https://doi.org/10.2352/ISSN.2470-1173.2020.4.MWSF-292

Analyzing the decoding rate of circular coding in a noisy transmission channel

Yufang Sun[1]***, Jan P. Allebach***[1]***;***
[1]***Electronic Imaging Systems Laboratory, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47906, U.S.A.*** {***sun361,allebach***}***@purdue.edu***

Abstract

Embedding information into a printed image is useful in many aspects, in which reliable channel encoding/decoding systems are crucial, since there is information loss and error propagation during transmission. Circular coding is a general two-dimensional channel coding method that allows data recovery with only a cropped portion of the code, and without the knowledge of the carrier image. While some traditional methods add redundancy bits to extend the length of the original massage length, this method embeds message into image rows in a repeated and shifted manner with redundancy, then uses the majority votes of the redundancy bits for recovery. In this paper, we developed a closed-form formula to predict its decoding success rate in a noisy channel under various transmission noise levels, using probabilistic modeling. The theoretical result is validated with simulations. This result enables the optimal parameter selection in the encoder and decoder system design, and decoding rate prediction with different levels of transmission error.

Introduction

In 1948, Shannon [1] demonstrated that by properly encoding of the information, errors induced by a noisy channel can be reduced to any desired level without sacrificing the rate of information transmission. Since then, much work has been done to find efficient encoding and decoding methods for error control in a noisy channel [2].

Circular coding is a channel encoding method [3] that was first invented in 2013 by Ulichney *et al.* [4], [5] to embed a binary message into a 2D halftone image. This is enabled by separating the message into payload and phase, and interleaving both types of information. The decoding process is based on the repeating data bits in a cropped window. Previously, Sun *et al.* [6] implemented the data transmission system, and then studied the payload decoding rate with different parameter settings (payload bit length, row-to-row shift, interleaving phase period, etc.) using a closed formed solution. So parameters can be selected for optimal decoding performance.

For the data embedding and retrieval in the coding channel, Tai *et al.* [7] quantified the modulation transfer function [8], halftone cluster size, blur, and contrast of the clustered-dot halftone patterns by searching for strong peaks in the frequency domain. And Zhao *et al.* [9] analyzed a frequency-based method to detect the scale, orientation, and location of the carrier image. In this paper, we will focus on the circular coding (channel encoder and decoder) module instead.

Generally speaking, the decoder will be able to successfully decode the message as long as we have a sufficient number of repeating bits. Unlike other rate-less channel encoding methods [10] such as the LT code [11] and the Raptor code [12], [13], the circular coding method encodes the data in two dimensions. And the decoder does not need to know where the message starts.

In this paper, the objective is to study the performance of the circular coding method being used in the halftone image in the presence of noise. We use a Binary Symmetric Channel (BSC) model for the transmission channel, and predict the decoding rate with different levels of transmission error rate for any given circular coding parameters.

We first model the transmission error as a stochastic random process. Then we develop a closed form solution for the payload decoding rate step by step, following the procedures of the decoding process. Finally, we design the simulation of the decoding process to validate the closed form solution.

Review of the communication channel

The pipeline of the circular encoding/decoding framework includes the following main procedures:

1. Encode the digital message $\mathbf{u}$ using the circular coding method to get the coded 2D data array $\mathbf{v}$;
2. Halftone the continuous-tone carrier image $I[m,n]$;
3. Shift dots within a selected subset of halftone cells corresponding to the metadata to be embedded into the image;
4. Print the encoded image;
5. Capture the printed image;
6. Decode the data array, denoted as $\hat{\mathbf{r}}$;
7. The recovered data array is then decoded to get the message back, denoted as $\hat{\mathbf{u}}$.

In this paper, we focus on the impact of random noise in the communication channel, so we will consider the coding channel as a whole to simplify the data transmission system. Please refer to Fig. 1.

Circular coding encoding process

Circular coding is a two-dimensional coding method that allows recovery of data from only a cropped portion of the code. The message $\mathbf{u}$ can be separated into two parts: (1) standard form S, which is the circularly shifted version of the binary payload $\mathbf{P}$ that has the smallest decimal value of the binary string; (2) the minimum bit shift from the standard form $\mathbf{S}$ to the original payload $\mathbf{P}$, denoted as C. Here, C is the binary representation of the bit shift; and we assume that the payload $\mathbf{P}$ has length B bits.

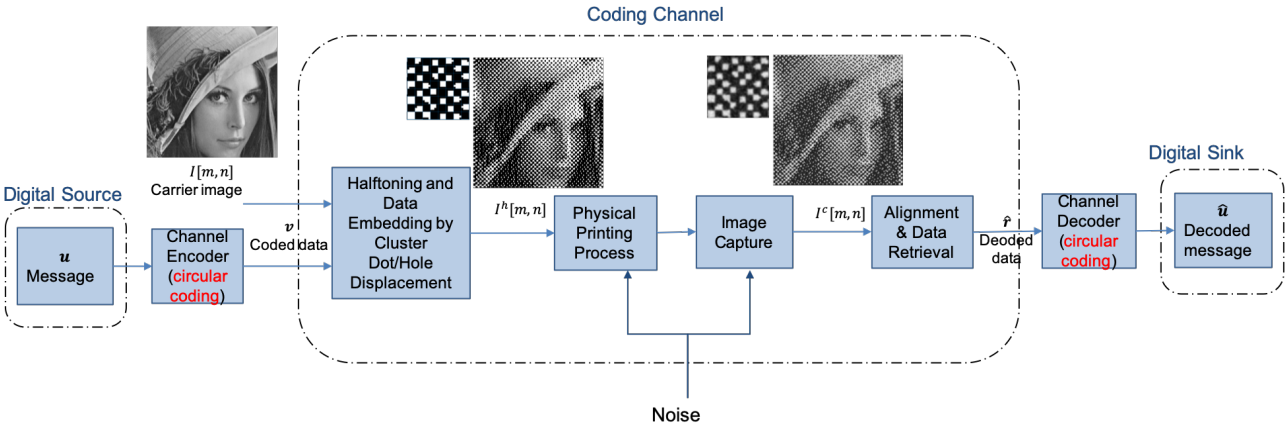


Figure 1: *Framework of the data transmission system. A message* $\mathbf{u}$ *is circularly coded and embedded in the carrier image* I*, and then transmitted in the coding channel, where noise may impact the result.*

The standard form $\mathbf{S}$ and the circular shift C are encoded separately in different rows of a two-dimensional data array, with a determined interleaving period of V. That is, for every $V-1$ rows of standard payload data array, there will be one row of data that encodes the circular shift C.

The B-bit length standard form $\mathbf{S}$ of the payload is repeated from row-to-row, with each row being circularly shifted by D positions relative to the previous row. In every V rows, a phase row is interleaved between the payload rows. It is also shifted in the same manner as the payload rows.

Then the encoded data array is embedded in a halftone image by shifting the dot-clusters within the halftone cells. The resulting image can be printed, and then captured with a mobile phone camera. The encoded data array is extracted from the captured halftone image by detecting the shifts in the dot-clusters.

Circular coding decoding process

The circular coding encoding and decoding methods are discussed in detail in [6]. Generally speaking, the decoder knows the parameters including the payload length B, the row-to-row shift D, and the interleaving phase period V. Also, a cropped portion of the data array will be the input to the decoder. But the decoder doesn't know in which row the phase row appears in the cropped data array. The decoder tries every possible case where that the first phase row could be, removes the assumed phase rows, and calculates the confidence that the remaining rows are pure payload rows. If the assumption of which rows are phase rows is correct (In other words, the assumption of payload rows is correct.), and if there is no error in the data array, then every bit will be repeated in its predefined position. So this will yield perfect consistency. Even if there are some bit errors, it will still have a high consistency. On the other hand, if the assumption of phase rows is incorrect, then the remaining payload rows will contain both payload rows and phase rows. For each bit and its repeating positions, it contains the value of the payload and phase, which will have a lower consistency. The higher the consistency of the repeating bits, the higher probability that these are the payload rows. So we can separate the payload and phase rows by selecting the starting phase row with the highest consistency.

Then, the decoder takes the majority bit value of each repeating bit position of the payload rows, to find the shifted version of the payload, denoted as $\mathbf{P}^{'}$. Similarly, by checking the majority bit value of each repeating bit position of the phase rows, we can find a shifted version of the phase, denoted as $\mathbf{U}^{'}$.

