $E \geq 2$ and any $k, t$, we have $r_f(k, t) \geq 2t$.

Proof. Since $E \geq 2$, it is guaranteed that there exist $u, u' \in \mathbb{Z}_2^k$ with $d(u, u') = 1$ and $f(u) \neq f(u')$. From Corollary 2 it follows that $r_f(k, t) \geq N(2, 2t - 2t)$. \hfill \Box

We now prove the existence of FCCs with small redundancy. We start by defining the distance between two function values.

Definition 6. The distance between two function values $f_1, f_2 \in \text{Im}(f)$ is defined as the smallest distance between two information vectors that evaluate to $f_1$ and $f_2$, i.e.,

$$d_f(f_1, f_2) \triangleq \min_{u_1, u_2 \in \mathbb{Z}_2^k} d(u_1, u_2) \text{ s.t. } f(u_1) = f_1 \wedge f(u_2) = f_2.$$

Note that the distance $d_f(f_1, f_1) = 0, \forall f_1 \in \text{Im}(f)$. The function-distance matrix of $f$ is thus defined as follows.

Definition 7. The function-distance matrix of a function $f$ is denoted by the $E \times E$ matrix $D_f(t, f_1, \ldots, f_E)$ with entries $D_f(t, f_1, \ldots, f_E)_{ij} = [2t + 1 - d_f(f_i, f_j)]^+$. One way to construct FCCs is to assign the same redundancy to all information vectors $u$ that evaluate to the same function value. This is not a necessity, however it gives rise to the following existence theorem.

Theorem 2. For any arbitrary function $f : \mathbb{Z}_2^k \to \text{Im}(f)$,

$$r_f(k, t) \leq N(E, D_f(t, f_1, \ldots, f_E)).$$

Proof. We describe how to construct an FCC. First of all, we choose the redundancy vectors to only depend on the function value of $u$, i.e., the encoding mapping is defined by $u \mapsto (u, p(f(u)))$ and denote by $p$, the redundancy vector appended to all $u$ with $f(u) = f_i$. Therefore, any two information vectors which evaluate to the same function value have the same redundancy vectors. We then choose $p_1, \ldots, p_E$ such that $d(p_i, p_j) \geq 2t + 1 - d_f(f_i, f_j)$. It follows that for any $u_i, u_j$ with $f(u_i) = f_i, f(u_j) = f_j, f_i \neq f_j$, we have $d(\text{Enc}(u_i), \text{Enc}(u_j)) = d(u_i, u_j) + d(p_i, p_j) \geq d_f(f_i, f_j) + 2t + 1 - d_f(f_i, f_j) = 2t + 1$. By the definition of $N(M, D)$ we can guarantee the existence of such parity vectors $p_1, \ldots, p_E$, if they have length $N(E, D_f(t, f_1, \ldots, f_E))$.

There are cases in which the bound in Theorem 2 is tight. We characterize one important case in the following corollary, which is an immediate consequence of Corollary 2 and Theorem 2.

Corollary 5. If there exist a set of representatives $u_1, \ldots, u_E$ with $\{f(u_1), \ldots, f(u_E)\} = \text{Im}(f)$ and $D_f(t, f_1, \ldots, f_E) = D_f(t, f_1, \ldots, f_E)$, then

$$r_f(k, t) = N(E, D_f(t, f_1, \ldots, f_E)).$$

However, even for the case when the bound in Theorem 2 is not necessarily tight, in many cases it is much easier to derive the function distance matrix $D_f(t, f_1, \ldots, f_E)$ and the corresponding value $N(E, D_f(t, f_1, \ldots, f_E))$, especially when $E$ is small. Having these general theorems, we now present a code construction that can be applied to any function $f$. The construction uses standard error-correcting codes, suitably adapted to correct function values with low redundancy.
Lemma 6. For any function \( f \), \( r_f(k, t) \leq N(E, 2t) \).

