cryptography

MDPI

Article

Generalized Concatenated Codes over Gaussian and Eisenstein Integers for Code-Based Cryptography

Johann-Philipp Thiers and Jürgen Freudenberger *

Institute for System Dynamics (ISD), HTWG Konstanz, University of Applied Sciences, 78462 Konstanz, Germany; jthiers@htwg-konstanz.de
* Correspondence: juergen.freudenberger@htwg-konstanz.de; Tel.: +49-7531-206-647

Abstract: The code-based McEliece and Niederreiter cryptosystems are promising candidates for post-quantum public-key encryption. Recently, q-ary concatenated codes over Gaussian integers were proposed for the McEliece cryptosystem, together with the one-Mannheim error channel, where the error values are limited to the Mannheim weight one. Due to the limited error values, the codes over Gaussian integers achieve a higher error correction capability than maximum distance separable (MDS) codes with bounded minimum distance decoding. This higher error correction capability improves the work factor regarding decoding attacks based on information-set decoding. The codes also enable a low complexity decoding algorithm for decoding beyond the guaranteed error correction capability. In this work, we extend this coding scheme to codes over Eisenstein integers. These codes have advantages for the Niederreiter system. Additionally, we propose an improved code construction based on generalized concatenated codes. These codes extend to the rate region, where the work factor is beneficial compared to MDS codes. Moreover, generalized concatenated codes are more robust against structural attacks than ordinary concatenated codes.

Keywords: public-key cryptography; McEliece cryptosystem; Niederreiter cryptosystem; maximum distance separable codes; concatenated codes

Citation: Thiers, J.-P.; Freudenberger, J. Generalized Concatenated Codes over Gaussian and Eisenstein Integers for Code-Based Cryptography. *Cryptography* **2021**, *5*, 33. https://doi.org/10.3390/cryptography5040033

Academic Editors: Edoardo Persichetti, Paolo Santini, Marco Baldi and Qiang Wang

Received: 1 November 2021
Accepted: 25 November 2021
Published: 29 November 2021

1. Introduction

Public-key cryptographic algorithms are important for today's cyber security. They are used for key exchange protocols or digital signatures, e.g., in communication standards like transport layer security (TLS), S/MIME, and PGP. Public-key encryption is based on a trapdoor-function which also defines the systems security. The most common public-key cryptosystems nowadays are the Rivest–Shamir–Adleman algorithm (RSA) and the elliptic curve cryptography (ECC). Those are based on the intractability of integer factorization and the elliptic curve discrete logarithm problem, respectively. Both problems can be solved using quantum algorithms [1,2]. Hence, large scale quantum computers threaten the security of today's RSA and ECC cryptosystems.

To cope with this issue, many post-quantum encryption methods were proposed [3], e.g., code-based cryptography. Code-based cryptography is based on the problem of decoding random linear codes, which is known to be NP-hard [4]. The best-known code-based cryptosystems are the McEliece system [5] and the Niederreiter system [6].

For the McEliece system, the public key is a permuted and scrambled version of the generator matrix of an error correcting code. The message is encrypted by encoding the information with the scrambled generator matrix and adding intentional errors. The private key is the original generator matrix and the matrices used for scrambling and permutation. Using the private key, the received vector can be decoded into the original message. Due to the scrambling of the generator matrix, it is not possible to obtain its structure without the knowledge of the private key. Hence, an attacker needs to decode the received vector for a random-looking linear code. The best-known decoding attacks

for code-based cryptosystems are based on information-set decoding (ISD) [7], which is therefore the most interesting attack scenario [8–11]. For information-set decoding, the attacker tries to find an error-free information set, which is then used to re-encode the codeword.

The Niederreiter system is comparable to the McEliece system. However, secure digital signature schemes are only known for the Niederreiter system [12]. Instead of the generator matrix, the scrambled parity check matrix is used as public key. For encryption, the message is encoded as an error vector, and the cypher text is the syndrome calculated with the public parity check matrix. The private key consists of the original parity check matrix, as well as the matrices used for scrambling. For decryption, a syndrome decoding algorithm is required, which recovers the error vector from the syndrome. As for the McEliece scheme, the most relevant attacks are based on ISD.

Different code families were proposed for those systems, e.g., Reed–Solomon (RS) codes [13,14], BCH codes [15], LDPC codes [16–19], or polar codes [20]. For some code families, there exist structural attacks, which make use of the structure of the codes, e.g., the attacks in [21,22].

In [23], product codes of outer RS codes and inner one-Mannheim error correcting (OMEC) codes were proposed for the McEliece system. Those codes are defined over Gaussian integers, which are complex numbers with integers as real and imaginary parts [24,25]. They are able to correct more errors than maximum distance separable (MDS) codes. MDS codes are linear block codes which are optimal codes for the minimum Hamming distance, i.e., they achieve equality in the Singleton bound. The codes over Gaussian integers achieve a higher error correction capability due to restriction of the error values. The used channel model allows only errors of Mannheim weight (magnitude) one. The work factor of ISD only depends on the number of errors, but not on their values. A higher error correction capability leads to a higher work factor for comparable parameters. On the other hand, the concatenated codes presented in [23] can be attacked with a combination of the structural attacks from [21,22].

In this work, we propose a new code construction based on generalized concatenated (GC) codes. This construction is motivated by the results in [26], which show that GC codes are more robust against structural attacks than ordinary concatenated codes. Furthermore, we adapt the code construction to Eisenstein integers. Eisenstein integers are complex numbers of the form $a + b\omega$, where a and b are integers and $\omega = -1/2 - \mathrm{i}\sqrt{3}/2$ is a third root of unity [27]. Eisenstein integers form a hexagonal lattice in the complex plain [28]. While the one-Mannheim error channel has four different error values, a similar channel model for Eisenstein integers has six different error values. In the Niederreiter cryptosystem, the message is encoded as an error vector. Hence, the representation with Eisenstein integers allows for longer messages compared with codes over Gaussian integers. In this work, we additionally derive and discuss the channel capacity of the considered weight-one channel over Eisenstein integers. Moreover, we extend the GC code construction to Eisenstein integers.

This publication is structured as follows. In Section 2, we review the McEliece and the Niederreiter system, as well as the attacks based on information-set decoding. In Section 3, we briefly explain Gaussian and Eisenstein integers, as well as the one-Mannheim error channel. We investigate the weight-one error channel for Eisenstein integers in Section 4. In Section 5, we adapt the product codes from [23] to Eisenstein integers. The new code construction based on generalized concatenated codes is discussed in Section 6. Finally, we conclude our work in Section 7.

2. Code-Based Cryptosystems

In this section, we review the basics of the McEliece and Niederreiter systems, as well as information-set decoding.

2.1. The McEliece System

The McEliece cryptosystem utilizes the problem of decoding random linear codes as trapdoor function. In the following, we will shortly explain the basic concept of this system.

Consider a q-ary code $\mathcal{C}(n,k,t)$ of length n, dimension k, and error correction capability t. The code can be represented by its generator matrix $\mathbf{G}$, and should enable an efficient decoding algorithm $\phi(\cdot)$ for up to t errors. The public key is the pair $(\mathbf{G}',t)$. The matrix $\mathbf{G}'$ is a scrambled generator matrix $\mathbf{G}' = \mathbf{SGP}$, with the random non-singular $k \times k$ scrambling matrix $\mathbf{S}$, and the $n \times n$ permutation matrix $\mathbf{P}$. The private key consists of the three matrices $(\mathbf{G},\mathbf{S},\mathbf{P})$.

For encrypting a message $\mathbf{u}$ of length k, the message is encoded using the public generator matrix $\mathbf{G}'$ and a random error vector $\mathbf{e}$, containing at most t non-zero error values added, i.e., $\mathbf{v} = \mathbf{uG}' + \mathbf{e}$. Using the private key, the message can be decrypted by first computing $\mathbf{r} = \mathbf{vP^{-1}} = \mathbf{uSG} + \mathbf{eP^{-1}}$. Note that $\mathbf{eP^{-1}}$ is a permuted error vector and the permutation does not change the number of errors. We decode $\mathbf{r}$ as $\phi(\mathbf{r}) = \phi(\mathbf{vP^{-1}}) = \mathbf{uS}$. Finally, the message can be calculated using the inverse scrambling matrix.

2.2. The Niederreiter System

The Niederreiter system is based on the parity check matrix. Consider a code $\mathcal{C}(n,k,t)$ with parity check matrix $\mathbf{H}$ and an efficient syndrome decoding algorithm $\phi(\cdot)$. The public key is $(\mathbf{H}',t)$. The scrambled parity check matrix is calculated as $\mathbf{H}' = \mathbf{SHP}$, where $\mathbf{S}$ is a random non-singular $(n-k) \times (n-k)$ scrambling matrix, and $\mathbf{P}$ is a random $n \times n$ permutation matrix. The private key consists of the three matrices $(\mathbf{H},\mathbf{S},\mathbf{P})$.

