Best binary equivocation code construction for syndrome coding

K Zhang
M Tomlinson
M Ambroze (School of Engineering, Computing and Mathematics)
MZ Ahmed
MRD Rodrigues

Let us know how access to this document benefits you

General rights
All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

Take down policy
If you believe that this document breaches copyright please contact the library providing details, and we will remove access to the work immediately and investigate your claim.

Follow this and additional works at: https://pearl.plymouth.ac.uk/secam-research

Recommended Citation

This Article is brought to you for free and open access by the Faculty of Science and Engineering at PEARL. It has been accepted for inclusion in School of Engineering, Computing and Mathematics by an authorized administrator of PEARL. For more information, please contact openresearch@plymouth.ac.uk.
Best binary equivocation code construction for syndrome coding

K. Zhang
Instituto de Telecomunicações
Dep. de Ciência de Computadores
Universidade do Porto
{kezhang}@dcc.fc.up.pt

M. Tomlinson, M. Z. Ahmed, M. Ambroze
University of Plymouth
United Kingdom
{M.Tomlinson, M.Ahmed, M.Ambroze}@plymouth.ac.uk

M. R. D. Rodrigues
Dep. of Electronic and Electrical Engineering
University College London
United Kingdom
{m.rodrigues}@ucl.ac.uk

Abstract

Traditionally, codes are designed for an error correcting system to combat noisy transmission channels and achieve reliable communication. These codes can be used in syndrome coding, but it is shown in this paper that the best performance is achieved with codes specifically designed for syndrome coding. In the view of the security of the communication, the best codes are the codes, which have the highest value of an information secrecy metric, the equivocation rate, for a given code length and code rate and are well packed codes. A code design technique is described, which produces the best binary linear codes for the syndrome coding scheme. An efficient recursive method to determine the
equivocation rate for the Binary Symmetric Channel and any linear binary code is also presented. A large online database of best equivocation codes for the syndrome coding scheme has been produced using the code design technique with some examples presented in the paper. The presented results show that the best equivocation codes produce a higher level of secrecy for the syndrome coding scheme than almost all best known error correcting codes. Interestingly, it is unveiled that some outstanding best known error correcting codes are also best equivocation codes.

I. INTRODUCTION

The wiretap channel, which was introduced by Wyner [1], is a physical layer model that captures the fundamentals of communication security. In this model, a transmitter, Alice, wishes to send confidential information to a legitimate receiver, Bob, in the presence of an eavesdropper, Eve. Wyner proposed a syndrome coding scheme to guarantee the security of the communication for a specific wiretap channel, where the main channel is error-free and the eavesdropper channel is a binary symmetric channel [2]. Wyner showed that as long as the size of the syndrome space is chosen to be smaller than the Shannon entropy of the binary symmetric channel, there exist codes that leak a vanishing proportion of information to the eavesdropper as the code length approaches infinity. In [3], Bennett strengthens Wyner’s result in the sense that the channels are more general and the estimate of the number of information bits leaked to Eve is stronger. Cohen and Zemor [4] also provide a more refined analysis of the information leakage of syndrome coding for the wiretap channel.

The syndrome coding scheme is an important physical layer scheme to guarantee the secure communication that has been widely studied for various applications. For example, Cohen and Zemor [5] studied a generalization of Wyner’s syndrome coding scheme which is applicable to noisy main and eavesdropper channels, for biometric applications. Reddy et al. [6] studied low-density parity-check (LDPC) codes for syndrome coding with applications to video coding systems. Suresh et al. [7] showed that duals of certain LDPC codes, when used in a syndrome coding scheme, provide strong secrecy over the binary erasure wiretap channel. Salim and Emina showed that network security can be achieved by using syndrome coding as an additional layer to a network code [8].

For the syndrome coding scheme, it is important to construct the good codes to guarantee the security of the communication. Numerous contributions have been made in designing good code
families such as BCH and Goppa codes [9] or by modifying good codes by concatenation [10], extension [11]–[14] or shortening [9]. Traditionally, good codes are designed for an error correcting system to combat noisy transmission channels and achieve reliable communication for different systems [15], [16]. A central property of an error correcting code is the minimum Hamming distance, \( d \), which determines the number of independent errors that can be corrected and the resulting reliability of communication. The main objective in error correcting code design is to optimise one of the parameters \( n \), \( k \) and \( d \) for given values of the other two [14]. The covering radius \( R_c \) is usually not considered a design parameter in traditional error correcting code design, but it is an indicator of codeword packing [9]. Tables of optimum binary codes are published in an online database by Grassl as Best Known Codes (BKC) in the form of tables of lower and upper bounds to \( d \) [17] for different code rates and code lengths up to 256 bits. With so much research attention historically, improvements to these tables are relatively rare.

