A Differential Privacy Encrypted Communication Method Based on Transmit Power Allocation

In the field of machine learning, a new concept called differential privacy was introduced in recent years. By adding a specific distribution of noise to the data, the deterministic values of the data will be indistinguishable. Since the distribution characteristics of the noise are known and regular, the statistical characteristics of the original data, such as mean and variance, can still be estimated. This concept can be applied to wireless communication as well. In some wireless communication scenarios, users want their data to be confidential to the endpoint, while the endpoint requires access to the user’s overall statistics for further calculation. In this study, we found that, by controlling the transmit power in a communication system, the data error caused by the channel and additive Gaussian white noise satisfies a specific distribution. Thus, the received data can be regarded as the sum of the original data and a specific distributed noise.

Typically, a critical evaluation metric in a wireless communication system is the bit error rate, or in other words, bit error probability. Usually, we define the bit error probability as , and it is the consequence of the effects of the channel and additive Gaussian white noise. To be more specific, in a BPSK system where transmitted symbols are "*a*" and "*-a*", if we transmit a symbol "*a*" through a wireless channel, the receiving detector has the probability to mistake it for the opposite symbol "*-a*".

Assume that a decimal digit is transmitted in a packet of 4 bits, i.e., one number is represented by four bits and the range of the digit is [0,15]. For example, If the transmitted number is 10, the actual bit stream transmitted by the system is 1010, and if the transmitted number is 0, the system will transmit 0000 instead of a single 0.

Define data error as , the purpose of our work is to study the distribution of data errors and see if there is any possibility of controlling the error distribution manually. Since the digits are regenerated after the digital receiver, it is clear that the error distribution is directly related to . According to wireless communication theory, is only related to the signal-to-noise ratio in a fixed-channel wireless communication system, and channel noise power usually remains constant over time. Therefore, the purpose of our study shifts to controlling the data error distribution by regulating the transmit power.

# Modeling Methods and Measurements

In the first section, we clarify the system architecture and noise evaluation metrics used in this project.

## Baseband Equivalent System Model

All the discussion below is based on the baseband equivalent wireless communication system model.