Proof. We will prove the lemma based on an explicit code construction. Denote by \( C(E, 2t) \) a code with cardinality \( E \), minimum distance 2t and optimum length \( N(E, 2t) \). Further let \( E: \text{Im} f \rightarrow C(E, 2t) \) be an encoding function of this code. Then, define the FCC \( \text{Enc}_1(u) = (u, E(f(u))) \). We can directly verify that this construction yields an FCC by proving that Definition 1 applies. Let \( u_1, u_2 \in \mathbb{Z}_2^k \) with \( f(u_1) \neq f(u_2) \). Then, \( d(\text{Enc}_1(u_1), \text{Enc}_1(u_2)) = d(u_1, u_2) + d(E(f(u_1)), E(f(u_2))) \). Since \( d(E(f(u_1)), E(f(u_2))) \geq 2t \), it follows that \( d(\text{Enc}_1(u_1), \text{Enc}_1(u_2)) \geq 1 + 2t \). Therefore, \( \text{Enc}_1 \) defines an FCC for the function \( f \). \( \square \)

While for the statement of Lemma 6 a code of optimum length has been chosen, in practice it is certainly possible to choose any code of cardinality \( E \) and minimum distance 2t.

By Corollary 4 this simple construction has optimal redundancy for any binary-valued function with \( E = 2 \), which is summarized in the following corollary.

Corollary 6. For any function \( f \) with \( E = 2 \), \( r_f(k, t) = 2t \).

However, for larger images \( E > 2 \), this construction is not necessarily optimal anymore. We will see that Corollary 6 is a special case of a broader class of functions, called locally binary-valued codes, which we will discuss in the sequel.

V. APPLICATION TO SPECIFIC FUNCTIONS

We now turn to discuss specific functions and give bounds on their optimal redundancy, which are tight in several cases. For several instances we additionally give explicit code constructions that can be encoded efficiently. The functions under discussion are locally binary-valued functions, the Hamming weight, the Hamming weight distribution, and the min-max function.

A. Locally Binary-Valued Functions

In the following we define a broad class of functions, called locally binary-valued functions. We derive their optimal redundancy and show how it can be obtained using a simple explicit code construction. This class of functions is defined next.

Definition 8. The function ball of a function \( f \) with radius \( \rho \) around \( u \in \mathbb{Z}_2^k \) is defined by
\[
B_f(u, \rho) = \{ f(u') : u' \in \mathbb{Z}_2^k \land d(u, u') \leq \rho \}.
\]

Locally binary-valued functions are defined as follows.

Definition 9. A function \( f: \mathbb{Z}_2^k \rightarrow \text{Im} f \) is called a \( \rho \)-locally binary-valued function, if for all \( u \in \mathbb{Z}_2^k, |B_f(u, \rho)| \leq 2 \).

Intuitively, a \( \rho \)-locally binary-valued function is a function, where the preimages of all function values are well spread in the sense that each information word is close to only one preimage of another function value. By this definition, any binary-valued function is also a \( \rho \)-locally binary-valued function for arbitrary \( \rho \). We can directly prove the following optimality.

Lemma 7. For any 2t-locally binary-valued function \( f \), \( r_f(k, t) = 2t \).

Proof. By Corollary 4, \( r_f(k, t) \geq 2t \). On the other hand, we can prove achievability using the following explicit code construction.

TABLE I

| \( f(u) \) | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 4 | 5 | 2 | 1 | 0 | 0 |
| 1 | 4 | 0 | 4 | 3 | 2 | 1 | 0 |
| 2 | 3 | 4 | 0 | 4 | 3 | 2 | 1 |
| 3 | 2 | 3 | 4 | 0 | 4 | 3 | 2 |
| 4 | 1 | 2 | 3 | 4 | 0 | 4 | 3 |
| 5 | 0 | 1 | 2 | 3 | 4 | 0 | 4 |
| 6 | 0 | 0 | 1 | 2 | 3 | 4 | 0 |

Let \( \text{Im} f = \{ f_1, \ldots, f_E \} \) and set w.l.o.g. \( f_i := i \). Let \( u \) be the information word to be encoded and define the following function,
\[
\omega_{2t}(u) = \begin{cases} 
1, & \text{if } f(u) = \max B_f(u, 2t) \\
0, & \text{otherwise.}
\end{cases}
\]

Now, use \( \text{Enc}_1(u) = (u, \omega_{2t}(u))^{2t} \), i.e. the 2t-fold repetition of \( \omega_{2t}(u) \). This gives an FCC for the function \( f \) due to the following.