For every payload, since the standard version is unique, there is a unique circular shift C that shifts from the standard version $\mathbf{S}$ to the original payload $\mathbf{P}$. We can find the standard version from $\mathbf{P}^{'}$, and figure out the circular shift $C^{'}$ that will take us from $\mathbf{P}^{'}$ to $\hat{\mathbf{S}}$. It will be the same shift that shifts the phase $\mathbf{U}^{'}$ to $\hat{\mathbf{U}}$. The circular shift $\hat{C}$ from the original payload to the standard form can be decoded from the phase $\hat{\mathbf{U}}$. This can then be used to predict the decoded payload $\hat{\mathbf{P}}$, as indicated in Fig. 2.

Methods

We use simulation combined with a theoretical framework to study the payload decoding rate with different levels of the transmission error rate.

The process of developing this methodology consists of three steps: 1. model the communication channel and transmission error; 2. represent the decoding process; 3. develop a closed form solution for the probability of success decoding rate;

We validate the closed form solution for the probability of successful decoding using the simulation results in the Results section.

Model the communication channel and transmission error

The message embedded in the image is memoryless, so we use a Binary Symmetric Channel to model the communication channel. It has binary input and output, with a probability of transmission error ε, i.e. the probability of switching values between 0 and 1. So the probability of successful transmission one bit is $P(\text{Sucess}) = 1-\varepsilon$, and the probability of failure of transmission of one bit is $P(\text{Fail}) = \varepsilon$. We assume that the probability of successful transmission at each position is independent and identically distributed (i.i.d.).

Represent the decoding process

As noted before, the crop window of the data array $\hat{\mathbf{v}}$ is mixed with payload rows and phase rows, with a fixed interleaving pe-

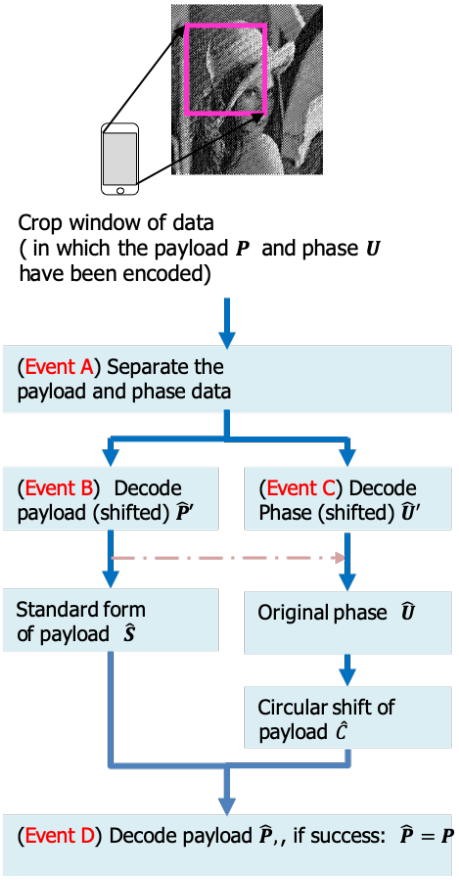


Figure 2: *Overview of the circular coding decoding process. The transmission error will impact the process of these four events: Event A: separating the payload and phase data; Event B: decoding the shifted sequence of the payload; Event C: decoding the shifted sequence of the phase; Event D: decoding the payload. All the other processes are deterministic.*

riod V. But the starting row index of the phase is unknown.

Here is how to decode the payload: First, we align each bit position index by circularly shifting back (shifting left) by D bit positions to the position of the first row in the crop window. So each column of the data array represents the same bit position index of the payload or phase.

Then we rearrange the data array so that each payload bit position is aligned in a column. So the number of columns is B, the payload bit length, and the number of rows is the bit position repeat count, denoted as R.

Among the R repeating bit positions in the data array, let the number of phase repeats be M. Then, the number of payload repeats is $R-M$.

In order to separate the phase from the payload, we try every possible starting row index of the phase (V possible starting row indices), and select the one with highest confidence. Once we try to remove M "phase" rows in the data array, either correctly or incorrectly, there will be $R-M$ rows of data left. There are only two possible cases:

- Case 1: get all the phase rows out, leave $R-M$ repeats of payload rows. There is only one chance over the V possible positions that this case will occur.
- Case 2: get none of the phase rows out, leave $R-M$ rows that are a mixture of payload and phase rows. There will be $V-1$ chances for this to occur. So the number of payload rows left is $R-2M$. Define $N=R-2M$.

Thus, the relationship between R, M, and N is illustrated in Fig. 3. And an example of the two cases is shown in Fig. 4

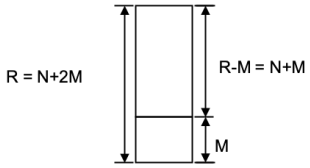


Figure 3: *Illustration of bit position repeat count. The total bit position repeat count in a crop window of data is R, the phase bit position repeat count in a crop window of data is M, and the payload bit position repeat count in a crop window of data is $R-M$. We define $N=R-2M$. Thus, $R-M=N+M$.*

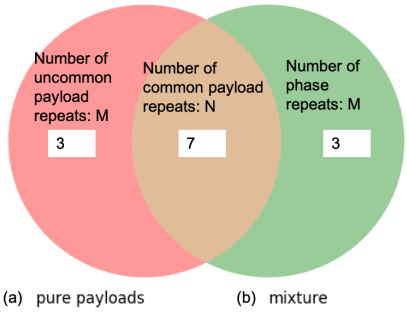


Figure 4: *Illustration of the (a) pure payload subset in Case 1 and (b) mixture of payload and phase subset in Case 2, using a Venn diagram. In this example $N=7$, $M=3$, and $R=13$.*

In either case, the remaining data array contains the common N rows of payload repeats, and the M rows of payload or phase repeats.

Model the status of change as a random variable

Once the data is transferred in the communication channel, the bit values might be switched due to a transmission error. We model the switch of a bit value as a random variable, and it does not depend on the bit position, so it is i.i.d. for each bit.

Then, for each case (Case 1 gets all the phase rows out, or Case 2 gets none of the phase rows out), let the sets $\mathscr{Y}_1$ and $\mathscr{Y}_2$ denote the random variables at each data array position, respectively. Here is the mathematical representation of the random variables:

- For each column of data (each bit position index j) and row of data (each bit repeat index i), let the status of switching its original value be the random variables $X_i^{(j)}$ and $Z_i^{(j)}$ for payload and phase, respectively. So $j=0,\cdots,B-1$, and $i=1,2,\cdots,M$ for Z; and $j=0,\cdots,B-1$ and $i=1,2,\cdots,R-M=N+M$ for X.

- $X_i^{(j)} = 0$: status NOT changed at bit position j and row position i. $P\{X_i^{(j)} = 0\} = 1 - \varepsilon$.
- $X_i^{(j)} = 1$: payload status changed at bit position j and row position i. $P\{X_i^{(j)} = 1\} = \varepsilon$.
- Similarly for $Z_i^{(j)}$, $Z_i^{(j)} = 1$: phase status changed at bit position j and row position i.