For encryption, a message is first encoded as an error vector $\mathbf{m}$ of length n and at most t non-zero symbols. The cyphertext is the syndrome calculated using the public parity check matrix, i.e., $\mathbf{s}^T = \mathbf{H}'\mathbf{m}^T$. The legitimate recipient receives $\mathbf{s}^T = \mathbf{H}'\mathbf{m}^T = \mathbf{SHPm}^T$ and computes $\mathbf{S}^{-1}\mathbf{s}^T = \mathbf{HPm}^T$. Applying the syndrome decoding algorithm $\phi(\cdot)$ results in the permuted error vector $\mathbf{Pm}^T$. Finally, the message $\mathbf{m}$ is obtained using the inverse permutation $\mathbf{P}^{-1}$. As for the McEliece system, this decoding is only feasible with the knowledge of the scrambling and permutation matrices $\mathbf{S}$ and $\mathbf{P}$.

2.3. Information-Set Decoding

The best known attacks on the McEliece system as well as the Niederreiter system are based on information-set decoding (ISD). Those attacks do not rely on any code structure except linearity, i.e., the attacks try to decode a random-looking linear code. Such attacks were proposed in [8,9], and more recently, some improvements were proposed in [10,11]. We only review the basic concept of attacks based on ISD.

For the McEliece system, the attacker tries to recover the information vector $\mathbf{u}' = \mathbf{uS}$ from the cyphertext $\mathbf{v} = \mathbf{uG}' + \mathbf{e}$. To achieve this, the attacker tries to guess k error-free positions $\mathbf{u}''$, such that the corresponding columns of the public generator matrix $\mathbf{G}'$ form a non-singular matrix $\mathbf{G}''$. If such positions are found, the attacker can use Gaussian elimination on the guessed positions of $\mathbf{G}'$ and re-encode a codeword $\mathbf{v}'' = \mathbf{u}''\mathbf{G}''$ agreeing with $\mathbf{v}$ in the guessed positions. If $\mathbf{v}''$ differs in at most t positions from $\mathbf{v}$, there are no errors in $\mathbf{u}''$, and the attacker obtains $\mathbf{u}' = \mathbf{u}''\mathbf{G}''^{-1}$.

For the Niederreiter system, the attacker tries to find an error vector $\mathbf{m}$ of weight t, such that $\mathbf{H}'\mathbf{m}^T = \mathbf{s}^T$. To achieve this, an attacker tries random permutations $\tilde{\mathbf{P}}$ on the public key $\mathbf{H}'$ and computes the systematic form as $\mathbf{H}'' = \mathbf{UH}'\tilde{\mathbf{P}} = (\mathbf{A}|\mathbf{I}_{n-k})$, where $\mathbf{U}$ is the matrix that produces the systematic form and $\mathbf{I}_{n-k}$ is the $(n-k) \times (n-k)$ identity matrix. The attacker searches for a permutation such that the permuted message vector $\tilde{\mathbf{P}}\mathbf{m}$ has all non-zeros in the rightmost $n-k$ positions. Such a permutation can be detected by the Hamming weight of the scrambled syndrome $\mathbf{Us}^T = \mathbf{H}''\mathbf{m}^T$. Due to the systematic form of $\mathbf{H}''$, the permuted message vector is $\tilde{\mathbf{P}}\mathbf{m} = \left(0,\ldots,0|\mathbf{Us}^T\right)$.

The complexity of information-set decoding attacks is determined by the expected number of trials required to find a permutation fulfilling those criteria. The probability for such a permutation is

$$P_s = \frac{\binom{n-k}{t}}{\binom{n}{t}} \tag{1}$$

and the expected number of trials is

$$N_{ISD} = \frac{1}{P_s} = \frac{\binom{n}{t}}{\binom{n-k}{t}}. \tag{2}$$

We use N_{ISD} to measure the work factor for ISD attacks.

3. Codes over Gaussian and Eisenstein Integers

Next, we review some properties of Gaussian and Eisenstein integers, as well as some known code constructions for these number fields.

3.1. Gaussian and Eisenstein Integers

Gaussian integers are a subset of complex numbers with integers as real and imaginary parts, i.e., of the form $a + b\mathrm{i}$, where a and b are integers. We denote the set of Gaussian integers by $\mathcal{G}$. The modulo operation in the complex plain is defined as

$$z \mod \pi = z - \left\lfloor \frac{z\pi^*}{\pi\pi^*} \right\rceil \cdot \pi, \tag{3}$$

where $\lfloor \cdot \rceil$ denotes rounding to the closest Gaussian integer, which is equivalent to rounding the real and imaginary parts individually. The set of Gaussian integers modulo $\pi \in \mathcal{G}$ with $p = \pi\pi^*$ elements is denoted by $\mathcal{G}_p$. For $\pi \in \mathcal{G}$, such that $p \mod 4 \equiv 1$, the set $\mathcal{G}_p = \mathcal{G} \mod \pi$ is a finite field which is isomorph to the prime field $\mathbb{F}_p$ [24].

We measure the weight $wt_M(z)$ of a Gaussian integer z as Mannheim weight which is the sum of the absolute values of its real and imaginary parts, i.e.,

$$wt_M(z) = \min_{a+b\mathrm{i} \in \mathcal{K}(z)} |a| + |b|, \tag{4}$$

where $\mathcal{K}(z)$ is the set of Gaussian integers z', such that $z = z' \mod \pi$. The Mannheim distance between two Gaussian integers is the weight of the difference

$$d_M(z, y) = wt_M(z - y). \tag{5}$$

The Mannheim weight of a vector is the sum of Mannheim weights of all elements of the vector. The same holds for the Mannheim distance between two vectors.

Eisenstein integers are similar to Gaussian integers, but of the form $x = a + b\omega$, where a and b are integers, and $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i}$ is a third root of unity. Eisenstein integers form a hexagonal structure in the complex plain and are denoted as $\mathcal{E}$. As for Gaussian integers, a finite field can be defined as the set $\mathcal{E}_p = \mathcal{E} \mod \pi$, where $\pi \in \mathcal{E}$ and $p = \pi\pi^*$. In contrast to Gaussian integers, the prime p has to fulfill $p \mod 6 \equiv 1$ due to the hexagonal structure. For such π, the field $\mathcal{E}_p$ is isomorph to the prime field $\mathbb{F}_p$ [27].

We measure the weight of an Eisenstein integer as a hexagonal weight, which is defined by the minimum number of unit steps in directions which are a multiples of 60°. An Eisenstein integer z can be written as $z = g_1\epsilon_1 + g_2\epsilon_2$, with $\epsilon_{1,2} \in \{\pm 1, \pm\omega, \pm(1+\omega)\}$. Note, that $(1+\omega)$ is a sixth root of unity and ω is a third root of unity. Hence, $\epsilon_{1,2}$ can take the six powers of the sixth root of unity. The weight is defined as

$$wt_{HX} = \min_{\{g_1, g_2 : g_1\epsilon_1 + g_2\epsilon_2 = z\}} |g_1| + |g_2|. \tag{6}$$

As for Gaussian integers, the weight of a vector is the sum of weights of the elements, and the distance between two Eisenstein integers is the weight of the difference.

3.2. One Error Correcting (OEC) Codes

One error correcting (OEC) codes over Gaussian as well as over Eisenstein integers fields were proposed in [24,27], respectively. The parity check matrix $\mathbf{H}$ is defined as

$$\mathbf{H} = \left(\alpha^0, \alpha^1, \alpha^2, \ldots, \alpha^{n-1}\right), \tag{7}$$

where α is a primitive element of the field. A vector $\mathbf{v} = (v_0, v_1, v_2, \ldots, v_{n-1})$ is a codeword if, and only if, $\mathbf{H}\mathbf{v}^T = 0$. For codes over Eisenstein integers, we have $v_i \in \mathcal{E}_p$, and the length of an OEC code satisfies $n \leq \frac{p-1}{6}$. For OEC codes over Gaussian integers, we have $n \leq \frac{p-1}{4}$ and $v_i \in \mathcal{G}_p$.

The dimension of an OEC code is $k = n - 1$, and the minimum Hamming distance is $d_H = 2$. The minimum hexagonal distance is $d_{HX} = 3$ for OEC codes over Eisenstein integers, and the minimum Mannheim distance is $d_M = 3$ for OEC codes over Gaussian integers. Hence, such codes can detect any single error of arbitrary weight and correct a single error of Mannheim weight one or hexagonal weight one, respectively.

3.3. Product Codes over Gaussian Integers

In [23] a product code construction from outer Reed–Solomon (RS) and inner one, error correcting (OEC) codes over Gaussian integers was proposed. In the following, we review this code construction. Later on, this construction is extended to codes over Eisenstein integers.

We consider an outer RS code $\mathcal{C}_o(n_o, k_o, d_o)$ over $\mathbb{F}_p$ and an inner OEC code $\mathcal{C}_i(n_i, k_i, d_i)$ over $\mathcal{G}_p$, where $p = \pi\pi^*$. Note that d_o denotes the minimum Hamming distance of the RS code, while d_i denotes the minimum Mannheim distance of the OEC code. The codeword of a product code can be represented as $(n_i \times n_o)$-matrix. For encoding, first, k_i codewords of the outer RS code are encoded and written to the first k_i rows of the codeword matrix. Next, the symbols are mapped from $\mathbb{F}_p$ to the isomorphic $\mathcal{G}_p$, and each column of the codeword matrix is encoded in the inner OEC code. The product code has length $n = n_o n_i$, dimension $k = k_o k_i = k_o$, and minimum Mannheim distance $d = d_o d_i$, as shown in [23].

