entropy

MDPI

Article

Optimizing Finite-Blocklength Nested Linear Secrecy Codes: Using the Worst Code to Find the Best Code

Morteza Shoushtari * and Willie Harrison

Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602, USA; willie.harrison@byu.edu

* Correspondence: morteza.shoushtari@byu.edu

Abstract: Nested linear coding is a widely used technique in wireless communication systems for improving both security and reliability. Some parameters, such as the relative generalized Hamming weight and the relative dimension/length profile, can be used to characterize the performance of nested linear codes. In addition, the rank properties of generator and parity-check matrices can also precisely characterize their security performance. Despite this, finding optimal nested linear secrecy codes remains a challenge in the finite-blocklength regime, often requiring brute-force search methods. This paper investigates the properties of nested linear codes, introduces a new representation of the relative generalized Hamming weight, and proposes a novel method for finding the best nested linear secrecy code for the binary erasure wiretap channel by working from the worst nested linear secrecy code in the dual space. We demonstrate that our algorithm significantly outperforms the brute-force technique in terms of speed and efficiency.

Keywords: wiretap channel; generalized Hamming weights; dimension/length profile; nested linear codes; equivocation; optimal secrecy code; two-edge LDPC codes; dual codes

check for updates

Citation: Shoushtari, M.; Harrison, W. Optimizing Finite-Blocklength Nested Linear Secrecy Codes: Using the Worst Code to Find the Best Code. *Entropy* **2023**, *25*, 1456. https://doi.org/10.3390/e25101456

Academic Editors: Predrag Ivanis and Goran Djordjević

Received: 5 September 2023
Revised: 3 October 2023
Accepted: 14 October 2023
Published: 17 October 2023

1. Introduction

The wiretap channel introduced by Wyner in [1] and later generalized by Csiszár and Körner in [2] is the most fundamental channel model that has been used to study broadcast security problems in the context of information theory. One version of this channel model is depicted in Figure 1, where the confidential communication occurring over a discrete memoryless main channel is observed by an eavesdropper who has access to a noisy version of the channel input. Later, in [3], Ozarow and Wyner introduced the wiretap channel type II, wherein the eavesdropper is able to select the positions of revealed bits, and they provided a secure coding technique based on coset codes. These channel models have been studied by many authors from the perspectives of security, reliability, and coding construction [4–6].

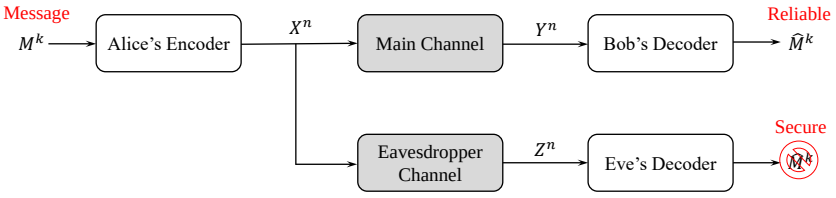

Figure 1. The wiretap channel model.

In recent years, coset coding has emerged as an important coding technique in the context of the finite-blocklength regime [7–9]. In this regime, the design of the code plays a critical role in achieving high communication rates while balancing the tradeoff between complexity and performance. The effectiveness of coset coding in the finite-blocklength

regime has been demonstrated in a wide range of applications, including wiretap channels, broadcast channels, and multiple-access channels [10–14].

The *nested linear code* construction was first presented in [15] to generate a diluted version of the original coset code. Later, in [16,17], the authors proposed a secure error-correcting code based on the nested code construction for the wiretap channel type II, and in [18] the authors considered nested codes based on low-density parity-check (LDPC) codes for the original wiretap channel when both the eavesdropper and main channels are binary erasure channels (BECs).

Generalized Hamming weights (GHW) and the dimension/length profile (DLP), which were first introduced in [19,20], respectively, were two of the first parameters of linear block codes that could be used to characterize the performance of the original linear coset codes, especially over a wiretap channel of type II. Numerous papers have investigated these parameters on various linear codes [21–26]. Later, the authors of [16,27] extended these two parameters to nested coding constructions and defined two new formats for them: the relative generalized Hamming weight (RGHW) and relative dimension/length profile (RDLP), which can be used to characterize the security and error-correction performance of nested linear codes. Further studies have shown that with the rank properties of generator and parity-check matrices, the performance of linear codes can also be precisely characterized [28–30]. In [29], we utilized rank properties to create and develop a tool for analyzing finite-blocklength wiretap codes based on coset coding over erasure channels, known as an equivocation matrix.

Designing the most secure nested linear codes, referred to as nested linear secrecy codes, to achieve optimal performance in the-finite blocklength regime is still a challenging task, and there is currently no single solution for creating the best codes in this scenario. Identifying these optimal codes would facilitate a comparison of the tradeoffs between complexity and performance for different codes, providing a benchmark for the optimality of other wiretap code designs.

1.1. Our Contributions

This paper explores the characteristics of nested linear codes in a wiretap channel model where both the main and eavesdropper channels are BECs. A novel approach is proposed to find the optimal nested linear secrecy codes by using a dual relationship between nested linear codes and their dual codes. Essentially, we demonstrate that instead of searching for the best code directly, it can be found by starting with the worst nested linear secrecy code from the dual space, which is easy to identify. The results demonstrate an efficient and fast technique for finding optimal nested linear secrecy codes.

The main contributions of this work can be summarized as follows:

1. New representation of RGHW: We introduce a new representation of the relative generalized Hamming weight (RGHW) by analyzing the rank properties of parity-check matrices. This innovative approach enables us to accurately predict the security performance of nested linear codes based on rank properties.
2. Equivalence condition evaluation: A comprehensive evaluation of the equivalence codes for nested linear codes is conducted, along with an exploration of its associated properties.
3. Exploration of equivocation curves: We explore and evaluate the equivocation curves of nested linear codes. Notably, we discover that these curves can exhibit both convex and concave characteristics simultaneously, a novel observation in the field. This discovery presents an exciting opportunity to concentrate on codes that effectively balance both secrecy and reliability constraints.
4. Efficient algorithm for best code identification: The main contribution of this paper lies in the development of an algorithm that efficiently identifies the best nested linear secrecy codes. This algorithm surpasses conventional methods in terms of speed and effectiveness.

These contributions not only enrich our understanding of nested linear codes but also enhance their design and deployment for diverse applications.

1.2. Organization of the Paper

The rest of the paper is organized as follows. Section 2 consists of preliminary details about the channel model and nested linear coding structure. Section 3 introduces a new expression of the generalized Hamming weight that can be used to quantify the performance of nested linear codes. We also present some properties of nested linear codes in this section. Section 4 describes the behavior of equivocation curves of the nested linear codes. The novel algorithm for finding the best nested linear secrecy codes is explained in Section 5. Section 6 presents a numerical example. In Section 7, we evaluate the complexity of our proposed algorithm. Finally, in Section 8, we present our conclusion.