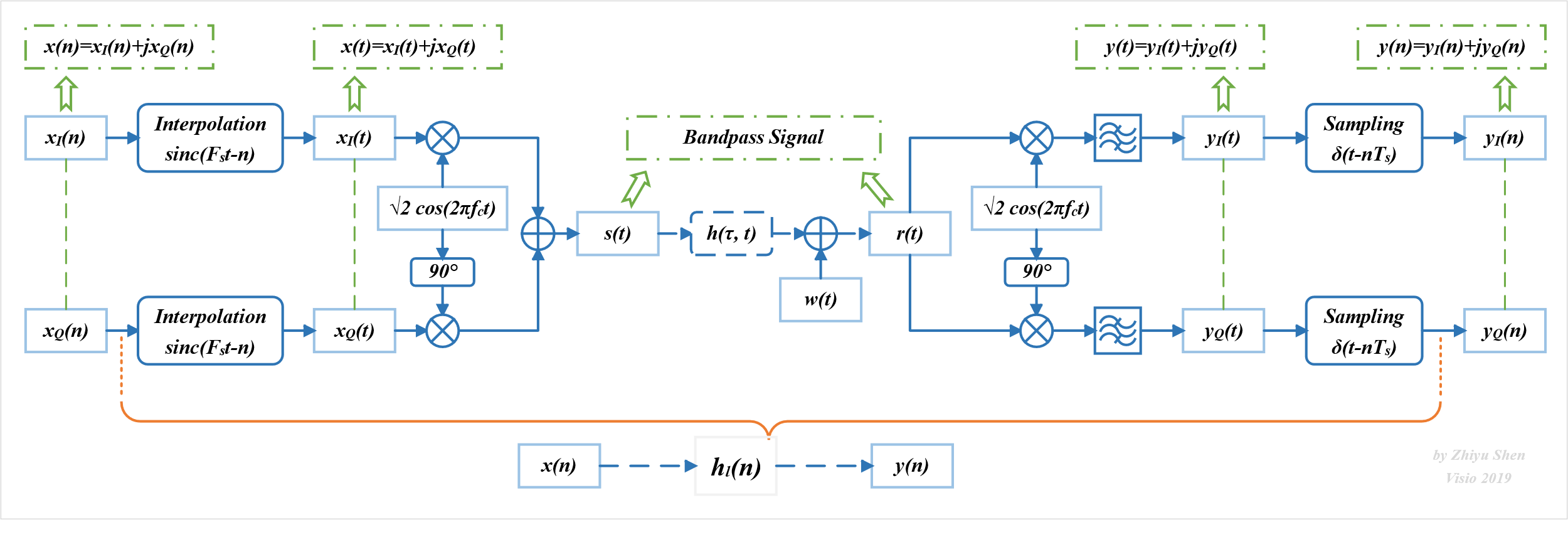


Figure Wireless communication system architecture

As is shown in Figure 1, the baseband equivalent model directly investigates the relationship between the baseband signals at the transmit and receive ends, making it unnecessary to consider the RF characteristic of the system, i.e., the up-conversion and down-conversion parts, in the simulation.

By introducing the baseband equivalent model, the transmission characteristic function of the system turns into the equation below.

|  |  |
| --- | --- |
|  | (1.1) |

Where is the channel coefficient matrix and is a circular symmetry complex Gaussian variable, .

## How to Evaluate the Noise Level

In our system, we adopt to measure the noise figure. Usually, in digital communication systems, we use to replace which is widely adopted in analog systems. The relationship between the two indicators is as follows.

|  |  |
| --- | --- |
|  | (1.2) |

Where is the bitrate of the signal and is the bandwidth of the signal. In most cases, , where is the sampling rate of the baseband signal.

It is also worth mentioning that an important measurement of a digital communication system is the curve.

## Channel Model

### AWGN Channel

To look into the relationship between data error distribution and transmission power, we need to clarify our channel model, which is the most critical factor that influences the BER of a communication system. As we have concluded in the section at the very beginning, the data error distribution is only related to the bit error probability . Therefore, the only thing we have to focus on in channel modeling is the bit error probability of a system with one particular channel model. Let us start with the AWGN channel model.

In a communication system with an AWGN channel, we can simplify Equation (1.1) to the following format:

|  |  |
| --- | --- |
|  | (1.3) |

For a system with BPSK modulation and AWGN channel, the theoretical bit error probability is as follows.

|  |  |
| --- | --- |
|  | (1.4) |

### Rayleigh Fading Channel

However, the channel model would be much more complex because of electromagnetic waves' reflection, diffraction, and scattering. Typically, channel fading is classified into large-scale fading and small-scale fading. The large-scale fading usually causes a logarithmic decrease in the amplitude of the transmission signal, and the small-scale fading brings about signal distortion both in amplitude and phase. Sometimes, the small-scale fading will also cause frequency shifts in the carrier wave, called Doppler Shift.

Researchers have built statistical models to describe small-scale fadings to better model and simulate the system. Some well-known models have been widely adopted, like the Rayleigh fading model, the Rician fading model, and the Nakagami fading model. The details of channel modeling for different statistical models are virtually the same, so here we look at the Rayleigh fading channel as an example.

In this project, we try not to make the channel modeling process too complicated. Assume that the fading channel model is a frequency non-selective fast fading channel, and the large-scale fading is temporarily ignored. On top of that, Equation (1.1) would turn into the following format:

|  |  |
| --- | --- |
|  | (1.5) |

Where is a random process following complex Gaussian distribution.

For a system with BPSK modulation and AWGN channel, the theoretical bit error probability is as follows:

|  |  |
| --- | --- |
|  | (1.6) |

When is quite large, usually larger than 10dB, Equation (1.6) turns into:

|  |  |
| --- | --- |
|  | (1.7) |

# Discussion on Data Error Distribution

Before planning a power allocation strategy, it is necessary to clarify the data error distribution of the communication system under study.

Apparently, for a system whose packet size is , the possible data error is different when different digits are transmitted, and the error distribution seems to be complicated. So, let us start with a simple case where the transmitted digits are uniformly distributed between the widest digit range, e.g., [0,15] for a 4-bit packet.

## Theoretical Foundation

Before simulating with MATLAB, it is necessary to figure out the theoretical basis of our work.

### The Error Distribution Is Symmetrical

In the discussion below, we adopt data with a packet size of 4, for example.

