Received 1 February 2023, accepted 8 February 2023, date of publication 10 February 2023, date of current version 16 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3244071

TOPICAL REVIEW

Identification Codes: A Topical Review With Design Guidelines for Practical Systems

CASPAR VON LENGERKE[1], ALEXANDER HEFELE[1], JUAN A. CABRERA[1], OLIVER KOSUT[2], (Senior Member, IEEE), MARTIN REISSLEIN[2], (Fellow, IEEE), AND FRANK H. P. FITZEK[1,3], (Senior Member, IEEE)

[1]Deutsche Telekom Chair of Communication Networks, Technische Universität Dresden, 01062 Dresden, Germany
[2]School of ECEE, Arizona State University, Tempe, AZ 85287, USA
[3]Centre for Tactile Internet with Human-in-the-Loop (CeTI), Technische Universität Dresden, 01062 Dresden, Germany

Corresponding author: Martin Reisslein (reisslein@asu.edu)

This work was supported in part by the Federal Ministry of Education and Research of Germany programs "Souverän. Digital. Vernetzt." and joint project 6G-life, Project Id 16KISK001K; in part by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) under Project 450566247; in part by the Germany's Excellence Strategy–EXC 2050/1 under Project 450566247; and in part by the DFG as part of Germany's Excellence Strategy – EXC 2050/1 – Project 390696704 – Cluster of Excellence "Centre for"; Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden.

ABSTRACT A wide range of information technology applications require the identification of a particular message or label that represents the identity of an object at a distance, e.g., over a wireless channel. Conventionally, the underlying information that represents the identity is transmitted over the channel, following the information-theoretic concept of message transmission. If the purpose of the interaction over the channel is only to verify (match) an identity, then the concept of identification over channels—utilizing the identification codes that have been developed by the information theory community—can provide an exponential efficiency gain over message transmission. This topical review article conducts for the first time a comprehensive detailed evaluation of the existing identification codes for the practically relevant regime of finite parameters. We examine essentially all published identification codes, including codes based on inner constant weight codes that are concatenated with outer linear block codes, such as Reed-Solomon and Reed-Muller codes. Specifically, we conduct a holistic identification code comparison based on the logarithm of the number of representable identities (in shannon), the size (in bit) of the transmitted cue that represents an identity, and the corresponding type II error probability bound for essentially all existing identification codes. Based on the resulting insights, we formulate guidelines for the design of practical (finite-parameter) identification codes. For instance, we find that a linear block code (without concatenation with a sophisticated inner constant-weight code) is sufficient for most practical identification code usages.

INDEX TERMS Beyond-Shannon communication, error probability, false-positive identification, goal-oriented communication, linear block code, identity verification, performance metrics.

I. INTRODUCTION

An identity or a set of identification information is the main message of interest in various distributed information technology and communication systems. For instance, radio-frequency identification (RFID) systems [1], [2], [3], [4], [5] identify physical objects via attached tags or smart labels [6], [7], [8], [9] via principles of wireless communication. Similarly, the connection establishment procedure in cellular wireless systems identifies the individual user equipment (UE) nodes to the corresponding base station via wireless transmissions [10], [11], [12], [13], [14]. Conventionally, the identification information is transmitted via communication channels from a source, e.g., smart label or UE, to a sink, e.g., warehouse controller or base station, based on Shannon's principles of message transmission (communication via channels). In many operational scenarios, the sink may have some prior history (e.g., inventory information) or prediction (e.g., UE mobility prediction) of the source identifiers that are to be expected. In such scenarios, the sink only needs to

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

verify whether the identity of the source matches the identity that is expected by the sink. This distributed identity verification can be conducted through *identification via channels (ID)* [15], [16]. With ID, the number of identifiable identities (messages) scales *double exponentially* with the block length and code rate, compared to the single-exponential scaling with message transmission as introduced by Shannon. Thus, ID is a specific approach within the wider Beyond-Shannon communication paradigm [17], [18], [19], [20], [21].

Similar identity verification tasks arise in distributed information systems that need to maintain consistency (synchronism). For instance, distributed databases [22], [23], [24] need to maintain consistency among the different locations. Also, in multi-access edge computing (MEC), the computing instances that execute a mobile computing application need consistent application state information as the computing instances are moved among MEC servers to remain in close vicinity of the mobile application [25], [26], [27], [28]. Similarly, in digital twin systems [29], [30], [31], [32], [33], [34], [35], the real physical system and the corresponding digital twin (simulation of the real physical system) need to stay synchronized. ID can achieve exponential gains in efficiency compared to conventional Shannon data transmission for such verifications whether distributed systems have consistent state information.

A. CONTRIBUTIONS OF THIS TOPICAL REVIEW

ID has to date mainly been studied by the information theory community; specifically, the focus has been on examining the fundamental information-theoretic limits (mainly in the scaling regime of parameters tending to infinity) and on developing ID codes that achieve these fundamental limits. We note that [36] and [37] have investigated the computational complexity of ID codes, specifically of concatenated PPM-RS2 ID codes (see Section III-A). Also, the Reed-Muller (RM) ID coding study [38] discusses the complexity. Thus, these studies consider some aspects of the finite-parameter regime, because they examine the scaling of the computational complexity. However, they do not conduct a holistic investigation of ID coding in the finite-parameter regime, e.g., they do not investigate which part of the concatenated code has which impact on the overall ID coding performance.

Our topical review conducts the first holistic comparison of essentially all existing ID codes in the *practical regime of finite parameters* for the ID codes and the various system characteristic. Thus, this topical review seeks to provide a comprehensive critical review of the state-of-the-art of ID coding so as to inform the usage of ID coding in practical application contexts. We compare essentially all existing ID codes; specifically, we compare all existing ID codes, except the recently proposed random linear ID code [39], in terms of the number of IDs that they can represent (i.e., the ID size), the amount of traffic that the ID code incurs (i.e., the cue size), and the type II error probability (both in the worst case and on average).

We emphasize that our holistic comparison does not consider complexity in detail; rather, we only briefly note elementary general complexity differences between PPM-RS2 and PPM-RM. Our holistic comparison considers the ID size, cue size, as well as the error probability (specifically, bound and mean of error probability) as ID coding performance metrics. Also, for each concatenated ID code, we identify the code element that governs the ID coding performance. We leave a unified comparison of the computational complexities of the different ID code elements to future research. We facilitate that future research with our finding that the commonly used CWC initialization in ID coding does not improve the ID coding performance. Therefore, investigating and optimizing the computational complexity of CWC initializations appears to be not helpful for advancing ID coding.

B. LITERATURE REVIEW METHODOLOGY

Given the relatively small set of existing published studies on identification codes, we collected all existing published ID coding studies for this topical review. We searched the published literature with the commonly used search tools for engineering and information systems, including Google Scholar and IEEE Xplore, with the search terms ''identification via channels'' as well as ''identification'' and ''information theory'' as well as ''identification'' and ''coding'' as well as ''identification code''. We also examined the references in each of the ID coding studies for additional published articles. We present the complete set of collected published ID coding studies in Sections II and III, except for the study [39].

We exclude the recently proposed random linear ID coding [39] from this topical review, since the random creation of the linear codewords fundamentally differs from the deterministic creation of the codewords of all other existing ID codes. More specifically, ID codes with deterministically created codewords employ a deterministically prescribed codebook, i.e., a set of codewords that is created according to a prescribed deterministic procedure, e.g., for creating codewords for Reed-Solomon or Reed-Muller coding. In contrast, the random linear ID code [39] independently uniformly randomly draws the codewords from a given considered field. Aside from the initial exploratory study in [39], the random linear ID coding concept has not yet been further investigated. We leave the topical review of random linear ID codes for future work.

We emphasize that we conduct a topical review of random ID coding which *randomly* selects a cue from a set of cues (i.e., a set of ID codewords). For the investigated random ID codes, the set of cues is determined *deterministically*; whereas, random linear ID coding [39] generates the set of cues randomly.

C. ARTICLE STRUCTURE

Section II briefly explains the fundamentals of general code constructions, identification theory, and specific ID code constructions. Section III describes the existing (published)

TABLE 1. Overall length n, overall dimension k, and overall minimum distance d of different linear block code types.

Lin. block code	Length n	Dim. k	Min. dist. d
RS (inner)	$q-1$	k_i	$q-k_\mathrm{i}$
RS (outer)	$q^{k_\mathrm{i}}-1$	k_o	$q^{k_\mathrm{i}}-k_\mathrm{o}$
RS2	$(q-1)(q^{k_\mathrm{i}}-1)$	$k_\mathrm{i}k_\mathrm{o}$	$(q-k_\mathrm{i})(q^{k_\mathrm{i}}-k_\mathrm{o})$
RM	q^m	$\binom{r+m}{r}$	$(q-r)q^{m-1}$

ID code types in a modular fashion based on the employed linear block code, and on the manner of constructing a constant-weight ID code based on the linear block code. Section IV conducts an in-depth analysis of the error probability bound and mean error probability. Section V evaluates the performance of ID codes in terms of the achieved ID size and the required cue size. Section VI holistically compares the performance tradeoffs of ID codes as functions of the various coding parameters. Finally, based on all findings, Section VII proposes heuristics for determining ID coding parameters for good performance. Section VIII summarizes this article.

II. FUNDAMENTALS

A. CODE CONSTRUCTIONS

In this subsection, we briefly describe the linear block codes, constant-weight codes, and concatenated codes that are relevant to this article. A Reed-Solomon (RS) code $\mathcal{C}_\mathrm{RS}$ is an $(n, k, d)_q$ linear block code [40], [41], [42]. For given symbol size q and dimension k of the linear block code, selecting the linear block code length $n = q-1$ and the minimum distance $d = n-k+1 = q-k$ fully characterizes the RS code. The result is a linear block code with the characteristics:

$$(q-1,\ k,\ q-k)_q\,. \tag{1}$$

We refer to such RS codes as (q, k) RS codes, with $k < q$.

A (q, m, r) Reed-Muller (RM) code $\mathcal{C}_\mathrm{RM}$ [43], [44] of symbol size q, generation m, and order r is an $(n, k, d)_q$ linear block code with the characteristics:

$$\left(q^m,\ \binom{r+m}{m},\ (q-r)q^{m-1}\right)_q\,. \tag{2}$$

For (q, m, r) RM codes, $0 < m \leq r < q$.

Two linear block codes can be concatenated. This requires the symbol size q_o of the outer code $\mathcal{C}^{(\mathrm{o})}$ to be a power of the symbol size q_i of the inner code $\mathcal{C}^{(\mathrm{i})}$ of dimension k_i, i.e., that $q_\mathrm{o} = q_\mathrm{i}^{k_\mathrm{i}}$. The result is a concatenated linear block code $\mathcal{C}^{(\mathrm{c})} = \mathcal{C}^{(\mathrm{i})} \circ \mathcal{C}^{(\mathrm{o})}$ of symbol size $q_\mathrm{i}^{k_\mathrm{i}}$, length $n_\mathrm{i}n_\mathrm{o}$, dimension $k_\mathrm{i}k_\mathrm{o}$, and distance $d_\mathrm{i}d_\mathrm{o}$. Concatenating an inner (q, k_i) RS code with an outer $(q^{k_\mathrm{i}}, k_\mathrm{o})$ RS code yields a concatenated RS code that we refer to as $(q, k_\mathrm{i}, k_\mathrm{o})$ RS2 code. We give an overview of the linear block code characterization of RS, RM, and RS2 codes in Table 1.

Binary constant-weight codes (CWCs) are codes for which all codewords in the codebook share the same Hamming weight W. In a binary CWC $\mathcal{C}_\mathrm{cw}$, each codeword of block length S consists of W ones and $S-W$ zeros. The dimension N of the codebook determines the number of codewords in the codebook. Finally, the upper bound K for the codeword overlap specifies in how many positions (at most) any two codewords take on identical values. In conclusion, an (S, N, W, K) CWC is characterized by its block length S, its dimension N, its Hamming weight W, and its upper bound K for the codeword overlap.

An inner binary $(S_\mathrm{i}, N_\mathrm{i}, W_\mathrm{i}, K_\mathrm{i})$ CWC $\mathcal{C}_\mathrm{cw}^{(\mathrm{i})}$ can be concatenated with an outer $(n_\mathrm{o}, k_\mathrm{o}, d_\mathrm{o})_{q_\mathrm{o}}$ linear block code $\mathcal{C}^{(\mathrm{o})}$. The result is a new concatenated CWC $\mathcal{C}_\mathrm{cw}^{(\mathrm{c})} = \mathcal{C}_\mathrm{cw}^{(\mathrm{i})} \circ \mathcal{C}^{(\mathrm{o})}$ [45]. For the resulting CWC $\mathcal{C}_\mathrm{cw}^{(\mathrm{c})}$, the upper bound K_c on the codeword overlap is not necessarily tight. The resulting CWC $\mathcal{C}_\mathrm{cw}^{(\mathrm{c})}$ is characterized by [45]:

$$S_\mathrm{c} = S_\mathrm{i}n_\mathrm{o}, \tag{3}$$

$$N_\mathrm{c} = N_\mathrm{i}^{k_\mathrm{o}}, \tag{4}$$

$$W_\mathrm{c} = W_\mathrm{i}n_\mathrm{o}, \tag{5}$$

$$K_\mathrm{c} = W_\mathrm{i}(n_\mathrm{o}-d_\mathrm{o}) + K_\mathrm{i}n_\mathrm{o}. \tag{6}$$

Günlü et al. [46, Lemma 3] tightened the bound on the codeword overlap to

$$K_\mathrm{c} = W_\mathrm{i}(n_\mathrm{o}-d_\mathrm{o}) + K_\mathrm{i}d_\mathrm{o}. \tag{7}$$

Since the resulting concatenated code is a CWC, recursive concatenation with another linear block code is possible. The concatenation of two outer linear block codes $\mathcal{C}^{(\mathrm{i})}$ and $\mathcal{C}^{(\mathrm{o})}$ with an inner CWC $\mathcal{C}_\mathrm{cw}^{(\mathrm{i})}$ results in the concatenated CWC $\mathcal{C}_\mathrm{cw}^{(\mathrm{c})} = \left(\mathcal{C}_\mathrm{cw}^{(\mathrm{i})} \circ \mathcal{C}^{(\mathrm{i})}\right) \circ \mathcal{C}^{(\mathrm{o})}$. Because the inner CWC $\mathcal{C}_\mathrm{cw}^{(\mathrm{i})}$ translates the outer linear block code(s) into a concatenated CWC, we refer to the inner CWC as the CWC initialization.

B. IDENTIFICATION

The fundamental theory of message identification (ID) was mainly introduced in [15], [47], and [48]. Identification is a communication problem, in which the source and the sink each consider a singular message from a finite set of possible messages. To highlight the difference between message identification and message transmission, messages are called ID messages or simply IDs in message identification. Based on an ID codeword that the source transmitted over a channel to the sink, the goal of the sink is to determine whether the source considers the same ID as the sink. The sink does not try to find out *which* message the source tries to convey to the sink, but rather whether the IDs of the source and sink are identical, or not. For example, when a traveler waits in a general waiting room for her number to be called so as to buy a ticket at one of the counters in a train station, the traveler is not really interested in which number is displayed on the screen, but rather whether it is the number printed on the piece of paper that the traveler drew from the queue number dispenser.

The ID problem can be solved using message transmission, however, [15] showed that the ID problem can be solved much more efficiently than the message transmission scheme allows for. While message transmission is agnostic to the goal of the communication, message identification includes knowledge of the goal of the sink, i.e., the goal of

identification, and can thus achieve exponential gains over message transmission.

For a codeword length n, an ID code is defined as [46, Definition 1]:

Definition 1 (Identification Code): An $(n, N, \lambda_1, \lambda_2)$ ID code is a collection of N probability distributions p_i on $\mathcal{X}^n$ and N decoding subsets $\mathcal{D}_i \subset \mathcal{Y}^n$ such that for a channel $W^n_{\text{ch}}(y^n|x^n)$, $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$ the following is true:

$$\sum_{y^n \in \mathcal{D}_i} \sum_{x^n \in \mathcal{X}^n} p_i(x^n) W^n_{\text{ch}}(y^n|x^n) \geq 1 - \lambda_1 \; \forall i \in [1 : N], \quad (8)$$

$$\sum_{y^n \in \mathcal{D}_i} \sum_{x^n \in \mathcal{X}^n} p_j(x^n) W^n_{\text{ch}}(y^n|x^n) \leq \lambda_2 \; \forall i, j \; i \neq j, \quad (9)$$

with

$$\lambda_2 = \frac{\sup_{i \neq j} |\mathcal{D}_i \cap \mathcal{D}_j|}{n}. \quad (10)$$

The upper type I error probability bound λ_1 depends on the level of distortion that the channel $W^n_{\text{ch}}(y^n|x^n)$ imposes on the transmitted codeword x^n. Following the common convention in ID studies [37], [38], [46], we only consider noiseless channels in this article, such that the upper type I error probability bound $\lambda_1 = 0$. Thus, we are only interested in the upper type II error probability bound λ_2, which we refer to as the *error probability bound* λ_2 for brevity for the remainder of this article. Type-II errors are false-positive verdicts at the sink, i.e., based on the received ID codeword y^n, the sink considers the ID j of the source to be identical to the ID i of the sink, even though $i \neq j$.