2. Preliminaries

2.1. Notation

In this paper, capital letters represent random variables and matrices, lowercase letters represent realizations of these random variables, and calligraphic letters indicate the discrete alphabets associated with the random variables. The distributions $p(x)$ and $p(y|x)$ are probability mass functions. The length of vectors is denoted by superscripts, and sets used as subscripts on matrices specify sub-matrices that include only the columns indexed in the set, i.e., $(H_1)_U$ is the sub-matrix of H_1 made up of only the columns with indices in the set U. All vectors are row vectors, and all codes are binary. The notation $[\![1, \gamma]\!]$ represents a series of integers ranging from 1 to γ, where $\gamma \geq 1$. The set $\mathcal{R}^n$ represents all possible revealed-bit patterns over n transmitted bits by containing all subsets of $[\![1, n]\!]$, whereas the set $J \backslash r$ indicates the set difference operation and is often read as J *delete* r.

2.2. Channel Models

Consider the wiretap channel model in Figure 1. The channels between Alice and Bob and Alice and Eve can be any discrete memoryless channels, but for the purposes of this work, we assume that both channels are BECs, with erasure probabilities of ϵ_m and ϵ_e for the main and eavesdropper channels, respectively. In this model, Alice wants to transmit a secret message M^k, which is assumed to be chosen uniformly at random from the alphabet $\mathcal{M} = \mathbb{F}_2^k$, to Bob through the main channel and wishes to keep it secret as much as possible from a passive eavesdropper (Eve). To achieve this, Alice converts M^k into an n-bit binary codeword X^n. The encoding is an invertible one-to-many mapping. This means that no more than one message can be mapped into the same codeword, but each message can be encoded to one of several possible codewords. Bob and Eve observe a noisy version of the transmitted codeword X^n through each of their channels, which are denoted by Y^n and Z^n, respectively. Thus, $\mathcal{Y}^n = \mathcal{Z}^n = \{0, 1, ?\}^n$.

There are two main constraints when utilizing coding over this type of wiretap channel.

1. Reliability constraint for Bob: $\Pr(M \neq \hat{M}) < \delta_r$;
2. Security constraint for Eve: $\mathbb{I}(M; Z^n) < \delta_s$.

Here, δ_r and δ_s are the desired secrecy and reliability levels, respectively, which can be defined by the system designer. Concisely, the encoding function that maps secret message M^k to codeword X^n should be such that Bob can decode M^k from Y^n reliably, and at the same time, Eve receives as little information as possible about M^k from Z^n. The level of secrecy achieved by a code can be quantified by either the average equivocation

$$\mathbb{H}(M^k|Z^n) = \sum_{z \in \mathcal{Z}} p(z^n)\mathbb{H}(M^k|Z^n = z^n), \tag{1}$$

or the average leakage

$$\mathbb{I}(M^k; Z^n) = \mathbb{H}(M^k) - \mathbb{H}(M^k|Z^n). \tag{2}$$

Both of these information-theoretic functions are used to evaluate the performance of a wiretap code in terms of its secrecy. As a result, in this scenario, it is preferable to minimize the average leakage $\mathbb{I}(M^k; Z^n)$ or maximize the average equivocation $\mathbb{H}(M^k|Z^n)$, while also enhancing Bob's error-correction capabilities, which can be achieved by reducing $\mathbb{H}(M^k|Y^n)$.

2.3. Nested Linear Codes

The fundamental concept behind the nested linear coding approach is to partition the main code into sub-codes and employ $n-k$ overhead bits to aid in secrecy or reliability as desired. The information rate between Alice and Bob is $R = k/n$. Let the number of overhead bits assigned to reliability and secrecy be α and l, respectively, and let

$$n = k + \alpha + l. \tag{3}$$

Let $\mathcal{C}_0$ be an $(n, k+l)$ linear block code and $\mathcal{C}_1$ be an (n, l) linear block code. Then, the nested linear code $(\mathcal{C}_0, \mathcal{C}_1)$ is defined, where $\mathcal{C}_0$ is a *fine code* with rate R_0 and $\mathcal{C}_1$ is a *coarse code* with rate R_1, where $R_1 \leq R_0$, satisfying

$$\mathcal{C}_1 \subseteq \mathcal{C}_0, \tag{4}$$

which means that each codeword of $\mathcal{C}_1$ is also a codeword of $\mathcal{C}_0$. Let the $l \times n$ matrix G_1 be the generator matrix, and the $(n-l) \times n$ matrix H_1 be the parity-check matrix for $\mathcal{C}_1$. The generator matrix G_0 is defined as follows:

$$G_0 = \begin{bmatrix} G' \\ \cdots \\ G_1 \end{bmatrix}, \tag{5}$$

where G' is comprised of k linearly independent rows from $\mathbb{F}_2^n$ that are not in $\mathcal{C}_1$ and make G_0 a full-rank matrix. The parity-check matrix H_1 also consists of two sub-matrices such that

$$H_1 = \begin{bmatrix} H' \\ \cdots \\ H_0 \end{bmatrix}, \tag{6}$$

where H_0 is $\alpha \times n$ and forms a basis for the dual space of the rowspace of G_0. The dimension of the sub-matrix H' is $k \times n$. It is important to note that according to the algebraic properties of nested linear codes, $G_0(H_0)^T = 0$ and $G_1(H_1)^T = 0$.

The encoding process begins by selecting an auxiliary message m' uniformly at random from $\mathbb{F}_2^l$ and then computing

$$x^n = \begin{bmatrix} m & m' \end{bmatrix} G_0 = \begin{bmatrix} m & m' \end{bmatrix} \begin{bmatrix} G' \\ G_1 \end{bmatrix} \tag{7}$$

$$= mG' \oplus m'G_1, \tag{8}$$

where m is a k-bit secret message. Now, the fine code $\mathcal{C}_0$ is randomly partitioned into 2^k disjoint subsets (cosets). The term mG' selects the coset, and the term $m'G_1$ selects the specific codeword from the corresponding coset at random.

Bob uses the following decoding approach to retrieve M^k from Y^n. First, Bob recovers as many erased bits as possible using the parity-check matrix H' and obtains an estimated version $\hat{X}^n$ of X^n [31]. Assuming $\hat{X}^n = X^n$, then Bob's decoder computes the syndrome S of $\hat{X}^n$ as

$$s = \hat{x}(H_0)^T = mG'(H_0)^T \oplus m'G_1(H_0)^T \tag{9}$$

$$= mG'(H_0)^T. \tag{10}$$

It is possible to choose matrices such that $G'(H_0)^T$ is the $k \times k$ identity; therefore, $s = m$ [32].

To achieve reliability and/or security, both codes $\mathcal{C}_0$ and $\mathcal{C}_1$ need to meet specific requirements. In this case, the fine code is primarily responsible for ensuring reliability, while the coarse code is utilized for security purposes. The following section will explore different properties of nested linear codes and examine several parameters that measure the performance of such codes.