For instance, consider the special case of the inner OEC codes of length $n_i = 2$ and minimum Mannheim distance $d_i = 4$ [23]. These codes are generated by a field element a of weight at least three. The parity check matrix is $\mathbf{H} = (1, a)$ and the generator matrix is $\mathbf{G} = (-a, 1)$. Depending on the choice of a, this can result in a code of minimum Mannheim distance $d_i \geq 4$. Note, that this does not change the Hamming weight, hence, only one error of arbitrary weight can be detected. Like the original OEC codes proposed in [24], this code can only correct one error of Mannheim weight one, but it can detect any error vector of weight two. The product code has length $n = 2n_o$ and minimum Mannheim distance $d = 4(n_o - k_o + 1)$.

In order to develop a low-complexity decoding algorithm that can decode up to half the minimum distance, a new channel model was considered in [23]. This one-Mannheim error channel is a discrete memoryless channel restricting the error values to Mannheim weight one [29]. Given an error probability ϵ, each error symbol is zero with probability $1 - \epsilon$. Error values are from the set $\{1, -1, \mathrm{i}, -\mathrm{i}\}$, which occur with probability $\epsilon/4$. Due to this restriction, the error vector in each inner codeword with $n_i = 2$ can have a Mannheim weight of at most two, and therefore can be detected by the inner OEC codes. While the inner decoder corrects any error vector of Mannheim weight one, it declares an erasure for each error vector of Mannheim weight two. Hence, all error positions are known for the outer RS decoder, and an erasure-only decoding method can be applied. Using the Forney algorithm, this erasure-only decoding can correct up to $n_o - k_o$ erasures.

The restriction of the error values allows for a guaranteed error correction capability of $t = 2(n_o - k_o) + 1 = n - 2k + 1$ errors, because $n_o - k_o$ erasures can be corrected, and

each erasure requires at least two errors. One additional error can be corrected in any inner codeword. For code rates $R < 1/3$, this error correction capability is higher than the error correction capability of MDS codes, i.e., $t_{MDS} = (n-k)/2$.

4. The Weight-One Error Channel

In this section, we extend the concept of the one Mannheim error channel from [23] to Eisenstein integers. Furthermore, we derive the capacity for this weight-one error channel and discuss its relation to code-based cryptosystems. These results demonstrate that codes over Eisenstein integers are attainable, where the expected number of errors exceeds the number of redundancy symbols $n-k$, which prevents error free information sets.

While the minimum Hamming distance of codes over Eisenstein integer fields is comparable with other code constructions, they may have a significantly higher minimum hexagonal distance. This leads to an increased error correction capability in terms of hexagonal-weight errors. Hence, a channel model, which restricts the error weight, is advantageous for such codes.

The weight-one error channel is a discrete memoryless channel, which restricts the error values to hexagonal weight one. Hence, only error values $e_i \in \{\pm 1, \pm\omega, \pm(1+\omega)\}$ are possible. Note that ω is a third root of unity and $1+\omega$ is a sixth root of unity. Hence, these six possible values form a hexagon in the complex plain.

Figure 1 illustrates the channel model of the weight-one error channel. For a given channel error probability ϵ, error-free transmission ($e_i = 0$) occurs with probability $1-\epsilon$, while each of the six errors has the same probability of $\frac{\epsilon}{6}$.

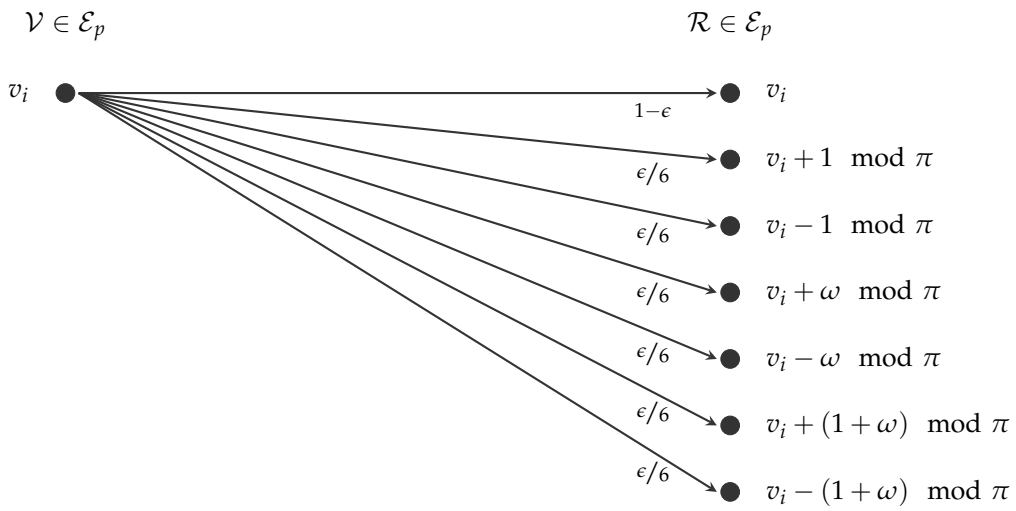

Figure 1. Channel model of the weight-one error channel.

Proposition 1. *The channel capacity of the weight-one error channel with transmitted symbols $v_i \in \mathcal{E}_p$ is*

$$C = \log_2(p) + (1-\epsilon) \cdot \log_2(1-\epsilon) + \epsilon \cdot \log_2\left(\frac{\epsilon}{6}\right). \tag{8}$$

Proof. The channel capacity of a symmetric discrete memory-less channel is [30]

$$C = \log_2(|\mathcal{R}|) - H(\mathbf{P}), \tag{9}$$

where $|\mathcal{R}|$ is the cardinality of the output alphabet $\mathcal{R} = \mathcal{E}_p$ and $H(\mathbf{P})$ the entropy of a row $\mathbf{P}$ of the transition matrix. The cardinality of the output alphabet $\mathcal{E}_p$ is p. Each row of the transition matrix has seven non-zero elements, one element $(1-\epsilon)$ for the case that no error happened, and six elements $\epsilon/6$ for the six equally probable error values. Hence, the entropy is

$$\begin{aligned} H(\mathbf{P}) &= -\sum_{i=0}^{p-1} P_i \cdot \log_2(P_i) \\ &= -(1-\epsilon)\cdot\log_2(1-\epsilon) - 6\cdot\left(\frac{\epsilon}{6}\cdot\log_2\left(\frac{\epsilon}{6}\right)\right) \end{aligned} \tag{10}$$

and thus follows (8). □

Example 1. *Figure 2 shows the relative channel capacity $C/\log_2(p)$ of the weight-one error channel. This relative capacity is the supremum of all achievable code rates R. Moreover, the line $1-\epsilon$ is shown, on which the expected relative number of errors is equal to the relative amount of redundancy $(n-k)/n = 1-R$. For the achievable rate region above this line, the expected number of errors ϵn surpasses $n-k$, and therefore no error-free information sets exist. As shown in Figure 2, codes which are able to correct more than $n-k$ errors are possible for code rates above 0.3 or 0.2 for $p = 43$ and $p = 97$, respectively.*

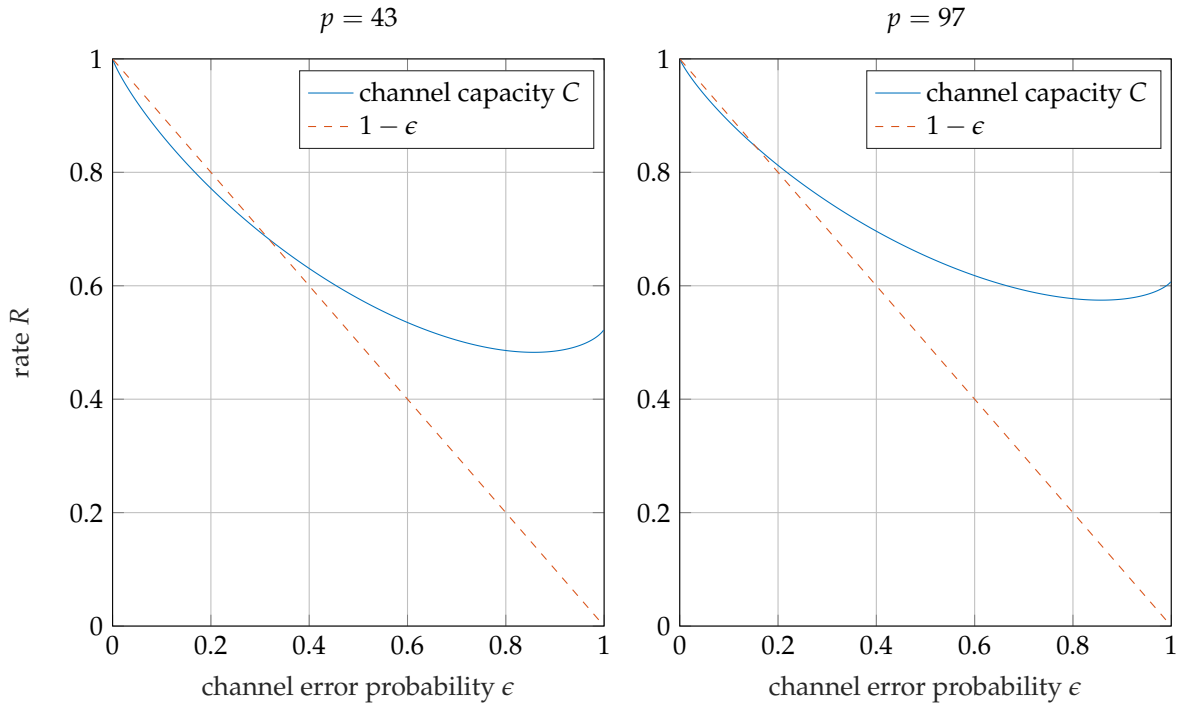

Figure 2. Capacity of the weight-one error channel for $p = 43$ and $p = 97$.