An ID scheme is able to successfully identify a maximum of $N(n, \lambda_1, \lambda_2)$ different ID messages with the definition of $N(n, \lambda_1, \lambda_2)$ in Definition 1. As opposed to the exponential scaling in the number of messages known from message transmission, for message identification, the number of IDs scales *double*-exponentially in the block length:

$$N(n, \lambda_1, \lambda_2) = \exp\exp nR \iff R = \frac{\log\log N}{n}. \quad (11)$$

The ID rate R is upper bounded by the channel's ID capacity C that is identical to the channel capacity for message transmission (whereby for a noiseless discrete memoryless channel $C = 1$).

C. IDENTIFICATION CODES

Similar to Shannon's channel coding theorem [49], Ahlswede and Dueck [15] only proved the existence of ID codes as defined in Definition 1, but did not provide any explicit construction of such ID codes. Based on CWCs that result from concatenating a CWC initialization with outer linear block codes, Verdú and Wei [50] suggested the first explicit ID code construction.

Let C_{id} be a codebook for an (S, N, W, K) ID CWC $\mathcal{C}_{\text{id}}$. Since the codebook C_{id} is constructed using a binary CWC, the codebook can be interpreted as a matrix of 1s and 0s with N rows and S columns.

$$C_{\text{id}} = \begin{pmatrix} c_{0,0} & \cdots & c_{0,S-1} \\ \vdots & \ddots & \vdots \\ c_{N-1,0} & \cdots & c_{N-1,S-1} \end{pmatrix} \in \mathbb{F}_2^{N \times S}. \quad (12)$$

From an ID coding perspective, every row in this codebook denotes a codeword c_i for an ID $m_i \in \mathcal{W}$ of index $i \in [0 : N-1]$. The ID m_i is encoded to the ID codeword x by randomly choosing a position $x \in [0 : S-1]$ from all symbols $c_{i,x}$ within the codeword c_i for which $c_{i,x} = 1$. We refer to this randomly chosen position x within the constant-weight codeword c_i as a *cue*. We call the set of selectable positions, for which $c_{i,x} = 1$, the ID codeword set, or set of cues, of size W. The source randomly determines a cue x, that is a lossy representation of the ID m_i, from the set of cues, and transmits the cue over the channel to the sink. The sink wants to verify whether the source transmitted a cue based on the ID m_j, i.e., whether $m_j = m_i$. For the verification of the ID at the sink, the sink selects the row associated with its chosen ID m_j, and checks at the position x (as determined by the received cue) whether $c_{j,x} = 1$ or not. If $c_{j,x} \neq 1$, then the sink concludes the IDs to be different. The verification of an ID is considered a binary hypothesis test that the source and sink perform jointly.

The number S of columns in the codebook C_{id} denotes the length of the constant-weight codeword c_i. The cue, that is the ID codeword transmitted over the channel, is a position in the constant-weight codeword where the constant-weight codeword takes on the value 1. Encoding this position requires

$$n_{\text{cue}} = \log S \quad \text{in bit.} \quad (13)$$

Since the cue size n_{cue} determines how much data is transmitted over the channel, the cue size n_{cue} corresponds to n in Definition 1. For the remainder of this article, we understand all logarithms log() as logarithms in base 2.

The number N of rows in the codebook C_{id} in Equation (12) determines the size of the message set $\mathcal{W}$, i.e., the number of IDs that the codebook can represent:

$$|\mathcal{W}| = N \quad \text{with} \quad \mathcal{W} = [0 : N-1]. \quad (14)$$

Since we investigate the use of ID codes instead of message transmission codes, we do not consider the transmission of full IDs over a channel, but only the transmission of a cue, which is a fractional representation of an ID. We determine the number of shannons required to represent a full ID, and call it ID size n_{id}.

$$n_{\text{id}} = \log N \quad \text{in shannon.} \quad (15)$$

We use the unit *shannon* for the ID size n_{id} to emphasize that the ID size does not directly influence the number of bits that are physically transmitted over the channel. Rather, the ID size is a measure of the information content of each ID. Instead of the full ID, only a fractional representation (cue) is transmitted, and we measure the cue size n_{cue} in bit.

Using Equation (15), we can rewrite the ID rate determined in Equation (11) as

$$R = \frac{\log\log N}{n_{\text{cue}}} = \frac{\log n_{\text{id}}}{n_{\text{cue}}}. \tag{16}$$

Next to the ID rate R, the error probability bound λ_2 is the most important theoretical metric to evaluate ID codes. For ID codes based on binary CWCs, the codeword overlap bound K determines the number of positions in which two constant-weight codewords overlap at most, while the Hamming weight W determines in how many positions the symbols of the codeword take on values of 1. Considering that the cue is selected uniformly randomly from all available positions that take on values of 1, the bound for the probability of selecting a cue that is shared by two constant-weight codewords is

$$\lambda_2 = \frac{K}{W}. \tag{17}$$

For optimality proofs, the error exponent E_2 is frequently used in the literature:

$$E_2 = \frac{-\log \lambda_2}{n_{\text{cue}}}. \tag{18}$$

A large error exponent E_2 corresponds to a small error probability bound λ_2 and is thus desirable. Ahlswede and Dueck [15] defined an upper bound for the achievable ID rate R based on the channel capacity C and error exponent E_2 [16, Theorem 13ii]:

$$R + 2E_2 \leq C. \tag{19}$$

Hence, for finite-length ID codes, achieving a high ID rate R limits the lowest achievable error probability bound λ_2 and vice versa. For a comparison of different ID coding schemes based on the ID rate R and the error exponent E_2, we refer to [46, Fig. 2]. In this article, we do not evaluate ID codes based on their error exponent E_2, but based on their error probability bound λ_2, as the error exponent is a function of the error probability bound λ_2, and the error probability bound λ_2 is closer related to practically relevant measures, such as the error probability.

To conclude this section, we briefly cover the optimality of ID codes. Since we consider identification over a noiseless channel, we consider the channel to be an error-free discrete memoryless channel of capacity $C = 1$. An ID code is considered optimal for ID, i.e., ID capacity-achieving, if its ID rate $R \to 1$, and its error probability bound $\lambda_2 \to 0$ for infinite block lengths S. Verdú and Wei [50] proposed the following requirements to validate the capacity-achieving properties of an ID code:

$$\lim_{S\to\infty} \frac{\log W}{\log S} \longrightarrow 1, \tag{20}$$

$$\lim_{S\to\infty} \frac{\log\log N}{\log S} = \lim_{S\to\infty} R \longrightarrow 1, \tag{21}$$

$$\lim_{S\to\infty} \frac{K}{W} = \lim_{S\to\infty} \lambda_2 \longrightarrow 0. \tag{22}$$

From a practical point of view, each of these three equations describes the asymptotic behavior of certain code characteristics. The *weight factor* states that a good and asymptotically optimal ID code should have a high weight W compared to the block length S, cf. Eq. (20). Since only codeword symbols that are equal to one can be chosen as cues, it is desirable not to be limited in the choice of those positions. Equation (16) denotes the *second order rate* and ensures asymptotically achieving the double exponential rate, cf. Eq. (21), which is stated in [15] as the main ID property. Finally, the *overlap fraction* asserts decreasing the error probability to zero, cf. Eq. (22).

III. EXPLICIT ID CODE CONSTRUCTIONS

Before evaluating the error probability bound in Section IV, and the ID size and cue size in Section V, this section briefly reviews the evaluated ID code types. We proceed chronologically: we begin by reviewing the first explicit ID code construction, and then describe a different form of representation of ID codes, that is equivalent to the CWC form of representation. We continue with a short investigation on the extension of RS codes. Finally, we review alternatives for the linear block codes and the CWC initialization used for ID codes.

A. FIRST EXPLICIT ID CODE: PPM-RS2 CODES

1) PPM-RS2 CODING CONCEPT

For the first explicit ID code construction, [50] proposed a three-layer concatenated binary CWC consisting of a pulse position modulation (PPM) initialization concatenated with an inner RS code $\mathcal{C}_{\text{RS}}^{(\text{i})}$ and an outer RS code $\mathcal{C}_{\text{RS}}^{(\text{o})}$. This concatenated code is optimal for ID, i.e., for infinite block lengths, the ID rate R approaches 1 while simultaneously guaranteeing that the error probability bound λ_2 approaches 0. We refer to these codes as PPM-RS2 codes for the remainder of this study; or as RS2 codes if it is clear from the context that the employed CWC initialization is the PPM initialization.

The PPM initialization is a $(q, q, 1, 0)$ CWC and does not have any impact on the code's performance in terms of the metrics considered in this study. The inner and outer RS codes define the ID coding performance. With the definition of an unextended Maximum Distance Separable (MDS) RS code in Equation (1), $\mathcal{C}_{\text{RS}}^{(\text{i})}$ and $\mathcal{C}_{\text{RS}}^{(\text{o})}$ are $(n, k, d)_q$ linear block codes determined by the parameters q, k_{i}, and k_{o}:

$$\mathcal{C}_{\text{RS}}^{(\text{i})} : \ (q-1,\ k_{\text{i}},\ q-k_{\text{i}})_q\,, \tag{23}$$

$$\mathcal{C}_{\text{RS}}^{(\text{o})} : \ \left(q^{k_{\text{i}}}-1,\ k_{\text{o}},\ q^{k_{\text{i}}}-k_{\text{o}}\right)_{q^{k_{\text{i}}}}. \tag{24}$$

We do not extend the RS codes, i.e., $n_{\text{i}} = q_{\text{i}} - 1$ and $n_{\text{o}} = q_{\text{o}} - 1$.

Concatenating a CWC, such as the PPM, with a linear block code, such as RS, creates a new CWC with parameters determined by Eqs. (3)–(7). The resulting PPM-RS CWC can be concatenated again with another linear block code,

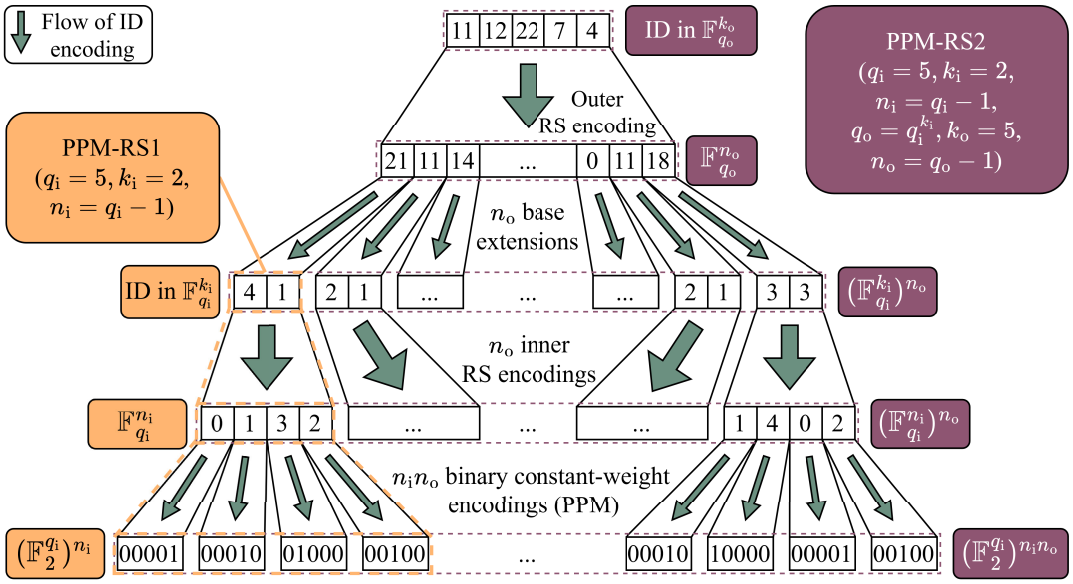

FIGURE 1. ID encoding example using a ($q = 5, k_i = 2, k_o = 5$) PPM-RS2 code that forms a concatenated binary constant-weight code.

yielding the full PPM-RS2 binary CWC with

$$S_{\text{PPM-RS2}} = q(q-1)(q^{k_i}-1), \tag{25}$$

$$N_{\text{PPM-RS2}} = q^{k_i k_o}, \tag{26}$$

$$W_{\text{PPM-RS2}} = (q-1)(q^{k_i}-1), \tag{27}$$

$$K_{\text{PPM-RS2}} = (q-1)(k_o-1) + (q^{k_i}-k_o)(k_i-1). \tag{28}$$

A PPM-RS2 CWC is thus characterized by the symbol size q, as well as the dimensions k_i and k_o. We denote these three defining parameters by writing (q, k_i, k_o) PPM-RS2 code.

2) PPM-RS2 CODING EXAMPLE

Figure 1 visualizes the encoding of an ID with a ($q = 5,\ k_i = 2,\ k_o = 5$) PPM-RS2 code. The choice of symbol size $q = q_i = 5$ and dimension $k_i = 2$ of the inner RS code determines the symbol size $q_o = q^{k_i} = 25$ of the outer RS code. We select one of $N_{\text{PPM-RS2}} = q^{k_i k_o} = 9,765,625$ possible IDs, that is represented in $\mathbb{F}_{25}^5$, namely (11, 12, 22, 7, 4) in the example illustrated in Figure 1. The ID is encoded by the outer RS code, yielding its error-correction codeword in $\mathbb{F}_{25}^{24}$. In order to enable further encoding by the inner RS code, we perform a base extension of each symbol in $\mathbb{F}_{25}$ of the error-correction codeword of the outer RS code. Specifically, each $\mathbb{F}_{25}$ symbol is cast to $\mathbb{F}_5^2$. For example, the symbol 21 is extended to (4, 1). Since the outer RS codeword consists of $n_o = 24$ symbols, the base extension is performed 24 times in parallel, once for each symbol. The base extension results in 24 codewords in $\mathbb{F}_5^2$.

Each of the 24 codewords can be interpreted as an ID in $\mathbb{F}_5^2$ that is to be encoded in PPM-RS1, as indicated on the left side of Figure 1. For the PPM-RS2 encoding illustrated on the right in Figure 1, each codeword in $\mathbb{F}_5^2$ is separately further encoded by the inner RS code, yielding an error-correction codeword in $\mathbb{F}_5^4$. Whereby, the resulting 24 inner RS codewords can be interpreted as a single codeword of the PPM-RS2 encoding in $(\mathbb{F}_5^4)^{24}$. This inner RS encoding step of the PPM-RS2 encoding corresponds to the first PPM-RS1 encoding step of 24 independent IDs in $\mathbb{F}_5^2$, which results 24 RS codewords in $\mathbb{F}_5^4$. Finally, each symbol in $\mathbb{F}_5$ is translated into its PPM representation, resulting in $W_{\text{PPM-RS2}} = n_i \cdot n_o = 4 \cdot 24 = 96$ binary codewords. One symbol of every PPM codeword is set to 1, while the other four symbols are set to 0. The overall result is a binary constant-weight codeword in $(\mathbb{F}_2^5)^{96}$.

The ID encoding is completed by selecting a random position (cue) within the binary constant-weight codeword, at which the corresponding symbol equals 1. In the example, counting (indexing) the bit positions from the right side (starting with 0) to the left side, possible cues include 2, 5, 14, and 16. Transmitting the cue to a receiver requires only $\log_2(S_{\text{PPM-RS2}}) = \log_2(q_i \cdot n_i \cdot n_o) = \log_2(5 \cdot 4 \cdot 24) \approx 8.9$ bit, while transmitting the full ID requires $\log_2(N_{\text{PPM-RS2}}) = \log_2(q_o^{k_o}) = \log_2(25^5) \approx 23.2$ bit. For larger coding parameters, the relative traffic reduction grows more pronounced.

3) PPM-RS2 CODING CONVENTIONS FOR THIS TOPICAL REVIEW

The CWC characteristics (S, N, W, K) of PPM RS2 codes stated in Eqs. (25)–(28) differ from the parameters reported in the study [50, Proposition 2] on PPM-RS2 codes in two regards. First, we do not investigate RS codes of length $n > q - 1$, as we explain in Section III-C. This changes

several terms from q to $q-1$ and from q^{k_i} to $q^{k_\mathrm{i}}-1$, respectively. Second, we do not use the estimation for the overlap bound K provided at the end of the proof for [50, Proposition 2], but the newer result from [46, Lemma 3], which yields a tighter bound K based on Eq. (28).

In order to achieve the optimality for ID, the choice of q, k_i, and k_o is limited to prescribed value ranges when constructing PPM-RS2 codes. We need to ensure a positive code distance $d = n - k + 1$ for RS codes via $k_\mathrm{i} < q$ and $k_\mathrm{o} < q^{k_\mathrm{i}}$. For PPM-RS2 codes to be optimal for ID, the following condition is required to fulfill the optimality requirements in Section II-C [50, Proposition 3]:

$$k_\mathrm{o} = q^t \quad \text{with} \quad t < k_\mathrm{i},\ t \in \mathbb{N}. \tag{29}$$

To allow for arbitrary dimensions k_o of the outer RS code, [46] provided an alternative set of optimality conditions for ID CWCs using concatenated RS codes in [46, Eqs. (19)–(22)], replacing the condition for k_o in Equation (29) by

$$\log k_\mathrm{o} \to \infty, \quad \frac{\log k_\mathrm{o}}{k_\mathrm{i}} \to 0, \quad \frac{k_\mathrm{i}}{q} \to 0, \quad \frac{k_\mathrm{o}}{q^{k_\mathrm{i}}} \to 0. \tag{30}$$

These conditions state the desirable limits of the corresponding expressions when using large coding parameters, also compare Eqs. (20)–(22).