Define the probability of digit being sent at the transmitter as . Since the transmitted digits are uniformly distributed between the widest range, different numbers within the range have the same probability of occurring and let it be . We can easily calculate that .

Define the probability of data error when transmitted digit as and the probability of data error in general is . According to the Bayesian formula:

|  |  |
| --- | --- |
|  | (2.1) |

If we send two digits that are 1's complement to each other, the error distribution at the receiving end is symmetrical because, under our assumption, the probability of the receiver misclassifying bit 0 as bit 1 is equal to that of misclassifying bit 1 as bit 0. For example, numbers 0 and 15 are 1's complement to each other in binary format: 0000 and 1111. The possible data error that will occur when transmitting digit 0 can be and the possible data error that will occur when transmitting digit 15 can be. We can quickly get that and it is the same with other digits with the same feature. The conclusion can be written as follows:

|  |  |
| --- | --- |
|  | (2.2) |

Where represents 1's complement of digit .

On this basis, we can summarize a law of data error distribution as follows.

|  |  |
| --- | --- |
|  | (2.3) |

The derivation above proves the first half of the earlier conclusion, while the second half is obvious.

If we set a different transmit power for each bit in a pack individually, we can get different bit error probability for different bits. Let us suppose the bit error probabilities for different bits in a pack of 4 are and the probability of data error 0 should be

|  |  |
| --- | --- |
|  | (2.4) |

In a general communication system, unless the signal-to-noise ratio is extremely low, in most cases, the bit error probability or BER (bit error rate) is not too large, usually much less than 1. Therefore, the value of the expression above is very close to one in a high SNR case and at least larger than the probability of other data errors in most cases.

##### Conclusion 2.1 The data error distribution is symmetrical

*The distribution of data errors is symmetric about zero, with the peak at zero point.*

|  |  |
| --- | --- |
|  | (2.5) |
|  | (2.6) |

*Where and is the number of bits in a packet.*

## Recursive Expression of Data Errors

In last section, we defined the bit error probability as . At the transmission end, uniformly distributed data is transmitted, the probability of transmitting 0 and 1 are both 1/2. Thus, at the receiving end, the error could be -1, 0, 1and their occurrence probabilities are as follow

|  |  |
| --- | --- |
|  | (2.7) |

Assume the system transmits data in a pack of . The decimal format of transmitted and received data can be expressed by their binary format in the following way:

|  |  |
| --- | --- |
|  | (2.8) |
|  | (2.9) |

Where is the th transmitted bit, is the th received bit.

The data error is the difference between Equation (2.8) and (2.9):

|  |  |
| --- | --- |
|  | (2.10) |

For each bit, it follows the error probability derived in Equation (2.7):

|  |  |
| --- | --- |
|  | (2.11) |

Next, look at each term in the expression of . Let , then the expression of in Equation (2.12) can be written as:

|  |  |
| --- | --- |
|  | (2.12) |

To rewrite the above equation into a recursive form, first define:

|  |  |
| --- | --- |
|  | (2.13) |

The recursive expression of can be easily derived from Equation (2.12):

|  |  |
| --- | --- |
|  | (2.14) |

Therefore, the distribution of random variable can be derived as:

|  |  |
| --- | --- |
|  | (2.15) |

Since s are independent, s are independent random variables. According to the Convolution Theorem, the probability distribution of can be written as:

|  |  |
| --- | --- |
|  | (2.16) |

To simplify the equation above, we first look at the valid value of random variable :

From the previous derivation, we have already known that the probability of any random variable is symmetric about 0. Thus, we can only look at the probability of .

First, let us deal with the case when , it can be simply written as:

Then it comes to the case when . Since , then . The minimal valid value of satisfies the equation , i.e., .

From the above derivation, it can be verified that: when , ; when , .

Similarly, when , ; when , .

##### Conclusion 2.2 The data error distribution can be expressed in a recursive form

*A communication system transmits data in a pack of and the error probability of each bit in a pack is . The data error distribution can be derived from the following recursive expression:*

(.)

*The cases where and the cases where are symmetric:*

(.)