5. Product Codes over Eisenstein Integers

In this section, we adapt the product code construction from [23] to Eisenstein integers for the Niederreiter system. The adaptation of the product code construction is trivial, i.e., we simply replace the inner codes over Gaussian integers by codes over Eisenstein integers. The restriction of applicable primes is different for Gaussian and Eisenstein integers. However, there are primes that fulfill both restrictions leading to the same code parameters. Nevertheless, Eisenstein integers have advantages for the Niederreiter cryptosystem, where the message is encoded as error vector of weight at most t. The information mapping consists of two parts. One part defines the error positions, and can take $\lfloor \log_2 \binom{n}{t} \rfloor$ bits of information. The other part defines the error values and can take $\lfloor \log_2 (m^t) \rfloor$ bits of information, where m is the number of possible error values. Codes over Eisenstein integers increase the message length compared to codes over Gaussian integers, because the number of possible error values m for Eisenstein integers is higher than for Gaussian integers.

The Niederreiter cryptosystem requires an adaptation of decoding method, because only the syndrome is available, and the decoding method needs to find the corresponding error vector. In the following, we devise such a syndrome decoding procedure.

5.1. Syndrome Decoding

For the syndrome decoding, we use look-up tables for the inner OEC codes and erasure decoding for the outer RS codes. We consider the private parity check matrix of the form

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{RS} & \mathbf{0} \\ \mathbf{I}_{n_o} & a \cdot \mathbf{I}_{n_o} \end{pmatrix}, \tag{11}$$

where $\mathbf{H}_{RS}$ is the parity check matrix of the outer RS code, and the lower part $(\mathbf{I}_{n_o} \;\; a \cdot \mathbf{I}_{n_o})$ is the Kronecker-product of the parity check matrix of the OEC codes and an $(n_o \times n_o)$ identity-matrix. With this definition, the first $n_o - k$ syndrome values correspond to the RS code, and the last n_o syndrome values belong to the inner OEC codes. The public key is a scrambled version of the parity check matrix, i.e., $\mathbf{H}' = \mathbf{SHP}$, where $\mathbf{S}$ is a random invertible scrambling matrix, and $\mathbf{P}$ is a random permutation matrix.

To decode the scrambled syndrome $\mathbf{s}^T = \mathbf{SHPm}^T$, one first unscrambles the syndrome as $\tilde{\mathbf{s}}^T = \mathbf{S}^{-1}\mathbf{s}^T = \mathbf{HPm}^T$, and then decodes the inner OEC codes using a look-up in a precomputed syndrome table. Since the inner codewords have a length of two, and the OEC codes have minimum hexagonal distance $d_i \geq 4$, any single error resulting from the weight-one error channel can be corrected, while any error vector of up to two errors can be detected. The precomputed syndrome table provides the error location and value for each correctable error pattern, i.e., each error pattern with only one error. For each error pattern with two errors, an erasure is declared. These erasures are resolved in the outer decoder. Since $\tilde{\mathbf{s}}^T = \mathbf{HPe}^T$, the inner decoder produces parts of the permuted error vector, which is denoted as $\mathbf{P}\hat{\mathbf{e}}^T$.

After the inner decoding, we update the residual syndrome for the outer decoder. The residual syndrome is the syndrome corresponding to an error vector $\mathbf{e} - \hat{\mathbf{e}}$ of lower weight. The syndrome to the partial error vector $\hat{\mathbf{e}}$ can be computed using the private matrices $\mathbf{H}$ and $\mathbf{P}$. This syndrome can be subtracted from the received syndrome

$$\tilde{\mathbf{s}}_{res}^T = \mathbf{HP}(\mathbf{e} - \hat{\mathbf{e}})^T = \tilde{\mathbf{s}}^T - \mathbf{HP}\hat{\mathbf{e}}^T. \tag{12}$$

The outer RS code is now decoded using the residual syndrome $\tilde{\mathbf{s}}_{res}$, as well as the erasure positions declared by the inner decoders. Since the inner decoders detected all error vectors, there are no unknown error positions, and erasure only decoding can be applied to the RS code. This is done using the Forney algorithm [31]. Using the positions j_i, $i = 1, \ldots, \nu$ corresponding to the ν erasures, the error location polynomial can be calculated as

$$\Lambda(x) = \prod_{i=1}^{\nu}(1 - xX_i). \tag{13}$$

This polynomial has roots at $X_1^{-1}, \ldots, X_\nu^{-1}$, with $X_i = \alpha^{j_i}$. Similarly, we represent the residual syndrome as polynomial, i.e., $S_{res}(x) = s_0 + s_1 x + \ldots + s_{n_o-k-1}x^{n_o-k-1}$ and calculate the error-evaluator polynomial $\Omega(x)$ using the key equation

$$\Omega(s) = S_{res}(x)\Lambda(x) \mod x^{n_o-k}. \tag{14}$$

The error values are determined as

$$\hat{e}_i = -\frac{\Omega(X_i^{-1})}{\Lambda'(X_i^{-1})}, \tag{15}$$

where $\Lambda'(x)$ is the derivative of $\Lambda(x)$.

The RS decoder is able to find all error values in the information digits of the OEC codewords if the number of erasures ν does not exceed $n_o - k$. Now, the step in (12) can be used again, with an updated error vector $\hat{\mathbf{e}}$. Hence, the syndrome decoding of the OEC codewords can be repeated to find all remaining errors. The inner codewords have a length of two. Consequently, after correcting one position using the outer code, only a single weight-one error can remain, which is corrected using the syndrome table for the inner code.

Next, we estimate the error correcting capability of this decoding procedure. A minimum of $2(n_o - k)$ channel errors is required to cause a decoding failure in the outer decoder, because $n_o - k$ erasures can be corrected by the outer decoder, and an erasure requires two errors in an inner codeword. Additionally, the OEC code corrects all single errors in the inner codewords. Therefore, at least $t = 2(n_o - k) + 1 = n - 2k + 1$ errors can be corrected. Depending on the error positions, this decoding procedure can correct some patterns with up to $2(n_o - k) + k = n - k$ errors. In comparison with MDS codes, which have an error correction capability of $(n-k)/2$, the proposed construction is advantageous for code rates $R < 1/3$.

5.2. Code Examples

Table 1 shows a comparison of the proposed code construction with MDS codes. The table provides the field size p, code length n, dimension k, and error correction capability t, as well as the work factor N_{ISD} for information-set decoding. The left-hand side of the table considers the proposed code construction, while the right-hand side illustrates comparable MDS codes. In all examples, the work factor for information-set decoding of the proposed construction is significantly higher than for MDS codes.

Table 1. Parameters of codes with work factors between 2^{89} and 2^{124}.

Codes over $\mathcal{E}_p$					**MDS Codes**				
p	n	k	t	N_{ISD}	p	n	k	t	N_{ISD}
139	276	55	167	2^{89}	277	276	55	111	2^{47}
157	312	63	187	2^{101}	313	312	63	124	2^{53}
193	384	77	231	2^{124}	389	384	77	153	2^{65}

Table 2 shows a comparison of the proposed code construction over Eisenstein integers, with the same construction over Gaussian integers from [23], where we compare the message lengths for a Niederreiter system. Note that the restrictions of the field sizes are different. For $p = 137$, we can construct only codes over Gaussian integers, whereas for $p = 139$, we can construct only codes over Eisenstein integers. However, the corresponding codes are comparable. For $p = 157$ and $p = 193$, Eisenstein and Gaussian integer fields exist. The message size with Eisenstein integers is notably increased. This results from the different channel models. Eisenstein integers allow for six different error values, instead of four with Gaussian integers. Due to the same code parameters, the work factor for information-set decoding is the same. Therefore, the codes over Eisenstein integers are only advantageous for Niederreiter systems.

Table 2. Comparison of Eisenstein integers with Gaussian integers.

p	n	k	t	**Message-Length [Bytes]**	
				$\mathcal{G}_p$	$\mathcal{E}_p$
137	272	55	163	73	-
139	276	55	167	-	86
157	312	63	187	84	97
193	384	77	231	103	120

6. Generalized Concatenated (GC) Codes over Gaussian and Eisenstein Integers

While the product code construction shows a significantly increased work factor for information-set decoding, the construction may not be secure against structural attacks. The attack proposed in [22] may allow one to produce the concatenated structure of the code construction. Afterwards, the attack proposed in [21] can produce the structure of the outer Reed–Solomon code.