B. REDUNDANCY OF PPM INITIALIZATION

The binary CWC initialization maps the output of the outer linear block code (or concatenated linear block codes) onto a vector of 0s and 1s. An ID is encoded by the linear block code into a length n codeword of symbol size q. If the CWC initialization is the PPM initialization, the PPM initialization maps this codeword to a binary vector of length nq. This is achieved by creating one-hot encodings of every value $\in \mathbb{F}_q$ in each of the n positions of the error correction codeword. In other words, a position in the binary vector is set to 1 if the position corresponds to a (symbol, position) tuple that is part of the error correction codeword.

For example, for a linear block code of symbol size $q = 5$, there exist $q = 5$ different values that a symbol can take on: 0, 1, 2, 3, 4. The PPM maps the block codeword symbol 0 to its binary one-hot encoded representation by setting the first element of the one-hot encoding vector of length $q = 5$ to 1, yielding the vector (00001), cf. lower left corner of Figure 1.

The PPM initialization does not add additional distance between the codewords. Rather, the PPM initialization is a bijective mapping of the block codeword to a binary representation thereof. The PPM initialization does not change the properties of the ID code; rather, the PPM initialization only translates the ID code into a CWC. When using the PPM initialization for the CWC initialization, it suffices to investigate the outer linear block code(s) to determine the ID coding performance.

The value of the symbol in each position of the linear block codewords has been called a *tag* in [36], [37], and [38]. Instead of discussing CWCs as ID codes, the studies [36], [37], [38] only consider the linear block codes and refer to the code construction as *tagging code*. A tagging code includes a high-distance linear block code without concatenating the block code with a binary CWC for the cue deduction. Thus, for the ID procedure with the tagging code representation, a (symbol, position) tuple [also called (tag, position) tuple] is transmitted rather than a cue. Thus, every cue of the CWC representation of ID coding corresponds to exactly one (tag, position) tuple of the tagging code representation of ID coding, and vice versa. That is, a cue and the corresponding (tag, position) tuple are two different representations of the same randomly chosen fractional information of a full block codeword. In this article, we review all ID codes in their CWC (cue) form of representation.

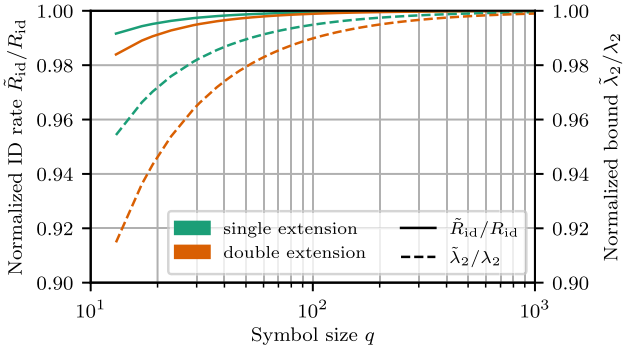

FIGURE 2. ID rate $\bar{R}_\mathrm{id}$ and error probability bound $\bar{\lambda}_2$ for single and double extended $(q, k_\mathrm{i}, k_\mathrm{o}) = (q, 2, q)$ PPM-RS2 codes, normalized by the respective rate R and bound λ_2 of the unextended $(q, 2, q)$ PPM-RS2 code. Extending the RS code decreases the error probability bound for low symbol sizes q, which is desirable, while moderately decreasing the ID rate, which is undesirable.

C. EXTENSION OF REED-SOLOMON

Some studies extend the utilized RS codes to improve the theoretical performance of the ID codes [46], [50]. The length n of an RS code is limited by the symbol size q of the code. Traditionally, the highest value that n can take on is $q-1$. By extending the RS code, n can take on values of q and even $q+1$. An increased length n can increase the distance d of an RS code and is therefore beneficial for generating high-distance codewords: the higher the distance between the codewords, the smaller the type-II error probability. The study [50] uses single-extended RS codes, while [46] proposes to use double-extended RS codes.

In practice, extending the RS code is only possible if the full RS codeword is determined during the encoding process, since the extension is a function of the full unextended RS codeword. However, in ID, determining the full RS codeword is computationally infeasible for all but very small symbol sizes [36]. To limit computational complexity, it suffices to compute only a part of the RS codeword [36]. Computing the RS codeword only partially increases the feasibility of RS codes in practical ID scenarios but disables the possibility of extending the RS code, because the unextended RS codeword is not determined in full anymore.

We observe from Figure 2 that the increased length of the extended code slightly decreases the ID rate, which

corresponds to a decrease in the ID size compared to the cue size, cf. Equation (16). The error probability bound is significantly smaller when using extended RS codes, especially for lower symbol sizes, since the relative symbol size gain is more pronounced for smaller symbols than for larger symbols. The impact of the field extension diminishes at larger symbol sizes q.

Due to the significantly increased computational complexity for small gains at large symbol sizes q, we limit our investigation to unextended RS codes of length $n = q - 1$ for the remainder of this article, as noted in Section III-A3. Nevertheless, in some use-cases it may be beneficial to use very small symbol sizes and to compute the full RS codeword; albeit at a relatively high computational cost, so as to enable the extension of the RS code. The small symbol sizes limit the computational complexity of determining the full RS codewords; thus, the tradeoff may be desirable to decrease the error probability bound by extending the RS codewords.

D. ALTERNATIVE LINEAR BLOCK CODES

Several ID code types that are optimal for ID in terms of their asymptotic behavior in ID rate and error probability bound use a concatenation with RS codes, such as the PPM-RS2 code that we reviewed in Section III-A. Because even the optimized computation for PPM-RS2 codes is expensive [36], it may be beneficial to use suboptimal block codes, instead. Thus, in this subsection, we review the concatenation of the CWC initialization with a singular RS code, and the ID code construction using Reed-Muller (RM) codes. As noted in Section I-B, we do not discuss the recent idea of random linear codes by Sidorenko and Deppe that is also capacity-achieving [39].

We give an overview of the linear block codes and their characteristics in Table 1. Specifically, the characteristics are the overall length n of the block code, the overall dimension k, and the overall minimum distance d. Note that we state the characteristics of ID codes consisting of two concatenated RS codes, i.e., the characteristics of RS2 codes, which are the tagging code representation of PPM-RS2 CWCs. We explained the equivalence of the CWC and tagging code representations in Section III-B. In the tagging code representation, the error probability bound and ID rate are determined from the characteristics of the concatenated linear block code. Note that in the CWC representation, the CWC initialization is *not* concatenated with the concatenated RS2 code in a single step. Instead, first, the CWC initialization is concatenated with the inner RS code, and then the resulting CWC is concatenated with the outer RS code, as illustrated in Figure 1. For RS2, the overall length n of the code, the overall dimension k, and the overall minimum distance d are each the product of the respective length n, dimension k, and minimum distance d of the inner and outer RS code, cf. Section II-A.

1) SINGLE REED-SOLOMON (RS1)

The PPM-RS2 codes in [50] implicitly include the ID codes that consist of the concatenation of the PPM initialization with a single RS code, i.e., only with RS (inner). Instead of concatenating a secondary outer RS code to the resulting inner CWC as described in Section III-A, the inner CWC [i.e., CWC initialization concatenated with RS (inner)] is taken as the final ID code. Explicitly, a CWC based on a single RS code is the concatenation of the PPM initialization that is a $(q, q, 1, 0)$ CWC with the RS code as parameterized in Eq. (23). The result is a (q, k) PPM-RS1 ID code that forms a $(q(q-1), q^k, q-1, k-1)$ CWC. This ID code has been applied to the use-case of watermarking in [51], and is not capacity-achieving. We do not further investigate RS1 in this study.

2) REED-MULLER (RM)

As an alternative to using RS codes, [38] proposes using an RM code as the linear block code for constructing an ID code. Based on RM codes being a linear block code similar to RS codes, a concatenation of multiple RM codes would be possible. Unlike the concatenation of RS codes, concatenating RM codes does not yield a better asymptotic behavior [38]. Thus, only a single RM code is used for constructing the RM ID code. While in [38], the RM ID code was proposed in the tagging code presentation form as opposed to the CWC form, we concatenate a PPM initialization with the RM codes, thus creating a (q, m, r) PPM-RM CWC. This way, the RM code is comparable to the other ID CWCs based on their respective CWC characteristics, cf. Table 2. The overall block-code characteristics of RM codes are shown in Table 1 along the other block codes.

3) HOW TO MAKE RS2 AND RM COMPARABLE

Since the parameters chosen for the linear block code have a significant impact on the properties of the resulting ID CWC, attention has to be paid to the parameter selection when comparing linear block codes. This is especially important for comparing the error probability, since an equal length n and overall dimension k of the used linear block code assures equality in cue size and ID size. Since both types of linear block codes (types of code constructions), i.e., both RS and RM codes, use different code parameters, it is necessary to determine relations that assert an equal length n and overall dimension k for the resulting block code, given an identical symbol size q. This is necessary since the two code types do not share any common code parameters apart from the symbol size q. Therefore, the overall length n and the overall dimension k of the linear block code are common parameters that allow comparing the two code types.

By equating the corresponding expressions shown in Table 1, given identical symbol size q, the lengths n of the RM code and the RS2 code are equal if

$$k_\mathrm{i} = \log_q\left(\frac{q^m}{q-1} + 1\right) \overset{q\gg 1}{\approx} m - 1. \tag{31}$$

Additionally, by equating the corresponding expressions shown in Table 1 and given identical symbol size q, the overall dimensions k of the RM code and the RS2 code are equal if

$$k_\mathrm{o} = \frac{1}{k_\mathrm{i}}\binom{r+m}{m} \overset{\text{Eq. (31)}}{=} \frac{1}{m-1}\binom{r+m}{m}. \tag{32}$$

TABLE 2. Characteristics of ID CWCs: block length *S*, dimension *N*, Hamming weight *W*, and upper bound for the codeword overlap *K*. The numerical investigations in this article focus on the ID CWC types in the bold font.

ID CWC type	Parameters	S	N	W	K-bound
PPM	(q)	q	q	1	0
OOC1	(q)	$q(q-1)$	q	$q-1$	0
PPM-RS1 [50], [51]	(q,k)	$q(q-1)$	q^k	$q-1$	$k-1$
OOC1-RS1 [46]	(q,k)	$q(q-1)^2$	q^k	$(q-1)^2$	$(q-1)(k-1)$
HF [52]	(q,k)	q^2	q^{k+1}	q	k
PPM-RS2 [50]	$(q,k_\mathrm{i},k_\mathrm{o})$	$q(q-1)(q^{k_\mathrm{i}}-1)$	$q^{k_\mathrm{i}k_\mathrm{o}}$	$(q-1)(q^{k_\mathrm{i}}-1)$	$(q-1)(k_\mathrm{o}-1)+(q^{k_\mathrm{i}}-k_\mathrm{o})(k_\mathrm{i}-1)$
OOC1-RS2 [46]	$(q,k_\mathrm{i},k_\mathrm{o})$	$q(q-1)^2(q^{k_\mathrm{i}}-1)$	$q^{k_\mathrm{i}k_\mathrm{o}}$	$(q-1)^2(q^{k_\mathrm{i}}-1)$	$(q-1)^2(k_\mathrm{o}-1)+(q-1)(q^{k_\mathrm{i}}-k_\mathrm{o})(k_\mathrm{i}-1)$
HF-RS [52]	$(q,k_\mathrm{i},k_\mathrm{o})$	$q^2(q^{k_\mathrm{i}}-1)$	$q^{k_\mathrm{i}k_\mathrm{o}+1}$	$q(q^{k_\mathrm{i}}-1)$	$q(k_\mathrm{o}-1)+(q^{k_\mathrm{i}}-1)k_\mathrm{i}$
PPM-RM [38]	(q,m,r)	q^{m+1}	$q^{\binom{m+r}{m}}$	q^m	$q^m r/q$

Using Equation (31) and (32), we can determine the RS2 parameters to match the length n and overall dimension k of an RM code with arbitrary parameters. Due to the binomial relation in Equation (32), the inverse for determining r based on k_o does not necessarily result in an integer for r. Thus, when calculating the order r of the RM code, we round r to the nearest integer. Based on the relations in Equation (31) and (32), we provide a parameter translation table for equal code length n and overall dimension k in Table 5 in the Appendix. Table 5 provides a corresponding order r of an RM code for a given m and an RS2 code with its parameters k_i and $k_\mathrm{o} = q$.

E. ALTERNATIVE CWC INITIALIZATIONS

Next to extending the RS codes as described in Section III-C, and using different outer linear block codes as described in Section III-D, another possibility to alter the performance of an ID code lies in using a different CWC initialization. We describe two CWC initializations that were proposed in [52] and [46] in concatenation with outer RS codes.

Figure 3 provides a visual overview of the combinations of CWC initializations with linear block codes that we investigate in this study, whereby the dashed arrows indicate the implications of the parameter settings of the CWC codes on the left-hand side of Figure 3 for the parameter settings of the linear block codes on the right-hand side. In particular, for the CWC initialization, the symbol size q has to be selected; and this CWC initialization symbol size q is also the symbol size of the overall concatenated ID code. This CWC initialization symbol size q implies the specific symbol size $q_\mathrm{i} = q$ of the first concatenated linear block code, i.e., either of the inner RS code, or of the RM code. Additionally, the CWC initialization symbol size q and the dimension k_i of the inner RS code imply the specific symbol size $q_\mathrm{o} = q^{k_\mathrm{i}}$ of the outer RS code.

1) OPTICAL ORTHOGONAL CODES (OOCs)

One method of initializing the CWC differently from the PPM initialization involves modified prime sequences based on optical orthogonal codes (OOCs) [46]. More specifically, the study [46] proposes two OOCs that are optimal for ID in terms of the asymptotic behavior of ID rate and error probability bound. We include the first proposed construction in the comparisons in this study. We refer to CWCs resulting from modified prime sequences based on the first OOC construction proposed in [46] as OOC1 for the remainder of this article.

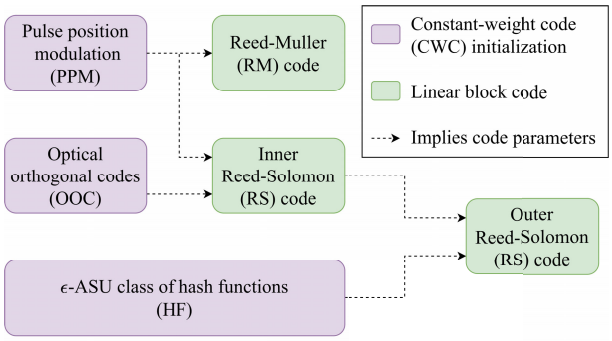

FIGURE 3. Overview of concatenation options of published concatenated constant-weight ID codes. The CWC initialization symbol size *q* has specific implications for the feasible parameter settings of the linear block codes; furthermore, the outer RS code symbol size depends on the CWC initialization and inner RS coding parameters.

Similar to the PPM initialization for PPM-RS2 codes, the OOC initialization is concatenated: first with an inner RS code, and then with an outer RS code, yielding an OOC1-RS2 code, cf. Figure 3. The CWC parameters of the OOC initialization, and of the CWCs resulting from single and double concatenation with RS codes are given in Table 2. The OOC1 initialization does not increase the distance between the obtained codewords. Thus, similar to the PPM initialization, the OOC initialization is a bijective relation that maps a (tag, position) tuple to a cue, as explained for the PPM initialization in Section III-B.

2) ϵ–Almost STRONGLY UNIVERSAL CLASS OF HASH FUNCTIONS (HFs)

The study [52] suggests an alternative CWC initialization based on ϵ–*almost strongly universal* hash functions. For brevity, we refer to a CWC based on ϵ–almost strongly universal hash functions as HF.

The HF initialization forms a larger CWC than the OOC and PPM initializations, cf. Table 2 and Figure 3. Similar to the other CWC initializations, the HF initialization can be concatenated with an outer RS code, yielding an HF-RS code that is comparable in its CWC parameters to OOC1-RS2 codes and PPM-RS2 codes. Table 2 compares the parameters of the HF CWC ID code (without and with concatenation

of HF with an outer RS code) to the parameters of the other CWC ID codes. Note that the dimensions N of the HF initialization and the HF-RS code exceed the dimensions N of comparable ID CWC types, by a factor of q. The outer RS code of the HF-RS code has $q^{k_\mathrm{i}k_\mathrm{o}}$ codewords. For each codeword of the outer RS code in an HF-RS code, you can additionally choose a y value from the field of size q (this leads to the extra factor of q). The ID codeword itself will then be the incidence vector for when a hash function applied to the RS codeword equals this y. Thus for each y, there is a different ID codeword, supporting a total of $q^{k_\mathrm{i}k_\mathrm{o}+1}$ ID codewords.

Therefore, the concatenation of an HF initialization with an outer RS code follows different concatenation rules than characterized in Eqs. (3)–(7). In particular, the concatenated overall HF-RS CWC is characterized by the following:

$$S_{\text{HF-RS}} = q^2 n_\mathrm{o}, \tag{33}$$
$$N_{\text{HF-RS}} = q q_\mathrm{o}^{k_\mathrm{o}} = q q^{k_\mathrm{i}k_\mathrm{o}}, \tag{34}$$
$$W_{\text{HF-RS}} = q n_\mathrm{o}, \tag{35}$$
$$K_{\text{HF-RS}} = q(n_\mathrm{o} - d_\mathrm{o}) + k_\mathrm{i} n_\mathrm{o}. \tag{36}$$

Similar to OOCs and PPMs, HFs do not introduce additional distance and thus also represent a bijective relation for mapping a (tag, position) tuple to a cue, as explained for PPMs in Section III-B.