Codes designed for error correction can also be used in syndrome coding, but it is shown below that the best performance is not generally achieved and better codes exist. The information rate of a syndrome coding scheme using an \((n, k, d)\) linear code is \( \frac{n-k}{n} \) and all possible binary vectors of length \( n \) may be transmitted. This is quite different from error correction coding applications where only codewords are transmitted and the information rate is \( k/n \). For the syndrome coding, performance is measured in terms of the level of information theoretic secrecy, the equivocation rate, \( R_e \) [2]. The best codes have the highest value of equivocation rate for a given code length and code rate. It is convenient to use the parameter \( m = n - k \), to represent the number of parity bits of the code. A best code for syndrome coding is represented by the functions: \( n(m, R_e) \) - the shortest code for a given \( R_e \) and \( m \) and \( R_e(n, m) \) - the highest equivocation rate for which a binary \((n, k = n - m)\) code exists. In contrast, a best code for error correction is represented by the function \( n(m, d, R_e) \) - the longest code for a given \( d \), \( m \) and \( R_e \). Based on a list of parity check matrices of best performing codes for the syndrome coding scheme for given \( m \) and \( n \), a Best Equivocation Codes (BEC) online database has been generated in the form of \( n \), \( m \) and \( R_e(n, m) \) [18].

In this paper we restrict the choice of codes to binary, linear codes. We present a numerical design technique to produce the best codes for the syndrome coding scheme as measured by the equivocation rate and compare these codes to the best known error correcting codes. The technique is limited practically to codes having around 35 parity bits, but this is not a major
limitation as near 100% secrecy is attainable with code lengths less than 256 bits [18]. The BKC error correcting code tables list codes less than 256 bits long, which make the comparisons possible. We also present an efficient recursive method to analyse the secrecy performance of any linear binary code, which is used in syndrome coding for the Binary Symmetric Channel (BSC). Finally, we present performance results of some best equivocation codes for the syndrome coding scheme and compare these to the best known codes. In almost all cases there is a substantial improvement.

II. PRELIMINARIES: SYNDROME CODING AND THE WIRETAP CHANNEL

The wiretap channel used in the syndrome coding scheme is shown as Figure 1 in which the main channel is noiseless and the eavesdropper channel is a binary symmetric channel with the crossover probability of $\alpha$. This represents an abstraction of a typical bugging situation where the legitimate transmitter, Alice, and the legitimate receiver, Bob, employs a robust channel, and the eavesdropper, Eve, is listening, using a wireless bugging device with limited transmission power and antenna gain, typically with non line of sight radio propagation conditions so that she receives the transmissions with errors due to the poor communication channel. The secrecy capacity of this wiretap channel [1] is:

$$C_s = -\alpha \cdot \log_2 \alpha - (1 - \alpha) \cdot \log_2 (1 - \alpha)$$  (1)

which is the highest transmission rate that can be achieved whilst keeping communications secure from eavesdropping for this channel.

Figure 1. The wiretap channel model
Alice wishes to convey a sequence of independent and uniformly distributed \( m \)-bit binary messages, \( M(1), \ldots, M(p) \), in which \( p \) is the block length of the messages, to the legitimate receiver. This sequence of \( m \)-bit messages, \( M(1), \ldots, M(p) \), is encoded into a sequence of \( n \)-bit codewords, \( V_T(1), \ldots, V_T(p) \). Bob observes the sequence of \( n \)-bit words, \( V_M(1), \ldots, V_M(p) \), where

\[
V_M(i) = V_T(i), \quad i = 1, \ldots, p
\]

and Eve observes the sequence of \( n \)-bit words, \( V_E(1), \ldots, V_E(p) \), where

\[
V_E(i) = V_T(i) + e(i), \quad i = 1, \ldots, p
\]

and \( e(i) \) represents a \( n \)-bit error vector arising from the binary symmetric channel. Bob produces the sequence of original messages, \( M(1), \ldots, M(p) \), from the main channel output sequence, \( V_M(1), \ldots, V_M(p) \), whereas Eve produces a sequence of estimated messages, \( \hat{M}(1), \ldots, \hat{M}(p) \), from the eavesdropper channel output sequence, \( V_E(1), \ldots, V_E(p) \).

Traditionally, the syndrome coding scheme uses a \((n, k, 2t + 1)\) linear block code, which is capable of correcting \( t \) errors, defined either by a \( k \times n \) generator matrix \( G \) or by a \( m \times n \) parity check matrix \( H \) [20]. It is a property of any linear block code that there exist \( 2^m \) distinct \( n \)-bit minimum weight error patterns, in which each pattern \( e_j \) produces a distinct syndrome \( S_j \) of the total \( 2^m \) syndromes, based on \( S_j = e_j \times H^T \), where \( 0 \leq j \leq 2^m - 1 \). These minimum weight error patterns, \( e_j \), are the coset leaders of the code and may be represented in a table of \( 2^m \) syndromes, \( S_j \), and associated minimum weight error patterns, \( e_j \). For perfect codes, the error patterns, \( e_j \), in the syndrome table are all of the \( n \)-bit binary error vectors with a weight \( w \leq t \). For non-perfect codes, some of the error patterns in the table are \( n \)-bit binary vectors with a weight \( w \leq t \) and the remaining error patterns are \( n \)-bit binary vectors with a weight \( w > t \). Regardless of whether or not the code is perfect, in syndrome coding all \( 2^m \) syndromes, \( S_j \), are used to send messages. In the traditional syndrome coding, it is necessary to store an error pattern-syndrome look up table, which can be accessed by Alice, Bob and Eve. However, it is shown below that such a look up table is unnecessary and the parity check matrix of the code is sufficient.