In [26], it was shown that generalized concatenated codes may withstand the aforementioned structural attacks. Furthermore, those codes enable higher code rates. In the following, we will discuss a generalize concatenated code construction, which may withstand the structural attacks, and has a higher work factor for information-set decoding than MDS codes, as well as the proposed product codes.

In this section, we propose a generalized concatenated code construction. First, we consider codes over Gaussian integers, which, in combination with the one-Mannheim error channel, is advantageous for use in code-based cryptosystems. We investigate a decoding procedure for those codes. Finally, we demonstrate that the GC construction can be extended to codes over Eisenstein integers.

6.1. Code Construction

Generalized concatenated (GC) codes are multilevel codes with one inner code $\mathcal{B}(n_i, k_i, d_i)$ and multiple outer codes $\mathcal{A}^{(l)}\left(n_o, k_o^{(l)}, d_o^{(l)}\right)$ with different dimensions. The basic idea of GC codes is to partition the inner code into multiple levels of subcodes, which are then protected by different outer codes. For the sake of clarity, we only consider GC codes with two outer codes $\mathcal{A}^{(0)}$ and $\mathcal{A}^{(1)}$ of same length n_o, but different dimensions. Again, we represent a codeword as a matrix, where each column is a codeword of the inner code $\mathcal{B}$.

Figure 3 shows the encoding of GC codewords, where first the outer encoder encodes the two codewords $\mathbf{a}_1 \in \mathcal{A}^{(1)}$ and $\mathbf{a}_0 \in \mathcal{A}^{(0)}$. Then, each column is encoded by the inner encoder to a codeword $\mathbf{b}_j \in \mathcal{B}$. The length of the GC code is $n = n_o n_i$, as can be seen from the construction. The dimension is the sum of the outer dimensions.

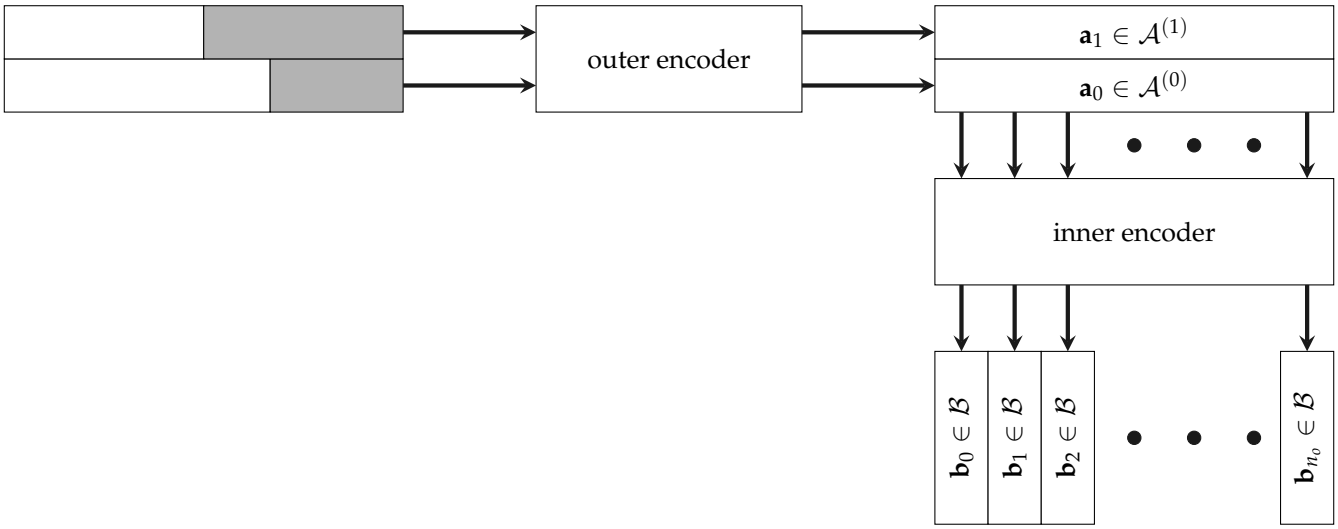

Figure 3. Encoding of GC codes.

For the inner codes, we consider codes over Gaussian integers which achieve a high error correction capability over the one-Mannheim error channel, and enable a partitioning into subcodes with increased minimum distance. Table 3 shows some examples for such inner codes, with their field size p, their modulus π, their generator matrix, as well as the minimum Mannheim distance d of the code and $d^{(1)}$ for the subcode. These codes are not

constructed from one-Mannheim error correcting codes, but found by computed search. The generator matrix of the code $\mathcal{B}$ is chosen in the form

$$\mathbf{G} = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \end{pmatrix}, \tag{16}$$

where a, b, and c are elements of $\mathcal{G}_p$. In this case, the first row is the generator matrix of a subcode $\mathcal{B}^{(1)}\left(3, 1, d^{(1)}\right) \subset \mathcal{B}$ with minimum Mannheim distance $d^{(1)}$ at least 7. This distance allows one to correct any possible error pattern introduced by the one-Mannheim error channel. Note that no codes with $d \geq 5$ were found for field sizes $p < 109$.

Table 3. Examples for inner codes.

p	π	$\mathbf{G}$	d	$d^{(1)}$
109	$10+3\mathrm{i}$	$\begin{pmatrix} 1 & -1-3\mathrm{i} & 3+4\mathrm{i} \\ 0 & 1 & 4-3\mathrm{i} \end{pmatrix}$	5	11
157	$11+6\mathrm{i}$	$\begin{pmatrix} 1 & -2-2\mathrm{i} & 4-3\mathrm{i} \\ 0 & 1 & 2+5\mathrm{i} \end{pmatrix}$	5	12
197	$14+\mathrm{i}$	$\begin{pmatrix} 1 & -4+3\mathrm{i} & -1-5\mathrm{i} \\ 0 & 1 & 6-\mathrm{i} \end{pmatrix}$	5	13

For the GC construction, we consider inner codes of length $n_i = 3$ and dimension $k_i = 2$, i.e., $\mathcal{B}(3, 2, d_i)$, where $d_i \geq 5$ is the minimum Mannheim distance. Those codes can correct up to two errors of Mannheim weight one. For the first level outer code $\mathcal{A}^{(0)}$, we apply a Reed–Solomon code of length n_o and dimension k_o. Since the subcodes in Table 3 are able to correct at least three errors of Mannheim weight one, the information digits of the second level need no further protection if the one-Mannheim error channel model is used.

The resulting GC code has length $n = 3n_o$ and dimension $k = n_o + k_o$, because the second outer level is uncoded. Figure 4 represents the encoding of a single column of the codeword. The outer code symbol $a_{j,0}$ is encoded with the second row of the generator matrix $\mathbf{G}$ of the inner code, which results in a codeword $\mathbf{b}_j^{(0)} \in \mathcal{B}$. The outer code symbol $a_{j,1}$ is encoded with the first row of $\mathbf{G}$, which is the generator matrix of the subcode, and results in $\mathbf{b}_j^{(1)} \in \mathcal{B}^{(1)}$. The codeword in the j-th column is the sum of two codewords, i.e., $\mathbf{b}_j = \mathbf{b}_j^{(0)} + \mathbf{b}_j^{(1)} \in \mathcal{B}$. Note, that the upper part of Figure 4 has the same form as the generator matrix (16), where the gray blocks represent the parity symbols.

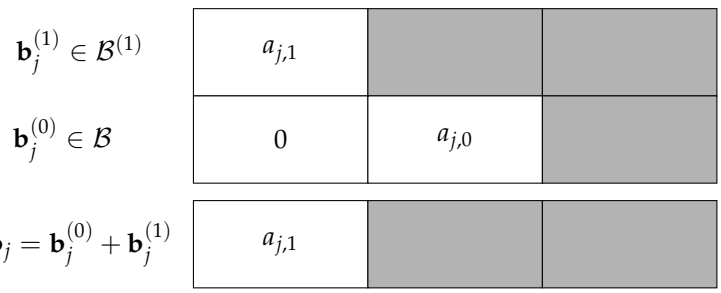

Figure 4. Inner encoding of GCC.

6.2. Decoding

For decoding the GC code, we first decode the inner codes $\mathcal{B}(3, 2, 5)$. While those codes are able to correct two errors of Mannheim weight one, we only correct one error, and therefore can detect any possible error pattern generated by the one-Mannheim error channel. A look-up table with precomputed syndromes is used for decoding all error

patterns with a single error. In cases where more errors occur, we declare an erasure, and store the erasure location. Note that all error patterns are detected. Hence, an erasure only decoding can be applied for the outer RS code.

Decoding the outer code $\mathcal{A}^{(0)}$ requires the code symbols $a_{j,0}$ for all positions where no erasure was declared. Note that the inner codeword in the j-th column is the sum of two codewords of the subcodes, i.e., $\mathbf{b}_j = \mathbf{b}_j^{(0)} + \mathbf{b}_j^{(1)}$. The first digit of $\mathbf{b}_j$ is the outer code symbol $a_{j,1}$ (cf. Figure 4), as the second row of $\mathbf{G}$ has a zero in the first position. Hence, this symbol can be used to determine the codeword $\mathbf{b}_j^{(1)}$ of the subcode $\mathcal{B}^{(1)}$. Subtracting $\mathbf{b}_j^{(1)}$ from $\mathbf{b}_j$ results in $\mathbf{b}_j^{(0)}$.