IV. EVALUATION OF ERROR PROBABILITY

Next, we compare the ID codes based on the error probability bound λ_2, which is the first metric we investigate. The error probability bound λ_2 is a more practical metric than the CWC characteristics (S, N, W, K) investigated in Section III since the error probability has a direct measurable impact on real systems.

First, we define the mean error probability $\overline{p_\mathrm{err}}$, since $\overline{p_\mathrm{err}}$ offers a point of reference for the error probability bound λ_2. While the error probability bound λ_2 characterizes the worst-case behavior, the mean error probability $\overline{p_\mathrm{err}}$ characterizes the average behavior of the error probability in an ID system [53]. The mean error probability $\overline{p_\mathrm{err}}$ can be determined by finding the average overlap between different codewords; in contrast, the maximum overlap of different codewords defines the error probability bound λ_2.

For uniformly distributed symbols, each codeword symbol of symbol size q has the probability of q^{-1} to take on any of the q possible values [37]. The probability of two codewords coinciding in a single symbol position is a form of the birthday problem with q possible values and two participants. Thus, the average probability for a randomly selected single codeword symbol to coincide with another randomly selected codeword symbol is the mean error probability

$$\overline{p_\mathrm{err}}(q) = q^{-1}. \tag{37}$$

This holds for all ID codes. The error probability bound λ_2 being the maximum error probability typically exceeds the mean error probability $\overline{p_\mathrm{err}}$.

In the following, we determine the impact of the choice of linear block code type, of the block codes' parameters, and of the CWC initialization onto the error probability bound λ_2.

A. LINEAR BLOCK CODES

The linear block code is the outer code in an ID CWC construction, and its overall distance d determines the overlap K of the resulting ID CWC. As described in Section III-B, we can use the PPM initialization to investigate the block codes in a standalone fashion, since the PPM initialization does not contribute to the properties of the ID code next to translating the ID code into a CWC. We investigate the impact of RS2 and RM codes on the error probability bound λ_2 in this subsection.

1) REED-SOLOMON CODES

The error probability bound of RS2 can be determined using $\lambda_2 = K/W$ from Equation (17) and the CWC parameters of PPM-RS2 codes from Table 2, yielding

$$\lambda_2^{\mathrm{RS2}}(q, k_\mathrm{i}, k_\mathrm{o}) = \underbrace{\frac{k_\mathrm{o} - 1}{q^{k_\mathrm{i}} - 1}}_{:=\alpha} + \underbrace{\frac{(k_\mathrm{i} - 1)(q^{k_\mathrm{i}} - k_\mathrm{o})}{(q-1)(q^{k_\mathrm{i}} - 1)}}_{:=\beta^{\mathrm{PPM}}}. \tag{38}$$

This result differs from the one reported in [50] since we consider slightly different CWC parameters, as explained in Section III-A. In particular, we differ in two regards from [50]: we do not extend the RS code, and we use the K-estimate in Equation (7) as proposed in [46]. Accordingly, for given block code parameters $(n, k, d)_q$, we obtain a slightly different set of CWC characteristics (S, N, W, K), see Eqs. (25)–(28), than [50].

As apparent from Equation (38), the error probability bound λ_2^{RS2} is influenced by both dimensions k_i and k_o of the RS2 code and by the symbol size q. Thus, we can consider the bound to be a function $\lambda_2^{\mathrm{RS2}}(q, k_\mathrm{i}, k_\mathrm{o})$. In general, smaller dimensions k_i and k_o of the RS2 code imply a lower λ_2^{RS2}, which is desirable. Additionally, however, the symbol sizes of the inner and the outer RS code are dependent on each other by $q_\mathrm{o} = q_\mathrm{i}^{k_\mathrm{i}}$, which is a condition for the concatenation, cf. Section II-A. This warrants further investigation of the judicious selection of the inner and the outer dimensions k_i and k_o of the RS codes to achieve a low error probability bound λ_2.

For a fixed symbol size q, Figure 4 shows the influence of k_i and k_o on λ_2^{RS2} for PPM-RS2 codes. Since λ_2 is the bound on an error probability, λ_2 can not exceed 0.5. For low k_o, PPM-RS2 codes exhibit a saturation phase that covers an increasing range of dimensions k_o for increasing k_i. This is caused by the first term of λ_2 approaching $\alpha \approx 0$ for dimensions $k_\mathrm{o} \ll q^{k_\mathrm{i}}$. In this $k_\mathrm{o} \ll q^{k_\mathrm{i}}$ regime, the outer RS code can achieve a high distance $d_\mathrm{o} = n_\mathrm{o} - k_\mathrm{o} + 1 = q^{k_\mathrm{i}} - k_\mathrm{o} + 1$, and thus provide a high distance code, which is desirable for ID codes.

The second term β^{PPM} is the dominant summand and almost solely determines the bound based on the values of k_i and q. Note that in β^{PPM}, the factor $q^{k_\mathrm{i}} - k_\mathrm{o} \approx q^{k_\mathrm{i}}$ for

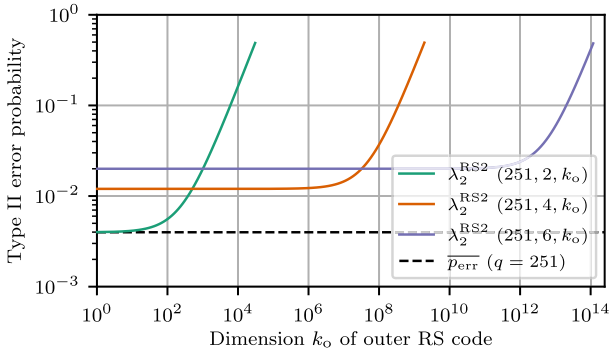

FIGURE 4. Error probability bound λ_2^{RS2} of $(q = 251, k_i, k_o)$ PPM-RS2 codes, for several dimensions k_i of the inner RS code, as a function of dimension k_o of the outer RS code. For reference, we also plot the mean error probability $\overline{p_{err}} = 1/q$. For a fixed dimension k_i, increasing dimension k_o increases λ_2^{RS2}, which is undesirable. For higher dimensions k_i, λ_2^{RS2} remains in a saturation regime for increasing dimension k_o, and only significantly increases once dimension k_o exceeds a certain threshold.

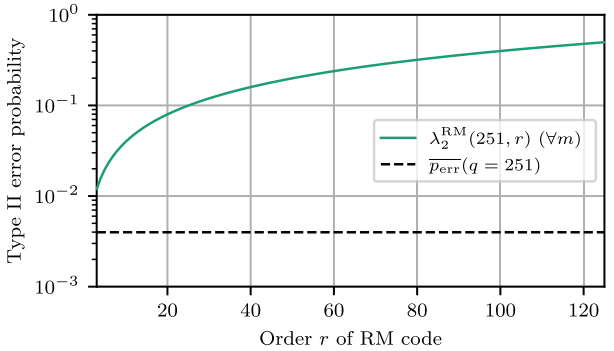

FIGURE 5. Error probability bound λ_2^{RM} of $(q = 251, m, r)$ PPM-RM, as a function of order r. For reference, we also plot the mean error probability $\overline{p_{err}} = 1/q$. Increasing order r increases λ_2^{RM}, which is undesirable.

dimensions $k_o \ll q^{k_i}$; thus, β^{PPM} is almost independent of k_o for dimensions $k_o \ll q^{k_i}$. In summary, with $\overline{p_{err}} = q^{-1}$, the error probability bound for small dimensions k_o of the outer RS code is

$$\lambda_2^{RS2} \approx \frac{k_i - 1}{q} = (k_i - 1)\overline{p_{err}} \quad \text{for} \quad k_o \ll q^{k_i}. \tag{39}$$

Therefore, the smallest achievable λ_2^{RS2} over all k_o increases for increasing dimensions k_i of the inner RS code. The range of the saturation regime is determined by the inner dimension k_i. Therefore, larger k_i allow a wider saturation range for k_o. The asymptotic error probability bound, however, increases for higher k_i.

For higher outer dimensions k_o, the α term increases and gains relevance, with λ_2^{RS2} asymptotically approaching 0.5, which makes the code with large k_o infeasible as an ID code. PPM-RS2 codes of different inner dimensions k_i share a similar increasing characteristic when reaching the maximum possible outer dimension of the RS code.

We investigate at which dimension k_o of the outer RS code the error probability bound λ_2 of a PPM-RS2 code $\mathcal{C}(q, k_i, k_o)$ intersects the error probability bound $\tilde{\lambda}_2$ of another PPM-RS2 code $\tilde{\mathcal{C}}(\tilde{q}, \tilde{k}_i, \tilde{k}_o)$ of higher dimension $\tilde{k}_i = k_i + 1$ of the inner RS code. Both codes share the same symbol size $\tilde{q} = q$ and the same dimension $\tilde{k}_o = k_o$ of the outer RS code. For a visual reference, in Figure 4, the error probability bounds of different PPM-RS2 codes intersect. When both codes $\mathcal{C}$ and $\tilde{\mathcal{C}}$ share the same symbol size q and dimension k_o, the error probability bound $\tilde{\lambda}_2$ of the code $\tilde{\mathcal{C}}$ is in its saturation regime, because $\tilde{k}_o \ll q^{\tilde{k}_i} = q^{k_i+1}$. Therefore, with Equation (39), $\tilde{\lambda}_2 \approx (\tilde{k}_i - 1)/q = k_i/q$. Given identical symbol size $q = \tilde{q}$, the error probability bound λ_2 of the code $\mathcal{C}$ intersects the error probability bound $\tilde{\lambda}_2$ of code $\tilde{\mathcal{C}}$ at dimension $k_o = \tilde{k}_o = q^{k_i-1}$, i.e.,

$$\lambda_2(k_i) = \tilde{\lambda}_2(\tilde{k}_i = k_i + 1) \text{ at } k_o = \tilde{k}_o = q^{k_i-1}. \tag{40}$$

Proof:

$$\begin{aligned}
\lambda_2(k_o = q^{k_i-1}) &\overset{\text{Eq. (38)}}{=} \frac{q^{k_i-1} - 1}{q^{k_i} - 1} + \frac{(k_i - 1)(q^{k_i} - q^{k_i-1})}{(q-1)(q^{k_i} - 1)} \\
&\overset{q^{k_i-1} \gg 1}{\approx} \frac{1}{q} + \frac{k_i - 1}{q-1}\frac{q-1}{q} = \frac{1}{q} + \frac{k_i - 1}{q} \\
&= \frac{k_i}{q} = \tilde{\lambda}_2(\tilde{k}_i = k_i + 1). \qquad \square
\end{aligned}$$

Above the threshold, i.e., for dimensions larger than $k_o = \tilde{k}_o = q^{k_i-1}$, the error probability bound λ_2 of code $\mathcal{C}$ exceeds the error probability bound $\tilde{\lambda}_2$ of code $\tilde{\mathcal{C}}$.

In conclusion, the dimension k_i of the inner RS code should be chosen to be as small as possible to achieve a small error probability bound λ_2. The dimension k_o of the outer RS code of PPM-RS2 codes should be in the saturation phase of k_o since the error probability bound is fairly constant. The inner dimension k_i and symbol size q determine the error probability bound λ_2^{RS2} for this regime. For a given maximum required dimension k_o of the outer RS code, the dimension k_i of the inner RS code should be chosen accordingly. Figure 4 illustrates that PPM-RS2 codes with $k_i = 2$ and $k_o \ll q^{k_i}$ are suitable to limit the error probability bound λ_2.

2) REED-MULLER CODES

The error probability bound λ_2 of RM codes can be determined using the CWC parameters of PPM-RM from Table 2, yielding

$$\lambda_2^{RM}(q, r) = \frac{K}{W} = \frac{r}{q} = r\overline{p_{err}}. \tag{41}$$

For an RM code-based construction, the error probability bound λ_2^{RM} depends on the order r of the RM code and the symbol size q. Thus, the order r of the RM code yields a multiple of the mean error probability $\overline{p_{err}}$ as the error probability bound λ_2^{RM}. The generation m does not influence λ_2^{RM}.

Figure 5 illustrates the dependency of the error probability bound λ_2^{RM} and the mean error probability $\overline{p_{err}}$ on the order r of the RM code. For a low error probability bound, the objective should be to achieve the desired dimension of the linear block code with the lowest possible order r.

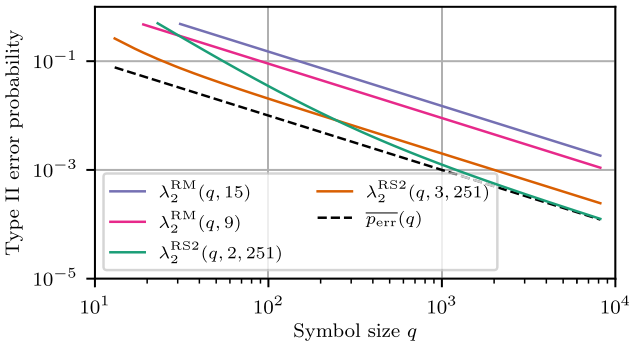

FIGURE 6. Error probability bound λ_2 for two $(q, k_\mathrm{i}, k_\mathrm{o} = 251)$ PPM-RS2 codes and $(q, m = 3, r = 9)$ and $(q, m = 4, r = 15)$ PPM-RM ID codes as a function of symbol size q. For comparison, the mean error probability $\overline{p_{\mathrm{err}}} = q^{-1}$ is also plotted. For increasing symbol size q, λ_2^{RS2} of the PPM-RS2 codes approaches the mean error probability, whereas λ_2^{RM} of the PPM-RM constructions remains significantly higher. The four specific ID codes are chosen to be comparable according to Section III-D3.

3) SYMBOL SIZE

The common property of all ID CWCs is the symbol size q, irrespective of the used linear block code or CWC initialization. The symbol size q denotes the base of the finite field $\mathbb{F}_q$ of the inner linear block code. Thus, for any ID CWC, before mapping the codeword to a sequence of binary symbols, the codeword is represented as a sequence of symbols in the finite field $\mathbb{F}_q$.

Figure 6 compares the impact of the symbol size q on the error probability bounds λ_2^{RS2} of two PPM-RS2 codes, as determined by Equation (38), and λ_2^{RM} of two PPM-RM codes from Equation (41), respectively. Given constant dimensions k_i and k_o and constant order r, both RS2 and RM constructions asymptotically approach a constant ratio $\lambda_2/\overline{p_{\mathrm{err}}}$ for increasing symbol size q. Specifically, for the RS2 code construction with $k_\mathrm{i} = 2$, the ratio $\lambda_2/\overline{p_{\mathrm{err}}}$ asymptotically approaches one, i.e., the bound λ_2 asymptotically approaches the mean error probability $\overline{p_{\mathrm{err}}}$, because for a fixed k_o and an increasing q, eventually the regime where $k_\mathrm{o} \ll q^{k_\mathrm{i}}$ holds is reached, cf. Equation (39). With increasing k_i (specifically for $k_\mathrm{i} = 3$), however, λ_2^{RS2} has an offset to $\overline{p_{\mathrm{err}}}$, but the λ_2^{RS2} for $k_\mathrm{i} = 3$ also saturates earlier compared to the RS2 code with $k_\mathrm{i} = 2$. On the other hand, the ID CWC constructions based on RM codes, have a constant $\lambda_2/\overline{p_{\mathrm{err}}}$ ratio, whereby a smaller order r results in a smaller ratio of λ_2^{RM} to $\overline{p_{\mathrm{err}}}$, cf. Equation (41). Therefore, only the error probability bound λ_2 of ID CWCs based on RS2 codes with small k_i and relatively small k_o compared to q^{k_i} is capable of approaching the mean error probability $\overline{p_{\mathrm{err}}}$.

B. CWC INITIALIZATION

The CWC initialization is the second component in an ID CWC—besides the linear block code—and forms the innermost CWC that maps the sequence of non-binary symbols of the linear block code to a sequence of binary symbols. The ID community suggested alternatives for the PPM initializations only for ID codes using RS codes. Thus, we investigate the performance of the PPM and OOC1

TABLE 3. Error probability bounds λ_2 of CWC ID codes. The numerical investigations in this article focus on the ID CWC types in the bold font.

ID CWC type	λ_2 bound
PPM-RS1, OOC1-RS1 HF	$(k-1)/(q-1)$ k/q
PPM-RS2 [50], **OOC1-RS2** [46] **HF-RS** [52]	$\frac{k_\mathrm{o}-1}{q^{k_\mathrm{i}}-1} + \frac{(k_\mathrm{i}-1)(q^{k_\mathrm{i}}-k_\mathrm{o})}{(q-1)(q^{k_\mathrm{i}}-1)}$ $\frac{k_\mathrm{o}-1}{q^{k_\mathrm{i}}-1} + \frac{k_\mathrm{i}}{q}$
PPM-RM [38]	r/q

initializations with double-concatenated outer RS codes, i.e., PPM-RS2 codes and OOC1-RS2 codes, and the performance of the HF initialization with a single concatenated outer RS code, i.e., HF-RS. We do not investigate alternative CWC initializations for RM codes. Note that there are no parameters for the CWC initializations themselves, as their parameters are a consequence of the parameter choice of the outer concatenated linear block codes, as shown in Table 2. Thus, we investigate the CWC initializations based on the linear block code parameters q, k_i, and k_o.