Thus, the two sets (pure payload set and mixture of payload and phase set) can be written as shown in in Eqs. 1 and 2, respectively.

$$\mathscr{S}_1 = \begin{bmatrix} X_1^{(0)} & X_1^{(1)} & \dots & X_1^{(j)} & \dots & X_1^{(B-1)} \\ X_2^{(0)} & X_2^{(1)} & \dots & X_2^{(j)} & \dots & X_2^{(B-1)} \\ X_3^{(0)} & X_3^{(1)} & \dots & X_3^{(j)} & \dots & X_3^{(B-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_N^{(0)} & X_N^{(1)} & \dots & X_N^{(j)} & \dots & X_N^{(B-1)} \\ X_{N+1}^{(0)} & X_{N+1}^{(1)} & \dots & X_{N+1}^{(j)} & \dots & X_{N+1}^{(B)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{N+M}^{(0)} & X_{N+M}^{(1)} & \dots & X_{N+M}^{(j)} & \dots & X_{N+M}^{(B-1)} \end{bmatrix} \quad (1)$$

$$\mathscr{S}_2 = \begin{bmatrix} X_1^{(0)} & X_1^{(1)} & \dots & X_1^{(j)} & \dots & X_1^{(B-1)} \\ X_2^{(0)} & X_2^{(1)} & \dots & X_2^{(j)} & \dots & X_2^{(B-1)} \\ X_3^{(0)} & X_3^{(1)} & \dots & X_3^{(j)} & \dots & X_3^{(B-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_N^{(0)} & X_N^{(1)} & \dots & X_N^{(j)} & \dots & X_N^{(B-1)} \\ Z_1^{(0)} & Z_1^{(1)} & \dots & Z_1^{(j)} & \dots & Z_1^{(B-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_M^{(0)} & Z_M^{(1)} & \dots & Z_M^{(j)} & \dots & Z_M^{(B-1)} \end{bmatrix} \quad (2)$$

Note that for a crop window of data $\hat{\mathbf{r}}$, we are able to calculate the bit repeat count for each bit position index. The bit repeat count for each bit position index might be slightly different, depending on the size and position of the crop window. Here, we assume that each bit position index has the same repeat number R.

Model the data array after corruption by transmission errors as a sequence of random variables

For the original payload P and the phase U both with length B bits, we can write their original values as the data arrays:

$$P = [P^{(0)}, P^{(1)}, \dots, P^{(B-1)}] \quad (3)$$

$$U = [U^{(0)}, U^{(1)}, \dots, U^{(B-1)}] \quad (4)$$

Note that we already defined the status of switching value as the random variables in Eqs. 1 and 2, so we can define the value of the data sets in Cases 1 and 2 at each position, as the original value XOR the random variable of a status change at that position, as shown in Eqs. 5 and 6.

$$\tilde{\mathscr{S}}_1 = \begin{bmatrix} \tilde{X}_1^{(0)} & \tilde{X}_1^{(1)} & \dots & \tilde{X}_1^{(j)} & \dots & \tilde{X}_1^{(B-1)} \\ \tilde{X}_2^{(0)} & \tilde{X}_2^{(1)} & \dots & \tilde{X}_2^{(j)} & \dots & \tilde{X}_2^{(B-1)} \\ \tilde{X}_3^{(0)} & \tilde{X}_3^{(1)} & \dots & \tilde{X}_3^{(j)} & \dots & \tilde{X}_3^{(B-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{X}_N^{(0)} & \tilde{X}_N^{(1)} & \dots & \tilde{X}_N^{(j)} & \dots & \tilde{X}_N^{(B-1)} \\ \tilde{X}_{N+1}^{(0)} & \tilde{X}_{N+1}^{(1)} & \dots & \tilde{X}_{N+1}^{(j)} & \dots & \tilde{X}_{N+1}^{(B)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{X}_{N+M}^{(0)} & \tilde{X}_{N+M}^{(1)} & \dots & \tilde{X}_{N+M}^{(j)} & \dots & \tilde{X}_{N+M}^{(B-1)} \end{bmatrix} \quad (5)$$

where $\tilde{X}_i^{(j)} = X_i^{(j)} \bigotimes P^{(j)}$, for $i = 1,2,\dots,N+M$ and $j = 0,1,...,B-1$.

Similarly,

$$\tilde{\mathscr{S}}_2 = \begin{bmatrix} \tilde{X}_1^{(0)} & \tilde{X}_1^{(1)} & \dots & \tilde{X}_1^{(j)} & \dots & \tilde{X}_1^{(B-1)} \\ \tilde{X}_2^{(0)} & \tilde{X}_2^{(1)} & \dots & \tilde{X}_2^{(j)} & \dots & \tilde{X}_2^{(B-1)} \\ \tilde{X}_3^{(0)} & \tilde{X}_3^{(1)} & \dots & \tilde{X}_3^{(j)} & \dots & \tilde{X}_3^{(B-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{X}_N^{(0)} & \tilde{X}_N^{(1)} & \dots & \tilde{X}_N^{(j)} & \dots & \tilde{X}_N^{(B-1)} \\ \tilde{Z}_1^{(0)} & \tilde{Z}_1^{(1)} & \dots & \tilde{Z}_1^{(j)} & \dots & \tilde{Z}_1^{(B-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{Z}_M^{(0)} & \tilde{Z}_M^{(1)} & \dots & \tilde{Z}_M^{(j)} & \dots & \tilde{Z}_M^{(B-1)} \end{bmatrix} \quad (6)$$

where $\tilde{X}_i^{(j)} = X_i^{(j)} \otimes P^{(j)}$, for $i = 1,2,\dots,N$ and $j = 0,1,...,B-1$.

And $\tilde{Z}_i^{(j)} = Z_i^{(j)} \otimes U^{(j)}$, for $i = 1,2,\dots,M$ and $j = 0,1,...,B-1$.

Here $\otimes$ denotes the XOR operation.

Separate the payload and phase and then decode their values

When we compare any two sets (the pure payload set in Case 1 and the mixture of payload and phase set in Case 2), we calculate their confidence values, and then select the one with higher value as the pure payload set.

For the pure payload data set in Eq. 1 and for the mixture of payload and phase data set in Eq. 2, the methods to estimate the bit value are the same. That is, we calculate the confidence value of each data set, and select the one with higher confidence as the pure payload set. It includes the following four steps:

1. Calculate the sample mean of each subset. Let $\bar{Y}_1^{(j)}$ denote the sample mean for the pure payload subset and $\bar{Y}_2^{(j)}$ for the mixture payload and phase subset:

$$\bar{Y}_1{}^{(j)} = \frac{1}{N+M} \left\{ \sum_{i=1}^{N} \tilde{X}_i^{(j)} + \sum_{i=N+1}^{N+M} \tilde{X}_i^{(j)} \right\} \quad (7)$$

$$\bar{Y_2}^{(j)} = \frac{1}{N+M}\left\{\sum_{i=1}^{N}\tilde{X}_i^{(j)} + \sum_{i=1}^{M}\tilde{Z}_i^{(j)}\right\} \tag{8}$$

2. Calculate the estimated bit value for each data set:

$$\hat{Y_k}^{(j)} = \begin{cases} 1, & \text{if } \bar{Y_k}^{(j)} > 0.5 \\ 0, & \text{if } \bar{Y_k}^{(j)} < 0.5 \\ \text{random choice of 0 or 1}, & \text{if } \bar{Y_k}^{(j)} = 0.5 \end{cases} \tag{9}$$

$$\Delta_k^{(j)} = \mid \hat{Y_k}^{(j)} - \bar{Y_k}^{(j)} \mid \tag{10}$$

for $k = 1, 2$.

For the values of bit position repeats, we define the minority bits as the bits with value that appear less frequently than bits with the other value. The other bits with the value that appeared more than half the time are called majority bits. Thus, $\Delta^{(j)}$ is actually calculating the proportion of the minority bits for bit position index j.

3. Calculate the confidence value for each data set:

$$\mathscr{C}_k = 1 - \frac{2}{B}\sum_{j=0}^{B-1}\Delta_k^{(j)} \tag{11}$$

for $k = 1, 2$.

Note that the confidence is in the range of 0 and 0.5, where the higher the confidence is, the more likely this selection is a pure payload data set. For the pure payload data set in Case 1, $k = 1$; for the mixture of payload and phase data set in Case 2, $k = 2$. And the confidence is negatively proportional to the sum of $\Delta^{(j)}$, $j = 0, 1, \ldots, B-1$.

Select the payload data set with the higher confidence value, and assign each bit its value that has been calculated in Step 2.

$$[\hat{Y}^0, \hat{Y}^1, ..., \hat{Y}^{B-1}] = \begin{cases} [\hat{Y}_1^0, \hat{Y}_1^1, ..., \hat{Y}_1^{B-1}], & \text{if } \mathscr{C}_1 > \mathscr{C}_2 \\ [\hat{Y}_2^0, \hat{Y}_2^1, ..., \hat{Y}_2^{B-1}], & \text{if } \mathscr{C}_1 < \mathscr{C}_2 \\ [\hat{Y}_1^0, \hat{Y}_1^1, ..., \hat{Y}_1^{B-1}], & \text{if } \mathscr{C}_1 = \mathscr{C}_2 \end{cases} \tag{12}$$

There will be totally V possible starting row positions for the phase rows. For a successful decoding, it is required that the pure payload data set has higher confidence than any mixture payload data set in each comparison.