## Zero-Centered Decreasing Distribution

This project aims to adjust the transmission power of different bits so that the data error forms a zero-centered decreasing distribution, which means the probability of smaller data errors should be higher than that of larger ones. A better scenario is to find a transmit power allocation scheme such that the received data error has a Laplace or Gaussian distribution.

# A Power Allocation Strategy to Form Laplace Distributed Data Error

## Prerequisite Assumptions

In our research, we specify the transmit gain of the highest bit in a pack as the unit gain and calculate the noise power with it. To be more specific, if we define the ratio between the transmit power of MSB and the th bit as , the transmit gain of each bit in a pack can be expressed as:

|  |  |
| --- | --- |
|  | (3.1) |

And the transmit power of each bit in a pack can be expressed as:

|  |  |
| --- | --- |
|  | (3.2) |

In addition, to simplify the calculation, we do not introduce baseband shaping process in this model, which means the bitrate of the signal is the same with its bandwidth, i.e., . Define the of MSB as . The power of additive Gaussian noise can be calculated according to Equation (1.2):

|  |  |
| --- | --- |
|  | (3.3) |

Since the noise power almost remains constant all the time, the real for each bit is different:

|  |  |
| --- | --- |
|  | (3.4) |

In Section 2.2, we obtained a recursive expression for the data error distribution with the BER for each bit as the independent variable, which is shown in Equation (2.17) and (2.18). Also, in Section 1.3, we derived the relation between and BER for both AWGN channel and Rayleigh fading channel, which is expressed by Equation (1.4) and (1.6). In theory, by combining these formulas with Equation (3.4), we can obtain the data error’s PDF (probability density function) with transmit power as independent variable. However, the derivation is complex. Instead, in our project, a MATLAB program is written to calculate the value of data error’s PDF, and by adjusting the transmit power of each bit, we attempted to find a power allocation strategy that can form Laplacian or Gaussian distributed data error.

## Exponentially Distributed Transmit Power

Through our simulation and analysis, it is found that if the transmit gain decays exponentially with the number of bits, i.e., , the data error distribution will approximate a Laplace distribution.

### Experiment 1: Simulation on Data Error with Constant Transmit Power

#### Experiment Condition

|  |  |
| --- | --- |
| Channel model | AWGN / Rayleigh |
| Modulation | BPSK |
| Noise power | 1 W (30 dBm) |
| Bitrate | 100 kHz |
| Number of data | 100,000 |
| Pack size | 4 |

Tabel Experiment condition of transmission of uniformly distributed random data

In this experiment, we transmit a large amount of data through a system with a specific channel and BPSK modulation. The data is in a pack of 4. The noise power is 1 W, i.e., 30 dBm and the transmission power remains constant all the time.