We stated the error probability bound λ_2^{RS2} of PPM-RS2 codes in Equation (38). To clarify the notation for comparison to the other CWC initializations, for this subsection we will write λ_2^{PPM} for the error probability bound of PPM-RS2 codes instead of λ_2^{RS2}. For the OOC1-RS2 and HF-RS codes, we can determine the analytic expressions for their respective λ_2 using Equation (17) and the CWC parameters from Table 2, yielding

$$\lambda_2^{\mathrm{OOC}} = \frac{K}{W} = \underbrace{\frac{k_\mathrm{o}-1}{q^{k_\mathrm{i}}-1}}_{=\alpha} + \underbrace{\frac{k_\mathrm{i}-1}{q-1}\frac{q^{k_\mathrm{i}}-k_\mathrm{o}}{q^{k_\mathrm{i}}-1}}_{=\beta^{\mathrm{PPM}}}, \tag{42}$$

$$\lambda_2^{\mathrm{HF}} = \frac{K}{W} = \underbrace{\frac{k_\mathrm{o}-1}{q^{k_\mathrm{i}}-1}}_{=\alpha} + \underbrace{\frac{k_\mathrm{i}}{q}}_{:=\beta^{\mathrm{HF}}}. \tag{43}$$

For clarity, we shorten the notation of the error probability bounds from $\lambda_2^{\mathrm{OOC1\text{-}RS2}}$ to λ_2^{OOC}, and from $\lambda_2^{\mathrm{HF\text{-}RS}}$ to λ_2^{HF}. For comparison, we show the error probability bounds of all investigated ID codes in Table 3.

We examined the performance of PPM-RS2 codes in detail in Section IV-A1. Since the error probability bound λ_2^{OOC} of OOC1-RS2 codes equals the error probability bound λ_2^{PPM} of PPM-RS2 codes, no additional evaluation of the error probability bound λ_2^{OOC} of OOC1-RS2 codes is necessary. The error probability bound λ_2^{HF} of HF-RS codes shares the first summand α with the error probability bound λ_2^{PPM} of PPM-RS2 codes, because that term is caused by the outermost RS code that PPM-RS2, OOC1-RS2, and HF-RS codes have in common. However, the second summand β^{HF} differs. For large symbol sizes q, the second summands β of the error probability bounds of PPM-RS2 and HF-RS codes are

$$\beta^{\mathrm{HF}} = \frac{k_\mathrm{i}}{q} = \frac{k_\mathrm{i}-1}{q} \cdot \left(1 + \frac{1}{k_\mathrm{i}-1}\right) \tag{44}$$

$$\beta^{\mathrm{PPM}} \overset{q \gg 1}{\approx} \frac{k_\mathrm{i}-1}{q}\frac{q^{k_\mathrm{i}}-k_\mathrm{o}}{q^{k_\mathrm{i}}} = \frac{k_\mathrm{i}-1}{q} \cdot \left(1 - \frac{k_\mathrm{o}}{q^{k_\mathrm{i}}}\right). \tag{45}$$

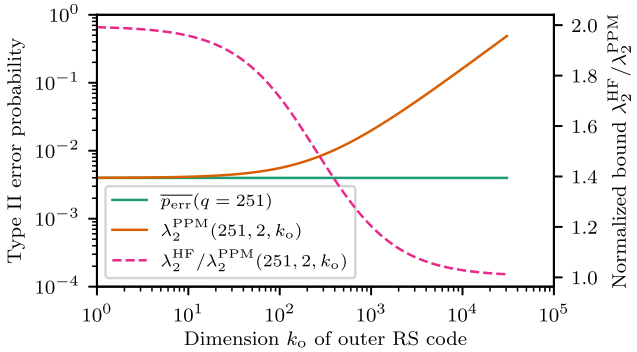

FIGURE 7. Absolute and normalized error probability bound λ_2 of $(251, 2, k_\mathrm{o})$ PPM-RS2 and $(251, 2, k_\mathrm{o})$ HF-RS codes, as a function of the dimension k_o of the outer RS code. An increased dimension k_o increases the error probability bound of both codes. The error probability bound of HF-RS codes exceeds the error probability bound of PPM-RS2 codes for all dimensions k_o.

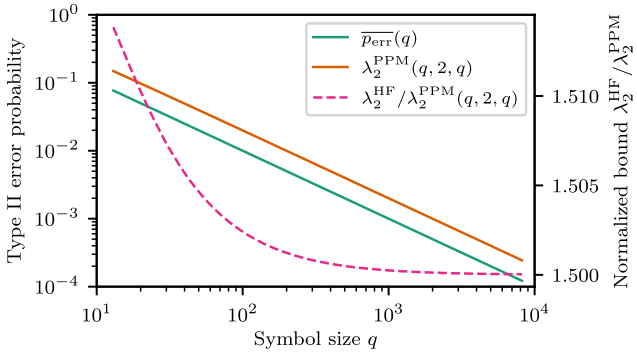

FIGURE 8. Absolute and normalized error probability bound λ_2 of $(q, 2, q)$ PPM-RS2 and $(q, 2, q)$ HF-RS codes, as a function of the symbol size q. An increased symbol size q decreases the error probability bound of both codes. The error probability bound of HF-RS codes exceeds the error probability bound of PPM-RS2 codes for all symbol sizes q.

Since $1 + 1/(k_\mathrm{i} - 1) > 1 - k_\mathrm{o}/q^{k_\mathrm{i}}$ for all positive parameters q, k_i, and k_o, the error probability bound of HF-RS codes is strictly worse than that of PPM-RS2 and OOC1-RS2 codes. However, outside the $k_\mathrm{o} \ll q^{k_\mathrm{i}}$ regime, the first summand α grows larger than the second summand β, such that

$$\lambda_2^{\mathrm{PPM}} \approx \lambda_2^{\mathrm{HF}} \quad \text{for} \quad k_\mathrm{o} \approx q^{k_\mathrm{i}}. \tag{46}$$

For large k_o, the error probability bound of all three investigated CWC initializations is almost identical.

We remark that [52] claims that HF-RS codes achieve a smaller error probability bound than PPM-RS2 codes. That claim is based on the comparison with the estimation of the loose K-bound of PPM-RS2 codes made in [50, Proposition 2], instead of the tighter K-bound shown in the proof of [50, Proposition 2]. Both K-bounds Equations (6) and (7) yield a K-bound that results in an error probability bound for HF-RS codes that is worse than the error probability bound of PPM-RS2 codes.

Next, we compare the error probability bound λ_2^{HF} of HF-RS codes with the error probability bound λ_2^{PPM} of PPM-RS2 codes for varying dimensions k_o of the outer RS code. In general, the linear block code with its high distance codewords is the main contributor to increasing the error probability bounds λ_2 in ID CWCs, and the HF CWC initialization results in a relatively minor performance loss (a further increase of the error probability bound) compared to the PPM initialization. Thus, for the comparison of the error probability bound λ_2^{HF} of HF-RS codes with the error probability bound λ_2^{PPM} of PPM-RS2 codes, we normalize λ_2^{HF} with λ_2^{PPM}.

The analysis in Section IV-A1 showed that with an increasing dimension k_o of the outer RS code, the error probability bound λ_2^{PPM} increases. Depending on the dimension k_i of the inner RS code, λ_2^{PPM} saturates for low k_o at $(k_\mathrm{i} - 1)\overline{p_\mathrm{err}}$. Figure 7 illustrates the normalized error probability bound of HF-RS, where the normalization is with respect to the corresponding λ_2^{PPM}. The error probability bound λ_2^{HF} of HF-RS codes exceeds the error probability bound λ_2^{PPM} of PPM-RS2 codes over the entire range of possible values for the dimension k_o of the outer RS code. For very large k_o, $\lambda_2^{\mathrm{HF}} \approx \lambda_2^{\mathrm{PPM}}$, but because the error probability bounds of both constructions are very high in that regime, it is undesirable to operate either code in that regime.

Finally, we compare the error probability bound λ_2^{HF} of HF-RS codes with the error probability bound λ_2^{PPM} of PPM-RS2 codes for varying dimensions k_o of the outer RS code for different symbol sizes q in Figure 8. Figure 8 indicates that the choice of symbol size q has only a minuscule impact on the normalized error probability bound λ_2^{HF} of HF-RS codes.

C. SUMMARY

We find that the error probability bound λ_2 is mainly determined by the linear block code that is part of the overall ID CWC. For all ID codes, a large symbol size q causes a small error probability bound λ_2, and simultaneously causes a small mean error probability $\overline{p_\mathrm{err}} = 1/q$. For PPM-RS2 codes, the dimension k_i of the inner RS code should be chosen as small as possible to achieve a small error probability bound λ_2. Depending on the dimension k_i of the inner RS code, increasing the dimension k_o of the outer RS code does not deteriorate the error probability bound λ_2 within a certain saturation regime. Only if the dimension k_o of the outer RS code exceeds a certain threshold (that depends on the dimension k_i of the inner RS code), then the error probability bound λ_2 increases significantly. The error probability bound λ_2 of PPM-RM codes increases with the order r of the RM code. Thus, to obtain a small error probability bound λ_2, the order r of the RM code should be chosen as small as possible.

For the CWC initializations, we find that the OOC1 initialization achieves the same error probability bounds λ_2 as the PPM initialization. The HF initialization deteriorates the error probability bound λ_2 compared to the PPM initialization.

V. CUE SIZE, ID SIZE, AND IDENTIFICATION RATE

The ID rate R measures the ratio of information per transmitted data and is upper-bounded by the capacity of the

TABLE 4. Overview of the ID size n_{id}, and the cue size n_{cue} for ID CWCs, based on the block length S and dimension N of the CWCs stated in Table 2; the corresponding ID rate is $R = (\log n_{\mathrm{id}})/n_{\mathrm{cue}}$. The numerical investigations in this article focus on the ID CWC types in the bold font.

ID CWC type	Parameters	Cue size n_{cue}	Cue size n_{cue} for $q \gg 1$	ID size n_{id}
Arbitrary CWC		$\log S$		$\log N$
PPM-RS1 [50], [51]	(q, k)	$\log[q(q-1)]$	$2 \log q$	$k \log q$
OOC1-RS1 [46]	(q, k)	$\log[q(q-1)^2]$	$3 \log q$	$k \log q$
HF [52]	(q, k)	$2 \log q$	$2 \log q$	$(k+1) \log q$
PPM-RS2 [50]	$(q, k_{\mathrm{i}}, k_{\mathrm{o}})$	$\log[q(q-1)(q^{k_{\mathrm{i}}}-1)]$	$(k_{\mathrm{i}}+2) \log q$	$k_{\mathrm{i}} k_{\mathrm{o}} \log q$
OOC1-RS2 [46]	$(q, k_{\mathrm{i}}, k_{\mathrm{o}})$	$\log[q(q-1)^2(q^{k_{\mathrm{i}}}-1)]$	$(k_{\mathrm{i}}+3) \log q$	$k_{\mathrm{i}} k_{\mathrm{o}} \log q$
HF-RS [52]	$(q, k_{\mathrm{i}}, k_{\mathrm{o}})$	$\log[q^2(q^{k_{\mathrm{i}}}-1)]$	$(k_{\mathrm{i}}+2) \log q$	$(k_{\mathrm{i}} k_{\mathrm{o}}+1) \log q$
PPM-RM [38]	(q, m, r)	$(m+1) \log q$	$(m+1) \log q$	$\binom{m+r}{m} \log q$

channel over which the data is transmitted. Specifically for ID CWCs, as outlined in Section II-C, the amount of data that is transmitted over a channel for each identification process is the cue size $n_{\mathrm{cue}} = \log S$ in bit, cf. Equation (13). We measure the amount of information represented by each identity via the ID size $n_{\mathrm{id}} = \log N$ in shannon, cf. Equation (15). Finally, the double-exponential ID rate R is defined as $(\log n_{\mathrm{id}})/n_{\mathrm{cue}}$, cf. Equation (16).

In this section, we investigate how the choice of CWC initialization and the choice of linear block code and its parameters influence the ID size n_{id}, the cue size n_{cue}, and the resulting ID rate R_{id}. We focus our investigation on the ID size n_{id}, and the cue size n_{cue} since they correspond to the amount of transmitted data and the number of IDs supported by the code, which are both practically relevant, whereas the ID rate R is more theoretical in nature. For an overview, we summarize the ID size n_{id} and the cue size n_{cue} for all ID codes investigated in this study in Table 4. Since, according to Equation (16), the ID rate R is the ratio between the ID CWC characteristics S and N (see Table 2), the impact of the linear block codes on the S and N characteristics is important in this section.

A. LINEAR BLOCK CODE

As we explained in Section III-B, we employ the PPM initialization to create comparable ID CWCs based on different linear block code types. Thus, we examine ID CWCs based on RS2 and RM codes using a PPM initialization for both of these two block codes types.

The block length S of CWCs based on either linear block code type is determined by the respective linear block code length n that we stated in Table 1. Applying the log-function to the respective block length S as given in Table 2 determines the cue size; for large symbol sizes $q \gg 1$:

$$n_{\mathrm{cue}} \overset{q \gg 1}{\approx} \begin{cases} (k_{\mathrm{i}}+2) \log q & \text{for RS2 codes,} \\ (m+1) \log q & \text{for RM codes.} \end{cases} \tag{47}$$

Thus, for RS2 codes and RM codes, the parameters k_{i} and m determine the cue size n_{cue}. Equation (47) reiterates the statement that $k_{\mathrm{i}} = m - 1$ is necessary to ensure the same cue size n_{cue} between RS2 and RM codes, as we originally noted in Equation (31). We state the exact cue size n_{cue} for arbitrary symbol sizes q in Table 4.

The dimensions N of CWCs based on either linear block code type are determined by the respective linear block code's overall dimension k that we stated in Table 1. Applying the log-function to the respective dimension N of the CWC as given in Table 2 determines the ID size:

$$n_{\mathrm{id}} = \begin{cases} k_{\mathrm{i}} k_{\mathrm{o}} \log q & \text{for RS2 codes,} \\ \binom{m+r}{m} \log q & \text{for RM codes.} \end{cases} \tag{48}$$

Therefore, for ID CWCs based on RS2 codes, the ID size n_{id} depends on the dimensions k_{i} and k_{o} of the inner and the outer RS code. For ID CWCs based on RM codes, n_{id} depends on the binomial coefficient of the parameters r and m.

To illustrate the dependence of the ID size n_{id} on the dimension k_{o} of the outer RS code and the order r of the RM code, we plot Equation (48) in Figure 9 for $q = 251$. The figure has two independent horizontal axes. The lower axis denotes the range for the dimension k_{o} of the outer RS code, whereas the upper axis denotes the range of orders r of the RM code at the given symbol size $q = 251$. The figure illustrates the ID size n_{id} for various common cue sizes n_{cue} of the ID CWC constructions using RS2 codes and RM codes. We select the parameters k_{i} for the RS code and m for the RM code according to Equation (47).

Figure 9 indicates that for small dimensions k_{o} of the outer RS code, the PPM-RM codes have larger ID sizes than the PPM-RS2 codes. For larger dimensions k_{o} of the outer RS

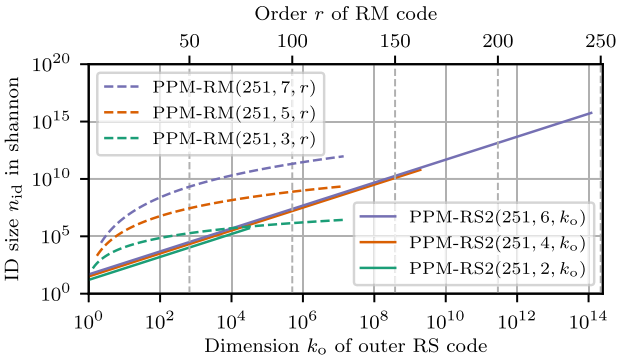

FIGURE 9. ID size n_{id} of three $(q = 251, k_{\mathrm{i}}, k_{\mathrm{o}})$ PPM-RS2 codes as a function of the dimension k_{o} of the outer RS code, and ID size n_{id} of three $(q = 251, m, r)$ PPM-RM codes as a function of the order r of the RM code. We choose the set of PPM-RS2 and PPM-RM codes such that pairs of PPM-RS2 and PPM-RM codes share the same cue size n_{cue}. Specifically, all $(251, 2, k_{\mathrm{o}})$ PPM-RS2 and $(251, 3, r)$ PPM-RM codes have a cue size $n_{\mathrm{cue}} = 31.9$ bit, all $(251, 4, k_{\mathrm{o}})$ PPM-RS2 and $(251, 5, r)$ PPM-RM codes have a cue size $n_{\mathrm{cue}} = 47.8$ bit, and all $(251, 6, k_{\mathrm{o}})$ PPM-RS2 and $(251, 7, r)$ PPM-RM codes have a cue size $n_{\mathrm{cue}} = 63.8$ bit. Increasing the parameters m or k_{i} increases the maximum achievable ID size.

code, the PPM-RS2 codes can obtain larger ID sizes than the PPM-RM codes.

The RM curves in Figure 9 are only plotted up to $r = 125$, because the error probability bound $\lambda_2 = r/q = 125/251 \approx 0.5$. In other words, for larger orders r the bound would exceed 0.5. The same holds for the RS2 curves, where $\lambda_2 \approx k_o/q^{k_i}$, such that k_o should not exceed $q^{k_i}/2$.

B. CWC INITIALIZATION

After comparing the influence of linear block codes on the ID size n_{id}, we now address the effect of the CWC initialization on the rate-related KPIs, namely the ID size n_{id} and cue size n_{cue}. Similar to our analysis in Section IV, we compare the CWC initializations in concatenation with RS codes for the linear block codes, as the CWC initializations next to the PPM initialization were proposed in combination with outer RS codes [46], [52].