Develop a closed form solution for the probability of successful decoding

Recall that the decoding process has three steps that involve the error estimation: 1. separate the phase from the payload; 2. decode the standard form of the payload; 3. decode the phase. We need to estimate the success rate of each of these three steps, and combine the success rates of these three steps to get the final decoding rate.

Step 1: Compute the probability of separating the phase from the payload

The phase rows and payload rows are selected based on the confidence estimation, see Eqs. 11 and 12. Note that in order to correctly decode the payload, we can conclude that at least half of the bits need to retain their original value during the transmission. So we can assume that the probability of transmission error $\varepsilon < 0.5$. Higher confidence is equivalent to lower uncertainty, and equivalent to fewer minority bits. Thus, the probability of separating the phase from the payload can be computed by summing the probabilities that the pure payload data set has fewer minority bits than the mixed phase and payload data set, for each possible number of bits that switched value in the pure payload data set.

We first consider the easier case that the payload only has one bit. Then, we extend the payload bit length to multiple bits.

Payload bit length of one: $B = 1$

Let the payload length be one bit only, i.e. $B = 1$. So the phase is also one bit. Note this cannot be done in the real case.

The original payload value is $P^{(1)}$, and the phase value is $U^{(1)}$. So there are two possible situations: the original payload value is either the same or different, compared with the phase value (i.e. $P^{(1)} = U^{(1)}$, or $P^{(1)} \neq U^{(1)}$).

After we rearrange the crop window of data array, let's assume that there is $M+N$ bits of repeats. Thus, the bit repeats in case 1 (get all phases out) and in case 2 (get none of the phases out) will be simplified as $\mathscr{S}_1|_{B=1}$ and $\mathscr{S}_2|_{B=1}$, defined in Equation 13.

$$\mathscr{S}_1|_{B=1} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_N \\ \\ X_{N+1} \\ \vdots \\ X_{N+M} \end{bmatrix} \text{ and } \mathscr{S}_2|_{B=1} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_N \\ \\ Z_1 \\ \vdots \\ Z_M \end{bmatrix} \tag{13}$$

Let $\mathbb{A}$ represent the data set in Case 1, i.e. the pure payload data set, and let $\mathbb{B}$ represent the data set in Case 2, i.e. the mixture of payload and phase.

For the comparison of the bit repeats in Cases 1 and 2, there always N bits of payload $X_1, X_2, \ldots, X_N$ shared in common. In other words, if any bits have switched value in the common payload bits area, they will be the same for the data sets $\mathbb{A}$ and $\mathbb{B}$. But the status of switching value in the remaining parts are different and independent of each other.

Our goal is to find the probability that the total number of minority bits in data set $\mathbb{A}$ is less than the number of minority bits in data set $\mathbb{B}$.

We define α as the number of minority bits in data set $\mathbb{A}$, and β as the number of minority bits in data set $\mathbb{B}$. And we assume that fewer than half of the bits have switched value. Then, the number of minority bits is the same as the number of bits that switched value in each data set. For each number of minority bits in data set $\mathbb{B}$, there should be fewer minority bits in data set $\mathbb{A}$, or

$\alpha < \beta$. Also, the number of minority bits α in data set $\mathbb{A}$ should be fewer than half of the total number of bit repeats. Thus, we define:

$$\mathscr{K} = \lfloor \frac{M+N}{2} - 1 \rfloor \tag{14}$$

So the number of minority bits α and β can be any number between zero and $\mathscr{K}$. The α bits of positions that switched value could be in either the N bits of common payload or the M bits of uncommon payload in data set $\mathbb{A}$.

By the nature of the circular encoding process, the number of payload bit repeats N is much larger than the number of phase bit repeats M in each data set, i.e. $N \gg M$. So if the number of bits whose values are switched in the data set $\mathbb{A}$ is no more than the number M, i.e $\alpha \in [0,M]$, then these bits can be anywhere in the payload or phase. Otherwise, if the number of bits with switched values is greater than M, there will be at maximum M bits with switched values in the uncommon payload in data set $\mathbb{A}$ or the phase in data set $\mathbb{B}$, and the remaining bits that switched value will be in the common payload.

As noted before, with the assumption that fewer than half of the bits switched value, the number that switched value in the pure payload data set is the minority bit number α. Among these α bits that switched value, let m denote the number of bits that switched value in the uncommon payload. Thus,

- Part 1: When $\alpha \in [0,M]$, then the number of bits m that switched value in the data set $\mathbb{A}$ in the uncommon payload could be any number between 0 and α, or $m \in [0,\alpha]$. And the number of bits that switched value in the data set $\mathbb{A}$ in the common payload is $\alpha - m$.

- Part 2: When $\alpha \in (M,\mathscr{K}]$, then m could be any number between 0 and α, or $m \in [0,M]$. And the number of bits that switched value in the data set $\mathbb{A}$ in the common payload is also $\alpha - m$.

Let k be the number of bits that switched value in the M bits of the phase in the data set $\mathbb{B}$. In order to successfully separate the pure payload data set, it is required that the minority bit number α in the data set $\mathbb{A}$ be less than the minority bit number β in data set $\mathbb{B}$.

First, if the original payload value is the same as the phase value ($P^{(1)} = U^{(1)}$), then the number α of bits that switched can take any value between 0 and half of the total number of bit repeats $\mathscr{K}$. So we sum the probability that data set $\mathbb{A}$ has fewer minority bits than data set $\mathbb{B}$ for each α value. It includes two parts. Part 1: $\alpha \in [0,M]$; Part 2: $\alpha \in (M,\mathscr{K}]$.

For the data set $\mathbb{A}$, the minority bits are the ones that switched value. There are α such bits. For the data set $\mathbb{B}$, the number of bits that switched value must be greater than α, but smaller than the total number of the bits in each data set minus α, or $M+N-\alpha$. In other words, in order to have the number of bits that switched value in the data set $\mathbb{A}$ be smaller than the number of bits that switched value in the data set $\mathbb{B}$, there should be fewer bits that switched value in the uncommon payload in data set $\mathbb{A}$ than the number that switched value in the payload in data set $\mathbb{B}$, given the condition that the payload and phase have the same original value. Thus, these five conditions need to be met:

$$\begin{cases} \alpha \in [0,M] \\ \alpha < \beta < M+N-\alpha \\ \beta = (\alpha - m) + k \\ m < k \\ m \in [0,\alpha] \end{cases}$$

From these five conditions, we can obtain the range of k:

$$m < k < M+N-2\alpha+m \tag{15}$$

For Part 2, $\alpha \in (M,\mathscr{K}]$. Since there are only M bits in the uncommon payload part, there will be a maximum of M bits that could switch value in the uncommon payload, the remaining $\alpha - m$ bits must switch value in the common payload. The conditions for data set $\mathbb{B}$ are the same as those as those in Part 1. Please refer to Table 1 for details.

Next, we consider the case where the original payload value is different from the phase value. This is very similar to the case where the original payload value is the same as the phase value, except that for the mixture payload and phase data set $\mathbb{B}$, the phase data will be originally treated as the minority bits. This is because the number of payload bits is usually much larger than the number of phase bits, or $N \gg M$. So for any bit position j, we expect that there will be fewer phase rows than payload rows. Also, we assume that fewer than half of the bits switched their values. So the minority bits contain the bits that switched value in the common payload part in data set $\mathbb{B}$ and the bits that retained their original value in the phase part of data set $\mathbb{B}$. Thus, we again have five conditions that need to be met:

$$\begin{cases} \alpha \in (M,\mathscr{K}] \\ \alpha < \beta < M+N-\alpha \\ \beta = (\alpha - m) + (M-k) \\ m < M-k \\ m \in [0,M] \end{cases}$$

From these five conditions, we can obtain the range of k:

$$2\alpha - m - N < k < M - m \tag{16}$$

The details are summarized in Table 2.

Now we can write out the probability that the pure payload set will be correctly distinguished from the mixture subset. First, let's define the probability mass function for the binomial distribution. We define the status of switching value at each bit position as a random variable with a binomial distribution. The probability of getting exactly a successes in A independent Bernoulli trials is given by the probability mass function:

$$P(A,a,\xi) = \begin{pmatrix} A \\ a \end{pmatrix} \xi^a (1-\xi)^{A-a} \tag{17}$$

This calculates the probability that for every A bit positions, a bits switch value, when the probability of switching value at each position is ξ.