With some elementary row and column operations, every parity check matrix can be represented in a reduced echelon form [9]. In the following it is assumed that the parity check matrix
of the code is in the reduced echelon form with an identity sub-matrix of \( m \) parity bits followed by \( k \) information bit columns.

The encoding and decoding processes are as follows.

### A. Encoder

Alice equates the \( m \)-bit messages to \( m \)-bit syndromes \( S_T(i) = M(i) \) and produces \( n \)-bit vectors \( V_T(i) \) from each \( m \)-bit message \( M(i) \) at time \( i \) such that \( V_T(i) \times H^T = M(i) \) in three steps:

- Alice generates a random \( n \)-bit codeword \( C_T(i) \) from a random, uniformly distributed \( k \)-bit vector \( D_R(i) \):
  \[
  C_T(i) = D_R(i) \times G
  \]  
  (4)
- Alice forms an \( n \)-bit zero padded message vector \( V(i) \), consisting of the \( m \)-bit message \( M(i) \) followed by \( k \) 0’s.
- Alice then generates the transmitted \( n \)-bit vector \( V_T(i) \) by adding the \( n \)-bit codeword \( C_T(i) \) to the \( n \)-bit zero padded message vector
  \[
  V_T(i) = C_T(i) + V(i)
  \]  
  (5)

In this way Alice produces \( n \)-bit vectors with the property that \( V_T(i) \times H^T = M(i) \).

**Proof:** The syndrome of \( V_T(i) \), \( S_T(i) \) satisfies

\[
S_T(i) = V_T(i) \times H^T = C_T(i) \times H^T + V(i) \times H^T
\]

\[
= 0 + M(i)
\]  

(6)

\( V(i) \times H^T \) produces \( M(i) \), which is due to the structure of \( V(i) \) and \( H \).

The information rate of the syndrome coding scheme is \( R = \frac{m}{n} \).

### B. Legitimate decoder

Bob calculates the syndromes of the received vectors, \( V_M(i) \) to determine the original messages:

\[
S_{Bob}(i) = V_M(i) \times H^T
\]  

(7)
As there are no transmission errors in Bob’s channel $V_M(i) = V_T(i)$ and

$$S_{Bob}(i) = M(i) \quad (8)$$

C. **Eavesdropper decoder**

Eve computes the syndromes associated with the vectors, $V_E(i)$, as follows:

$$S_{Eve}(i) = V_E(i) \times H^T = V_T(i) \times H^T + e(i) \times H^T = M(i) + S_e(i) \quad (9)$$

Eve’s estimate of the message is given by:

$$\hat{M}(i) = S_{Eve}(i) = M(i) + S_e(i) \quad (10)$$

### III. **Calculation of the equivocation rate of a syndrome coding scheme**

In syndrome coding scheme, the secrecy is measured by the eavesdropper decoder output equivocation, which can be calculated as follows:

$$H(M(i)|\hat{M}(i)) = H(M(i), \hat{M}(i)) - H(\hat{M}(i))$$

$$= H(M(i)) - H(\hat{M}(i)) + H(\hat{M}(i)|M(i))$$

$$= H(M(i)) - H(M(i) + S_e(i)) + H(M(i) + S_e(i)|M(i))$$

$$= m - m + 0 + H(S_e(i)|M(i))$$

$$= H(S_e(i)) \quad (11)$$

$$= - \sum_{j=0}^{2^m-1} p(S_j) \log_2 p(S_j) \quad (12)$$

in which, $S_j$ denotes one of the total $2^m$ syndromes. The simplifications in equations (11) and (12) are due to $M(i)$ being uniformly distributed and independent of $S_e(i)$. The equivocation is calculated after deriving the probability mass function of the syndromes due to errors from the BSC, $p(S_j)$ and is a function of the parity check matrix of the code.