The choice of CWC initialization has an impact on the block length S and the dimension N of the CWC, cf. Table 2. The resulting differences between the ID sizes n_{id} of the three CWC initializations are minor, whereas the cue sizes n_{cue} of the three CWC initializations deviate slightly more from each other, cf. Table 4.

We first investigate the cue sizes n_{cue}. The difference between n_{cue}^{HF} and n_{cue}^{PPM} is minuscule due to the minor difference in their block lengths S and gradually vanishes for large symbol sizes q, see Table 2. Note that even for small symbol sizes, such as $q = 13$, the cue size n_{cue}^{HF} of RS-HF codes is very close to the cue size n_{cue}^{PPM} of PPM-RS2 codes. Specifically, for symbol size $q = 13$, $n_{cue}^{HF}/n_{cue}^{PPM} = 1.008$. The OOC initialization, however, generates significantly larger cue sizes. For large field sizes $q \gg 1$, the cue sizes for PPM-RS2, HF-RS, and OOC1-RS2 codes are:

$$n_{cue} = \begin{cases} (k_i + 2)\log q & \text{for PPM-RS2, HF-RS codes,} \\ (k_i + 3)\log q & \text{for OOC1-RS2 codes.} \end{cases} \tag{49}$$

We summarize the cue sizes n_{cue} for arbitrary field sizes q in Table 4.

To visualize these cue sizes n_{cue}, Figure 10 depicts the relations of Equation (49) over a range of symbol sizes q. We plot the cue size n_{cue} for different dimensions k_i of the inner RS code ($k_i = 2, 3, 4$), given a fixed dimension $k_o = q$ of the outer RS code. Figure 10 indicates that the cue size n_{cue} increases with increasing symbol size q; and, that an increasing dimension k_i of the inner RS code increases the cue size n_{cue} for all symbol sizes q. The cue size n_{cue}^{OOC1} of OOC1-RS2 codes is larger than the cue sizes of PPM-RS2 and HF-RS codes by $\log q$, given identical parameters q, k_i, k_o. Thus, given identical parameters, q, k_i, k_o, transmitting a cue generated by OOC1-RS2 codes requires additional traffic compared to transmitting a cue based on PPM-RS2 or HF-RS codes.

Next, we investigate the ID size n_{id} of the different CWC initializations. The ID size n_{id} of PPM-RS2 and

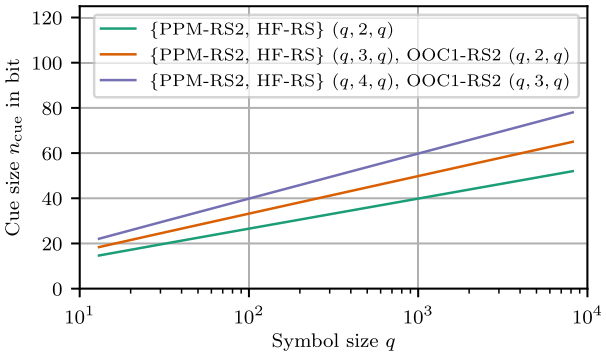

FIGURE 10. Cue size n_{cue} of (q, k_i, k_o) PPM-RS2, (q, k_i, k_o) OOC1-RS2, and (q, k_i, k_o) HF-RS codes as a function of the symbol size q, for dimensions $k_i = 2, 3, 4$; based on Equation (49).

OOC1-RS2 codes equals $n_{id} = k_i k_o \log q$, cf. Table 4 and Equation (48). For HF-RS codes, the ID size $n_{id} = (k_i k_o + 1)\log q$ is larger by a factor of $\log q$, i.e., HF-RS codes have more codewords than PPM-RS2 and OOC1-RS2 codes, and can thus represent more IDs, given the same parameters q, k_i, k_o. The difference diminishes for larger field sizes q and is small even for small parameters q, k_i, k_o. Specifically, the ratio of the ID size n_{id}^{HF} of HF-RS codes, and the ID size n_{id}^{PPM} of PPM-RS2 codes given the parameters ($q = 13, k_i = 2, k_o = 13$) is $n_{id}^{HF}/n_{id}^{PPM} = 1.038$. For small parameter sets, HF-RS thus allows for a moderate increase in the number of representable IDs.

C. SUMMARY

We defined the cue size n_{cue} and the ID size n_{id} as KPIs. The cue size n_{cue} determines the amount of data that is transmitted for each identification process. The ID size n_{id} determines how many IDs can be represented by the ID code.

We find that it is mainly the choice of the linear block code that determines the cue size n_{cue} and the ID size n_{id}. The cue size n_{cue} is mainly a result of the overall length n of the linear block code. To reduce the cue size n_{cue} in order to reduce the data sent over the channel, a small overall length n of the linear block code is desirable. The ID size n_{id} is mainly determined by the overall dimension k of the linear block code. To increase the number n_{cue} of representable IDs, a large overall dimension k of the linear block code is desirable.

To make RS2 and RM codes comparable, we compare them for matching overall length n and dimension k. Therefore, the cue size n_{cue} of RS2 and RM codes is identical due to this choice for a basis of comparison. The ID sizes n_{id} of RS2 and RM codes show different behavior in terms of their growth in the block code parameters. Overall, neither code type exhibits larger ID sizes n_{id} over the full range of parameters.

Choosing an alternative to the PPM initialization only has a minor impact on the ID size n_{id} of ID CWCs. The OOC1 initialization exhibits a larger cue size n_{cue} than the PPM and HF initializations and is thus to be avoided when trying to limit the amount of transmitted data.

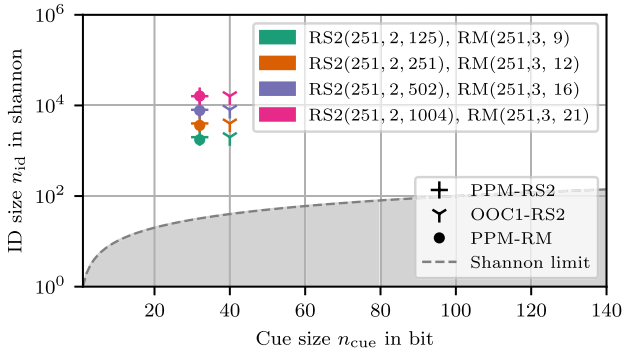

FIGURE 11. ID size n_{id} as a function of the cue size n_{cue} for ($q = 251$, $k_{\mathrm{i}} = 2$, k_{o}) {PPM, OOC1}-RS2 codes of dimensions $k_{\mathrm{o}} \in \{q/2,\ q,\ 2q,\ 4q\}$ of the outer RS code, and for ($q = 251$, $m = 3$, r) PPM-RM codes of orders $r \in \{9,\ 12,\ 16,\ 21\}$. ID codes of the same color share the same overall block code length n and overall block code dimension k, such that corresponding PPM-RS2 and PPM-RM codes share the same ID size n_{id} and cue size n_{cue}, except for rounding errors to the next integer parameter value. The shaded area marks (ID size n_{id}, cue size n_{cue}) pairs that are achievable by the traditional message transmission scheme, where $n_{\mathrm{id}} \leq n_{\mathrm{cue}}$. (ID size n_{id}, cue size n_{cue}) pairs in the unshaded area are only achievable using ID codes [15]. Increasing the dimension k_{o} of the outer RS code increases the ID size n_{id} without increasing the cue size n_{cue}. OOC1-RS2 codes require a larger cue size n_{cue} than PPM-RS2 codes to achieve the same ID size n_{id}.

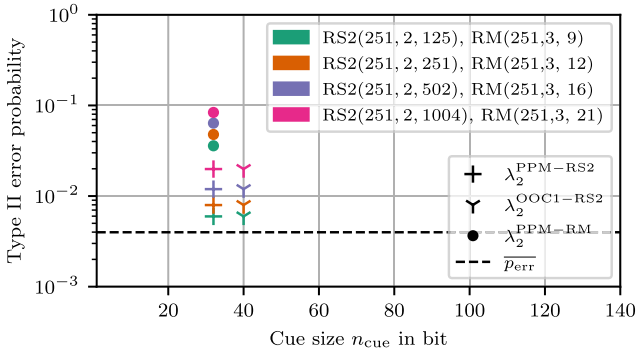

FIGURE 12. Error probability bound λ_2 as a function of the cue size n_{cue} for the twelve ID codes in Figure 11 all of which have the mean error probability $\overline{p_{\mathrm{err}}} = 1/q = 1/251$. Increasing the dimension k_{o} of the outer RS code increases the error probability bound λ_2, cf. Figure 4. OOC1-RS2 codes require a larger cue size n_{cue} than PPM-RS2 codes to achieve the same error probability bound λ_2. Increasing the order r of the RM code increases the error probability bound λ_2, which is undesirable. The error probability bound λ_2 of ID codes that include an RM code is significantly higher than for the ID codes that include RS codes.

VI. HOLISTIC ID CODE COMPARISON

In Sections IV and V, we analyzed how the choice of the ID code type affects the individual KPIs, namely error probability bound λ_2, cue size n_{cue}, and ID size n_{id}, based on the choice of the parameters, namely dimensions k_{i} and k_{o} for codes based on RS2 codes, order r and generation m for codes based on RM codes, and symbol size q for all codes. In contrast to the standalone investigations of individual KPIs in the previous sections, we investigate the tradeoffs between different KPIs in this section. For example, to achieve a low error-probability bound λ_2 for finite parameters, a larger cue size n_{cue} needs to be accepted. That is because by only sending the fragmentary representation of the full ID, i.e., by sending only a cue, Shannon's limit is exceeded, thus introducing the error probability. The further Shannon's limit is exceeded, the more IDs can be identified, but also the higher the error probability bound λ_2, cf. Equation (19).

We focus our investigation in this section on PPM-RS2 codes, because we found the other code types to be inferior in Sections IV and V: Specifically, we found that compared to PPM-RS2 codes, for identical parameters, OOC1-RS2 codes achieve the same ID sizes n_{id}, mean error probabilities $\overline{p_{\mathrm{err}}}$, and error probability bounds λ_2, but require larger cue sizes n_{cue} for otherwise identical performance. OOC1-RS2 codes are, therefore, always inferior to PPM-RS2 codes in terms of the KPIs we investigate.

For HF-RS codes, we found that, compared to PPM-RS2 codes, HF-RS codes require the same cue sizes n_{cue} to achieve minusculely larger ID sizes n_{id}, but also exhibit higher error probability bounds λ_2. In scenarios that require very small dimensions $k_{\mathrm{i}}, k_{\mathrm{o}}$ of the inner and outer RS codes, using HF-RS codes instead of PPM-RS2 codes could be beneficial to increase the ID size n_{id}, even if this incurs higher error probability bounds λ_2. However, in general, we find HF-RS codes inferior to PPM-RS2 codes in terms of the KPIs we investigate.

Overall, we find that both alternative CWC initializations (OOC1, HF) exhibit less favorable KPIs than the PPM initialization. This points to investigating other linear block codes for their performance as ID codes rather than investigating alternative CWC initializations in future ID code research.

RM codes are an alternative to using RS2 codes as the linear block code. In Sections IV and V, we explained that PPM-RM codes, for similar overall lengths n and overall dimensions k of the linear block codes, exhibit significantly higher error probability bounds λ_2 than PPM-RS2 codes. Additionally, the respective cue sizes n_{cue} are almost identical, while overall, the maximally achievable ID sizes n_{id} are smaller for PPM-RM codes. Thus, RM codes are worse than RS2 codes in terms of the KPIs that we investigate. However, determining cues (or tags) from RM codes requires significantly less computational power than the cue (or tag) determination from RS2 codes, which is a key benefit of RM codes [38].

In the following, we investigate the tradeoffs between cue sizes n_{cue}, ID sizes n_{id}, mean error probabilities $\overline{p_{\mathrm{err}}}$, and the error probability bounds λ_2, mainly for PPM-RS2 codes, for symbol sizes q, as well as the dimensions k_{i} and k_{o} of the inner and outer RS codes. We also include the results of OOC1-RS2 codes and PPM-RM codes for comparison. We exclude HF-RS codes for visual clarity because their performance is closest (but still overall inferior) to PPM-RS2 codes.

A. DIMENSION OF OUTER RS CODE

In Figure 11, we plot the relationship between the achieved ID size n_{id} and the required cue size n_{cue} for achieving this desired ID size n_{id} for several ID codes, for varying dimensions k_{o} of the outer RS code, symbol size $q = 251$, and dimension $k_{\mathrm{i}} = 2$ of the inner RS code. Figure 11 illustrates

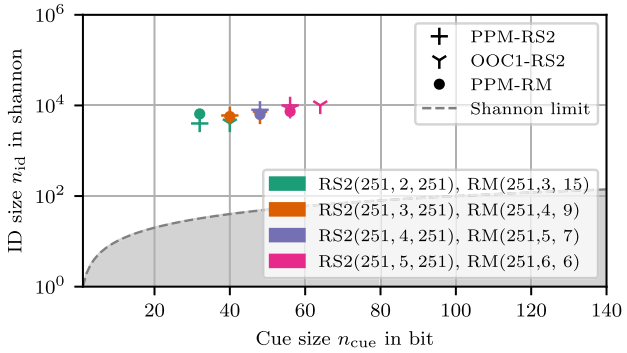

FIGURE 13. ID size n_{id} as a function of the cue size n_{cue} for ($q = 251$, k_{i}, $k_{\text{o}} = 251$) {PPM, OOC1}-RS2 codes of dimensions $k_{\text{i}} \in \{2,\ 3,\ 4,\ 5\}$ of the outer RS code, and for ($q = 251$, m, r) PPM-RM codes of parameter pairs $(m, r) \in \{(3, 15),\ (4, 9),\ (5, 7),\ (6, 6)\}$. ID codes of the same color share the same overall block code length n and overall block code dimension k. Increasing the dimension k_{i} of the inner RS code causes a minor increase in the ID size n_{id}, while increasing the cue size n_{cue} significantly.

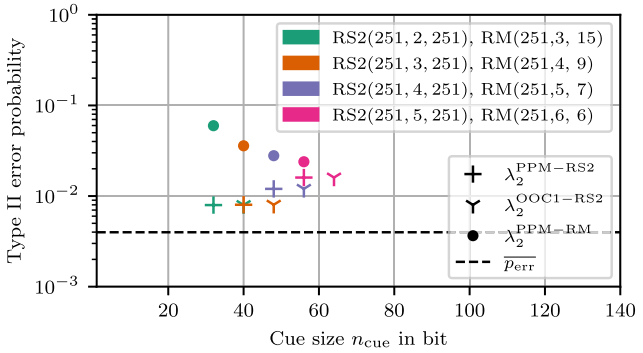

FIGURE 14. Error probability bound λ_2 as a function of the cue size n_{cue} for the twelve ID codes in Figure 13 with mean error probability $\overline{p_{\text{err}}} = 1/q = 1/251$. Increasing the dimension k_{i} of the inner RS code increases both the cue size n_{cue} and the error probability bound λ_2, which is both undesirable. Reducing the order r of the RM code reduces the error probability bound λ_2, which is desirable. The error probability bound λ_2 of ID codes that include an RM code is significantly higher than for the ID codes that include RS codes.

that the cue size n_{cue} of ID codes based on RS2 codes is invariant to the dimension k_{o}, and that OOC1-RS2 codes have larger cue sizes n_{cue} than PPM-RS2 codes. We choose the parameters of the plotted PPM-RM codes such that they match the ID size n_{id} and cue size n_{cue} of the plotted PPM-RS2 codes, cf. Section III-D3. For reference, we also plot the area corresponding to (ID size, cue size) tuples that are achievable by traditional message transmission codes, as opposed to ID codes that are the focus of this article. All ID codes shown in Figure 11 (as well as the subsequent ID size Figures 13 and 15) exceed the line $n_{\text{id}} = n_{\text{cue}}$, which is called Shannon limit, at the cost of an error probability, that is inherent to exceeding the Shannon limit.

Figure 12 depicts the relation between the cue size n_{cue} and the error probability bound λ_2, for the same ID codes considered in Figure 11. As the dimension k_{o} of the outer RS code increases, the K-bound of the PPM-RS2 code increases, thus increasing the error probability bound λ_2, as we already visualized in Figure 4 in Section IV-A1. The increase of the error probability bound λ_2 is large because the outer RS code's dimension k_{o} is not in its saturation phase, i.e., $k_{\text{o}} > q^{k_{\text{i}}-1}$, cf. Section IV-A1.

The PPM-RS2 codes exhibit a significantly lower error probability bound λ_2 than the PPM-RM codes. The error probability bounds λ_2 of the PPM-RM codes increase for an increasing order r of the RM codes. The OOC1-RS2 codes require larger cue sizes to achieve the same error probability bound λ_2 as the PPM-RM codes. All codes in Figure 12 share a common mean error probability $\overline{p_{\text{err}}} = 1/q$, since the symbol size q is shared by all considered codes.

As we stated in Section V, increasing the dimension k_{o} of the outer RS code only affects the ID size n_{id}, but not the cue size n_{cue}, cf. Table 2. From a practical perspective, the dimension k_{o} of the outer RS code determines how many IDs are allocated into the limited codeword set $\mathcal{X}^n \subseteq \mathbb{F}_q^n$. For an ID code based on the PPM initialization, the overall length n of the employed linear block code determines the cue size $n_{\text{cue}} = \log S = \log q + \log n$. Increasing the dimension k_{o} of the outer RS code creates a denser population of IDs within a stagnant codeword set $\mathcal{X}^n$. Consequentially, the overlap of codewords increases, and, therefore, the error probability increases.