3. Performance Parameters

This section explores practical metrics to measure the performance of nested linear codes and examines their properties. Consider the $(n, n-l)$ and $(n, n-k-l)$ dual codes of $\mathcal{C}_1$ and $\mathcal{C}_0$ and call them $\mathcal{C}_1^\perp$ and $\mathcal{C}_0^\perp$, respectively. $\mathcal{C}_1^\perp$ uses H_1 as the generator matrix and G_1 as the parity-check matrix. Hence, the nested linear code $(\mathcal{C}_1^\perp, \mathcal{C}_0^\perp)$ is the dual code of $(\mathcal{C}_0, \mathcal{C}_1)$. In the dual space, $\mathcal{C}_1^\perp$ serves as the fine code, and $\mathcal{C}_0^\perp$ is the coarse code. The information rate of the nested linear code in both spaces will not change and remains k/n. However, the secrecy and reliability overhead bits will change in the different spaces. In the dual space, α and l represent the number of security and reliability bits, respectively [30].

3.1. RGHW and RDLP

As previously stated, RGHW and RDLP are extended versions of the GHW and DLP, which can be utilized to characterize the security performance of nested linear codes over the wiretap channel of type II. Let J be a subset of $[\![1, n]\!]$. A new representation for the RGHW of the nested linear codes can be given as follows.

Proposition 1. *The τth relative generalized Hamming weight of the nested code $(\mathcal{C}_0, \mathcal{C}_1)$ can be written as*

$$M_\tau = \min_{1 \le \tau \le R_0 - R_1} \{|J| : \operatorname{rank}((H_1)_J) - \operatorname{rank}((H_0)_J) \ge \tau\} \tag{11}$$

$$= n - \max\{|r(z^n)| : \log_2 N_0[r(z^n)] - \log_2 N_1[r(z^n)] \ge \tau\}, \tag{12}$$

where $r(z^n)$ is a revealed-bit pattern over the erasure channel and $N_0[r(z^n)]$ and $N_1[r(z^n)]$ are the number of codewords in $\mathcal{C}_0$ and $\mathcal{C}_1$, respectively, that have zeros for all bit locations in the indexed set $r(z^n)$.

Proof. In [33], we showed that

$$\mathbb{H}(M|Z^n = z^n) = \log_2 N_0[r(z^n)] - \log_2 N_1[r(z^n)], \tag{13}$$

Since $|r(z^n)|$ represents the maximum number of bits that can be revealed while still maintaining at least τ bits of equivocation, the total number of bits minus the maximum revealed bits must equal the minimum number of bits that must be leaked to reveal at least τ bits of information, and the expression (12) is valid. Thus, Equations (11) and (12) represent two equivalent expressions for the τth relative generalized Hamming weight of the nested linear code. □

3.2. Rank Properties and the Equivocation Matrix

Let $r(z^n) = \{i : z_i \neq ?\}$, where z_i is the observation of the ith bit of the codeword x over the eavesdropper's BEC and "?" denotes an erased bit. Also, let $I = [\![1, n]\!]$. According to the results of [30,33], we showed that the exact equivocation for the observation z^n over a binary erasure channel (BEC), given the coding scheme presented in Section 2.3, is

$$\mathbb{H}(M|Z^n = z^n) = k - \operatorname{rank}\left[(G_0)_{r(z^n)}\right] + \operatorname{rank}\left[(G_1)_{r(z^n)}\right] \tag{14}$$

$$= \operatorname{rank}\left[(H_1)_{I \setminus r(z^n)}\right] - \operatorname{rank}\left[(H_0)_{I \setminus r(z^n)}\right]. \tag{15}$$

Thus, in terms of code design for security and reliability, a revealed-bit pattern $r(z^n)$ is secure if and only if $\operatorname{rank}((G_0)_{r(z^n)}) = \operatorname{rank}((G_1)_{r(z^n)})$. Furthermore, for reliability,

the message information is obtained if and only if $\text{rank}((G_0)_{r(z^n)}) - \text{rank}((G_1)_{r(z^n)}) = k$. The following definition is from [30].

Definition 1. *The $(k+1) \times (n+1)$ equivocation matrix A for the linear block code $\mathcal{C}$ is a matrix where each entry $(a_{e,\mu})$ counts the number of revealed-bit patterns of size μ that maintain e bits of equivocation.*

There are $\binom{n}{\mu}$ different patterns that can be used to reveal μ bits of n transmitted codeword bits over the erasure channel, and the bottom left entry of A is $a_{0,0}$.

3.3. Equivalence of Nested Linear Codes

Lemma 1. *Let $(\mathcal{C}_0^*, \mathcal{C}_1^*)$ and $(\mathcal{C}_0^{\circledast}, \mathcal{C}_1^{\circledast})$ be two nested linear codes with generator matrices G_0^* and $G_0^{\circledast}$, respectively. These two nested linear codes are equivalent if there exist two invertible scrambling matrices F_1 and F_2 and permutation matrix P, such that*

$$G_0^{\circledast} = \left[\begin{array}{c|c} F_1 & \underline{0} \\ \hline \underline{0} & F_2 \end{array}\right] \times G_0^* \times P, \tag{16}$$

where F_1 and F_2 are $k \times k$ and $l \times l$ full-rank matrices, respectively, and $\underline{0}$ is a zero matrix.

Note that, in general, codes are equivalent if the sets of codewords are the same up to the permutation of bit order in the codewords.

Proof. We know that the space spanned by the rows of G_1^* is the same as the space spanned by the rows of $F_2 G_1^*$ (and similarly for the space spanned by the rows of G'^* and $F_1 G'^*$). The multiplication by P changes only the order of bits in codewords and the mapping of specific messages to specific codewords but achieves equivalence. □

Lemma 2. *If generator matrices G_0^* and $G_0^{\circledast}$ correspond to respective equivalent nested codes $(\mathcal{C}_0^*, \mathcal{C}_1^*)$ and $(\mathcal{C}_0^{\circledast}, \mathcal{C}_1^{\circledast})$, then the RGHW, RDLP, and equivocation matrices for the two codes are identical.*

Proof. According to Lemma 1, two nested linear codes $(\mathcal{C}_0^*, \mathcal{C}_1^*)$ and $(\mathcal{C}_0^{\circledast}, \mathcal{C}_1^{\circledast})$ are equivalent if G_0^* can be converted into $G_0^{\circledast}$ using simple linear operations over rows and/or column pivots. These basic operations produce the same set of codewords from the new generator matrices up to a consistent bit permutation in the codeword sets. Thus, (11) and (12) are the same for both codes for all τ, and the equivalence for the RDLP is similarly trivial. For the equivalence of equivocation matrices, every $r(z^n)$ for code $(\mathcal{C}_0^*, \mathcal{C}_1^*)$ maps to a unique revealed-bit pattern for $(\mathcal{C}_0^{\circledast}, \mathcal{C}_1^{\circledast})$ of the same size such that (14) is equivalent. □

Previous research, including [16,17], has analyzed the bounds on the RGHW and RDLP of nested linear codes $(\mathcal{C}_0, \mathcal{C}_1)$ to aid in constructing nested linear secrecy codes. Furthermore, studies such as [29,30] enable comparisons between the performance of nested linear codes on specific sizes. Even with these results, the challenge of finding optimal nested linear secrecy codes remains unsolved and requires a brute-force search.