A. **Packed integer representation of the parity check matrix**

Historically many code design methods for linear codes, which have been presented in the literature, are based on construction of the parity check matrix, $H$ [11]–[13]. In the following, the best equivocation codes are also constructed from the parity check matrix of the code.
Any column of a parity check matrix of a \((n, k, 2t + 1)\) code, \(H\), can be represented by an integer, \(b_i\), in the range 0 to \(2^{n-k} - 1\), in which \(i\) is the index of the column and \(0 \leq i < n\). The parity check matrix of a \((n, k, 2t+1)\) code is defined by \(n\) integers, referred to as packed integers, and these integers can be in any order since the corresponding codes will all be equivalent \[9\]. Any code design of length \(n\) may be represented by these \(n\) packed integers. Since any parity check matrix can be put into reduced echelon form by elementary row and column matrix operations, the first \(m\) packed integers are fixed and the code design reduces to the determination of \(n-m\) packed integers. Let us start with a general binary, reduced echelon parity check matrix:

\[
H = \begin{pmatrix}
1 & 0 & \cdots & 0 & a_{m0} & \cdots & a_{(n-1)0} \\
0 & 1 & \cdots & 0 & a_{m1} & \cdots & a_{(n-1)1} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & a_{m(m-1)} & \cdots & a_{(n-1)(m-1)}
\end{pmatrix}
\]

(14)

in which \(a_{i,j}\) takes a value of 0 or 1, \(i\) and \(j\) denote the index of the rows and the columns respectively, \(0 \leq j \leq m - 1\) and \(m \leq i \leq n - 1\). Each column may be represented as a packed integer defined as 
\[
b_i = \sum_{j=0}^{m-1} a_{ij} \cdot 2^j
\]

The systematic packed integer form of the parity check matrix is \([1, 2, \cdots, 2^{m-1}, b_m, \cdots, b_{n-1}]\). For example, the parity check matrix of the \((15,7,5)\) BCH code is:

\[
H = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1
\end{pmatrix}
\]

(15)

which in packed integer representation is:

\[
H = [1, 2, 4, 8, 16, 32, 64, 128, 27, 47, 117, 202, 166, 71, 149]
\]

(16)

\[B. \text{ Analysis on the syndrome probability distribution}\]

Equation \[(13)\] shows that it is necessary to get the probability mass function of the syndromes, \(p(S_j)\), for the calculation of the equivocation in BSC. There are two ways to evaluate \(p(S_j)\), in
which \( 0 \leq j \leq 2^m - 1 \).

1) Traditional evaluation method of the syndrome probability distribution: There are \( 2^n \) possible ways in which error patterns, \( e(i) \), occur in each transmitted \( n \)-bit vector. These error patterns occur with probability:

\[
p(e(i)) = \alpha^{w(i)} \cdot (1 - \alpha)^{n-w(i)}
\]

where \( w(i) \) is the weight of \( e(i) \). Each error pattern results in one of the \( 2^m \) syndromes, which is produced as follows:

\[
S_e(i) = e(i) \times H^T
\]

Any error pattern, which is a codeword, has a syndrome equal to zero. As the code is linear, for each particular syndrome, \( S_j \), there are \( 2^k \) different error patterns, \( e(i) \), that produce the same syndrome and the probability of each syndrome due to all possible such error patterns is given by

\[
p(S_j) = \sum_{i=0}^{2^n-1} p(e(i))\delta(S_e(i) - S_j)
\]

where \( S_j \) denotes a specific syndrome vector in all of the \( 2^m \) syndromes, \( 0 \leq j \leq 2^m - 1 \), and \( \delta() \) is the Dirac function. We have

\[
H(S_e(i)) = -\sum_{j=0}^{2^m-1} p(S_j) \log_2 p(S_j)
\]

This method for evaluating the equivocation is computationally manageable for short codes \( (n < 40) \), but for the long codes it is not practical because it involves the evaluation of \( 2^n \) error patterns, an exponential function of the code length. However the probability distribution of the syndromes may be determined recursively and this also provides some insight into code construction by code extension.

2) Recursive evaluation of the syndrome probability distribution: Let us represent the parity check matrix of the code in the packed integer format. The packed integers directly correspond to syndrome values in that a single bit error in a transmitted codeword results in a syndrome equal to the packed integer of the column corresponding to the bit error position. As the codes are linear, any combination of bit errors produces a syndrome equal to the modulo 2 sum of the packed integers corresponding to each column of the parity check matrix in which a bit error occurs.
The following theorem shows that the probability mass function of the syndrome, $p(S_j)$, is a function of the parity check matrix of the code and the crossover probability of the BSC.

**Theorem 1:** The probability mass function (pmf) of $S_j$ for $j = 0$ to $2^m - 1$ may be defined as $p(S_j) = \beta(j)$ where $\beta(j)$ are coefficients of the probability generating function using the Z transform, $p_z(S)$ and $p_z(S)$ only depends on the columns of the parity check matrix and $\alpha$.

\[
p_z(S) = \sum_{j=0}^{2^m-1} \beta(j)Z^j = \prod_{i=0}^{n-1} ((1 - \alpha) + \alpha Z^{b_i}) \tag{21}
\]

where $b_i$ are the packed integer representations of the columns of the parity check matrix and exponent sums of powers of $Z$ are modulo 2 sums.