B. DIMENSION OF INNER RS CODE

Figure 13, similar to Figure 11 in the preceding Section VI-A, depicts the relation between the ID size n_{id} and the required cue size n_{cue}, for several PPM-RS2 codes of varying dimension k_{i} of the inner RS2 code, for symbol size $q = 251$ and dimension $k_{\text{o}} = q$ of the outer RS code. Figure 13 illustrates that increasing the dimension k_{i} of the inner RS code increases the cue size n_{cue} significantly, while the ID size n_{id} experiences only a minor increase. The OOC1-RS2 codes require larger cue sizes to achieve the same ID size n_{id} as PPM-RS2 codes, cf. Equation (49).

The parameters of the PPM-RM codes are again chosen to match their overall lengths n and their overall dimensions k to the respective corresponding PPM-RS2 code, cf. Section III-D3. The differences in ID size n_{id} between the corresponding (PPM-RS2 code, PPM-RM code) pairs are only caused by the rounding to integer values of the RM code parameters (m, r) to match the overall block code length n and the overall dimension k of the RS2 codes.

For the same ID codes as considered in Figure 13, Figure 14 illustrates the relation between the cue size n_{cue} and the error probability bound λ_2 for varying dimensions k_{i} of the inner RS code of PPM-RS2 codes. Increasing the dimension k_{i} of the inner RS code generally increases the error probability bound λ_2 of PPM-RS2 codes. The error probability bound λ_2 of the ($q = 251$, $k_{\text{i}} = 2$, $k_{\text{o}} = 251$) and the ($q = 251$, $k_{\text{i}} = 3$, $k_{\text{o}} = 251$) PPM-RS2 codes are similar, because the chosen dimension $k_{\text{o}} = 251$ of the outer RS code puts the ($q = 251$, $k_{\text{i}} = 2$, $k_{\text{o}} = 251$) PPM-RS2 code outside of the saturation regime of its error probability bound λ_2, see Section IV-A1 and Figure 4. For the other three PPM-RS2 codes, the chosen dimension $k_{\text{o}} = 251$ of the outer RS code is within the saturation regime.

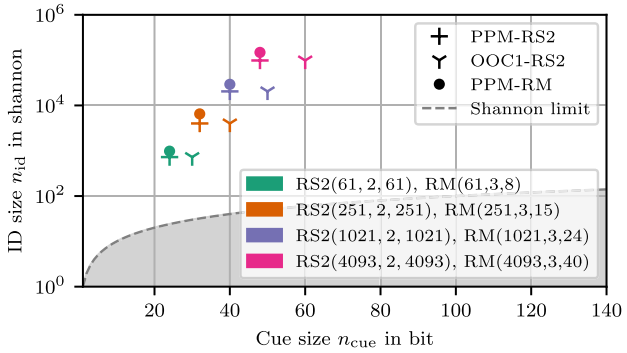

FIGURE 15. ID size n_{id} as a function of the cue size n_{cue} for $(q, k_{\mathrm{i}} = 2, k_{\mathrm{o}} = q)$ {PPM, OOC1}-RS2 codes of symbol size $q \in \{61, 251, 1021, 4093\}$, and for $(q, m = 3, r)$ PPM-RM codes of parameter pairs $(q, r) \in \{(61, 8), (251, 15), (1021, 24), (4093, 42)\}$. ID codes of the same color share the same overall block code length n and overall block code dimension k. Increasing the symbol size q increases the ID size n_{id} significantly while also increasing the cue size n_{cue}.

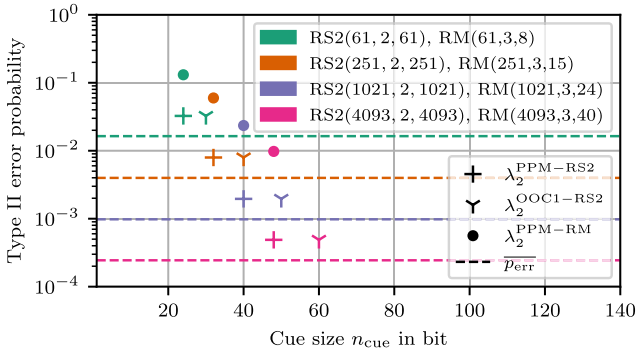

FIGURE 16. Error probability bounds λ_2 as a function of the cue size n_{cue} for ID codes in Figure 15. We plot the mean error probability $\overline{p_{\mathrm{err}}} = 1/q$ of codes of different symbol sizes q in the corresponding color of the ID codes. Increasing the symbol size q increases the cue size n_{cue}, but decreases the error probability bound λ_2 significantly. When the increase in the order r of the RM code is smaller than the increase in symbol size q of the RM code, the error probability bound λ_2 decreases overall, which is desirable.

For the PPM-RM codes, the error probability bound $\lambda_2 = r/q$ increases with an increasing order r of the RM codes. To match the overall block code length n and the overall dimension k of the RS2 codes, both parameters (m, r) of the RM codes differ between the four plotted PPM-RM codes. Thus, the decrease in the error probability bound λ_2 of the PPM-RM codes as the cue size n_{cue} increases cannot be attributed to a single parameter of the RM codes.

Unlike our investigation for the dimension k_{o} of the outer RS code, increasing the dimension k_{i} of the inner RS code of PPM-RS2 codes increases the cue size n_{cue}. Increasing the dimension k_{i} of the inner RS code increases the overall length n and the overall dimension k of the resulting linear block code, therefore also increasing the cue size n_{cue} and the ID size n_{id}.

The dimension k_{i} of the inner RS code restricts the values for the dimension k_{o} of the outer RS code, because the error probability bound λ_2 of PPM-RS2 codes as given in Table 3 is lower bounded by

$$\lambda_2^{\mathrm{PPM-RS2}} > \frac{k_{\mathrm{o}} - 1}{q^{k_{\mathrm{i}}} - 1}. \tag{50}$$

Consequently, when increasing the dimension k_{o} of the outer RS code with the goal of increasing the ID size n_{id}, the dimension k_{i} of the inner RS code eventually has to be increased as well in order to mitigate the impact of a growing dimension k_{o} of the outer RS code on the error probability bound λ_2. Because increasing the dimension k_{i} of the inner RS code increases the cue size n_{cue}, the dimension k_{i} of the inner RS code should be chosen as small as possible.

For practical ID CWC deployments, this investigation of the dimension k_{i} of the inner RS code indicates that increasing k_{i} is not advantageous, neither for the cue size n_{cue} nor for the error probability bound λ_2. However, increasing the dimension k_{i} of the inner RS code enables increasing values for k_{o}, which is beneficial to increase the ID size n_{id}.

C. SYMBOL SIZE

Similar to Figure 11 and Figure 13, in Figure 15, we plot the achieved ID size n_{id} and the required cue size n_{cue} for achieving the ID size n_{id}, for varying symbol sizes q, for dimension $k_{\mathrm{o}} = q$ of the outer RS code and dimension $k_{\mathrm{i}} = 2$ of the inner RS code. Figure 15 illustrates that increasing the symbol size q and the dimension $k_{\mathrm{o}} = q$ of the outer RS code increases both the cue size n_{cue} and the ID size n_{id} of ID codes based on RS2 codes. Again, we choose the parameters of the PPM-RM codes such that they match in ID size n_{id}, and cue size n_{cue}, cf. Section III-D3.

Figure 16 depicts the relation between the cue size n_{cue} and the error probability bound λ_2, for the same ID codes considered in Figure 15. As the symbol size q increases, the error probability bound λ_2 decreases, as we first observed in Figure 6. The PPM-RS2 codes exhibit a significantly lower error probability bound λ_2 than the PPM-RM codes. The error probability bounds $\lambda_2 = r/q$ of the PPM-RM codes decrease for increasing symbol sizes q. Generally, increasing the order r of the RM codes increases the error probability bound $\lambda_2 = r/q$ of PPM-RM codes. For the PPM-RM codes in Figure 16, the symbol size q grows faster than the order r of the RM codes. Accordingly, we observe from Figure 16 that the fast-increasing symbol size q dominates over the slowly-increasing order r of the RM codes, resulting overall in decreasing error probability bounds λ_2 for the successively increasing symbol sizes q and orders r.

Finally, a major advantage of increasing the symbol size q lies in the corresponding reduction of the mean error probability $\overline{p_{\mathrm{err}}} = 1/q$. Since the mean error probability $\overline{p_{\mathrm{err}}}$ is univariately determined by the symbol size q, increasing the symbol size q is the only method to reduce the mean error probability $\overline{p_{\mathrm{err}}}$.

In conclusion, increasing the symbol size q increases the cue size n_{cue}, thus increasing the traffic required for the ID process. However, increasing the symbol size q also increases the ID size n_{id}, while simultaneously reducing the error probability bound λ_2, and the mean error probability $\overline{p_{\mathrm{err}}}$.

Therefore, a large symbol size q is highly desirable. Overall, ID codes in general, and PPM-RS2 codes in particular, work best for large symbol sizes q. Because large symbol sizes q incur large cue sizes n_{cue}, the symbol size q cannot be chosen arbitrarily large, but needs to be moderated according to the use case requirements.

VII. PARAMETER SELECTION FOR ID CODES

Based on the findings of Section VI, in this section, we provide a set of heuristics for selecting parameters of ID codes so as to achieve good performance. First, we define the following goal: from every ID code we aim to obtain a prescribed number N of different ID messages (and corresponding ID size $n_{id} = \log N$), while achieving a low error probability and requiring the transmission of cues of small size n_{cue}. In other words, we examine ID codes for scenarios that require the ID code to be able to represent a prescribed ID size n_{id}. The goal is to identify this fixed number of IDs (represented by the ID size n_{id}) with as few errors as possible (low error probability), while transmitting as little data as possible (small cue size). Thus, we examine which cue sizes n_{cue} are required to achieve a certain ID size n_{id} at the cost of an introduced error probability for false-positive verification. The error probability is characterized by the mean error probability $\overline{p_{\text{err}}}$, as well as the error probability bound λ_2.

A. HEURISTICS FOR PPM-RS2 CODES

Through the evaluations in the preceding Section VI, we investigated the impact of the RS2 code parameters q, k_{i}, and k_{o} on the KPIs of ID codes. We proceed to formulate heuristics for selecting practical PPM-RS2 codes that strike a balance between small cue sizes n_{cue}, large ID sizes n_{id}, and small error probability bounds λ_2.

1) CHOICE OF SYMBOL SIZE

The evaluations in the preceding Section VI indicate that large symbol sizes q contribute to low error probability bounds λ_2 and large ID sizes n_{id} compared to the corresponding cue sizes n_{cue}. Additionally, with an increase in symbol size q, the mean error probability $\overline{p_{\text{err}}}$ decreases. Therefore, the symbol size q fundamentally determines the error probability. The dimensions k_{i} and k_{o} of the inner and outer RS codes, respectively, have a relatively minor impact, cf. Figure 12 and 14. Hence, a large symbol size q achieves a small error probability bound λ_2, while increasing the ID size n_{id} to the desired value with a moderate increase in the cue size n_{cue}.

As a limiting factor for large symbol sizes q when choosing parameters for ID codes, the increase in computational complexity for large symbol sizes should be considered [36]. Hence, for designing practical ID CWCs, the symbol size q should be chosen as large as possible without disregarding acceptable computational complexity with optimized and hardware-accelerated encoding algorithms.

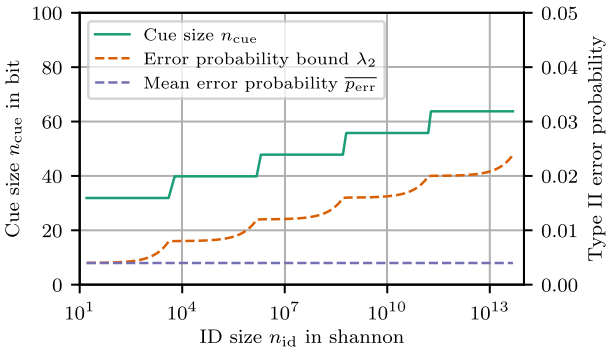

FIGURE 17. Cue size n_{cue} and error probability bound λ_2 as function of the ID size n_{id} of ($q = 251, k_{\text{i}}, k_{\text{o}}$) PPM-RS2 codes with the dimensions $k_{\text{i}}, k_{\text{o}}$ based on the heuristics in Section VII. When increasing the ID size n_{id}, the cue size n_{cue} is increased stepwise once the error probability bound λ_2 begins to increase significantly. By increasing the cue size n_{cue}, the error probability bound is returned to the saturation regime, cf. Equation (39).

2) CHOICE OF RS CODE DIMENSIONS

After selecting an appropriate symbol size q, the required ID size n_{id} can be achieved with the determination of suitable dimensions k_{i} and k_{o} of the inner and outer RS codes, respectively. Both dimension parameters, k_{i} and k_{o}, increase the ID size n_{id}, while increasing the error probability bound λ_2. In the preceding Section VI, we found that increasing the dimension k_{o} of the outer RS code does not increase the cue size n_{cue}. In addition, we observed from Figure 4 that the error probability bound λ_2 of PPM-RS2 codes saturates for $k_{\text{o}} \ll q^{k_{\text{i}}}$ at $\lambda_2 \approx (k_{\text{i}} - 1)/q$, cf. Equation (39). However, the error probability bound λ_2 of PPM-RS2 codes increases significantly for larger dimensions k_{o} of the outer RS code as the summand term $(k_{\text{o}} - 1)/(q^{k_{\text{i}}} - 1)$ in the λ_2 expression grows significantly, i.e., the lower bound for the error probability bound λ_2, cf. Equation (50), grows significantly. Thus, a required ID size n_{id} should be achieved by increasing the dimension k_{o} of the outer RS code while maintaining the smallest possible dimension k_{i}, for which $k_{\text{o}} \leq q^{k_{\text{i}}-1}$, cf. Equation (40). This ensures the smallest possible error probability bound λ_2, and the smallest possible cue size n_{cue}.

3) CUE SIZE AND ERROR PROBABILITY AS A FUNCTION OF ID SIZE

For PPM-RS2 codes of symbol size $q = 251$, Figure 17 visualizes the KPIs when following the heuristics we introduced for desired ID sizes n_{id}. We plot the required cue size n_{cue} and the error probability characteristics of the different PPM-RS2 codes. For the error probability characteristics, Figure 17 illustrates the diverging behavior of the error probability bound λ_2 from the mean error probability $\overline{p_{\text{err}}}$ for an increasing ID size n_{id}. The cue size n_{cue} follows the stepwise increases of the dimension $k_{\text{i}} \in [2, 3, 4, 5, 6]$ of the inner RS code. We increase the dimension k_{i} of the inner RS code if the dimension k_{o} otherwise (if k_{i} is not increased) exceeds the threshold $q^{k_{\text{i}}-1}$ that we determined in Equation (40).

B. DISCUSSION OF RM CODES

PPM-RM codes are parameterized by the generation m, the order r, and the symbol size q. To achieve a small error probability bound $\lambda_2 = r/q$, large symbol sizes q and small orders r are advisable. Increasing the symbol size q incurs a higher computational complexity of the cue determination and increases the cue size n_{cue} that needs to be transmitted over the channel. Therefore, the symbol size q should be selected as large as possible (so as to reduce the error probability) but as small as necessary (to be computationally feasible). To increase the ID size n_{id}, the generation m can be increased. Since $m \leq r$ is a condition for the parameter selection of RM codes, significant increases of the ID size n_{id} via the generation m also require an increase in the order r, which causes an undesirable increase in the error probability bound λ_2. Additionally, the generation m increases the cue size n_{cue}. In conclusion, the parameters m and r should be selected as small as possible to achieve a sufficient ID size n_{id}, while maintaining a small cue size n_{cue} and a small error probability bound λ_2.

VIII. CONCLUSION

This topical review article compared all existing ID codes, except for random linear ID codes [39], in the practical finite-parameter regime, based on three metrics: the cue size, the ID size, and the error probability (specifically, the mean and the upper bound of the error probability). Based on these metrics, we found that compared to the PPM initialization for constant-weight codes for ID, the alternative CWC initializations based on ϵ–almost strongly universal hash functions [52] and optical orthogonal codes [46] have limited usefulness for constructing ID codes. ID codes based on ϵ–almost strongly universal hash functions exhibit higher error probability bounds λ_2, and the cue size n_{cue} of the transmitted data is larger for ID codes based on optical orthogonal codes, compared to ID codes based on the PPM initialization. Conversely, we found that it is *not* the CWC initialization but rather the employed linear block code that primarily determines the performance of an ID code in terms of the three investigated metrics. Concatenated Reed-Solomon codes approach the ID capacity for infinite block lengths [50] and perform well for finite coding parameters. A computationally cheaper alternative are ID codes based on Reed-Muller codes [38]. However, ID codes based on Reed-Muller codes exhibit significantly higher error probability bounds than ID codes based on concatenated Reed-Solomon codes.

Additionally, we investigated the choice of coding parameters for good performance. ID codes work best for large symbol sizes q, which decrease the error probability bound and mean error probability, while also increasing the ID size. For the dimensions $k_{\text{i}}, k_{\text{o}}$ of concatenated Reed-Solomon codes, we proposed a heuristic to achieve a large ID size, while limiting the error probability bound and cue size to low levels. We also discussed the choice of parameters for ID codes based on Reed-Muller codes. In conclusion, concatenated Reed-Solomon codes achieve good performance for finite parameters, but their computational complexity is very high [36]. Reed-Muller codes provide a computationally cheaper alternative, at the cost of worse performance.