Now, we can decode the row consisting of the symbols $a_{j,0}$; $j = 0, \ldots, n_o - 1$, which we obtained by re-encoding. We apply an erasure decoding to the Reed–Solomon code [31,32], which is based on the Forney algorithm, as explained for the outer RS code in Section 5.1. This method can correct up to $n_o - k_o$ erasures.

The outer decoding determines all symbols $a_{j,0}$ in the codeword of the outer code $\mathcal{A}^{(0)}$. With these symbols, we can calculate the inner codewords $\mathbf{b}_j^{(0)}$ for all columns with erasures. Furthermore, we can determine the inner codewords $\mathbf{b}_j^{(1)} = \mathbf{b}_j - \mathbf{b}_j^{(0)} \in \mathcal{B}^{(1)}$ in the subcode. Finally, we can decode the resulting codewords in the subcode $\mathcal{B}^{(1)}$, which has a minimum distance $d^{(1)} \geq 7$, and can correct all remaining errors.

We summarize the GC code parameters and the properties of the proposed decoding algorithm in the next proposition.

Example 2. *Consider the code over $\mathcal{E}_{109}$ with $\pi = 10 + 3\mathrm{i}$, as given in the first row of Table 3. In this example, we focus on the decoding of the j-th inner codeword. Let us assume $a_{j,1} = 2 - 4\mathrm{i}$ and $a_{j,0} = -1 + 3\mathrm{i}$ as information symbols of the inner codeword. The codewords encoded with the two individual rows of the generator matrix are $\mathbf{b}_j^{(1)} = (2 - 4\mathrm{i},\ -4 + \mathrm{i},\ -1)$ and $\mathbf{b}_j^{(0)} = (0,\ -1 + 3\mathrm{i},\ -2 + 2\mathrm{i})$. The inner codeword is now $\mathbf{b}_j = \mathbf{b}_j^{(0)} + \mathbf{b}_j^{(1)} = (2 - 4\mathrm{i},\ -2 - 6\mathrm{i},\ -3 + 2\mathrm{i})$.*

We distinguish two cases for the decoding. First, consider the case where at most one error was introduced in the inner codeword. In this case, the inner codeword can be corrected and no errors remain for outer decoding. The first symbol of $\mathbf{b}_j$ is equal to the information symbol $a_{j,1} = 2 - 4\mathrm{i}$. This symbol can be used to re-encode the codeword $\mathbf{b}_j^{(1)} = (2 - 4\mathrm{i},\ -4 + \mathrm{i},\ -1)$. Subtracting $\mathbf{b}_j^{(1)}$ from $\mathbf{b}_j$ gives $\mathbf{b}_j^{(0)} = (0,\ -1 + 3\mathrm{i},\ -2 + 2\mathrm{i})$, which has $a_{j,0} = -1 + 3\mathrm{i}$ as its second symbol. This symbol is used by the outer RS decoder to correct the symbols corresponding to the erasures.

In the second case, more than one error was introduced in the inner codeword. In this case, the inner decoder results in an erasure. Note that the RS decoder does not need any symbol value for the erasure positions. Hence, no re-encoding is required to obtain $a_{j,0}$. The value of $a_{j,0}$ is determined by the outer RS decoder. If the RS decoder is successful, we obtain the correct information digit $a_{j,0} = -1 + 3\mathrm{i}$. This value can be used to re-encode the codeword $\mathbf{b}_j^{(0)} = (0,\ -1 + 3\mathrm{i},\ -2 + 2\mathrm{i})$, which is subtracted from the received vector $\mathbf{b}_j + \mathbf{e}$, resulting in the vector $\mathbf{b}_j^{(1)} + \mathbf{e}$. The error vector is still the error introduced by the channel with restricted values. The error vector of Mannheim weight of at most three can be corrected in this code, because the inner subcode $\mathcal{B}^{(1)}$ has minimum Mannheim distance $d^{(1)} = 11$.

Proposition 2. *The generalized concatenated code with outer Reed-Solomon code $\mathcal{A}^{(0)}\left(n_o, k_o, d_o^{(0)}\right)$ and inner code $\mathcal{B}(3, 2, 5)$ over $\mathcal{G}_p$ with subcode $\mathcal{B}^{(1)}\left(3, 1, d^{(1)} \geq 7\right)$ can correct*

$$t \geq 2(n_o - k_o) + 1 \tag{17}$$

errors of Mannheim weight one.

Proof. Let $\mathbf{b} \in \mathcal{B}(3,2,5)$ be a transmitted codeword of the inner code and $\mathbf{e}$ a length three error vector with up to three errors of Mannheim weight one. For any codeword $\mathbf{b}' \in \mathcal{B}(3,2,5)$, the Mannheim distance to the received sequence is lower bounded by

$$d_M(\mathbf{b}', \mathbf{b}+\mathbf{e}) = wt_M(\mathbf{b}'-\mathbf{b}-\mathbf{e}) \geq d - wt_M(\mathbf{e}) \geq 2 \tag{18}$$

Hence, any error pattern of a Mannheim weight one can be corrected, and any error pattern of Mannheim weight two or three can be detected. For error patterns of weight greater than one, an erasure is declared. The outer Reed–Solomon code can correct up to $n_o - k_o$ erasures [32], and each erasure requires at least two errors. Hence, $2(n_o - k_o)$ errors can be corrected in the erasure positions, and at least one additional error in any position. This results in (17) for the first level. If the first level decoding is successful, the second level is decoded in the inner subcode $\mathcal{B}^{(1)}\left(3,1,d^{(1)}\right)$ with $d^{(1)} \geq 7$. Note that this subcode is able to correct any possible error pattern with up to three errors, thus no outer decoding is required in the second level. The decoding procedure only fails if the first level fails, i.e., if more than $n_o - k_o$ erasures happen, which requires more than $2(n_o - k_o) + 1$ errors. □

The maximum number of errors, which can be corrected by this decoding procedure, is $3(n_o - k_o) + k_o$. For this we assume, that each erasure results from three errors and each of the k_o inner codewords, which does not result in an erasure with exactly one error. On the other hand, this requires a very specific distribution of the errors. Nevertheless, the decoder is able to decode many error patterns with more than $2(n_o - k_o) + 1$ errors. This is demonstrated in Section 6.4.

6.3. GC Code Examples

The guaranteed error correction capability of the proposed code construction is $t = 2(n_o - k_o) + 1$, which for code rates $R \leq 5/9$, is higher than the error correction capability $(n-k)/2$ of MDS codes. We compare the proposed code construction with the product code construction from [23], as well as MDS codes with respect to the work factor for information-set decoding, created according to (2).

Table 4 shows a comparison of the proposed GC codes with comparable MDS codes. We compare the codes with varying code rate R for constant code length $n = 312$. For low code rates, a significant gain is achieved, which decreases for higher code rates. This effect is also shown in Figure 5, where the work factors for ISD of GC codes and MDS codes are plotted over the code rate R for different code length n.

In Table 5, we compare the proposed code construction with product codes over Gaussian integers proposed in [23], since those codes are constructed for the same channel model. Note that those product codes are only applicable for low code rates, and have a higher work factor than MDS codes only for code rates $R < 1/3$. Hence, we compare rate 0.2 product codes with rate 0.5 GC codes with comparable lengths. While the error correction capability is significantly higher for the product codes, due to the lower code rate, the work factor is much lower.

Table 4. Comparison of proposed GC codes with MDS codes.

Reference	p	n	R	t	N_{ISD}
proposed	157	312	0.34	207	2^{283}
proposed	157	312	0.45	127	2^{182}
proposed	157	312	0.55	95	2^{104}
MDS	313	312	0.34	103	2^{79}
MDS	313	312	0.45	86	2^{92}
MDS	313	312	0.55	75	2^{100}

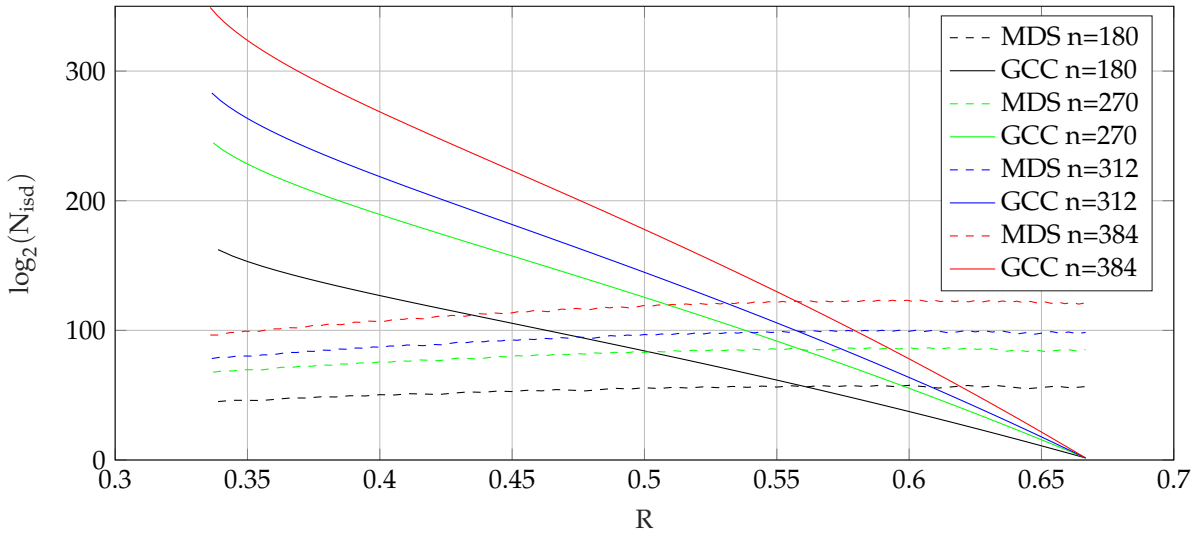

Figure 5. Work factor for information-set decoding over code rate.