Assume \( (u, p) = \text{Enc}_1(u) \) has been transmitted and \((u', p')\) has been received. The decoder first computes \( B_f(u', t) \). Notice that \( f(u) \in B_f(u', t) \subseteq B_f(u, 2t) \). If \( |B_f(u', t)| = 1 \), then it contains the correct function value \( f(u) \). Otherwise, \( B_f(u', t) = B_f(u, 2t) \), since \( |B_f(u', t)| > 1 \) and, by the definition of 2t-locally binary-valued functions, \( B_f(u, 2t) \leq 2 \). The decoder performs a majority decision over the 2t + 1 bits \( \omega_{2t}(u') \) (p') and obtains correctly \( \omega_{2t}(u) \), as at most t out of these 2t + 1 bits are erroneous. Finally, the receiver decides for max \( B_f(u', t) \), if \( \omega_{2t}(u) = 1 \) and for min \( B_f(u', t) \), otherwise. \( \square \)

In Section V-C we will present an explicit example for a locally binary-valued function. Other examples are variable-size threshold functions or codeword indicator functions.

B. Hamming Weight Function

Let \( f(u) = \text{wt}(u) \), where \( u \in \mathbb{Z}_2^k \). Note that the number of distinct function values is \( E = k + 1 \). We start by showing that for this function it is possible to achieve optimal redundancy by an encoding function which only depends on the function value, i.e., the Hamming weight of \( u \).

Lemma 8. Let \( f(u) = \text{wt}(u) \). Consider the \((k + 1) \times (k + 1)\) matrix \( D_{\text{wt}}(t) \) with entries \( |D_{\text{wt}}(t)|_{ij} = 0 \) and \( |D_{\text{wt}}(t)|_{ij} = \max\{2t + 1 - |i - j|, 0\} \) for \( i \neq j \). Then,
\[
r_{\text{wt}}(k, t) = N(k + 1, D_{\text{wt}}(t)).
\]

Proof. First, the function values of the Hamming weight function are in \( \{0, 1, \ldots, k\} \) and we define \( f_i = i - 1, i \in [k + 1] \). Therefore, \( d(f_i, f_j) = |i - j| \). It follows from Theorem 2 that \( r_{\text{wt}}(k, t) \leq N(k + 1, D_{\text{wt}}(t)) \). On the other hand, using \( u_i = (1^{i-1}0^{k-i+1}) \), \( i \in \{1, 2, \ldots, k + 1\} \), we directly obtain that \( d(u_i, u_j) = |i - j| \) and applying Corollary 2 gives \( r_{\text{wt}}(k, t) \geq N(k + 1, D_{\text{wt}}(t)) \). \( \square \)

The distance matrix \( D_{\text{wt}}(2) \) for \( k = 6 \) is depicted in Table I. Based on Lemma 8, we can infer a lower bound on the redundancy using the Plotkin-like bound in Lemma 1.