4. Concavity and Convexity of Equivocation Curves

The equivocation quantifies Bob and Eves' uncertainty about the secret message M^k after observing Y^n and Z^n, respectively. We may want to maximize the equivocation for security constraints or minimize it for reliability limitations, depending on the system requirements. In the noiseless main channel model where reliability constraints are not considered and all overhead bits are allocated for security purposes, the equivocation of the nested linear code $(\mathcal{C}_0, \mathcal{C}_1)$ is always a concave function of ϵ [32]. However, when both reliability and security are important, such as in the case of a noisy main channel, our simulation results indicate a different behavior compared to the noiseless main channel case.

Lemma 3. *Consider the nested linear code $(\mathcal{C}_0, \mathcal{C}_1)$ of rate R_0 and R_1, respectively. Assume that this pair of linear codes is used to transmit a k-bit message m over the binary erasure wiretap channel. The equivocation curve that can be achieved by nested coding construction can be concave, convex, or both as a function of ϵ.*

The proof follows directly from Theorem 2.7.4 [34] and is included here for completeness.

Proof. $\mathbb{H}(M)$ is a concave function of $p(m)$, and

$$\mathbb{I}(M^k; Z^n) = \mathbb{I}(Z^n; M^k) = \mathbb{H}(Z^n) - \mathbb{H}(Z^n|M^k) \tag{17}$$

$$= \mathbb{H}(Z^n) - \sum_m p(m)\mathbb{H}(Z^n|M^k = m). \tag{18}$$

If $p(z|m)$ is fixed, then $p(z)$ is a linear function of $p(m)$; hence, $\mathbb{H}(Z^n)$ is also a concave function of $p(m)$, and the second term of (17) is a linear function of $p(m)$. The difference is then a concave function of $p(m)$. Moreover, the conditional entropy $\mathbb{H}(Z^n|M^k)$ of $p(z|m)$ for a fixed $p(m)$ will be concave, and the difference of two concave functions can either be concave, convex, or both. □

Simulation results show that there are indeed three distinct equivocation curve behaviors for nested linear codes $(\mathcal{C}_0, \mathcal{C}_1)$, as follows:

- Convex equivocation curve: These codes are appropriate for situations when δ_r is small; thus, Alice may purposefully use a nested linear code of this nature to improve Bob's ability to correct errors.
- Concave equivocation curve: If δ_s is small, these codes give Alice the ability to keep data as secure as possible from the eavesdropper.
- Convex/concave equivocation curve: The more desirable and interesting codes are those that provide both reliability for Bob and confusion for Eve, in scenarios where Bob and Eve experience erasure with different rates. These codes can effectively balance both constraints as required, resulting in a convex/concave equivocation curve.

Simulations of (1) were completed using (14) and considering all possible erasure patterns $r(z^n)$. Curves were plotted as a function of the erasure probability ϵ, noting that $p(z^n) = \epsilon^{n-|r(z^n)|}(1-\epsilon)^{|r(z^n)|}$. This examination of the behavior of equivocation curves enhanced our comprehension of nested linear codes, unveiling the dual relationship between error correction capabilities and security attributes. Furthermore, our simulations demonstrated that the number of overhead bits allocated to security or reliability can have a significant impact on the shape and number of the equivocation curves. In particular, increasing the number of overhead bits allocated to security (l) can lead to an increase in the number of concave curves. This observation is consistent with the fact that adding more security overhead bits to the code will result in a higher level of confusion for the eavesdropper. Similarly, increasing the number of overhead bits allocated to reliability (α) can lead to an increase in the number of convex curves and affect their shape, as more reliability overhead bits will provide Bob with better error correction capabilities. Overall, our results highlight the importance of carefully balancing the allocation of overhead bits between security and reliability to achieve the desired level of secrecy and reliability for the system. Additionally, we showed that there exist codes that can balance both restrictions effectively (codes with convex/concave equivocation curves).

These types of codes are represented, respectively, with red, green, and blue equivocation curves in Figure 2 for the $n = 5$, $k = 2$, $l = 2$, and $\alpha = 1$ case, and in Figure 3 for the corresponding dual case, where l and α change their responsibilities, which means that the number of overhead bits allocated to security will be $\alpha = 2$, and the number of overhead bits allocated to reliability will be $l = 1$. This change in the allocation of overhead bits results in a different set of equivocation curves. The probability of erasure, ϵ, refers to both ϵ_m and ϵ_e to show performance for all users on the same plot.

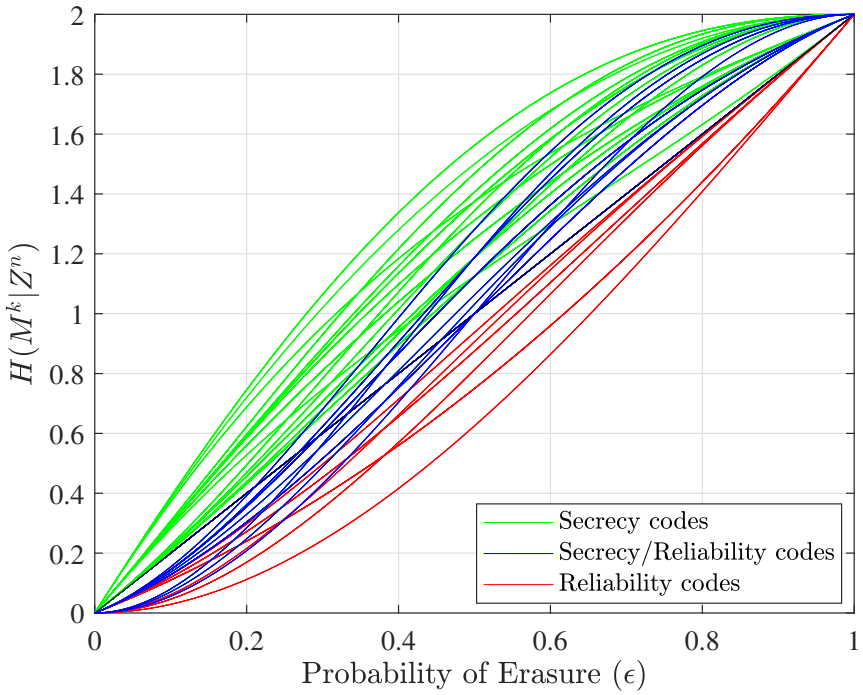

Figure 2. Equivocation curves of all nested linear codes versus ϵ (the nested linear codes designed for $n = 5$, $k = 2$, $l = 2$, and $\alpha = 1$). The green pair of codes are good for security purposes, and the red pair of codes are suitable for reliability.