Table 5. Comparison of proposed GC codes with product codes from [23].

Reference	p	n	R	t	N_{ISD}
proposed	109	270	0.5	91	2^{125}
proposed	157	312	0.5	105	2^{144}
proposed	197	384	0.5	129	2^{177}
[23]	137	272	0.2	163	2^{88}
[23]	157	312	0.2	187	2^{101}
[23]	193	384	0.2	231	2^{124}

6.4. Decoding beyond the Guaranteed Error Correction Capability

The guaranteed error correction capability of the proposed generalized concatenated codes is given in (17). Up to this bound, all possible error patterns can be corrected, but also, some error patterns with more errors are correctable. In this section, we discuss the error correction capability for decoding beyond the guaranteed error correction capability.

Example 3. *Figure 6 shows the residual word error rate (WER) versus the channel error probability ϵ, with decoding beyond the guaranteed error correction capability. We compare the proposed decoding method with bounded distance decoding up to the guaranteed error correction capability for the GC code of length $n = 270$ and rate $R = 0.5$. As can be seen, the proposed decoding method achieves a significant gain.*

On the other hand, decoding beyond the guaranteed error correction capability leads to a residual error rate. Note that this is the case for many decoders, which were proposed for McEliece systems [16–19]. While in some cases this may be undesirable, this allows for an increased number of errors, and therefore an increased work factor for information-set decoding.

Example 4. *As an example, we compare the work factor for information-set decoding with the guaranteed error correction capability, with an expected number of errors such that the residual error rate is at most 10^{-5}. Consider the code for length $n = 270$ from Example 3. The proposed decoding allows for 35% of errors, which corresponds to about 95 errors. According to (2), this results in a work factor of 2^{133}. The work factor for the guaranteed error correction capability is only 2^{125}, as shown in Table 5. Note that the work factor increases if a higher residual error rate is allowed. For instance, the work factor is increased to about 2^{144} for a residual error rate of 10^{-4}.*

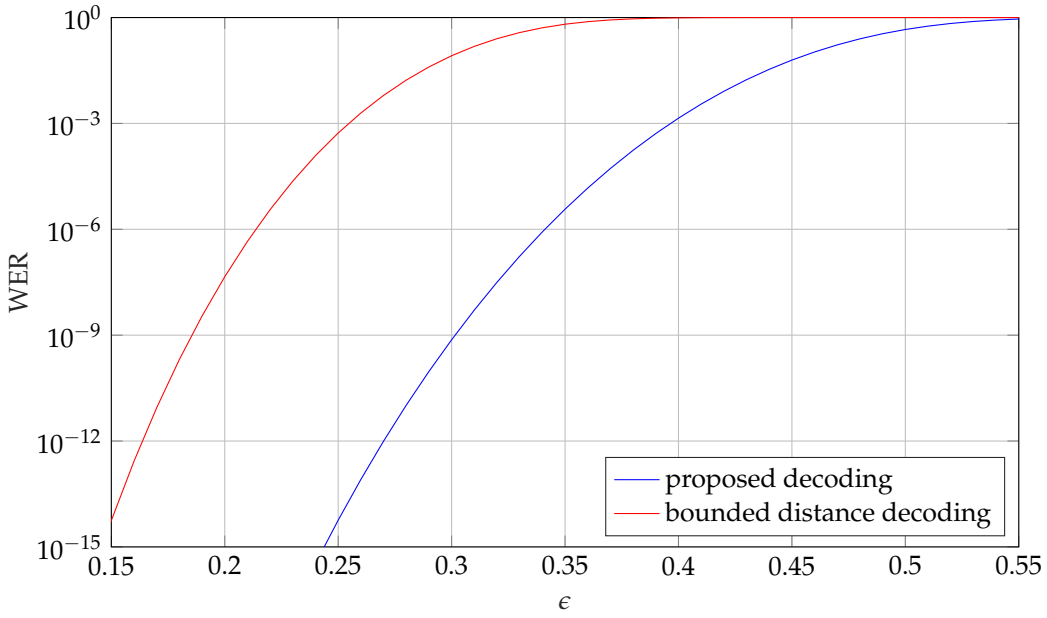

Figure 6. WER over channel error probability in comparison to bounded minimum distance decoding.

6.5. Adaptation to Eisenstein Integers

As for the product code construction over Eisenstein integers, which was adapted from the product code construction over Gaussian integers proposed in [23], the generalized concatenated code construction can also be applied to codes over Eisenstein integers. While the restrictions for the primes are different, using the same field size leads to the same code parameters, and therefore the same error correction capability. Hence, for the McEliece systems, this would result in the same work factor for information-set decoding-based attacks. However, for the Niederreiter system, the increased number of different error values leads to an increased message length. The adaption of the GC code construction to Eisenstein integers is straightforward given the partitioning of the inner codes. Table 6 shows some possible inner codes over Eisenstein integer fields, which were found by computed search. For primes less than 223, no codes with $d \geq 5$ were found.

Table 6. Examples for inner codes over Eisenstein integer fields.

p	π	$\mathbf{G}$	d	$d^{(1)}$
223	$11+17\omega$	$\begin{pmatrix} 1 & 1 & 1-7\omega \\ 0 & 1 & -4-7\omega \end{pmatrix}$	5	8
229	$12+17\omega$	$\begin{pmatrix} 1 & 2-4\omega & -2-5\omega \\ 0 & 1 & -6+2\omega \end{pmatrix}$	5	10
271	$10+19\omega$	$\begin{pmatrix} 1 & -4+5\omega & -6 \\ 0 & 1 & -3+2\omega \end{pmatrix}$	5	11
277	$12+19\omega$	$\begin{pmatrix} 1 & -2+4\omega & -2-6\omega \\ 0 & 1 & -6+\omega \end{pmatrix}$	5	10

Example 5. *For a comparison of the message length, we consider codes over fields of size $p = 229$, because this field size allows for inner codes over Gaussian as well as Eisenstein integers. Using the outer RS code $\mathcal{C}_o(80, 1, 80)$ of rate $R = 1/80$ leads to GC codes of length $n = 3n_o = 240$ and rate $R = 0.34$. Those codes can correct at least $t = 2 \cdot (n_o - k_o) + 1 = 159$ errors of Mannheim weight one or hexagonal weight one, respectively. The number of bits that can be mapped to the error vector for the Gaussian integer code is*

$$t \cdot \log_2(4) + \log_2 \binom{n}{t} \approx 535. \tag{19}$$

These bits are mapped to the error positions and to the error values. For the code over Eisenstein integers, the error values can take $t \cdot \log_2(6)$ bits of information. Hence, the overall number of bits that can be mapped to the error vector is 628*, which is about* 17% *higher than for Gaussian integers. To use the increased message length, the error values cannot be mapped independently, but as a vector of length t, where each component can take six different values.*

7. Conclusions

In this work, we have proposed a code construction based on generalized concatenated codes over Gaussian and Eisenstein integers for their use in code-based cryptosystems. These GC codes can be decoded with a simple decoding method that requires only table look-ups for the inner codes and erasure decoding of the outer Reed–Solomon codes. The proposed construction is a generalization of the ordinary concatenated codes proposed in [23]. The GC codes enable higher code rates. While the number of correctable errors is lower than with the concatenated codes, the work factor for information-set decoding (ISD) is increased with GC codes. For rates $R \leq 5/9$, the generalized concatenated codes can correct more errors than MDS codes. Very high work factors are achievable with short codes.

Codes over Eisenstein integers are advantageous for the Niederreiter system due to the increased message length. An investigation of the channel capacity of the weight-one error channel was performed. Capacity achieving codes over Eisenstein integers can correct more than $n - k$ errors, leading to increased security against information-set decoding attacks.

While we have adapted the GC code construction to Eisenstein integers, the syndrome decoding for the corresponding Niederreiter system, is still an open issue. An investigation of suitable decoding methods would be an interesting topic for further research.

The value of the proposed GC code construction can be seen when compared to the classic McEliece key encapsulation mechanism (KEM), which is among the finalists of the NIST standardization [3]. For example, the security against ISD attacks for the parameter set *McEliece 348864* is $N_{ISD} = 2^{143}$ (according to (2)) and the public-key size is about 261 kByte. A GC code over $p = 109$, of length $n = 159$ and dimension $k = 54$, results in the same work factor for ISD attacks, but its public-key size is only 7.3 kByte, which is about 3% of the key size for the classic McEliece system. For the longer code *McEliece 6688128*, the work factor is about $N_{ISD} = 2^{262}$, and the public-key size approximately 1045 kByte. A comparable GC code over $p = 109$ has length $n = 291$, dimension $k = 98$, work factor $N_{ISD} = 2^{264}$, and public-key size of only 24.4 kByte. However, the classic McEliece KEM uses Goppa codes, as originally proposed by McEliece in 1984. Goppa codes are still considered to be secure, as no structural attacks on these codes were found. On the other hand, the proposed GC code construction has no complete security analysis against structural attacks, such as the attacks proposed in [21,26,33]. This security analysis is subject to future work.

