【注意】本文档中所有蓝色文字，是教师对内容的要求或补充说明。学生在提交报告前，应将它们全部删除。

**目录**

[1. 系统设计 2](#_Toc9346)

[1.1. 设计目标 2](#_Toc4243)

[1.2. 设计方案 2](#_Toc6454)

[2. 系统实现 2](#_Toc10810)

[2.1. 信源 2](#_Toc529)

[2.2. 信道 3](#_Toc30422)

[2.3. 信源编解码 3](#_Toc3293)

[2.4. 信道编解码 4](#_Toc29236)

[3. 系统仿真与分析 5](#_Toc32408)

[3.1. 理想情况下 5](#_Toc19489)

[3.2. 信源非理想情况下 6](#_Toc21400)

[3.3. 信道非理想情况下 6](#_Toc21610)

[3.4. 一般非理想情况下 7](#_Toc5634)

[4. 系统改进【可选】 7](#_Toc10245)

[5. 总结 7](#_Toc17607)

[5.1. 设计完成情况 7](#_Toc12429)

[5.2. 经验与收获 8](#_Toc14280)

[6. 附录——相关技术简介 8](#_Toc20583)

**要求**：定稿后更新目录，以保证目录的正确

# 系统设计

## 设计目标

使用计算机编程语言，完整实现一个包含信源、信道、信源编解码、信道编解码等模块的信息传输系统模型，并通过与理论推导的结果进行比对，验证系统实现的正确性。再进一步通过多种参数组合下的仿真，模拟不同通信条件下的信息传递，掌握信息论与编码理论的理论和实践基础。

以上“设计目标”段落文字可以直接复制使用。

## 设计方案

**要求**

* 整体系统的原理方框图、文字说明
* 每个模块的文字说明

# 系统实现

**要求**

* 每个模块是一个独立可执行的程序，模块间通过文件进行数据交互
* 数据文件的有效载荷(payload)，必须经过比特打包(bit-packing)
* 报告正文中，每个模块给出关键的代码“片段”，并提供简要的文字解说

## 信源

### 二元离散无记忆信源（DMS）模块

**要求**

* 输入参数：信源消息概率分布*P*(0)
* 输入参数：消息序列的长度（字节为单位，一般至少10K字节长）
* 模块输出：信源输出消息序列文件

程序API如下：

byteSource.exe INPUT LEN OUTPUT

– INPUT：信源消息概率分布P(0)

– LEN：消息序列的长度（字节为单位，一般至少10K字节长）

– OUTPUT：信源输出消息序列文件的路径

例如：调用格式byteSource.exe 0.5 64 t:\data\byteSource.output.txt

### 信源指标计算模块

**要求**

* 模块输入：信源输出消息序列文件
* 模块输出：（输出形式自定）
  + 数据比特概率分布（即二元DMS的概率分布统计）：*P*(0)
  + 字节概率分布（即二元DMS的8次扩展信源的概率分布统计）：*P*(0), ..., *P*(255)
  + 二元DMS的信息熵（信息比特/二元消息）
  + 二元DSM的信源冗余度

程序API如下：

calcByteSource.exe INPUT OUTPUT

– INPUT：信源输出消息序列文件的路径

– OUTPUT：存放计算结果的路径

其中，OUTPUT文件最好是CSV文件。每次运行calcByteSource.exe之后，将会在OUTPUT文件后添加计算结果（注意：不是覆盖），格式为

"二元DMS的概率分布统计","二元DMS的信息熵","二元DSM的信源冗余度"

“二元DMS的8次扩展信源的概率分布统计”

例如：当调用命令行calcByteSource.exe t:\data\byteSource.output.txt t:\data\calcByteSource.output.csv时，t:\data\calcByteSource.output.csv文件应会增加一行类似以下的结果：

0.5,7.9999,0.00001

0,0.1

1,0.3

2,0.0

3,0.2

4,0.4

5,0

...

255,0

### 单元测试

**要求（同时适用于下文相同部分，有特殊要求的再另行补充）**

* 理论推导
  + 为各输入参数假设至少1个数值，推导出相关指标的结果
  + 详细描述假设条件、推导过程、推导结果
* 测试结果
  + 按上述理论推导中的假设条件设定该模块的输入参数
  + 收集相关指标的实测结果
* 对比理论推导的结果和测试结果，分析异同及原因
* byteSource部分：

理论推导：

设信源消息概率分布，消息序列的长度为1MB，则，即二元离散无记忆信源生成0和1的概率相等。而每个字节相当于二元DMS的8次扩展，设该二元DMS的8次扩展信源为Y，则Y的概率空间为：