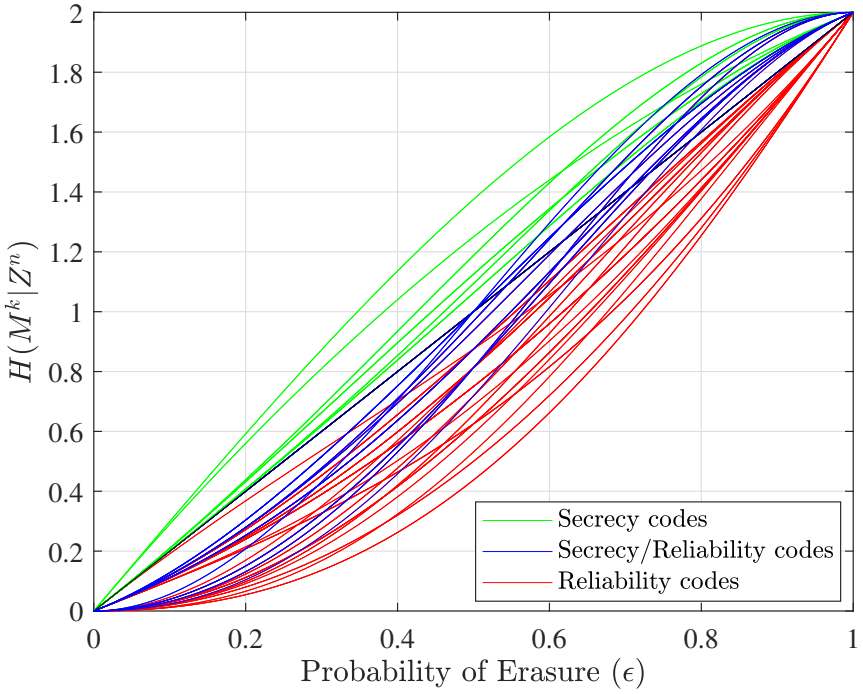

Figure 3. Equivocation curves of all nested linear codes when $n = 5$, $k = 2$, $l = 1$, and $\alpha = 2$ (referring to the dual nested linear codes).

5. Finding the Best Nested Linear Secrecy Codes

In this section, we propose a coding construction algorithm to generate the best nested linear secrecy code according to the equivocation by taking advantage of the dual relationship between nested linear codes. In essence, we show that the difficult search for the best code can be computed instead by the easy search for the worst nested linear secrecy

code in the dual space. The concept of the *worst* and *best* refers to nested linear codes with the lowest and highest security level, respectively, among all possible nested linear codes for a particular size. The general algorithm is provided here, and an example is given in Section 6.

Algorithm 1. *This algorithm demonstrates how to construct the best nested linear secrecy codes* $(\mathcal{C}_0, \mathcal{C}_1)$ *through the construction of the worst nested linear secrecy codes* $(\mathcal{C}_1^{\perp}, \mathcal{C}_0^{\perp})$ *using the subsequent steps:*

- *The first phase:*
  - *Generate the worst secrecy code* $(\mathcal{C}_0^{\perp}(n, \alpha))$ *with generator matrix* H_0*. The general schematic of the worst* H_0 *can be found in (30) in Section 6.*
  - *Generate* H' *by searching k random vectors from* $\mathbb{F}_2^n$ *with the following considerations:*
    - *For most patterns of revealed bits* $r(z^n)$*, the rank of* $(H_1)_{r(z^n)}$ *should be as large as possible compared to the rank of* $(H_0)_{r(z^n)}$*.*
    - H_1 *should remain a full-rank matrix.*
- *The second phase:*
  - *The best generator matrix* G_1 *for security code* $\mathcal{C}_1(n, l)$ *is equal to the basis of the dual space of the rowspace of* H_1*.*
  - *Choose k rows from a basis of the dual space of* H_0 *as* G'*, with a consideration of the following:*
    - *For most patterns of* $r(z^n)$*, the rank of* $(G_0)_{r(z^n)}$ *should be equal to the rank of* $(G_1)_{r(z^n)}$*.*
    - G_0 *remains a full-rank matrix.*

Proof. According to our result in [30], it can be deduced that minimizing the equivocation in the dual space of the nested linear codes leads to the maximization of equivocation in the original space of the nested linear code.

In particular, we have:

$$\mathbb{H}(M) = \underbrace{\operatorname{rank}\left[(G_0)_{r(z^n)}\right] + \operatorname{rank}\left[(G_1)_{r(z^n)}\right]}_{\mathbb{I}(M^k; Z^n = r(z^n))} + \tag{19}$$

$$\underbrace{\operatorname{rank}\left[(H_1)_{I\setminus r(z^n)}\right] + \operatorname{rank}\left[(H_0)_{I\setminus r(z^n)}\right]}_{\mathbb{H}(M^k | Z^n = r(z^n))}, \tag{20}$$

or, equivalently,

$$\mathbb{H}(M) = \underbrace{\operatorname{rank}\left[(G_0)_{r(z^n)}\right] + \operatorname{rank}\left[(G_1)_{r(z^n)}\right]}_{\mathbb{H}(M^k | Z^n = I\setminus r(z^n))} + \tag{21}$$

$$\underbrace{\operatorname{rank}\left[(H_1)_{I\setminus r(z^n)}\right] + \operatorname{rank}\left[(H_0)_{I\setminus r(z^n)}\right]}_{\mathbb{I}(M^k; Z^n = I\setminus r(z^n))}. \tag{22}$$

Therefore, by constructing the nested linear code that minimizes the equivocation in the dual space, we can generate the best nested linear secrecy code within the code space. □

This algorithm can be better understood with reference to an example; hence, the results of this algorithm for a specific size are shown in Figure 4, and the details of this example are given in the next section.

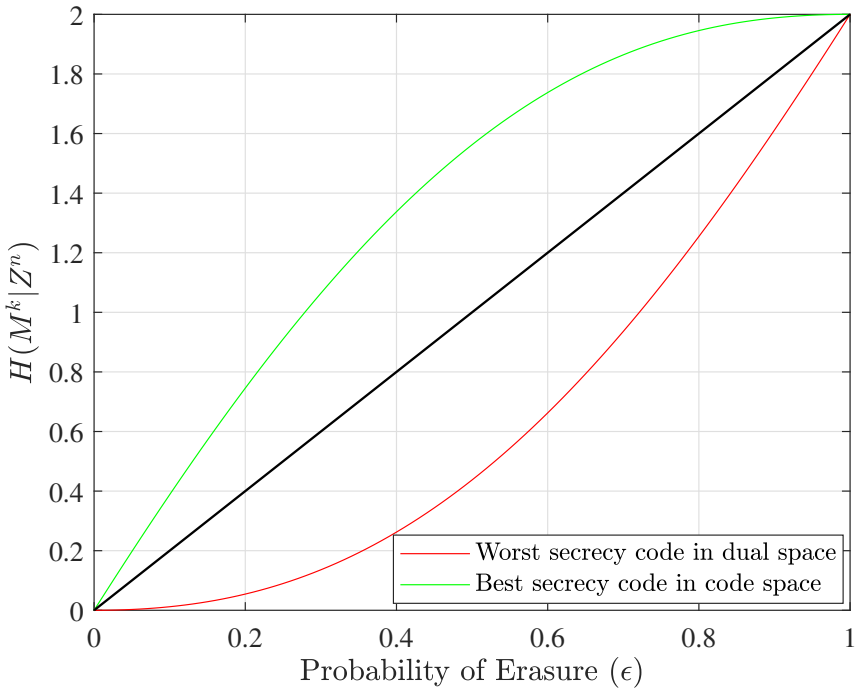

Figure 4. The best nested linear secrecy code when $n = 5$, $k = 2$, $l = 2$, and $\alpha = 1$, along with the worst nested linear code for secrecy in the dual space.