Table 1: The conditions under which the pure payload selection $\mathbb{A}$ has fewer minority bits than the mixture of payload and phase selection $\mathbb{B}$, when the phase original value is the same as the payload value. $\mathscr{K} = \lfloor \frac{M+N}{2} - 1 \rfloor$.

Original payload value is **same** as the phase value		
	$\mathbb{A}$ has α minority bits	$\mathbb{B}$ has β minority bits , $\beta > \alpha$
Part 1: $\alpha \in [0, M]$	m bits switch value in the uncommon payload, and $\alpha - m$ bits switch value in the common payload, where $m \in [0, \alpha]$	the same $\alpha - m$ bits switch value in the common payload, and k bits switch value in the phase, where $\beta = \alpha - m + k$, $\alpha < \beta < (M + N - \alpha)$
Part 2: $\alpha \in (M, \mathscr{K}]$	m bits switch value in the uncommon payload, and $\alpha - m$ bits switch value in the common payload, where $m \in [0, M]$	the same $\alpha - m$ bits switch value in the common payload, and k bits switch value in the phase, where $\beta = \alpha - m + k$, $\alpha < \beta < (M + N - \alpha)$

Under the assumption that the original payload value is the same as the phase value, i.e $P^{(1)} = U^{(1)}$, the probability that the number of minority bits α in the data set $\mathbb{A}$ being less than the number of minority bits β in the data set $\mathbb{B}$ is $P\{\alpha < \beta | P^{(1)} = U^{(1)}\}$. It is the cumulative distribution that for every possible number of bits i that switched value, where $i \in ([0, M-1] \cup [M, \mathscr{K}])$, the number of bits α that switched value in the data set $\mathbb{A}$ is less than the number of bits β that switched value in the data set $\mathbb{B}$.

From Table 1, we add the probabilities for first and second parts. For the first part, $i \in [0, M-1]$, the probability that m bits switched value in the M bits of uncommon payload is $P(M, m, \varepsilon)$. For the remaining bits that switched value in the common payload area, the probability is $P(N, i-m, \varepsilon)$; and the probability that k bits switched value in the phase data set is $P(M, k, \varepsilon)$. So the Part 1 conditional probability $P\{(\alpha < \beta) \cap (0 \le \alpha \le M) | B = 1, P^{(0)} = U^{(0)}\}$ that there are fewer minority bits in the pure payload data set $\mathbb{A}$ than in the mixture of payload and phase data set $\mathbb{B}$, when the original payload value is the same as phase value, is shown in Eq. 18. It is similar for Part 2, which is shown in Eq. 19. See Table 1 for the definition of Parts 1 and 2.

$$P\{(\alpha < \beta) \cap (0 \le \alpha \le M) | B = 1, P^{(0)} = U^{(0)}\} = \sum_{i=0}^{M-1} \left\{ \sum_{m=0}^{i} P(M, m, \varepsilon) \cdot P(N, i-m, \varepsilon) \cdot \sum_{k=m+1}^{M+N-2i+m-1} P(M, k, \varepsilon) \right\} \tag{18}$$

$$P\{(\alpha < \beta) \cap (M < \alpha \le \mathscr{K}) | B = 1, P^{(0)} = U^{(0)}\} = \sum_{i=M}^{\lfloor \frac{M+N}{2} - 1 \rfloor} \left\{ \sum_{m=0}^{M} P(M, m, \varepsilon) \cdot P(N, i-m, \varepsilon) \cdot \sum_{k=m+1}^{M+N-2i+m-1} P(M, k, \varepsilon) \right\} \tag{19}$$

Table 2: The conditions under which the pure payload selection $\mathbb{A}$ has fewer minority bits than the mixture of payload and phase selection $\mathbb{B}$, when the phase original value is different from the payload value. $\mathscr{K} = \lfloor \frac{M+N}{2} - 1 \rfloor$.

Original payload value is **different** from the phase value		
	$\mathbb{A}$ has α minority bits	$\mathbb{B}$ has β minority bits , $\beta > \alpha$
Part 1: $\alpha \in [0, M]$	m bits switch value in the uncommon payload, and $\alpha - m$ bits switch value in the common payload, where $m \in [0, \alpha]$	the same $\alpha - m$ bits switch value in the common payload, and $M - k$ bits switch value in the phase, where $\beta = (\alpha - m) + (M - k)$, $\alpha < \beta < (M + N - \alpha)$
Part 2: $\alpha \in (M, \mathscr{K}]$	m bits switch value in the uncommon payload, and $\alpha - m$ bits switch value in the common payload, where $m \in [0, M]$	the same $\alpha - m$ bits switch value in the common payload, and $M - k$ bits switch value in the phase, where $\beta = (\alpha - m) + (M - k)$, $\alpha < \beta < (M + N - \alpha)$

Then, we sum the probabilities of these two parts to get the probability that the data set $\mathbb{A}$ has fewer minority bits than $\mathbb{B}$ when the original phase value is the same as payload value, defined as $P\{\alpha < \beta | B = 1, P^{(0)} = U^{(0)}\}$ in Eq. 20.

$$\begin{aligned} &P\{\alpha < \beta | B = 1, P^{(0)} = U^{(0)}\} \\ &= P\{(\alpha < \beta) \cap (0 \le \alpha \le M) | B = 1, P^{(0)} = U^{(0)}\} \\ &+ P\{(\alpha < \beta) \cap (M < \alpha \le \mathscr{K}) | B = 1, P^{(0)} = U^{(0)}\} \end{aligned} \tag{20}$$

Under the assumption that the original payload value is different from the phase value, i.e $P^{(0)} \neq U^{(0)}$, the calculation of the probability that the data set $\mathbb{A}$ has fewer minority bits than the data set $\mathbb{B}$ is very similar to the assumption of $P^{(0)} = U^{(0)}$, except that we need to count the number of bits that retain their original value in the phase data in the data set $\mathbb{B}$ as the minority bits. The probability that the data set $\mathbb{A}$ has fewer minority bits than set $\mathbb{B}$ is calculated in Eqs. 21 and 22 for Parts 1 and 2, respectively.

$$P\{(\alpha < \beta) \cap (0 \le \alpha \le M) | B = 1, P^{(0)} \neq U^{(0)}\} = \sum_{i=0}^{M-1} \left\{ \sum_{m=0}^{i} P(M, m, \varepsilon) \cdot P(N, i-m, \varepsilon) \cdot \sum_{k=2i-m-N+1}^{M-m-1} P(M, k, \varepsilon) \right\} \tag{21}$$

$$P\{(\alpha < \beta) \cap (M < \alpha \le \mathscr{K}) | B = 1, P^{(0)} \neq U^{(0)}\} = \sum_{i=M}^{\lfloor \frac{M+N}{2} - 1 \rfloor} \left\{ \sum_{m=0}^{M} P(M, m, \varepsilon) \cdot P(N, i-m, \varepsilon) \cdot \sum_{k=2i-m-N+1}^{M-m-1} P(M, k, \varepsilon) \right\} \tag{22}$$

We sum the probabilities of these two parts to get the probability that the data set $\mathbb{A}$ has fewer minority bits than set $\mathbb{B}$, defined as $P\{\alpha < \beta | B = 1, P^{(0)} \neq U^{(0)}\}$ in Eq. 23.

$$
\begin{aligned}
&P\{\alpha<\beta|B=1,P^{(0)}\neq U^{(0)}\}\\
&=P\{(\alpha<\beta)\cap(0\leq\alpha\leq M)|B=1,P^{(0)}\neq U^{(0)}\}\\
&+P\{(\alpha<\beta)\cap(M<\alpha\leq\mathscr{K})|B=1,P^{(0)}\neq U^{(0)}\}
\end{aligned}
\tag{23}
$$

Then, the total probability that the data set $\mathbb{A}$ has fewer minority bits than the data set $\mathbb{B}$ is calculated in Eq. 24 based on the total probability law.