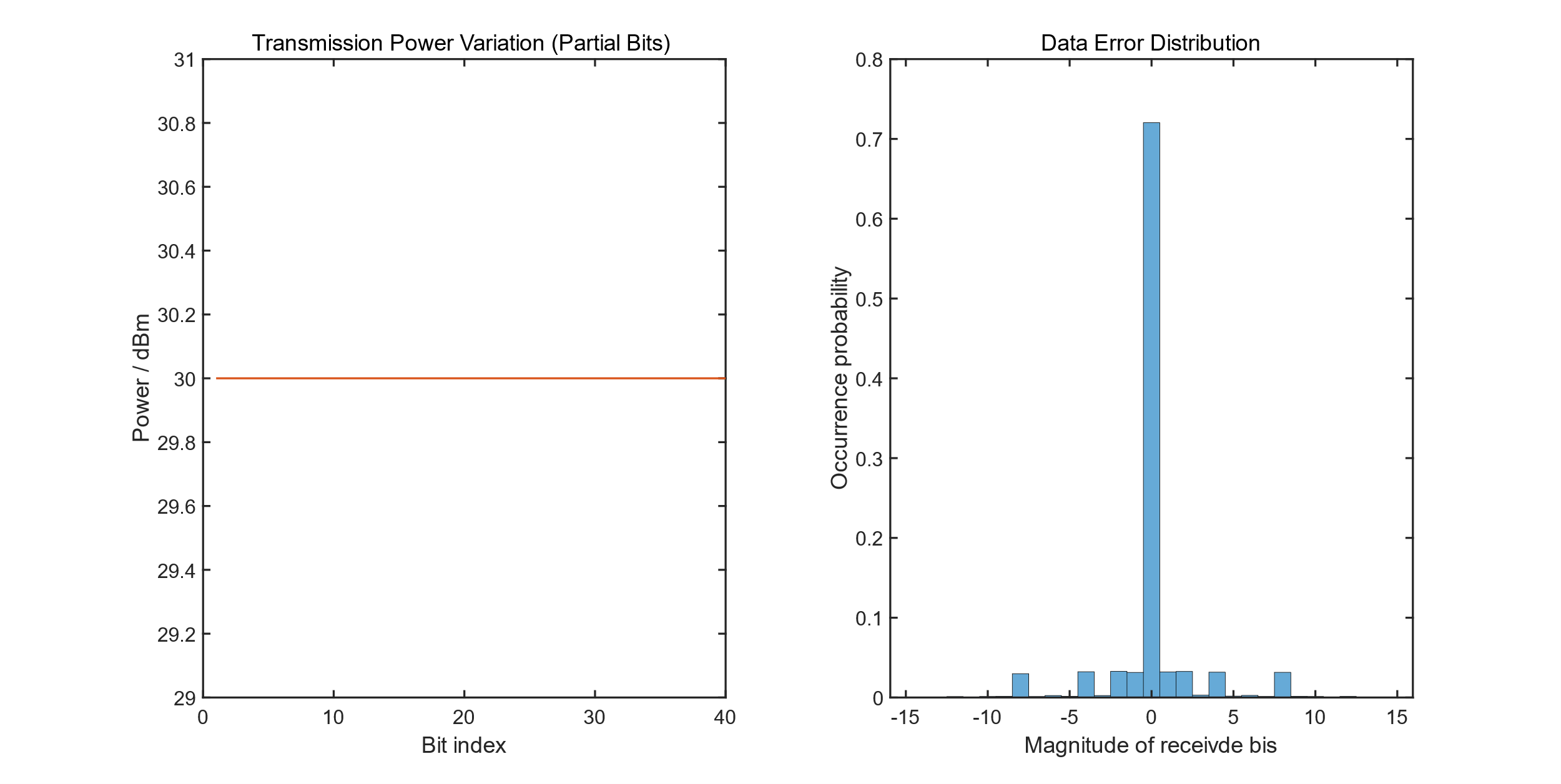
#### Transmission through an AWGN channel

In this part, the transmission power remains constant all the time. Table 2 shows the parameters related to the signal-to-noise ratio for each bit in a pack.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bit Index | Signal Power |  | Theoretical BER | Measured BER |
| 1 (LSB) | 30.00 dBm | 0 dB |  |  |
| 2 | 30.00 dBm | 0 dB |  |  |
| 3 | 30.00 dBm | 0 dB |  |  |
| 4 (MSB) | 30.00 dBm | 0 dB |  |  |

Tabel SNR-related parameter of constant Tx power experiment

Figure 2(a) shows the variation of the transmission power, and Figure 2(b) shows the data error distribution.



(a) Transmission power variation (b) Data error distribution

Figure Result of constant power transmission

Because the transmission power for each bit in a pack is the same, the data error is concentrated at zero.

#### Reduce the Power of one Particular Bit

In this part, we reduce the transmission power of one particular bit to create a dominant bit. The test conditions for the LSB to be the dominant bit are shown in Table 3.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bit Index | Signal Power |  | Theoretical BER | Measured BER |
| 1 (LSB) | Adjustable | Adjustable | Adjustable | Adjustable |
| 2 | 30.00 dBm | 0 dB |  |  |
| 3 | 30.00 dBm | 0 dB |  |  |
| 4 (MSB) | 30.00 dBm | 0 dB |  |  |

Table SNR-related parameter of experiment with one dominant bit

Figure 3 shows the data error distribution of the experiment.