其中分别由8位二元离散无记忆信源符号（0和1）组成，取十进制值0~255；为对应的概率，设中二元离散无记忆信源符号0的个数为，则

故当二元离散无记忆信源生成0和1的概率相等时，其8次扩展信源的每一个值的取值概率都相等，即信源输出消息序列文件的字节概率是均匀分布的。

测试结果：

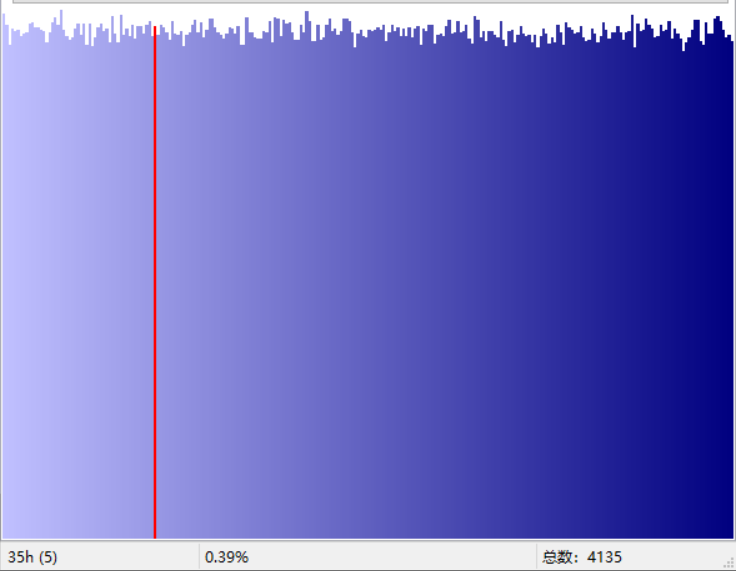
在cmd中输入命令：

python byteSource.py 0.5 1048576 unit-test\P0=0.5.LEN=1MB.OUTPUT.dat

生成的信源输出消息序列文件如下：



可以看到文件大小确实为1MB，在16进制文件编辑器HxD中打开该文件，使用统计功能显示字节概率分布如下：



可以看到字节概率分布基本均匀，均在0.39%~0.40%左右，与理论推导结果比较接近；而不能完全符合理论推导的均匀分布的原因是byteSource程序是根据蒙特卡罗法随机生成字节的，因此在一定文件大小下实际字节概率分布与理论会有微小的差距，而随着文件大小的增加，实际字节概率分布会趋于理论的均匀分布。

* byteSource\_calc部分：

理论推导：

根据byteSource部分设定的输入参数，信源指标理论计算结果应如下：

1. 数据比特概率分布（即二元DMS的概率分布统计）：
2. 字节概率分布（即二元DMS的8次扩展信源的概率分布统计）：
3. 二元DMS的信息熵：
4. 二元DSM的信源冗余度：，其中为信源的实际熵，则；为信源最大熵，此二元离散无记忆信源生成0和1的概率相等，故信息熵即为最大熵，，故

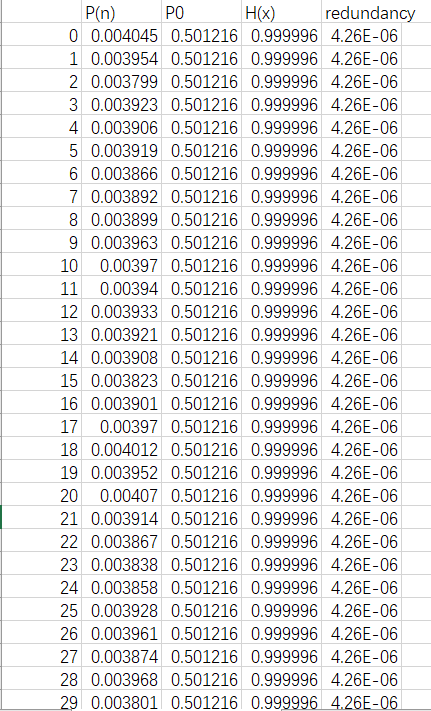
测试结果：

在cmd中输入命令：

python byteSource\_calc.py unit-test\P0=0.5.LEN=1MB.OUTPUT.dat unit-test\P0=0.5.LEN=1MB.calc.CSV

生成的信源指标计算结果如下：





其中第一列为信源输出消息序列文件字节取值，第二列为字节概率分布，第三列为数据比特概率分布，第四列为二元DMS的信息熵，第五列为二元DSM的信源冗余度。可以看到四个实际信源指标均与理论推导结果相接近，在一定误差允许范围内可以认为byteSource\_calc程序能正确计算信源指标，通过单元测试。

## 信道

### 二元对称信道（BSC）模块

**要求**

* 模块输入：信道输入消息序列文件
* 输入参数：错误传递概率*p*
* 模块输出：信道输出消息序列文件

程序API如下：

byteChannel.exe INPUT p OUTPUT

例如：调用格式byteChannel.exe 0.5 t:\data\byteChannel.output.txt

### 信道指标计算模块

**要求**

* 模块输入：信道输入消息序列文件、信道输出消息序列文件
* 模块输出：（输出形式自定）
  + 输入消息序列的信息熵（信息比特/二元消息）
  + 输出消息序列的信息熵（信息比特/二元消息）
  + 平均互信息量（信息比特/二元消息）