$$
\begin{aligned}
P\{\alpha<\beta|B=1\}=&\\
&P\{\alpha<\beta|B=1,P^{(0)}=U^{(0)}\}\cdot P\{P^{(0)}=U^{(0)}\}\\
&+P\{\alpha<\beta|B=1,P^{(0)}\neq U^{(0)}\}\cdot P\{P^{(0)}\neq U^{(0)}\}
\end{aligned}
\tag{24}
$$

Extend the number of payload bits from $B=1$ to $B>1$

When the payload length B is greater than one bit, the confidence can be calculated in Eq. 11. The confidence is the sum of Δ^j, $j\in(0,B-1)$. So the probability that the confidence is greater in the data set $\mathbb{A}$ than in the data set $\mathbb{B}$ is the same as the probability that the sum of Δ^j in the data set $\mathbb{A}$ is smaller than in the data set $\mathbb{B}$.

$$
P\{\mathscr{C}_1>\mathscr{C}_2\}=P\left\{\sum_{j=0}^{B-1}\Delta_1^j<\sum_{j=0}^{B-1}\Delta_2^j\right\}
\tag{25}
$$

Note that we already discussed that for bit payload length $B=1$, if we assume that fewer than half of the bits switched value for a given number of bit repeats R, the confidence is just the proportion of how many bits switched value over the total number of bit repeats. Now the payload is more than one bit in length, and we assume that fewer than half of the bits switched value for each bit position. Thus, the confidence, calculated as the sum of Δ^j, $j\in(0,B-1)$ for the data set $\mathbb{A}$ and the data set $\mathbb{B}$, is the sum of the number of bits that switched value among the total number of bit positions.

In Eq. 24, we already developed a closed form solution for each bit position assuming that the pure payload set has fewer minority bits α than the mixture of payload and phase, which has β minority bits. It is a function of the common payload bit repeat count N, the phase bit repeat count M in the mixture of payload and phase data sets, and the bit error probability ε. So for $B=1$ we can define it as in Eq. 26:

$$
P\{\alpha<\beta|B=1\}\equiv f(\alpha<\beta;N,M,\varepsilon)
\tag{26}
$$

For $B>1$, with the assumption that fewer than half of the bits switched value in each bit position index, the number of minority bits in each bit position index is the same as the number of bits that switched value in that bit position index. In addition, from our experiments, we found that when the payload bit length is large (i.e. $B>63$), the probability that the payload value is the same as the phase value at each bit position index is about 0.5. Thus, the sum of the minority bits for the whole payload bit length B can be separated into two parts: one for those bit position indices where the phase bit has the same original value as the payload, and the other for those bit position indices where the phase original value is different from the payload. Since the random variables that determine whether or not the bits switch value are i.i.d., at each bit repeat, the minority bit difference mainly comes from the first part: those bits where the payload and phase have different original values.

So, we can rewrite the probability $P(\mathscr{C}_1>\mathscr{C}_2)$ using the formula developed for payload bit length $B=1$, but with the new variables, as shown in Eq. 27. This approximation is validated using simulation.

$$
P\{\mathscr{C}_1>\mathscr{C}_2\}=P\{\alpha<\beta\}\approx f(\alpha<\beta;\frac{N}{2}B,\frac{M}{2}B,\varepsilon)
\tag{27}
$$

Step 2: Compute the conditional probability of successfully decoding the payload

Now assume that we already correctly separated the phase from the payload; so for each bit position, we have $N+M$ bit repeats for data sets $\mathbb{A}$ and $\mathbb{B}$. We will compute the probability of successfully decoding the payload conditioned on the event $\mathscr{C}_1>\mathscr{C}_2$.

The conditional probability that the detected bit value $\hat{\mathbf{P}}^{'(j)}$ is the same as the original bit value $\mathbf{P}^{'(j)}$ is equal to the conditional probability that fewer than half of the bits switched value. It can we calculated as:

$$
\begin{aligned}
&P\{\hat{\mathbf{P}}^{'(j)}=\mathbf{P}^{'(j)}|\mathscr{C}_1>\mathscr{C}_2\}\\
&\quad=\sum_{k=0}^{\mathscr{K}}\left(\begin{array}{c}(N+M)\\k\end{array}\right)(\varepsilon)^k(1-\varepsilon)^{(N+M)-k}
\end{aligned}
\tag{28}
$$

where $\mathscr{K}$ is half of the number of bit repeats for the payload:

$$
\mathscr{K}=\left\lfloor\frac{(N+M)}{2}-1\right\rfloor
\tag{29}
$$

The conditional probability that the entire payload is correctly decoded is the joint conditional probability that every bit in the payload is correctly decoded. Recall that we model the transmission error as identical and independent at each bit position, so the joint probability of successfully decoding the entire payload is just the product of the probabilities of successfully decoding at each bit position.

$$
\begin{aligned}
&P\{\hat{\mathbf{P}}^{'}=\mathbf{P}|\mathscr{C}_1>\mathscr{C}_2\}\\
&=\prod_{j=0}^{B-1}P\{\hat{\mathbf{P}}^{(j)}=\mathbf{P}^{(j)}|\mathscr{C}_1>\mathscr{C}_2\}\\
&=\left\{\sum_{k=0}^{\mathscr{K}}\left(\begin{array}{c}(N+M)\\k\end{array}\right)(\varepsilon)^k(1-\varepsilon)^{(N+M)-k}\right\}^B
\end{aligned}
\tag{30}
$$

Step 3: Compute the conditional probability of successfully decoding the phase

Note that for the phase encoding, we will encode the minimum number of bit shifts to go from the standard form S to the payload P. This number is denoted as C. We transfer the decimal

value C to a binary string, denoted as $\dot{\mathbf{U}}$. The maximum number of bits c needed to represent the decimal value C can be calculated in Eq. 31. The method we use to to encode the phase row is to repeat the string $\dot{\mathbf{U}}$ until all the B bits of length are used to form the phase. So for a phase row $\mathbf{U}$ with bit length B, the actual bit-repeat count $\dot{M}$ can be calculated as shown in Eq. 31, since in each phase row, we repeat each bit in the binary representation of C approximately $\frac{B}{c}$ times.

$$c = \lceil \log_2 B \rceil \tag{31}$$

$$\dot{M} \approx \frac{B}{c} M \tag{32}$$

Let's denote the decoded circularly shifted phase as $\hat{\dot{\mathbf{U}}}'$; so we have:

$$\hat{\mathbf{U}}' = [\hat{\mathbf{U}}'^{(0)}, \hat{\mathbf{U}}'^{(1)}, \ldots, \hat{\mathbf{U}}'^{(c-1)}] \tag{33}$$

$$\hat{\dot{\mathbf{U}}}' = [\hat{\dot{\mathbf{U}}}'^{(0)}, \hat{\dot{\mathbf{U}}}'^{(1)}, \ldots, \hat{\dot{\mathbf{U}}}'^{(c-1)}] \tag{34}$$

Similar to the requirement to correctly decode the payload, in order to decode the phase, we require that fewer than half of the phase bits change their value during the transmission. Note that here we assume that each phase bit has the same number M of repeat rows. Thus,

$$\dot{\mathscr{K}} = \left\lfloor \frac{\dot{M}}{2} - 1 \right\rfloor \tag{35}$$

$$\begin{aligned} P\{\hat{\dot{\mathbf{U}}}' = \dot{\mathbf{U}}' | \mathscr{C}_1 > \mathscr{C}_2\} &= \prod_{j=0}^{c-1} P\{\hat{\dot{\mathbf{U}}}'^{(j)} = \dot{\mathbf{U}}'^{(j)} | \mathscr{C}_1 > \mathscr{C}_2\} \\ &= \left\{ \sum_{k=0}^{\dot{\mathscr{K}}} \binom{M'}{k} (\varepsilon)^k (1-\varepsilon)^{M'-k} \right\}^c \end{aligned} \tag{36}$$

Step 4: Compute the final decoding rate

Without correctly separating the payload and phase data array, the chance to correctly decode the payload and phase bits is very low. Thus, we can approximate the probability of successfully decoding the payload as the product of the conditional probability of successfully decoding the payload and phase given that the payload and phase data are successfully separated, and the probability of successfully separating the payload and phase data.