**Proof:** Any error pattern may be represented as a sum of single bit error events:

\[
e(i) = [e^1 \ e^2 \ \cdots \ e^n] \\
= [e^1 \ 0 \ \cdots \ 0] + [0 \ e^2 \ \cdots \ 0] + \cdots + [0 \ 0 \ \cdots \ e^n] \tag{22}
\]

in which each term denotes a single bit error event where $e^i = 1$ with probability $\alpha$ and $e^i = 0$ with probability $1 - \alpha$.

The linearity of the syndrome coding scheme means that the syndrome resulting from any error pattern is the sum of the syndromes for each bit error position:

\[
S_e(i) = e(i) \times H^T = [e^1 \ e^2 \ \cdots \ e^n] \times H^T \tag{23}
\]

\[= b_1\delta(e^1 - 1) \oplus b_2\delta(e^2 - 1) \cdots \oplus b_n\delta(e^n - 1) \tag{24}
\]

in which $\oplus$ denotes the modulo 2 sum. Since the probabilities of $e^1, e^2, \ldots,e^n$ are independent, the probability of $S_e(i)$ is the product of the probabilities of $n$ separate error events. Adding together the coefficients of same powers of $Z$ results in the coefficients, $\beta_j$ and reduces the number of terms from $2^n$ to $2^m$.

Theorem 1 gives us some insight into code construction by code extension.

**C. Code construction by code extension**

There is a well known relationship between the parity check matrix and the minimum Hamming distance of the code from first studies of error correction coding [19]:

**Property 1:** A linear code has $d$ minimum distance if and only if its parity check matrix has $d$ linearly dependent columns but no set of $d - 1$ or fewer, linearly dependent columns.
Based on Property\ref{eq:property1} any packed integer of the parity check matrix cannot be equal to the modulo 2 sum of any combination of $d-2$ packed integers or fewer integers of the parity check matrix. This is apparent by observing that the syndrome of a codeword is equal to zero. If any packed integer of the parity check matrix were equal to the modulo 2 sum of any combination of $d-2$ packed integers or fewer integers then the resulting combination would have a syndrome equal to zero and therefore be a codeword of weight less than $d$, an obvious contradiction.

A corollary of Property\ref{eq:property1} is that every error pattern of weight less than $t+1 = \lfloor \frac{d+1}{2} \rfloor$ has a distinct syndrome. For syndrome coding this means that the syndrome probability mass function considering only error events with weight less than $t+1$ will be maximally flat. However as higher weight error events all the way up to weight $n$ can occur, other properties of the code other than $d$ are involved.

The code construction method for obtaining codes with good equivocation is based on the observation that the syndrome probability mass function of a code extended in length is a function of the probability mass function of the original code and good equivocation codes produce good extended codes and this is elaborated below. BKC’s have also been obtained by code extension. Bouyukliev’s code construction method extends a code by excluding all linear combinations of up to $d-2$ columns of the parity check matrix \cite{bouyukliev1994} exploiting Property\ref{eq:property1}. The inverting construction $Y_1$, proposed by Edel adds an additional parity bit and extends a $(n, k, d)$ code to a $(n + x, k + x - 1, d)$ code, by extending the parity check matrix \cite{edel2006} similar to Bouyukliev.

Theorem\ref{eq:thm1} gives us some insight into code construction by code extension. If the columns of the parity check matrix of the shortened code with a length $r$ are taken consecutively from $i = 0$ to $r - 1$, in which $i$ is the index of the column, and the Z transform the probability generating function of the shortened code is

$$p_z(S_r) = \prod_{i=0}^{r-1} ((1 - \alpha) + \alpha Z^{b_i}) \quad (25)$$

then the Z transform of the probability generating function the code of length $r + 1$ is

$$p_z(S_{r+1}) = \prod_{i=0}^{r} ((1 - \alpha) + \alpha Z^{b_i}) = p_z(S_r)((1 - \alpha) + \alpha Z^{b_r}) \quad (26)$$

By denoting the coefficients of the shortened code of length $r$, as $\beta_r(i)$ then

$$p_z(S_r) = \sum_{j=0}^{2^m-1} \beta_r(j) Z^j \quad (27)$$
and
\[ p_z(S_{r+1}) = (1 - \alpha) \sum_{j=0}^{2^m-1} \beta_r(j)Z^j + \alpha \sum_{j=0}^{2^m-1} \beta_r(j)Z^{j\oplus b_i} \quad (28) \]
which simplifies to
\[ p_z(S_{r+1}) = (1 - \alpha)p(S_r) + \alpha \sum_{j=0}^{2^m-1} \beta_r(j)Z^{j\oplus b_i} \quad (29) \]
It is apparent that the syndrome probability generating function in Z transform of the code of length \( r + 1 \) is equal to the syndrome probability generating function in Z transform of the shortened code of length \( r \), weighted by \( 1 - \alpha \) plus a permuted syndrome probability generating function in Z transform of the shortened code of length \( r \), weighted by \( \alpha \). The permutation arises from the results of the modulo 2 additions \( j \oplus b_i \). In general, the more uniform the syndrome pmf of a code is, the more uniform the syndrome pmf of the extended code will be.