Corollary 7. For any \( k > t \),
\[
r_{\text{wt}}(k, t) \geq \frac{10t^2 + 30t^2 + 20t + 12}{3t^2 + 12t + 12}.
\]

Proof. Let \( \{p_1, \ldots, p_{k+1}\} \) be a \([k + 1, D_{\text{wt}}(t)]\) code. We will prove the corollary by applying the Plotkin-type bound on a subcode of \( p_1, \ldots, p_{k+1} \). Consider the first \( t + 2 \) codewords \( p_1, \ldots, p_{t+2} \). By Lemma 8, we have that \( |D_{\text{wt}}(t)|_{ij} = 2t + 1 - \)}
and thus $|D_{wt}(t)|_{12} + |D_{wt}(t)|_{13} + |D_{wt}(t)|_{23} = 6t - 1$. However, since $d(p_1, p_2) + d(p_1, p_3) + d(p_2, p_3)$ must be an even value, it follows that $d(p_1, p_2) + d(p_1, p_3) + d(p_2, p_3) \geq 6t$. With this strengthened bound, the sum of the pairwise distances in Lemma 1 can be increased by one and we obtain

$$r_{wt}(k, t) \geq \frac{4}{(t + 2)^2} \left( 1 + \sum_{i=1}^{t+2} \sum_{j=i+1}^{t+2} |D_{wt}(t)|_{ij} \right)$$

which inequality (a) follows from Lemma 1, with an additional summand of 1 due to the fact that $d(p_1, p_2) + d(p_1, p_3) + d(p_2, p_3)$ must be even, as explained above. Eq. (b) follows from summing over the diagonals of $D_{wt}(t)$.

For the following results, we require the shifted modulo function, which is defined as follows.

**Definition 10.** We define the shifted modulo operator by

$$a \bmod b \triangleq ((a - 1) \mod b) + 1 \in \{1, 2, \ldots, b\}.$$  

We now describe a way to construct an FCC for $wt(u)$.

**Construction 1.** Set $Enc_{wt}(u) = (u, p_{wt(u)+1})$ with $p_i$ as follows. For $t = 1$, set $p_1 = (000)$, $p_2 = (110)$ and $p_3 = (011)$. Then set $p_i = p_i \bmod 3$ for $i \geq 4$.

For $t = 2$, set $p_1 = (000000)$, $p_2 = (110011)$, $p_3 = (001111)$. Then set $p_i = p_{i-4} \bmod (000011)$ for $i \in \{5, 6, 7, 8\}$ and $p_i = p_{i-8} \bmod 8$ for $i \geq 9$.

For $t \geq 3$, let $p_1, \ldots, p_{2t+1}$ be a code with minimum distance $2t$, i.e., $d(p_i, p_j) \geq 2t$ for all $i, j \leq 2t + 1$, $i \neq j$ and define $Enc_{wt}(u) = (u, p_{wt(u)} \bmod 2t+1)$.

We can use this bound to narrow down the optimal redundancy of FCCs for the Hamming weight function as follows.

**Lemma 9.** For any $k > 2$, $r_{wt}(k, 1) = 3$ and $r_{wt}(k, 2) = 6$.

Further, for $t \geq 5$ and $k > t$,

$$\frac{10t}{3} - \frac{10}{3} \leq r_{wt}(k, t) \leq \frac{4t}{1 - 2\sqrt{\ln(2t)}/(2t)}.$$  

**Proof.** We start with the case $t = 1$. It is quickly verified that $d(p_i, p_j) \geq |D_{wt}(1)|_{ij}$ for all $i \neq j$, $i, j \leq k + 1$ and thus giving a valid FCC. Further, Corollary 7 gives $r_{wt}(k, 1) \geq 3$. For the case $t = 2$, it can be verified that $d(p_i, p_j) \geq |D_{wt}(2)|_{ij}$. Again, Corollary 7 gives $r_{wt}(k, 1) \geq 6$, proving optimality of the proposed code. For $t \geq 3$, we obtain $d(p_i, p_j) \geq |D_{wt}(t)|_{ij}$ as desired. The lower and upper bound on $r_{wt}(k, t)$ follow from Corollary 7 and Lemma 4.

Recall that using a standard error-correcting code with minimum distance $2t+1$, e.g., a BCH code, results in a redundancy of roughly $t \log n$. Therefore, FCCs improve the redundancy scaling by a factor of $\log n$. While we find the optimal redundancy exactly for $t = 1$ and $t = 2$, there is still a gap for $t \geq 3$ narrowing down the optimal redundancy between roughly $10t/3$ and $4t$.