For future research, our findings point towards prioritizing investigations on other linear block codes to enhance ID code performance at acceptable computational complexity, rather than examining alternative CWC initializations. One specific direction for investigating linear block codes could be to explore polar codes [54], [55] for the purpose of ID coding. Also, it should be noted for future research that ID coding requires generally only efficient encoding, but effectively no decoding as only cues (or tags) are compared for the "ID decoding", i.e., the verification of an identity match. Thus, future ID codes could be based on linear block codes that are computationally feasible for encoding, but computationally prohibitive for conventional message decoding. Concomitantly, future research should examine hardware acceleration [56], [57], [58] to address the computational complexities of existing and future ID codes so as to advance ID coding as a practical means of goal-oriented communication over networks.

Another avenue for future research is to examine ID coding in the context of specific application contexts, e.g., in the smart labels [6], [7], [8], [9] context or the digital twin context [29], [30], [31], [32], [33], [34], [35]. These future examinations should consider typical operational scenarios with data from real-life systems in operational practice, e.g., smart label data exchanges from operational large-scale warehouses, or digital twin data exchanges from operational robots in industrial production lines. Such real-life system data could be utilized to evaluate a broader range of ID coding performance metrics, e.g., detailed stochastic characterizations of the false-positive identification errors during operational scenarios. Such detailed stochastic error characteristics could inform further ID code developments and evaluations as well as the development of communication protocols for the identification via channels paradigm.

APPENDIX: TRANSLATION TABLE FROM RS CODE DIMENSION TO RM CODE ORDER

See Table 5.

TABLE 5. Translation table for determining the order r of a (q, m, r) RM code of generation m based on a $(q, k_{\text{i}}, k_{\text{o}})$ RS2 code of dimension k_{i} of the inner RS code, and dimension $k_{\text{o}} = q$ of the outer RS code, for symbol size q, and a common overall code dimension k and length n of the linear block codes.

q	k_{i} / m	2 / 3	3 / 4	4 / 5	5 / 6	6 / 7
13		4	3	2	2	2
31		6	4	4	3	3
61		8	6	4	4	3
127		11	7	6	5	4
251		15	9	7	6	5
509		19	11	8	7	6
1021		24	14	10	8	7
2039		31	17	12	9	8
4093		40	21	14	11	9
8191		51	25	17	13	10
16381		65	30	20	15	12
32749		82	37	23	17	14
65521		104	44	27	19	15

REFERENCES

[1] Y. Chen, F. Zheng, T. Kaiser, and A. H. Vinck, "An information-theoretic approach to the chipless RFID tag identification," *IEEE Access*, vol. 7, pp. 96984–97000, 2019.

[2] P. Fraga-Lamas, J. Varela-Barbeito, and T. M. Fernández-Caramés, "Next generation auto-identification and traceability technologies for industry 5.0: A methodology and practical use case for the shipbuilding industry," *IEEE Access*, vol. 9, pp. 140700–140730, 2021.

[3] S. Gabsi, Y. Kortli, V. Beroulle, Y. Kieffer, A. Alasiry, and B. Hamdi, "Novel ECC-based RFID mutual authentication protocol for emerging IoT applications," *IEEE Access*, vol. 9, pp. 130895–130913, 2021.

[4] J. Song, S. He, and H. Yao, "TMIA: A tree-based multi-reader interactive anti-collision algorithm for RFID tag identification," *IEEE Access*, vol. 8, pp. 81594–81605, 2020.

[5] Z. Yang, X. Liu, Z. Li, B. Yuan, and Y. Zhang, "RF-Eletter: A cross-domain English letter recognition system based on RFID," *IEEE Access*, vol. 9, pp. 155260–155273, 2021.

[6] T. M. Fernández-Caramés and P. Fraga-Lamas, "A review on human-centered IoT-connected smart labels for the industry 4.0," *IEEE Access*, vol. 6, pp. 25939–25957, 2018.

[7] M. A. A. Khan, X. Lian, I. K. Mirani, and L. Tan, "Research on key technologies of electronic shelf labels based on LoRa," *J. Big Data*, vol. 3, no. 2, pp. 49–63, 2021.

[8] D. Mueller and F. Vogelsang, "Towards smart manufacturing logistics: A case study of potentials of smart label data in electronics manufacturing," *Proc. CIRP*, vol. 104, pp. 1741–1746, Jan. 2021.

[9] J. Y. Yang and S. R. Lee, "A study on introduction of IoT infrastructure based on BSC and AHP: Focusing on electronic shelf label," *J. Soc. e-Bus. Stud.*, vol. 22, no. 3, pp. 57–74, Aug. 2018.

[10] M. Agiwal and H. Jin, "Directional paging for 5G communications based on partitioned user ID," *Sensors*, vol. 18, no. 6, pp. 1845:1–1845:13, Jun. 2018.

[11] M. S. Ali, E. Hossain, and D. I. Kim, "LTE/LTE-A random access for massive machine-type communications in smart cities," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 76–83, Jan. 2017.

[12] L.-C. Kao and W. Liao, "5G intelligent A+: A pioneer multi-access edge computing solution for 5G private networks," *IEEE Commun. Standards Mag.*, vol. 5, no. 1, pp. 78–84, Mar. 2021.

[13] W. T. Toor, A. Basit, N. Maroof, S. A. Khan, and M. Saadi, "Evolution of random access process: From legacy networks to 5G and beyond," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 6, pp. e3776:1–e3776:15, Jun. 2022.

[14] R. R. Tyagi, F. Aurzada, K.-D. Lee, S. G. Kim, and M. Reisslein, "Impact of retransmission limit on preamble contention in LTE-advanced network," *IEEE Syst. J.*, vol. 9, no. 3, pp. 752–765, Sep. 2015.

[15] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 15–29, Jan. 1989.

[16] R. Ahlswede, "Identification via channels," in *Identification and Other Probabilistic Models: Rudolf Ahlswede's Lectures on Information Theory 6* (Foundations in Signal Processing, Communications and Networking), R. Ahlswede, A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm, Eds. Cham, Switzerland: Springer, 2021, pp. 3–43.

[17] J. A. Cabrera, H. Boche, C. Deppe, R. F. Schaefer, C. Scheunert, and F. H. Fitzek, "6G and the post-Shannon theory," in *Shaping Future 6G Networks: Needs, Impacts, and Technologies*. Hoboken, NJ, USA: Wiley, 2021, pp. 271–294.

[18] J. Dai, P. Zhang, K. Niu, S. Wang, Z. Si, and X. Qin, "Communication beyond transmitting bits: Semantics-guided source and channel coding," *IEEE Wireless Commun.*, early access, Aug. 8, 2022, doi: 10.1109/MWC.017.2100705.

[19] Q. Lan, D. Wen, Z. Zhang, Q. Zeng, X. Chen, P. Popovski, and K. Huang, "What is semantic communication? A view on conveying meaning in the era of machine intelligence," *J. Commun. Inf. Netw.*, vol. 6, no. 4, pp. 336–371, 2021.

[20] E. C. Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Comput. Netw.*, vol. 190, May 2021, Art. no. 107930.

[21] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assad, G. Liva, A. Munari, B. Soret, and K. H. Johansson, "Semantic communications in networked systems: A data significance perspective," *IEEE Netw.*, vol. 36, no. 4, pp. 233–240, Jul./Aug. 2022.

[22] A. Datta and F. Oggier, "Concurrency control and consistency over erasure coded data," *IEEE Access*, vol. 10, pp. 118617–118638, 2022.

[23] A. Krechowicz, S. Deniziak, and G. Lukawski, "Highly scalable distributed architecture for NoSQL datastore supporting strong consistency," *IEEE Access*, vol. 9, pp. 69027–69043, 2021.

[24] M. Pedrosa, R. Lebre, and C. Costa, "A performant protocol for distributed health records databases," *IEEE Access*, vol. 9, pp. 125930–125940, 2021.

[25] F. Barbarulo, C. Puliafito, A. Virdis, and E. Mingozzi, "Extending ETSI MEC towards stateful application relocation based on container migration," in *Proc. IEEE 23rd Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2022, pp. 367–376.

[26] T. V. Doan, G. T. Nguyen, M. Reisslein, and F. H. P. Fitzek, "FAST: Flexible and low-latency state transfer in mobile edge computing," *IEEE Access*, vol. 9, pp. 115315–115334, 2021.

[27] X. Fan, H. Xu, H. Huang, and X. Yang, "Real-time update of joint SFC and routing in software defined networks," *IEEE/ACM Trans. Netw.*, vol. 29, no. 6, pp. 2664–2677, Dec. 2021.

[28] M. A. Hathibelagal, R. G. Garroppo, and G. Nencioni, "Experimental comparison of migration strategies for MEC-assisted 5G-V2X applications," *Comput. Commun.*, vol. 197, pp. 1–11, Jan. 2023.

[29] M. Groshev, C. Guimarães, A. De La Oliva, and R. Gazda, "Dissecting the impact of information and communication technologies on digital twins as a service," *IEEE Access*, vol. 9, pp. 102862–102876, 2021.

[30] H. Laaki, Y. Miche, and K. Tammi, "Prototyping a digital twin for real time remote control over mobile networks: Application of remote surgery," *IEEE Access*, vol. 7, pp. 20325–20336, 2019.

[31] C. von Lengerke, A. Hefele, J. A. Cabrera, and F. H. P. Fitzek, "Stopping the data flood: Post-Shannon traffic reduction in digital-twins applications," in *Proc. IEEE/IFIP Netw. Oper. Manage. Symp. (NOMS)*, Apr. 2022, pp. 1–5.

[32] S. Khan, T. Arslan, and T. Ratnarajah, "Digital twin perspective of fourth industrial and healthcare revolution," *IEEE Access*, vol. 10, pp. 25732–25754, 2022.

[33] S. R. Newrzella, D. W. Franklin, and S. Haider, "Methodology for digital twin use cases: Definition, prioritization, and implementation," *IEEE Access*, vol. 10, pp. 75444–75457, 2022.

[34] Y. Qamsane, J. R. Phillips, C. Savaglio, D. Warner, S. C. James, and K. Barton, "Open process automation- and digital twin-based performance monitoring of a process manufacturing system," *IEEE Access*, vol. 10, pp. 60823–60835, 2022.

[35] J. Wen, B. Gabrys, and K. Musial, "Toward digital twin oriented modeling of complex networked systems and their dynamics: A comprehensive survey," *IEEE Access*, vol. 10, pp. 66886–66923, 2022.

[36] S. Derebeyoğlu, C. Deppe, and R. Ferrara, "Performance analysis of identification codes," *Entropy*, vol. 22, no. 10, p. 1067, Sep. 2020.

[37] R. Ferrara, L. Torres-Figueroa, H. Boche, C. Deppe, W. Labidi, U. Mönich, and A. Vlad-Costin, "Implementation and experimental evaluation of Reed–Solomon identification," in *Proc. Eur. Wireless*, 2022, pp. 1–6.

[38] M. Spandri, R. Ferrara, and C. Deppe, "Reed–Müller identification," 2021, *arXiv:2107.07649*.

[39] V. R. Sidorenko and C. Deppe, "Identification based on random coding," 2022, *arXiv:2207.03413*.

[40] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. Soc. Ind. Appl. Math.*, vol. 8, no. 2, pp. 300–304, Jun. 1960.

[41] R. S. Elagooz, A. Mahran, S. Gasser, and M. Aboul-Dahab, "Efficient low-complexity decoding of CCSDS Reed–Solomon codes based on Justesen's concatenation," *IEEE Access*, vol. 7, pp. 49596–49603, 2019.

[42] P. Liu, Z. Pan, and J. Lei, "Parameter identification of Reed–Solomon codes based on probability statistics and Galois field Fourier transform," *IEEE Access*, vol. 7, pp. 33619–33630, 2019.

[43] I. Reed, "A class of multiple-error-correcting codes and the decoding scheme," *Trans. I.R.E. Prof. Group Electron. Comput.*, vol. 4, no. 4, pp. 38–49, Sep. 1954.

[44] D. E. Müller, "Application of Boolean algebra to switching circuit design and to error detection," *Trans. I.R.E. Prof. Group Electron. Comput.*, vol. EC-3, no. 3, pp. 6–12, Sep. 1954.

[45] K. Eswaran. (2005). *Identification Via Channels and Constant-Weight Codes*. Accessed: Dec. 13, 2022. [Online]. Available: https://people.eecs.berkeley.edu/~ananth/229BSpr05/Reports/KrishEswaran.pdf

[46] O. Gunlu, J. Kliewer, R. F. Schaefer, and V. Sidorenko, "Code constructions and bounds for identification via channels," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 1486–1496, Mar. 2022.

[47] R. Ahlswede and G. Dueck, "Identification in the presence of feedback—A discovery of new capacity formulas," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 30–36, Jan. 1989.

[48] T. S. Han and S. Verdú, "New results in the theory of identification via channels," *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 14–25, Jan. 1992.
[49] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
[50] S. Verdu and V. K. Wei, "Explicit construction of optimal constant-weight codes for identification via channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 30–36, Jan. 1993.
[51] P. Moulin and R. Koetter, "A framework for the design of good watermark identification codes," *Proc. SPIE*, vol. 6072, pp. 565–574, Feb. 2006.
[52] K. Kurosawa and T. Yoshida, "Strongly universal hashing and identification codes via channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2091–2095, Sep. 1999.
[53] C. von Lengerke, A. Hefele, J. A. Cabrera, M. Reisslein, and F. H. P. Fitzek, "Beyond the bound: A new performance perspective for identification via channels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2023.
[54] V. Bioglio, C. Condo, and I. Land, "Design of polar codes in 5G new radio," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 29–40, 1st Quart., 2020.
[55] Z. B. K. Egilmez, L. Xiang, R. G. Maunder, and L. Hanzo, "The development, operation and performance of the 5G polar codes," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 96–122, 1st Quart., 2020.
[56] L. Linguaglossa, S. Lange, S. Pontarelli, G. Rétvári, D. Rossi, T. Zinner, R. Bifulco, M. Jarschel, and G. Bianchi, "Survey of performance acceleration techniques for network function virtualization," *Proc. IEEE*, vol. 107, no. 4, pp. 746–764, Apr. 2019.
[57] G. S. Niemiec, L. M. S. Batista, A. E. Schaeffer-Filho, and G. L. Nazar, "A survey on FPGA support for the feasible execution of virtualized network functions," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 504–525, 1st Quart., 2020.
[58] P. Shantharama, A. S. Thyagaturu, and M. Reisslein, "Hardware-accelerated platforms and infrastructures for network functions: A survey of enabling technologies and research studies," *IEEE Access*, vol. 8, pp. 132021–132085, 2020.

CASPAR VON LENGERKE received the bachelor's and master's degrees in electrical engineering from RWTH Aachen University, Germany, in 2017 and 2019, respectively. He joined the Deutsche Telekom Chair of Communication Networks, TU Dresden, in 2021. His research interests include goal-oriented communication, including message identification and common randomness generation.

ALEXANDER HEFELE received the Diploma (Dipl.-Ing.) degree in electrical and computer engineering from the Technical University Dresden, Germany, in 2022. He is currently pursuing the Ph.D. degree with the Deutsche Telekom Chair of Communication Networks. His research interests include semantic and goal-oriented communication with regard to its application in distributed robotic systems, digital twins, and cyber-physical systems.

JUAN A. CABRERA received the Dr.-Ing. degree from TU Dresden, Germany, in 2022. He works at the Deutsche Telekom Chair of Communication Networks, TU Dresden, where he leads the Research Group on semantic and goal-oriented communications. His research interests include semantic and goal-oriented communications, functional compression, message identification, common randomness generation, network coding, and in-network distributed storage and computing.

OLIVER KOSUT (Senior Member, IEEE) received the B.S. degree in electrical engineering and mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2004, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2010. Since 2012, he has been a Faculty Member at the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, USA, where he is currently an Associate Professor. Previously, he was a Postdoctoral Research Associate at the Laboratory for Information and Decision Systems, MIT, from 2010 to 2012. His research interests include information theory, particularly with applications to security and machine learning and power systems. He received the NSF CAREER Award in 2015. He is an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

MARTIN REISSLEIN (Fellow, IEEE) received the Ph.D. degree in systems engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 1998. He is currently a Professor with the School of Electrical, Computer, and Energy Engineering, Arizona State University (ASU), Tempe, AZ, USA. He is also an Associate Editor of IEEE ACCESS, IEEE TRANSACTIONS ON EDUCATION, IEEE TRANSACTIONS ON MOBILE COMPUTING, and IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT.

FRANK H. P. FITZEK (Senior Member, IEEE) received the Ph.D. (Dr.-Ing.) degree in electrical engineering from the Technical University Berlin, Germany, in 2002. He is currently a Professor and the Head of the Deutsche Telekom Chair of Communication Networks, TU Dresden. He is also the Spokesman of the DFG Cluster of Excellence CeTI and the 6G-life Hub in Germany. He became an Adjunct Professor at the University of Ferrara, Italy, in 2002. In 2003, he joined Aalborg University as an Associate Professor and later became a Professor. His current research interests include 5G/6G communication networks, in-network computing, network coding, compressed sensing, post-Shannon theory, quantum, molecular communication, and human–machine interaction in the virtual worlds.