Thus, the final decoding rate can be computed as the product of Eqs. 25, 30, and 36:

$$\begin{aligned} &P\{\hat{\mathbf{P}} = \mathbf{P}\} = P\{\hat{\mathbf{P}}' = \mathbf{P}'\} \cdot P\{\hat{\dot{\mathbf{U}}}' = \dot{\mathbf{U}}'\} \\ &= P\{\hat{\mathbf{P}}' = \mathbf{P}' | \mathscr{C}_1 > \mathscr{C}_2\} \cdot P\{\hat{\mathbf{U}}' = \dot{\mathbf{U}}' | \mathscr{C}_1 > \mathscr{C}_2\} \cdot P\{\mathscr{C}_1 > \mathscr{C}_2\} \end{aligned} \tag{37}$$

Results: Validate the closed form solution

The closed form solution is validated with experimental results. Here is the procedure of the experiment:

1. Randomly generate one sequence of payload data with a given length.
2. Encode the data array (circular coding).
3. Generate i.i.d. sequence of error values.
4. Crop the data array.
5. Decode the payload.
6. Calculate the average decoding success rate.

The number of bit repeats for each position might be different due to the crop window size and location. To simplify the calculation, we set the crop window height H to be an integer, which is a multiple of the interleaving phase period V. Thus, among the H rows of data in the crop window, there are H/V rows of phase, and $H \cdot (V-1)/V$ rows of payload. In addition, we assume that the number of columns W is also an integer multiple of the payload length B. So for each row in the crop window, there will be the same number of occurrences for each value of the bit position index j.

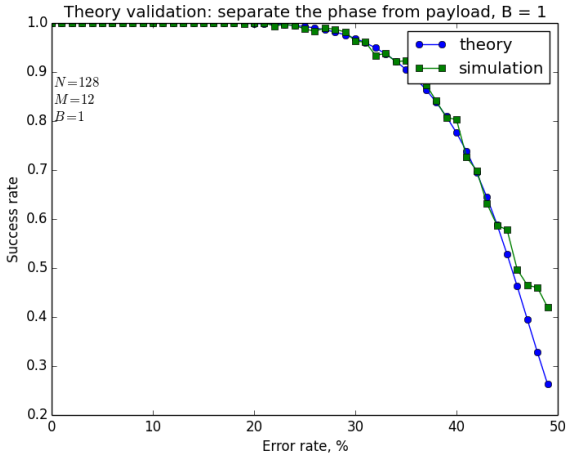
We start from the simplest case, $B = 1$, and then extend it to many bits, i.e. $B > 1$. The transmission error ε is randomly generated, and sampled from 1% to 50%, with a step size of 1%. The validation is done for each major equation, including:

- Probability of separating payload and phase data set in Eq. 27 (refer to Fig. 5). The simulated curve closely matches the theoretical curve, which proves the closed-form formula is well approximated.
- Conditional probability of decoding the payload in Eq. 30. This part was validated by Sun *et al*. [6].
- Conditional probability of decoding the phase in Eq. 36. The closed form solution and validation process is very similar to the Eq. 30.
- The final probability of decoding the original payload in Eq. 37. The validation of the simulation result with theoretical solution is shown in Fig. 7. Here it also shows the theatrical and simulated results are well matched.

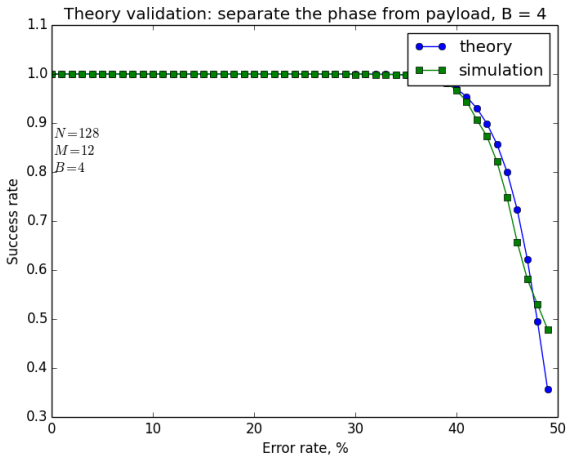
Note that for the simulation, when the confidence is exactly 50%, the decoder will select the first phase row. But in the formula, we require that fewer than half of the repeating bits switch value. To accommodate this situation, we take an average of the floor and ceiling operations of the bits for which the confidence is 50%. The simulation result is shown in Fig. 6

The Euclidean distance between the simulated and theoretical results will decrease when the number of simulation trials increases (10k $\rightarrow$ 40k $\rightarrow$ 100k), which is shown in Fig. 8. In other words, the simulation approaches the theory asymptotically.

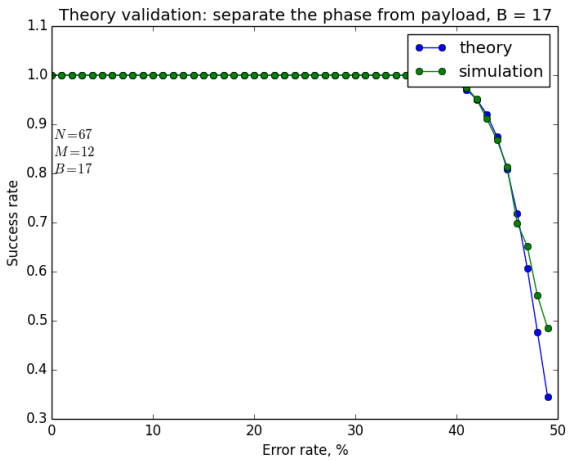
When the bit length becomes very large ($B \geq 67$), the probability P_1 of separating the payload and phase becomes very high compared to the probability P_2 and P_3 of decoding the payload and the phase, respectively. So we can use $P_2 P_3 \approx P_1 P_2 P_3$ to simplify the computation (refer to Fig. 9).



(a)



(b)



(c)

Figure 5: *Comparison of probability of successfully separating payload and phase bits based on theory and simulation for (a)* $N = 128$, $M = 12$, $B = 1$; *(b)* $N = 128$, $M = 12$, $B = 4$; *(c)* $N = 128$, $M = 12$, $B = 7$.

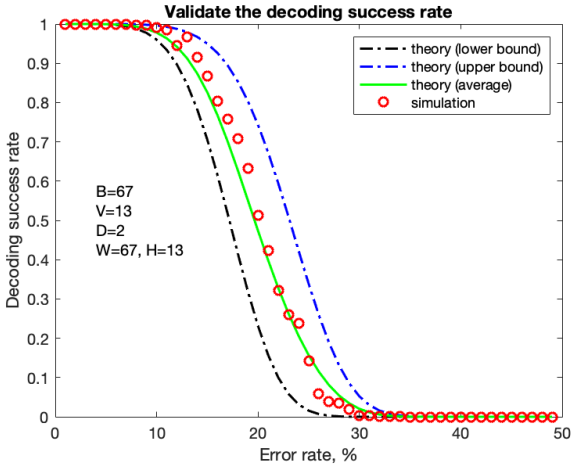


Figure 6: *Validation of the theory by simulation,* $B = 67$, $V = 13$, $D = 2$, $W = 67$, $H = 13$, $V = 13$.

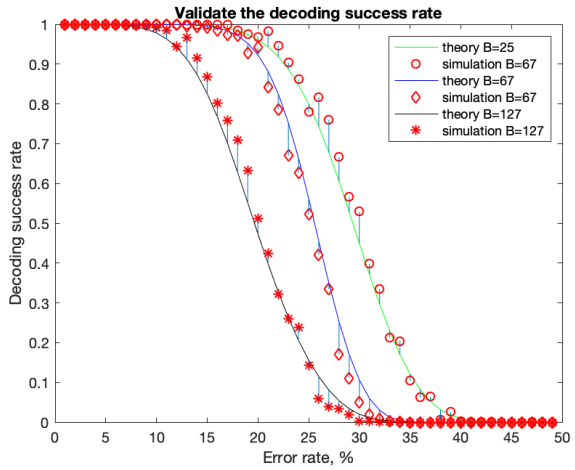


Figure 7: *Validation of the theory by simulation: the final decoding rate. The simulated decoding success rate is the average of 40k different samples of error at each transmission error rate. For the first comparison group,* $B = 25$, $N = 15$, *and* $M = 4$; *for the second comparison group,* $B = 67$, $N = 14$, *and* $M = 2$; *for the last comparison group,* $B = 128$, $N = 22$, *and* $M = 1$.