**C. Hamming Weight Distribution Function**

Assume for simplicity that $E$ divides $k + 1$ and define $T \triangleq \frac{k + 1}{E}$. Consider the function $f(u) = \Delta(u) \triangleq \frac{|\text{wt}(u)|}{T}$. This function defines a step threshold function with $E - 1$ steps based on the Hamming weight of $u$. The threshold values, where the function values increase by one, are at integer multiples of $T$. We restrict to the case where $2t + 1 \leq T$ and will give an optimal construction with redundancy $r_{\Delta}(k, t) = 2t$ in this regime. Note that, when $4t + 1 \leq T$, we can show that $\Delta(u)$ is $2t$-locally binary-valued, as two consecutive thresholds have distance at least $4t + 1$. Consequently, $r_{\Delta}(k, t) = 2t$ by Lemma 7. We now turn to the more general case, where $2t + 1 \leq T$. We start by describing the encoding function.

**Construction 2.** Let $p_i = \left(1^{t-1}02t-i+1\right)$ for $i \in \{2t + 1\}$ and $p_i = \left(1^{2t}\right)$ for $i \in \{2t + 2, \ldots, T\}$. Then, we construct $Enc_{\Delta}(u) = (u, p_{wt(u) + 1} \bmod T)$, with the shifted modulo operator from Def. 10.

We show that this encoding function gives an FCC for $\Delta(u)$.

**Lemma 10.** For any $2t + 1 \leq \frac{k + 1}{E} \leq T$, $r_{\Delta}(k, t) = 2t$.

**Proof.** By Corollary 4, $r_{\Delta}(k, t) \geq 2t$. We now argue that Construction 2 is an FCC of redundancy $2t$ by showing that $d(Enc_{\Delta}(u_1), Enc_{\Delta}(u_2)) \geq 2t + 1$ for all $u_1, u_2 \in Z_2^{k}$ with $f(u_1) \neq f(u_2)$. Let $u_1, u_2 \in Z_2^{k}$ with $f(u_1) \neq f(u_2)$. Note that, if $d(u_1, u_2) \geq 2t + 1$, we automatically have $d(Enc_{\Delta}(u_1), Enc_{\Delta}(u_2)) \geq 2t + 1$. Let therefore $d(u_1, u_2) \leq 2t$ and $(m - 1)T \leq wt(u_1) \leq mT$ and $mT < wt(u_2) \leq (m + 1)T$ for some $m \in \mathbb{N}$. By Construction 2, $d(p_{T-t}, p_j) \geq 2t + 1 - (i + j)$ for any $i \geq 0, j \geq 1$ with $i + j \leq 2t + 1$. Thus, $d(Enc_{\Delta}(u_1), Enc_{\Delta}(u_2)) = d(u_1, u_2) + d(p_{wt(u_1)+1} \bmod T, p_{wt(u_2)+1} \bmod T) \geq wt(u_2) - wt(u_1) + 2t + 1 - (wt(u_2) - wt(u_1)) = 2t + 1$.  

**D. Min-Max Functions**

Assume that $k = \ell \ell$ for some $w, \ell \in \mathbb{N}$. In this section, we consider $u$ to be formed of $w$ parts $u^{(1)}, \ldots, u^{(w)}$, each of length $\ell$. The function of interest is the min-max function defined next.

**Definition 11.** The min-max function is defined by

$$ww(u) = (\arg \min_{1 \leq i \leq w} u^{(i)}), \arg \max_{1 \leq i \leq w} u^{(i)}),$$

where $u = (u^{(1)}, \ldots, u^{(w)})$, $u^{(i)} \in \mathbb{Z}^{\ell}_2$ with $k = \ell \ell$ and the ordering $<$ between the $u^{(i)}$’s is primarily lexicographical (the left-most bit is the most significant) and secondarily, if $u^{(i)} = u^{(j)}$, according to ascending indices.