This leads to the conclusion that the syndrome pmf of the code may be obtained recursively, starting with the generating function \( p_z(S_1) \) determining \( p_z(S_2) \) then \( p_z(S_3) \) through to \( p_z(S_n) \). It is also apparent that good equivocation codes will also produce good equivocation codes when extended in length.

IV. CODE DESIGN FOR SYNDROME CODING

We are now in a position to consider code design for a syndrome coding scheme for the wiretap channel. To construct the best codes for the syndrome coding scheme, we want the equivocation of the eavesdropper to be as high as possible, i.e. the pmf of the syndromes should be as uniform as possible. Since the eavesdropper channel is a binary symmetric channel, for relatively low values of the error probability, \( \alpha \), the equivocation is dominated by error events of low weight. We note the following design considerations:

1) For the BSC, low weight error events have higher probabilities, so they dominate the pmf of the syndrome. If the low weight error events produce distinct syndromes, this helps the pmf of the syndromes to become more uniform. This has a connection to the minimum Hamming distance of the code, \( d \), because error events of weight up to \( \frac{d-1}{2} \) all have distinct syndrome sums. Best known codes, which have the highest \( d \), are likely to give good performance in the syndrome coding scheme, but are not necessarily the best codes,
because all error patterns, regardless of weight, contribute towards the syndrome pmf of the code.

2) With the parity check matrix in reduced echelon form, that is in systematic packed integer form, the packed integers of any of the information bits cannot have a weight less than \(d - 1\); otherwise the codeword formed from that information bit alone will have weight less than \(d\), where \(d\) is the minimum Hamming distance of the code.

3) There is at least one syndrome that is produced by a weight \(R_c\) error event and no lower weight error event, where \(R_c\) is the covering radius of the code. A good equivocation code will tend to have low covering radius compared to the minimum Hamming distance, \(d\). Perfect codes have low covering radius and are good equivocation codes.

4) A shortened best known code has a bounded covering radius \(R_c \geq (d - 1)\), because there exists at least one coset leader of the shortened code of weight \(d - 1\) or greater; otherwise the extended code formed by adding this coset leader to the generator matrix of the code will produce a code with minimum Hamming distance less than \(d\), a contradiction. It is unlikely that shortened best known codes will be best equivocation codes.

5) If columns of the parity matrix are repeated, a weight 2 error event will produce a zero syndrome, which makes the pmf of the syndrome non-uniform. To get a good equivocation code, the parity check matrix should have no repeated integers, unless the code is very long compared to the number of parity bits.

Following these observations, a code design algorithm has been formulated, that produces best equivocation codes as follows:

**INPUT:** \(C_{in}\)-Set of \(l\) inequivalent\(^1\) highest equivocation rate \((n, m)\) codes with parity check matrices in systematic packed integer form \(H\).

**OUTPUT:** \(C_{out}\)- Set of all extended inequivalent \((n + 1, m)\) codes, which are ranked by equivocation in descending order.

For each code \(C\) in \(C_{in}\), the extension steps are as follows:

1) Preset a status array \(a[i] = 1\), where \(1 \leq i \leq 2^{n-k}\).

2) If \(i\) is equal to any integer in \(H\), set \(a[i] = 0\).

\(^1\)If the codes have different equivocation rates, they are inequivalent codes.
3) For all \( j \) such that \( a[j] = 1 \), extend \( H \) with one integer, equal to \( j \), to \( H = [H, j] \). Eliminate all equivalent codes and evaluate the equivocation\(^2\) for each remaining code.

4) Rank the inequivalent codes by their equivocation in descending order, \( C_{out} \), and select the first \( l \) codes. This is termed the best codes subset. These are used as the INPUT for the next extension round.

For a given \( m \), all possible \((m+1, 1)\) codes are used as the initial INPUT with the parity check matrices: \([1, 2, \ldots, 2^{m-1}, i]\), for all \( i \), \( 1 \leq i \leq 2^{n-k} \). For each round as \( n \) is incremented, the best equivocation codes are obtained, provided that the best codes subset is sufficiently large. The syndrome probability generating function in Z transform of each \((n, m)\) code, \( p_z(S)_{in} \), is stored and the syndrome probability generating function in Z transform for each extended \((n+1, m)\) code, \( p_z(S)_{out} \), is determined using equation (29) which makes for a fast algorithm.

Usually, the best extended code will be derived from the highest ranked code from the previous round but this is not always the case. The size of the best codes subset, \( l \), needs to be large enough that no better extended code is missed. For short codes it is possible to store all inequivalent codes in the best codes subset and it is proven that the highest ranked code is the best \((n, m)\) equivocation code.