6. Numerical Example

Consider the nested code $(\mathcal{C}_0, \mathcal{C}_1)$ with the rate $R_0 = 4/5$ and $R_1 = 2/5$, respectively. In other words, $n = 5$, $k = 2$, $l = 2$, and $\alpha = 1$. In this example, the information rate is equal to $R = 2/5$. Let

$$G_0 = \left[\begin{array}{c} G' \\ \hdashline G_1 \end{array}\right] = \left[\begin{array}{ccccc} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ \hdashline 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{array}\right], \tag{23}$$

and

$$H_1 = \left[\begin{array}{c} H' \\ \hdashline H_0 \end{array}\right] = \left[\begin{array}{ccccc} 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ \hdashline 0 & 0 & 1 & 0 & 0 \end{array}\right]. \tag{24}$$

The RGHW and equivocation matrix for this nested code are equal to

$$M_\tau(\mathcal{C}_0, \mathcal{C}_1) = \{1,\ 2\} \tag{25}$$

$$A = \begin{bmatrix} 1 & 5 & 9 & 5 & 0 & 0 \\ 0 & 0 & 1 & 5 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \tag{26}$$

and for the dual nested code are equal to

$$M_\tau(\mathcal{C}_1^\perp, \mathcal{C}_0^\perp) = \{2,\ 4\} \tag{27}$$

$$A^\perp = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 4 & 5 & 1 & 0 & 0 \\ 0 & 0 & 5 & 9 & 5 & 1 \end{bmatrix}. \tag{28}$$

Figure 2 illustrates the equivocation curves of all unique nested linear codes in this specific size. It should be noted that the number of green equivocation curves is greater than the number of red equivocation curves because in this example we assume that two

of the three overhead bits are assigned to security ($l = 2$) and just one bit is assigned to reliability ($\alpha = 1$). If we evaluate nested linear codes in the dual space when $l = 1$ and $\alpha = 2$, we can see that the number of equivocation curves for nested codes that provide an error-correction capability is greater than the number for nested codes that offer a security capability (green equivocation curves). Figure 3 shows the equivocation curves for dual nested linear codes.

We now aim to determine how we can identify the optimal nested linear secrecy code $(\mathcal{C}_0, \mathcal{C}_1)$ at this specific size using the algorithm outlined in Algorithm 1. To start, we need to build the generator matrix H_0 for code $\mathcal{C}_0^{\perp}$ with the worst rank properties. We know that H_0 must be a full-rank matrix with the most zero columns, which results in a zero rank in most collections of columns. Hence, H_0 is

$$H_0 = [1\,0\,0\,0\,0]. \tag{29}$$

In general,

$$H_0 = [V_{\alpha\times\alpha}\ \underline{0}_{\alpha\times n-\alpha}], \tag{30}$$

where V and $\underline{0}$ are the identity and zero matrices, respectively. Then, we need to generate H' by searching k random vectors from $\mathbb{F}_2^n$, with the consideration of the specific criteria as mentioned in Algorithm 1. Let

$$H_1 = \begin{bmatrix} H' \\ \hdashline H_0 \end{bmatrix} = \left[\begin{array}{ccccc} 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ \hdashline 1 & 0 & 0 & 0 & 0 \end{array}\right]. \tag{31}$$

In the second phase, we can generate G_1 from the dual space of H_1, which is equal to

$$G_1 = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}. \tag{32}$$

Now, we need to select k rows from the basis of the dual space of H_0 as follows:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{33}$$

Additionally, we must be sure that G_0 is a full-rank matrix and

$$\text{rank}((G_0)_{r(z^n)}) = \text{rank}((G_1)_{r(z^n)}) \tag{34}$$

as much as possible for most patterns of $r(z^n)$, so

$$G_0 = \begin{bmatrix} G' \\ \hdashline G_1 \end{bmatrix} = \left[\begin{array}{ccccc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \hdashline 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{array}\right]. \tag{35}$$

The generator matrix G_0 outperforms other generator matrices in terms of security performance. The equivocation matrix of this example is equal to (26), and Figure 4 depicts the equivocation curves of the worst nested linear secrecy codes in the dual space and the best nested linear secrecy codes in the code space.

In the following section, we will analyze the computational complexity of our proposed algorithm for finding the optimal nested linear secrecy code and compare it with the computational complexity of traditional approaches (the brute-force method).

7. Computational Complexity Analysis

The number of distinct generator matrices G_0 that can be chosen such that the resulting matrix is full-rank can be calculated as

$$\prod_{i=0}^{(n-\alpha)-1} (2^n - 2^i), \tag{36}$$

where n is the number of codeword bits and α is the number of overhead bits allocated to the reliability. This equation gives the total number of different nested linear codes $(\mathcal{C}_0, \mathcal{C}_1)$ for a given size, which for the example explored in Section 6 is 624,960. However, not all of these nested linear codes are unique. Based on Lemma 1, some nested linear codes may be equivalent and have the same performance, while others are unique and cannot be transformed into each other through equivalent operations on generator matrices. Therefore, the total number of unique nested linear codes can be lower than the total number of different nested linear codes. For the example in the previous section, the number of unique generator matrices G_0 is 256, which is much smaller than the total number of different nested linear codes. However, finding equivalent codes itself is a complex problem, and it is not guaranteed that we can always identify all equivalent codes.

The traditional approach to finding the best nested linear secrecy codes involves a brute-force search over all possible generator matrices G_0, which are $(k + l) \times n$. This means that for a given size of nested linear code, all possible generator matrices must be formed and their performance calculated. Then, all codes must be compared based on their equivocation to find the best nested linear secrecy code. Using this approach, $2^{(k+l)n}$ generator matrices must be formed. Lemma 1 can be used to identify equivalent generators, but all of them must be examined at some level. In contrast, our proposed approach fixes the matrix H_0 in the dual space and only requires a search for different patterns of the matrix H', which is $k \times n$, with the consideration of the two restrictions explained in Algorithm 1. We can throw out a number of H' candidates due to the fixed form of H_0, e.g., H' matrices with any number of zero columns and/or H' matrices that do not result in a full-rank H_1. This significantly reduces the search space and computational complexity compared to the traditional approach. Fewer than 2^{kn} H' matrices must be compared.