For $w = 2$, the function is a binary-valued function and we have an optimal solution from Corollary 6. For $w \geq 3$, we derive an upper and a lower bound on the redundancy of an FCC designed for the min-max function as follows.

**Lemma 11.** For $w \geq 3$ and $\ell \geq 2$, the optimal redundancy $r_{\text{mm}}(k, t)$ is bounded from below by

$$r_{\text{mm}}(k, t) \geq \frac{4t(w^2 - w - 1) - 3w^2 + 7w - 5}{(w - 1)w}.$$  

**Lemma 12.** For $w \geq 3$ and $\ell \geq 3$, the optimal redundancy $r_{\text{mm}}(k, t)$ is bounded from above by

$$r_{\text{mm}} \leq N(w(w - 1), D_{\text{mm}}(t, f_1, \ldots, f_E)) \leq \min \{ r : \Phi(r) > 0 \},$$

where $\Phi(r) \triangleq 2r - (w^2 - w - 1) \Phi(r, 2t - 2) + (4w - 8)(2t - 1)$.  

The proofs are omitted and can be found in [13].
REFERENCES

[1] K. Mazooji, F. Sala, G. Van den Broeck, and L. Dolecek, “Robust channel coding strategies for machine learning data,” in 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 609–616, 2016.

[2] S. Kabir, F. Sala, G. Van den Broeck, and L. Dolecek, “Coded machine learning: Joint informed replication and learning for linear regression,” in 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1248–1255, IEEE, 2017.

[3] K. Huang, P. H. Siegel, and A. Jiang, “Functional error correction for robust neural networks,” IEEE Journal on Selected Areas in Information Theory, 2020.

[4] H. Garcia-Molina, J. D. Ullman, and J. D. Widom, Database Systems, vol. Pearson. Princeton University Press, 2008.

[5] R. M. Roth, “Fault-tolerant dot-product engines,” IEEE Transactions on Information Theory, vol. 65, no. 4, pp. 2046–2057, 2018.

[6] R. M. Roth, “Analog error-correcting codes,” IEEE Transactions on Information Theory, 2020.

[7] E. Dupraz and L. R. Varshney, “Noisy in-memory recursive computation with memristor crossbars,” in 2020 IEEE International Symposium on Information Theory (ISIT), pp. 804–809, 2020.

[8] H. Jeong and P. Grover, “Energy-adaptive error correcting for dynamic and heterogeneous networks,” Proceedings of the IEEE, vol. 107, no. 4, pp. 765–777, 2019.

[9] N. A. Khooshemehr and M. A. Maddah-Ali, “Fundamental limits of distributed encoding,” in IEEE International Symposium on Information Theory (ISIT), pp. 798–803, IEEE, 2020.

[10] M. Plotkin, “Binary codes with specified minimum distance,” IEEE Trans. Inf. Theory, vol. 6, pp. 445–450, Sept. 1960.

[11] E. N. Gilbert, “A comparison of signalling alphabets,” Bell System Technical Journal, vol. 31, pp. 504–522, May 1952.

[12] R. Varshamov, “Estimate of the number of signals in error correcting codes,” Dokl. Akad. Nauk SSSR, vol. 117, pp. 739–741, 1957.

[13] A. Lenz, R. Bitar, A. Wachter-Zeh, and E. Yaakobi, “Function-correcting codes,” CoRR, vol. abs/2102.03094, 2021.

[14] F. J. MacWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes, vol. 16. North Holland, 1983.

[15] K. J. Horadam, Hadamard Matrices and Their Applications. Princeton University Press, 2007.

[16] H. Lin, S. M. Moser, and P. Chen, “Weak flip codes and their optimality on the binary erasure channel,” IEEE Transactions on Information Theory, vol. 64, no. 7, pp. 5191–5218, 2018.

[17] R. B. Ash, Information Theory. Dover Publications, 1990.