Author Contributions: The research for this article was exclusively undertaken by J.-P.T. and J.F. Conceptualization and investigation, J.-P.T. and J.F.; writing—review and editing, J.-P.T. and J.F.; writing—original draft preparation, J.-P.T.; supervision, project administration, and funding acquisition J.F. All authors have read and agreed to the published version of the manuscript.

Funding: The German Federal Ministry of Research and Education (BMBF) supported the research for this article (16ES1045) as part of the PENTA project 17013 XSR-FMC.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Shor, P.W. Algorithms for quantum computation: Discrete logarithms and factoring. In Proceedings of the 35th Annual Symposium on Foundations of Computer Science, Washington, DC, USA, 20–22 November 1994; pp. 124–134. [CrossRef]
2. Proos, J.; Zalka, C. Shor's Discrete Logarithm Quantum Algorithm for Elliptic Curves. *Quantum Inf. Comput.* **2003**, *3*, 317–344. [CrossRef]
3. Alagic, G.; Alperin-Sheriff, J.; Apon, D.; Cooper, D.; Dang, Q.; Kelsey, J.; Liu, Y.K.; Miller, C.; Moody, D.; Peralta, R.; et al. *Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process*; Nistir 8309; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020.
4. Berlekamp, E.; McEliece, R.; van Tilborg, H. On the inherent intractability of certain coding problems. *IEEE Trans. Inf. Theory* **1978**, *24*, 384–386. [CrossRef]
5. McEliece, R. A public-key cryptosystem based on algebraic coding theory. *DSN Prog. Rep.* **1978**, *42–44*, 114–116.
6. Niederreiter, H. Knapsack-type cryptosystems and algebraic coding theory. *Probl. Control Inf. Theory* **1986**, *15*, 159–166.
7. Prange, E. The use of information sets in decoding cyclic codes. *IRE Trans. Inf. Theory* **1962**, *8*, 5–9. [CrossRef]
8. Lee, P.J.; Brickell, E.F. An Observation on the Security of McEliece's Public-Key Cryptosystem. In *Advances in Cryptology—EUROCRYPT '88, Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques, Davos, Switzerland, 25–27 May 1988*; Barstow, D., Brauer, W., Brinch Hansen, P., Gries, D., Luckham, D., Moler, C., Pnueli, A., Seegmüller, G., Stoer, J., Wirth, N., Eds.; Springer: Berlin/Heidelberg, Germany, 1988; pp. 275–280.
9. Stern, J. A method for finding codewords of small weight. In *Coding Theory and Applications*; Cohen, G., Wolfmann, J., Eds.; Springer: Berlin/Heidelberg, Germany, 1989; pp. 106–113.
10. Bernstein, D.J.; Lange, T.; Peters, C. Attacking and Defending the McEliece Cryptosystem. In *Post-Quantum Cryptography*; Buchmann, J., Ding, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 31–46.
11. May, A.; Meurer, A.; Thomae, E. Decoding Random Linear Codes in $\mathcal{O}(2^{0.054n})$. In *Advances in Cryptology—ASIACRYPT 2011, Proceedings of the 17th International Conference on the Theory and Application of Cryptology and Information Security, Seoul, South Korea, 4–8 December 2011;* Lee, D.H., Wang, X., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 107–124.
12. Courtois, N.T.; Finiasz, M.; Sendrier, N. How to Achieve a McEliece-Based Digital Signature Scheme. In *Advances in Cryptology—ASIACRYPT 2001, Proceedings of the 7th International Conference on the Theory and Application of Cryptology and Information Security, Gold Coast, Australia, 9–13 December 2001*; Boyd, C., Ed.; Springer: Berlin/Heidelberg, Germany, 2001; pp. 157–174.
13. Wieschebrink, C. Two NP-complete Problems in Coding Theory with an Application in Code Based Cryptography. In Proceedings of the IEEE International Symposium on Information Theory, Seattle, WA, USA, 9–14 July 2006; pp. 1733–1737. [CrossRef]
14. Berger, T.P.; Cayrel, P.L.; Gaborit, P.; Otmani, A. Reducing Key Length of the McEliece Cryptosystem. In *Progress in Cryptology—AFRICACRYPT, Proceedings of the Second International Conference on Cryptology in Africa, Gammarth, Tunisia, 21–25 June 2009*; Preneel, B., Ed.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 77–97.
15. Le Van, T.; Hoan, P.K. McEliece cryptosystem based identification and signature scheme using chained BCH codes. In Proceedings of the International Conference on Communications, Management and Telecommunications (ComManTel), DaNang, Vietnam, 28–30 December 2015; pp. 122–127. [CrossRef]
16. Monico, C.; Rosenthal, J.; Shokrollahi, A. Using low density parity check codes in the McEliece cryptosystem. In Proceedings of the 2000 IEEE International Symposium on Information Theory, Sorrento, Italy, 25–30 June 2000; p. 215. [CrossRef]
17. Shooshtari, M.K.; Ahmadian, M.; Payandeh, A. Improving the security of McEliece-like public key cryptosystem based on LDPC codes. In Proceedings of the 11th International Conference on Advanced Communication Technology, Gangwon-Do, Korea, 15–18 February 2009; Volume 2, pp. 1050–1053.
18. Baldi, M.; Bianchi, M.; Maturo, N.; Chiaraluce, F. Improving the efficiency of the LDPC code-based McEliece cryptosystem through irregular codes. In Proceedings of the IEEE Symposium on Computers and Communications (ISCC), Split, Croatia, 7–10 July 2013; pp. 000197–000202. [CrossRef]
19. Moufek, H.; Guenda, K.; Gulliver, T.A. A New Variant of the McEliece Cryptosystem Based on QC-LDPC and QC-MDPC Codes. *IEEE Commun. Lett.* **2017**, *21*, 714–717. [CrossRef]
20. Hooshmand, R.; Shooshtari, M.K.; Eghlidos, T.; Aref, M.R. Reducing the key length of McEliece cryptosystem using polar codes. In Proceedings of the 11th International ISC Conference on Information Security and Cryptology, Tehran, Iran, 3–4 September 2014; pp. 104–108. [CrossRef]
21. Sidelnikov, V.M.; Shestakov, S.O. On insecurity of cryptosystems based on generalized Reed-Solomon codes. *Discret. Math. Appl.* **1992**, *2*, 439–444. [CrossRef]
22. Sendrier, N. On the Concatenated Structure of a Linear Code. *Appl. Algebra Eng. Commun. Comput.* **1998**, *9*, 221–242. [CrossRef]
23. Freudenberger, J.; Thiers, J.P. A new class of q-ary codes for the McEliece cryptosystem. *Cryptography* **2021**, *5*, 11. [CrossRef]
24. Huber, K. Codes over Gaussian integers. *IEEE Trans. Inf. Theory* **1994**, Volume 40, 207–216.
25. Freudenberger, J.; Ghaboussi, F.; Shavgulidze, S. New Coding Techniques for Codes over Gaussian Integers. *IEEE Trans. Commun.* **2013**, *61*, 3114–3124. [CrossRef]

26. Puchinger, S.; Müelich, S.; Ishak, K.; Bossert, M. Code-Based Cryptosystems Using Generalized Concatenated Codes. In *Applications of Computer Algebra*; Kotsireas, I.S., Martínez-Moro, E., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 397–423.
27. Huber, K. Codes over Eisenstein-Jacobi Integers. In *Contemporary Mathematics*; American Mathematical Society: Providence, RI, USA, 1994; Volume 168, pp. 165–179.
28. Conway, J.; Sloane, N. *Sphere Packings, Lattices and Groups*, 3rd ed.; Springer: New York, NY, USA; Berlin/Heidelberg, Germany, 1999.
29. Rohweder, D.; Freudenberger, J.; Shavgulidze, S. Low-Density Parity-Check Codes over Finite Gaussian Integer Fields. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 481–485. [CrossRef]
30. Gallager, R.G. *Information Theory and Reliable Communication*; John Wiley & Sons, Inc.: New York, NY, USA, 1968.
31. Neubauer, A.; Freudenberger, J.; Kühn, V. *Coding Theory: Algorithms, Architectures and Applications*; John Wiley & Sons: New York, NY, USA, 2007.
32. Bossert, M. *Channel Coding for Telecommunications*; Wiley: New York, NY, USA, 1999.
33. Fabsic, T.; Hromada, V.; Stankovski, P.; Zajac, P.; Guo, Q.; Johansson, T. A reaction attack on the QC-LDPC McEliece cryptosystem. In Proceedings of the Post-Quantum Cryptography—8th International Workshop (PQCrypto), Utrecht, The Netherlands, 26–28 June 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 51–68. [CrossRef]