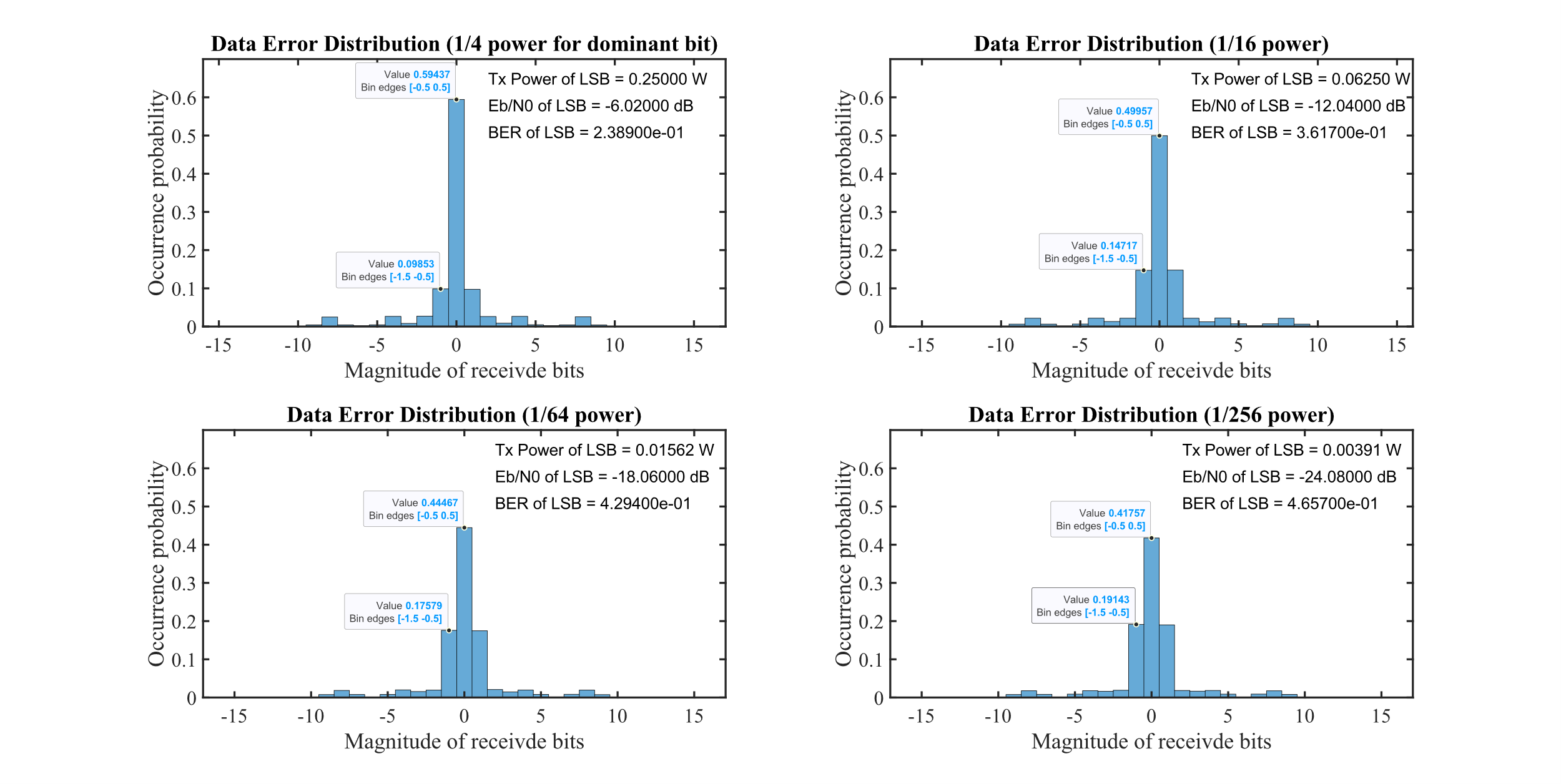


Figure Data error distribution when LSB is the dominant bit in a pack

The transmission power of the LSB is much smaller than the other bits, and as the power decreases, the effect of the dominant bit becomes more significant.