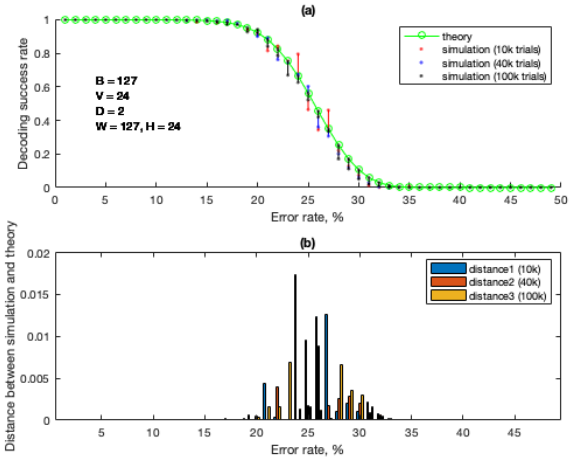


Figure 8: *Effect of increasing the number of simulation trials on the match between the theoretical ad simulation results. (a) Decoding success rate as a function of error rate. (b) The Euclidean distance between the simulated and theoretical results.*

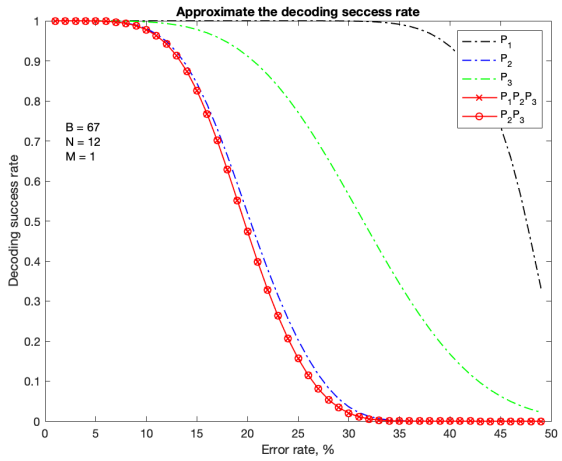


Figure 9: *The approximation of the final decoding rate. P_1: the probability of separating payload and phase defined in Eq. 27; P_2: the conditional probability of decoding the payload in Eq. 30; P_3: the conditional probability of decoding the phase in Eq. 36; $P_1P_2P_3$: the final probability of decoding the original payload in Eq. 37. P_2P_3: the approximation of the final probability of decoding the original payload in Eq. 37. The simulated decoding success rate is the average of 40k different samples of the error at each transmission error rate.*

Conclusion

To the best of our knowledge, this is the first paper to analyze the performance of the circular coding method in a noisy channel. A closed form solution is developed to calculate the decoding success rate for a given message payload length and bit position repeat count under different transmission error rates. This closed form solution is also validated by simulating the decoding process with noisy samples. With this decoding rate prediction, we can design the encoding/decoding system with the desired performance under different given transmission error rates. On the other hand, for a given encoding/decoding system, we will have the expected success rate as a measure of confidence for users.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] R. G. Gallager, *Information Theory and Reliable Communication*, vol. 2. Springer, 1968.

[3] S. Lin and D. J. Costello, *Error Control Coding*, vol. 2. Prentice Hall, 2001.

[4] R. Ulichney, M. Gaubatz, and S. Simske, "Circular Coding for Data Embedding," in *NIP & Digital Fabrication Conference*, vol. 2013, pp. 142–147, Society for Imaging Science and Technology, 2013.

[5] R. Ulichney, M. Gaubatz, and S. Simske, "Circular Coding with Interleaving Phase," in *Proceedings of the 2014 ACM Symposium on Document Engineering*, pp. 21–24, ACM, 2014.

[6] Y. Sun, R. Ulichney, M. Gaubatz, S. Pollard, S. Simske, and J. P. Allebach, "Analysis of a Visually Significant Bar Code System Based on Circular Coding," *Color Imaging XXIII: Displaying, Processing, Hardcopy, and Application*, vol. 2018, 29 Janurary - 2 February 2018.

[7] C.-J. Tai, R. Ulichney, and J. P. Allebach, "Effects on Fourier Peaks Used for Periodic Pattern Detection," *Color Imaging XXIII: Displaying, Processing, Hardcopy, and Application*, vol. 2016, no. 13, pp. 14–18, 2016.

[8] G. D. Boreman, *Modulation Transfer Function in Optical and Electro-optical Systems*, vol. 10. SPIE press Bellingham, WA, 2001.

[9] Z. Zhao, R. Ulichney, M. Gaubatz, S. Pollard, and J. P. Allebach, "Advances in the Decoding of Data-Bearing Halftone Images," in *NIP & Digital Fabrication Conference*, vol. 2019, pp. 162–167, Society for Imaging Science and Technology, 2019.

[10] R. Palanki and J. S. Yedidia, "Rateless Codes on Noisy Channels," in *ISIT*, p. 37, Citeseer, 2004.

[11] M. Luby, D. Fountain, A. Shokrollahi, M. Watson, D. Fountain, and T. Stockhammer, "Raptor Forward Error Correction Scheme for Object Delivery," in *IETF RMT Working Group, Work in Progress*, 2007.

[12] D. J. MacKay, "Fountain Codes," *IEE Proceedings-Communications*, vol. 152, no. 6, pp. 1062–1068, 2005.

[13] M. Luby, A. Shokrollahi, M. Watson, T. Stockhammer, and L. Minder, "Raptor Forward Error Correction Scheme for Object Delivery," tech. rep., RFC 5053 (Proposed Standard), 2007.

Author Biography

Yufang Sun received her BS in Electrical Engineering from the University of Jilin from China (2004). She is currently a PhD student, working as an image processing and data analysis research assistant with Prof. Jan Allebach, in the School of Electrical and Computer Engineering at Purdue University. Her research interests are in image information embedding, decoding error analysis, etc. She has been working on the projects of circular coding and stegaframe detection, both sponsored by HP Labs.

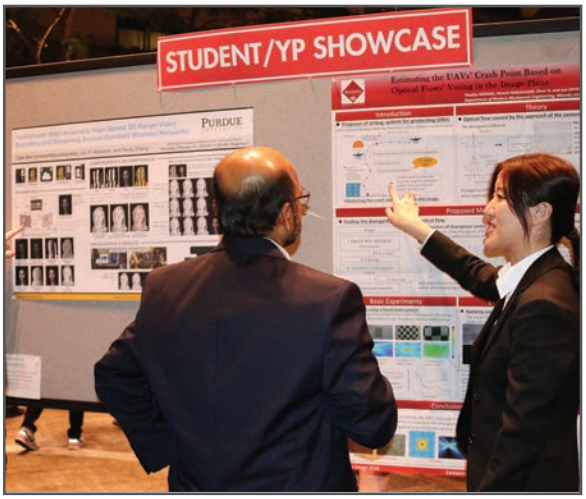
JOIN US AT THE NEXT EI!

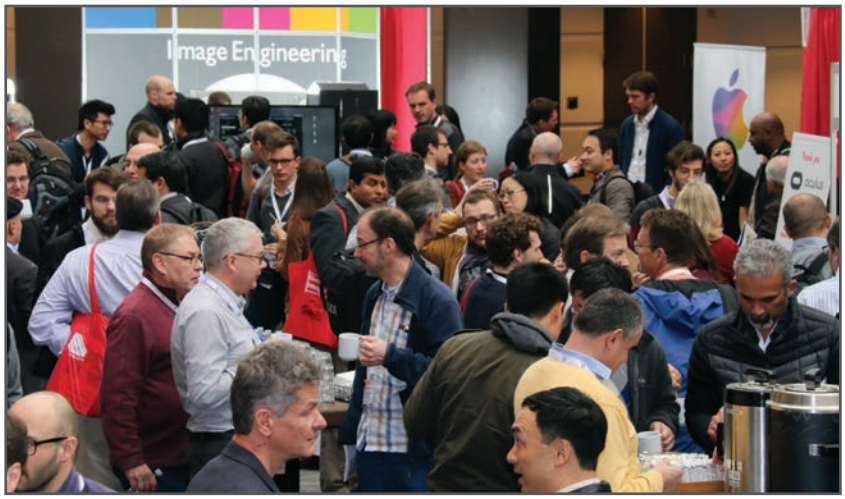
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org