In summary, our proposed approach of searching for the worst code instead of the best code was shown to be easier and more efficient, requiring fewer resources. This is because the generator matrix of the worst linear code $\mathcal{C}_0^{\perp}$ has as many zero columns as possible, making it easier to construct. This improvement in efficiency compared to the full brute-force search method could have important implications for the design of reliable and secure communication systems in practical settings.

On a personal laptop, it is possible to find best codes up to blocklength 12 with little issue, and we show in Figure 5 the results for the best and worst nested linear secrecy codes with $n = 12$, $k = 6$, $l = 3$, and $\alpha = 3$. The best and worst equivocation matrices for this example are given in Figure 6.

Note that although there is a marked increase in efficiency for identifying best codes by first finding worst codes in the dual space, Algorithm 1 still requires a brute-force search in choosing the elements of H'. Thus, for larger code sizes, we still have limitations in finding best codes. In Figure 7, we present performance curves for one set of candidate codes when $n = 40$, $k = 20$, $l = 10$, and $\alpha = 10$. The candidate was found by choosing random columns to fill out H' and checking for full rank, as depicted in Algorithm 1. We leave the identification of large optimal codes as an open problem.

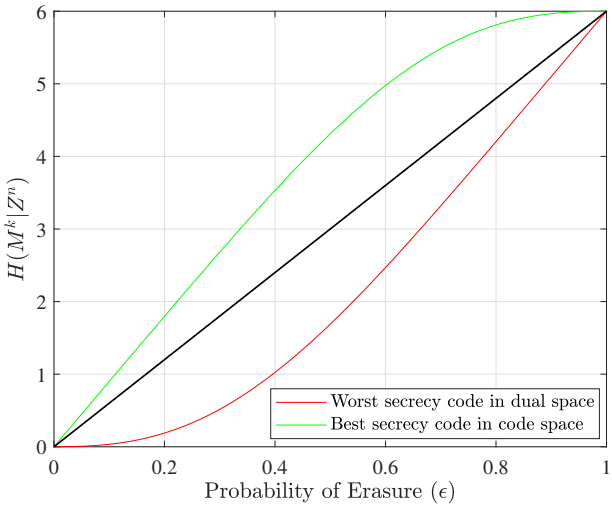

Figure 5. The best nested linear secrecy code when $n = 12$, $k = 6$, $l = 3$, and $\alpha = 3$, along with the worst nested linear secrecy code in the dual space.

$$A_{\text{worst}} = \begin{bmatrix} 1 & 3 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 9 & 27 & 27 & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 36 & 108 & 108 & 36 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 84 & 253 & 255 & 87 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 125 & 384 & 402 & 159 & 9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 117 & 379 & 437 & 207 & 34 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 56 & 202 & 279 & 186 & 64 & 12 & 1 \end{bmatrix},$$

$$A_{\text{best}} = \begin{bmatrix} 1 & 12 & 64 & 186 & 279 & 202 & 56 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 34 & 207 & 437 & 379 & 117 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 9 & 152 & 402 & 384 & 125 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 87 & 255 & 253 & 84 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 36 & 108 & 108 & 36 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 27 & 27 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3 & 3 & 1 \end{bmatrix}.$$

Figure 6. The equivocation matrices A_{worst} and A_{best} for the worst and best nested linear secrecy codes, respectively, when $n = 12$, $k = 6$, $l = 3$, and $\alpha = 3$.

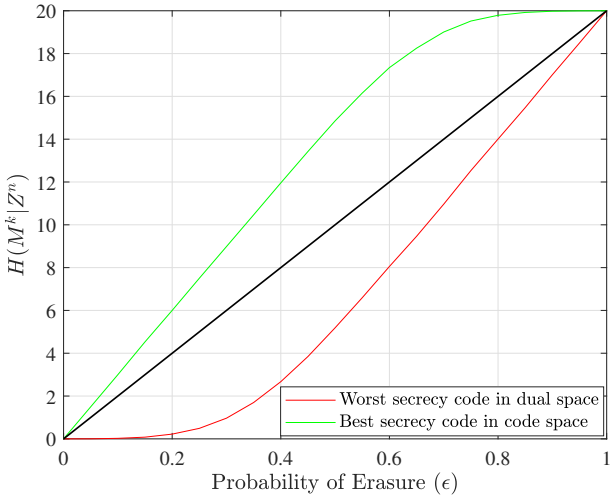

Figure 7. The best nested linear secrecy code for the parameters $n = 40$, $k = 20$, $l = 10$, and $\alpha = 10$, as well as the worst nested linear secrecy code in the dual space.

8. Conclusions and Future Study

In this study, we analyzed the properties of nested linear codes in the presence of a noisy wiretap channel model and derived a new expression for the relative generalized Hamming weight of these codes. We showed that there are three distinct behaviors in terms of equivocation in this coding scheme. Moreover, we proposed a code design algorithm to find the worst nested linear secrecy code, which is constructed by identifying the code with the lowest security in the dual space. Our results demonstrated that this approach is more efficient and quicker in producing optimal nested linear secrecy codes compared to brute-force methods.

Overall, the findings of this paper contribute to the development of reliable and secure communication systems in practical settings. The ability to efficiently design secure nested linear codes can enhance the privacy and security of communication channels, which is of great importance in various applications, such as wireless communication, network security, and cryptography. Future work could explore the applicability of our proposed algorithm to larger blocklengths and investigate its performance in other channel models.

Author Contributions: Conceptualization, M.S. and W.H.; methodology, M.S.; visualization, M.S.; validation, M.S. and W.H.; formal analysis, M.S.; investigation, M.S.; writing—original draft preparation, M.S.; writing—review and editing, W.H.; supervision, W.H.; funding acquisition, W.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation, grant number 1910812.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BEC	Binary erasure channel
LDPC	Low-density parity-check
GHW	Generalized Hamming weights
DLP	Dimension/length profile
RGHW	Relative generalized Hamming weight
RDLP	Relative dimension/length profile