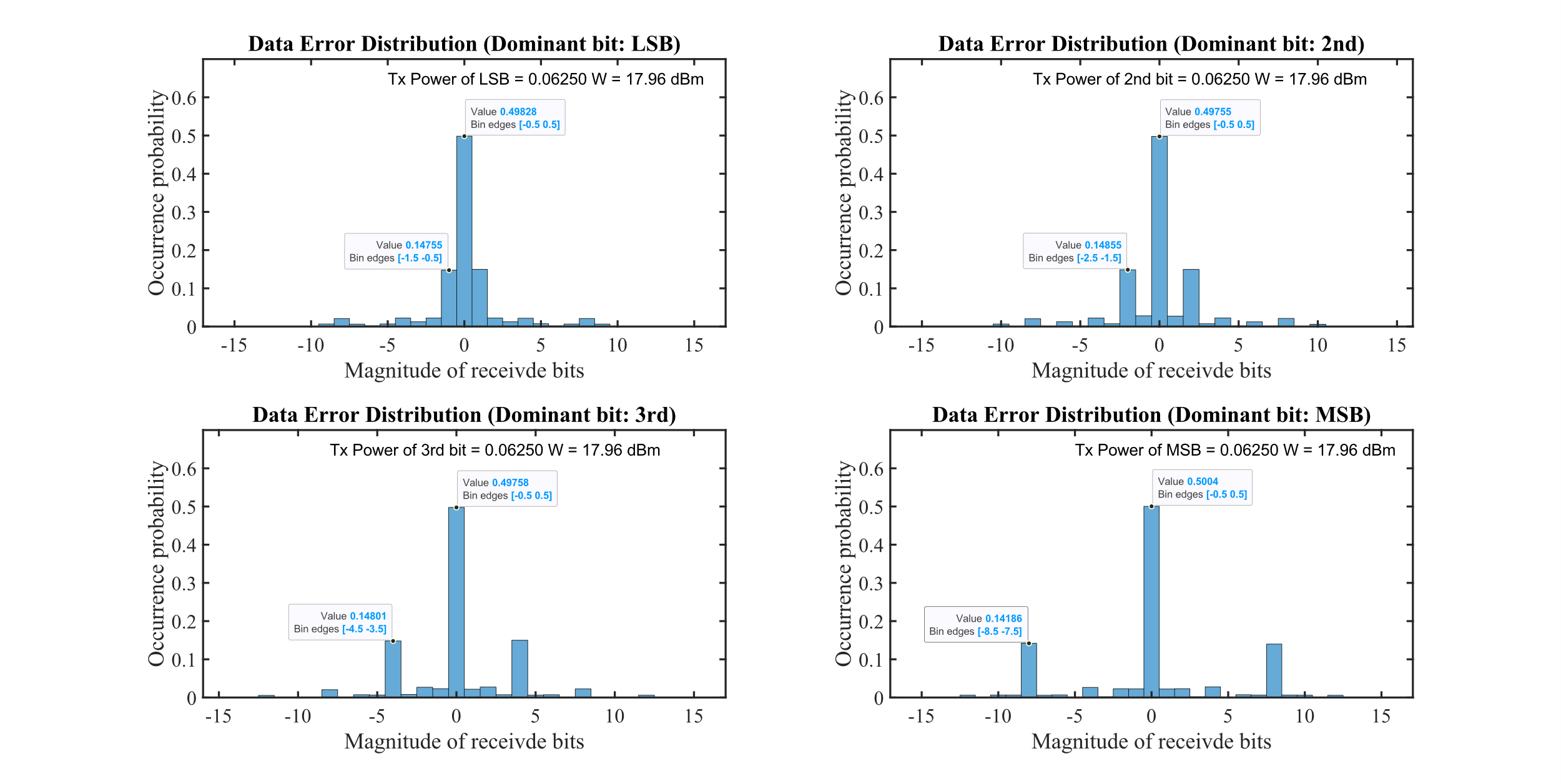
Then we change the dominant bit's position, reducing the transmission power of the second bit, the third bit, and MSB, respectively. Figure 4 shows the data error distribution of the experiment.

Figure Data error distribution when setting different bits as the dominant bit

### Experiment 2: Simulation on Zero-Centered Decreasing Distribution

#### Experiment Condition

|  |  |
| --- | --- |
| Channel model | AWGN |
| Modulation | BPSK |
| Noise power | 1 W (30 dBm) |
| Bitrate | 100 kHz |
| Number of data | 100,000 |
| Pack size | 4 |

Table Experiment condition of transmission of uniformly distributed random data

In this experiment, we transmit a large amount of data through a system with an AWGN channel and BPSK modulation. The data is in a pack of 4. The noise power is 1 W, i.e., 30 dBm.