To estimate the size required of the best codes subset, for a given number of parity bits, \( l \) is assigned (e.g. \( l = 20000 \)) and the code is extended in length by several bits. For the highest ranked code of length \( n \), the ranked position \( p_{n-1} \) is determined of its parent code of length \( n - 1 \) from the previous round. Statistics of \( p_{n-1} \) are collected for each code length and the subset size \( l \) is maintained to be greater by a significant margin to the maximum value of \( p_{n-1} \), i.e. \( l \geq \max p_{n-1} \). For \( m = 6 \) and \( m = 7 \), for all code lengths of \( n \leq 70 \), the maximum \( p_{n-1} \) values are 25 and 275 respectively. It is observed that the greater the length of the code, the higher the possibility of \( p_{n-1} = 1 \), i.e. the extended best equivocation code of length \( n \) produces the best equivocation code of length \( n + 1 \).

V. Results

Following the procedure above, the best equivocation codes have been determined for a given number of parity bits and code length. There are too many codes to present, so we only present

\(^2\)The equivocation is evaluated for the binary symmetric eavesdropper channel for a given error probability.
here some representative codes. The codes are listed in an on-line database \cite{18} in packed integer format. The minimum Hamming distance, covering radius and equivocation rate for a BSC error probability of $\alpha = 0.05$ are given for each code.

As examples, Table I lists the highest rate codes, in packed integer format, which provide at least 90% secrecy when used in syndrome coding for a BSC error probability of $\alpha = 0.05$. Interestingly, in general nearly all of the best equivocation codes do not coincide with the best known codes. However these best equivocation codes do have a respectable minimum Hamming distance and covering radius.

With reference to the on-line database \cite{18}, there are a few best known codes that are also best equivocation codes, notably all of the perfect codes. The perfect codes are the repetition codes, the Hamming codes and the Golay code. The reason for this coincidence is that these codes have the highest minimum Hamming distance, $d$, and are optimally packed codes with the lowest possible covering radius. It is notable that there is a small number of other best known codes that are also best equivocation codes and we term these best known codes as well packed codes. A feature of these well packed codes is that for a given number of parity bits there is no known longer code with the same minimum Hamming distance, $d$. Table II gives the code parameters of the well packed BKC for various values of $m$. We have observed that some of the shorter well packed BKC coincide with the best equivocation codes for given $m$ and given $k$.

It is also interesting to note that each well packed BKC of length $n$ is followed by a poorly packed BKC of length $n + 1$. These BKC’s are termed the worst packed BKC’s. It is found that the corresponding BEC’s have much higher equivocation rates than that of the worst packed BKC’s.

Figure 2 compares the equivocation rate of best known codes and best equivocation codes for increasing code rate, for codes with 26 parity bits at different values of $\alpha$. It shows that the BECs have higher equivocation than BKC’s not only for $\alpha = 0.05$ but also for other values of $\alpha$. The dashed line gives the secrecy capacity of the channel. It is readily apparent that the BEC’s outperform the BKC’s particularly in the lower rate region.

As the number of parity bits increase, codes get longer for the same information rate and the equivocation moves closer to the secrecy capacity limit, $C_s$, which is applicable to infinite length codes. This is shown in Figure 3 for $m = 6$ through to $m = 26$. 
### Table I

<table>
<thead>
<tr>
<th>α</th>
<th>Exact Code</th>
<th>90% Secrecy</th>
<th>Syndrome Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
</tr>
<tr>
<td>0.8</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
</tr>
<tr>
<td>0.7</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
</tr>
<tr>
<td>0.6</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
</tr>
<tr>
<td>0.5</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>1/2 8 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
</tr>
</tbody>
</table>

The best equivocation codes that achieve 90% secrecy in syndrome coding for α = 0.916 are shown in Table I.
### Table II