References

1. Wyner, A.D. The Wire-Tap channel. *Bell Syst. Tech. J.* **1975**, *54*, 1355–1387. [CrossRef]
2. Csiszár, I.; Körner, J. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory* **1978**, *24*, 339–348. [CrossRef]
3. Ozarow, L.H.; Wyner, A.D. Wire-tap channel II. *AT&T Bell Lab. Tech. J.* **1984**, *63*, 2135–2157. [CrossRef]
4. Wei, Y.P.; Ulukus, S. Polar Coding for the General Wiretap Channel With Extensions to Multiuser Scenarios. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 278–291. [CrossRef]
5. Harrison, W.K.; Shoushtari, M. On Caching with Finite Blocklength Coding for Secrecy over the Binary Erasure Wiretap Channel. In Proceedings of the Wireless Telecommunications Symposium (WTS), San Francisco, CA, USA, 21–23 April 2021; pp. 1–6. [CrossRef]
6. Shoushtari, M.; Arabian, F.; Harrison, W.K. On the Crucial Role of Information Theory in the Metaverse. In Proceedings of the 2023 Intermountain Engineering, Technology and Computing (IETC), Provo, UT, USA, 12–13 May 2023; pp. 77–82. [CrossRef]
7. Sheng, Z.; Tuan, H.D.; Nasir, A.A.; Poor, H.V. PLS for Wireless Interference Networks in the Short Blocklength Regime with Strong Wiretap Channels. In Proceedings of the IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6. [CrossRef]
8. Yang, W.; Schaefer, R.F.; Poor, H.V. Finite-blocklength bounds for wiretap channels. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 3087–3091. [CrossRef]
9. Taleb, K.; Benammar, M. On the information leakage of finite block-length wiretap polar codes. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Virtual Event, 12–20 July 2021; pp. 61–65. [CrossRef]
10. Harrison, W.K.; Bloch, M.R. Attributes of Generators for Best Finite Blocklength Coset Wiretap Codes over Erasure Channels. In Proceedings of the International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 827–831. [CrossRef]

11. Padakandla, A.; Pradhan, S.S. Achievable Rate Region for Three User Discrete Broadcast Channel Based on Coset Codes. *IEEE Trans. Inf. Theory* **2018**, *64*, 2267–2297. [CrossRef]
12. Shoushtari, M.; Harrison, W. Secrecy coding in the integrated network enhanced telemetry (iNET). In Proceedings of the International Telemetering Conference (ITC), International Foundation for Telemetering, Las Vegas, NV, USA, 25–28 October 2021.
13. Harrison, W.K.; Almeida, J.; Bloch, M.R.; McLaughlin, S.W.; Barros, J. Coding for Secrecy: An Overview of Error-Control Coding Techniques for Physical-Layer Security. *IEEE Signal Process. Mag.* **2013**, *30*, 41–50. [CrossRef]
14. Shoushtari, M.; Arabian, F.; Harrison, W.K. Post-Quantum Cryptography Based on Codes: A Game Changer for Secrecy in Aeronautical Mobile Telemetry. In Proceedings of the International Telemetering Conference (ITC), International Foundation for Telemetering, Phoenix, AZ, USA, 25–28 October 2022.
15. Zamir, R.; Shamai, S.; Erez, U. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Inf. Theory* **2002**, *48*, 1250–1276. [CrossRef]
16. Luo, Y.; Mitrpant, C.; Vinck, A.; Chen, K. Some new characters on the wire-tap channel of type II. *IEEE Trans. Inf. Theory* **2005**, *51*, 1222–1229. [CrossRef]
17. Liu, R.; Liang, Y.; Poor, H.V.; Spasojevic, P. Secure Nested Codes for Type II Wiretap Channels. In Proceedings of the IEEE Information Theory Workshop (ITW), Solstrand, Norway, 1–6 July 2007; pp. 337–342. [CrossRef]
18. Thangaraj, A.; Dihidar, S.; Calderbank, A.R.; McLaughlin, S.W.; Merolla, J. Applications of LDPC Codes to the Wiretap Channel. *IEEE Trans. Inf. Theory* **2007**, *53*, 2933–2945. [CrossRef]
19. Wei, V.K. Generalized Hamming weights for linear codes. *IEEE Trans. Inf. Theory* **1991**, *37*, 1412–1418. [CrossRef]
20. Forney, G. Dimension/length profiles and trellis complexity of linear block codes. *IEEE Trans. Inf. Theory* **1994**, *40*, 1741–1752. [CrossRef]
21. Heijnen, P.; Pellikaan, R. Generalized Hamming weights of q-ary Reed-Muller codes. *IEEE Trans. Inf. Theory* **1998**, *44*, 181–196. [CrossRef]
22. Cheng, J.; Chao, C.C. On generalized Hamming weights of binary primitive BCH codes with minimum distance one less than a power of two. *IEEE Trans. Inf. Theory* **1997**, *43*, 294–298. [CrossRef]
23. Feng, G.; Tzeng, K.; Wei, V. On the generalized Hamming weights of several classes of cyclic cods. *IEEE Trans. Inf. Theory* **1992**, *38*, 1125–1130. [CrossRef]
24. Rajaraman, V.; Thangaraj, A. EG-LDPC codes for the erasure wiretap channel. In Proceedings of the National Conference on Communications (NCC), Chennai, India, 29–31 January 2010; pp. 1–5. [CrossRef]
25. Bras-Amoros, M.; Lee, K.; Vico-Oton, A. New Lower Bounds on the Generalized Hamming Weights of AG Codes. *IEEE Trans. Inf. Theory* **2014**, *60*, 5930–5937. [CrossRef]
26. Yang, M.; Li, J.; Feng, K.; Lin, D. Generalized Hamming Weights of Irreducible Cyclic Codes. *IEEE Trans. Inf. Theory* **2015**, *61*, 4905–4913. [CrossRef]
27. Kurihara, J.; Uyematsu, T.; Matsumoto, R. Secret Sharing Schemes Based on Linear Codes Can Be Precisely Characterized by the Relative Generalized Hamming Weight. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2012**, *E95.A*, 2067–2075. [CrossRef]
28. Al-Hassan, S.; Ahmed, M.; Tomlinson, M. Extension of the parity check matrix to construct the best equivocation codes for syndrome coding. In Proceedings of the Global Information Infrastructure and Networking Symposium (GIIS), Montreal, QC, Canada, 15–19 September 2014; pp. 1–3. [CrossRef]
29. Harrison, W.K.; Bloch, M.R. On Dual Relationships of Secrecy Codes. In Proceedings of the Allerton Conf. Communication, Control, Computing, Monticello, IL, USA, 2–5 October 2018; pp. 366–372. [CrossRef]
30. Shoushtari, M.; Harrison, W.K. New Dual Relationships for Error-Correcting Wiretap Codes. In Proceedings of the IEEE Information Theory Workshop (ITW), Kanazawa, Japan, 17–21 October 2021; pp. 1–6. [CrossRef]
31. Richardson, T.; Urbanke, R. *Modern Coding Theory*; Cambridge University Press: Cambridge, UK, 2008. [CrossRef]
32. Pfister, J.; Gomes, M.A.C.; Vilela, J.P.; Harrison, W.K. Quantifying equivocation for finite blocklength wiretap codes. In Proceedings of the IEEE International Conference on Communications (ICC), Paris, France, 21–25 May 2017; pp. 1–6. [CrossRef]
33. Harrison, W.K. Exact Equivocation Expressions for Wiretap Coding Over Erasure Channel Models. *IEEE Commun. Lett.* **2020**, *24*, 2687–2691. [CrossRef]
34. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2006.