**Well packed BKC**

<table>
<thead>
<tr>
<th>m</th>
<th>Well packed BKC, (n,k,d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>(5,1.5),(8,4.4),(15,11,3)</td>
</tr>
<tr>
<td>5</td>
<td>(6,1.6),(16,11,4),(31,26,3)</td>
</tr>
<tr>
<td>6</td>
<td>(7,1.7),(8,2.5),(32,26,4)</td>
</tr>
<tr>
<td>7</td>
<td>(8,1.8),(9,2.6),(11,4.5),(64,57,4)</td>
</tr>
<tr>
<td>8</td>
<td>(9,1.9),(10,2.6),(11,3.5),(128,120,4),(255,247,3)</td>
</tr>
<tr>
<td>9</td>
<td>(10,1.10),(11,2.7),(18,9.6),(23,14,5)</td>
</tr>
<tr>
<td>10</td>
<td>(11,1.11),(12,2.8),(15,5.7),(24,14,6),(33,23,5)</td>
</tr>
<tr>
<td>11</td>
<td>(12,1.12),(16,5.8),(23,12,7),(34,23,6),(47,36,5)</td>
</tr>
<tr>
<td>12</td>
<td>(13,1.13),(14,2.9),(24,12,8),(48,36,6),(65,53,5)</td>
</tr>
<tr>
<td>13</td>
<td>(14,1.14),(15,2.10),(25,12,8),(27,14,7),(66,53,6),(81,68,5)</td>
</tr>
<tr>
<td>14</td>
<td>(15,1.15),(16,2.10),(17,3.9),(28,14,8),(31,17,7),(82,68,6),(128,114,5)</td>
</tr>
<tr>
<td>15</td>
<td>(16,1.16),(17,2.11),(18,3.10),(20,5.9),(32,17,8),(37,22,7),(129,114,6),(151,136,5)</td>
</tr>
<tr>
<td>16</td>
<td>(17,1.17),(18,2.12),(21,5.10),(23,7.9),(47,31,7),(152,136,6)</td>
</tr>
<tr>
<td>17</td>
<td>(18,1.18),(19,2.12),(20,3.11),(24,7.10),(27,10.9),(48,31,8),(63,46,7)</td>
</tr>
<tr>
<td>18</td>
<td>(19,1.19),(20,12.13),(21,3.12),(23,5.11),(28,10.10),(31,13.9),(64,46.8),(68,50.7)</td>
</tr>
<tr>
<td>19</td>
<td>(20,1.20),(21,14.14),(24,5.12),(26,7.11),(32,13.10),(35,16.9),(69,50.8),(88,69.7)</td>
</tr>
<tr>
<td>20</td>
<td>(21,1.21),(22,2.14),(27,7.12),(31,11.11),(36,16.10),(41,21.9),(89,69.8),(95,75.7)</td>
</tr>
<tr>
<td>21</td>
<td>(22,1.22),(23,2.15),(24,3.13),(32,11.12),(33,12.11),(42,21.10),(45,24.9),(96,75.8),(128,107.7)</td>
</tr>
<tr>
<td>22</td>
<td>(23,1.23),(24,2.16),(25,3.14),(27,5.13),(34,12.12),(36,14.11),(46,24.10),(49,27.9),(129,107.8),(155,133.7)</td>
</tr>
<tr>
<td>23</td>
<td>(24,1.24),(25,2.16),(28,5.14),(29,6.13),(37,14.12),(47,24.11),(50,27.10),(54,31.9),(156,133.8),(162,139.7)</td>
</tr>
<tr>
<td>24</td>
<td>(25,1.25),(26,2.17),(27,3.15),(30,6.14),(32,8.13),(48,24.12),(55,31.10),(64,40.9),(163,139.8)</td>
</tr>
<tr>
<td>25</td>
<td>(26,1.26),(27,2.18),(28,3.16),(31,6.15),(33,8.14),(34,9.13),(49,24.12),(65,40.10),(72,47.9)</td>
</tr>
<tr>
<td>26</td>
<td>(27,1.27),(28,2.18),(32,6.16),(35,9.14),(38,12.13),(50,24.12),(51,25.11),(73,47.10),(77,51.9)</td>
</tr>
<tr>
<td>27</td>
<td>(28,1.28),(29,2.19),(33,6.16),(35,8.15),(39,12.14),(40,13.13),(52,25.12),(63,36.11),(78,51.10),(94,67.9)</td>
</tr>
<tr>
<td>28</td>
<td>(29,1.29),(30,2.20),(31,3.17),(36,8.16),(37,9.15),(41,13.14),(44,16.13),(64,36.12),(69,41.11),(95,67.10),(128,100.9)</td>
</tr>
<tr>
<td>29</td>
<td>(30,1.30),(31,2.20),(32,3.18),(38,9.16),(40,11.15),(45,16.14),(46,17.13),(70,41.12),(71,42.11),(129,100.10),(135,106.9)</td>
</tr>
<tr>
<td>30</td>
<td>(31,1.31),(32,2.21),(33,3.18),(34,4.17),(41,11.16),(44,14.15),(47,17.14),(48,18.13),(72,42.12),(136,106.10),(142,112.9)</td>
</tr>
</tbody>
</table>

### VI. Conclusions

We considered the code design problem for syndrome coding from the information theoretic security view and presented a code construction method for best equivocation codes. In this way, an online table of constructed BEC’s in packed integer format has been produced. Code examples have been given of the highest rate BEC’s that achieve 90% secrecy to an eavesdropper using
the BSC with an error probability of 0.05. Unlike the traditional case, the presented encoding method does not require a syndrome look up table. This encoding method also provides a new interpretation of syndrome coding in which the message is impressed on the parity bits of a random codeword. In effect the $m$ parity bits of an $n$-bit random codeword are used to scramble an $m$-bit message necessitating the error free recovery of all $n$ bits for an eavesdropper in order to determine the $m$ bit message.

An efficient recurrent method was also presented for the calculation of the probability mass function of the syndromes of a code from the parity check matrix of the code, which in turn enables the secrecy achieved by the code, the equivocation rate, to be determined. It was shown that apart from the perfect codes, the best equivocation codes rarely coincide with the best known error correcting codes and that usually a best equivocation code will provide greater secrecy than
Figure 3. Equivocation rate vs. information rate for Best Equivocation Codes as a function of $m$

using a best known error correcting code with the same parameters.

REFERENCES