程序API如下：

calcByteChannel.exe INPUT1 INPUT2 OUTPUT

– INPUT1：信道输入消息序列文件的路径

– INPUT2：信道输出消息序列文件的路径

– OUTPUT：信道计算结果文件的路径

格式为

"输入消息序列的信息熵","输出消息序列的信息熵","平均互信息量"

例如：当调用这个API时，输出结果的文件中会追加以下内容：

"7.98888","7.898998","7.09099090"

### 单元测试

* byteChannel部分：

理论推导：

设错误传递概率，信道输入消息X为的概率分布，信道输出消息为Y，重用【实验3.1 二元对称信道（BSC）仿真】中制作的信道指标计算excel表格，在自变量中输入，得到理论推导结果如下红框所示：



其中输入消息序列的信息熵H(X)=1，输出消息序列的信息熵H(Y)=1，X和Y的联合熵H(XY)=2，损失熵H(X|Y)=1，噪声熵H(Y|X)=1，平均互信息量I(X;Y)=0。

测试结果：

运行byteChannel.py程序，在cmd中输入如下命令：

python byteChannel.py unit-test\P0=0.5.LEN=1MB.OUTPUT.dat 0.5 unit-test\p=0.5.P0=0.5.LEN=1MB.Channel\_OUTPUT.dat

生成的信道输出消息序列文件如下：



重用【实验3.1 二元对称信道（BSC）仿真】中教师提供的信道指标计算程序calcBSCInfo.exe，其程序API如下：

calcBSCInfo.exe X Y OUTPUT

X path to the channel input file

Y path to the channel output file

OUTPUT path to the output file to append results

在cmd中输入如下命令：

calcBSCInfo.exe ..\P0=0.5.LEN=1MB.OUTPUT.dat ..\p=0.5.P0=0.5.LEN=1MB.Channel\_OUTPUT.dat ..\Channel\_calcBSCInfo\_result.CSV

程序运行后信道指标计算结果如下：



可以看到实际测试的计算结果和理论推导的结果相接近，在一定误差允许范围内可以认为是正确的，byteChannel程序单元测试完成。

* byteChannel\_calc部分：

理论推导：

同上byteChannel部分。

测试结果：

运行byteChannel\_calc.py程序，将在cmd中输入如下命令：

python byteChannel\_calc.py unit-test\P0=0.5.LEN=1MB.OUTPUT.dat unit-test\p=0.5.P0=0.5.LEN=1MB.Channel\_OUTPUT.dat unit-test\Channel\_calc\_result.CSV

得到如下CSV文件和结果：





该计算结果与使用程序calcBSCInfo.exe计算得到的结果相同，且与理论推导的结果相接近，可以认为该结果是正确的，byteChannel\_calc单元测试完成。

## 信源编解码

### 霍夫曼编码模块

**要求**

* 模块输入：编码前的文件
* 输入参数：编码前的文件的字节概率分布
* 模块输出：编码后的文件，包含码书、码字序列

程序API如下：

byteSourceCoder.exe PMF INPUT OUTPUT

PMF path to probability mass function CSV file

INPUT path to the encoder input file

OUTPUT path to the encoder output file

### 霍夫曼解码模块

**要求**

* 模块输入：解码前的文件（格式与霍夫曼编码模块的输出相同）
* 模块输出：解码后的文件

程序API如下：

byteSourceDecoder.exe INPUT OUTPUT

INPUT path to the decoder input file

OUTPUT path to the decoder output file

### 信源编码指标计算模块

**要求**

* 模块输入：编码前的文件、编码后的文件
* 模块输出：（输出形式自定）
  + 压缩比（编码前文件长度:编码后文件长度）
  + 平均码长（码字数据比特/信源字节）
  + 编码效率
  + 编码前的文件的信息熵（信息比特/字节）
  + 编码后的文件的信息熵（信息比特/字节）

程序API如下：

calcByteSourceCoder.exe INPUT1 INPUT2 OUTPUT

– INPUT1：编码前的文件的路径

– INPUT2：编码后的文件的路径

– OUTPUT：信源编码计算结果文件的路径

格式为

"压缩比","平均码长","编码效率"," 编码前的文件的信息熵","编码后的文件的信息熵"

例如：当调用这个API时，输出结果的文件中会追加以下内容：

"7.98888","7.898998","7.09099090","7.898998","7.09099090"

### 单元测试

* byteSourceEncoder/ byteSourceDecoder部分：

理论推导：

注意测试时避免将未经验证的channelEncoder和channelDecoder程序串联起来自我验证。

测试结果：

待编码的原文件如下：



1. 编码测试：

在cmd中输入如下命令进行5长的重复码编码：

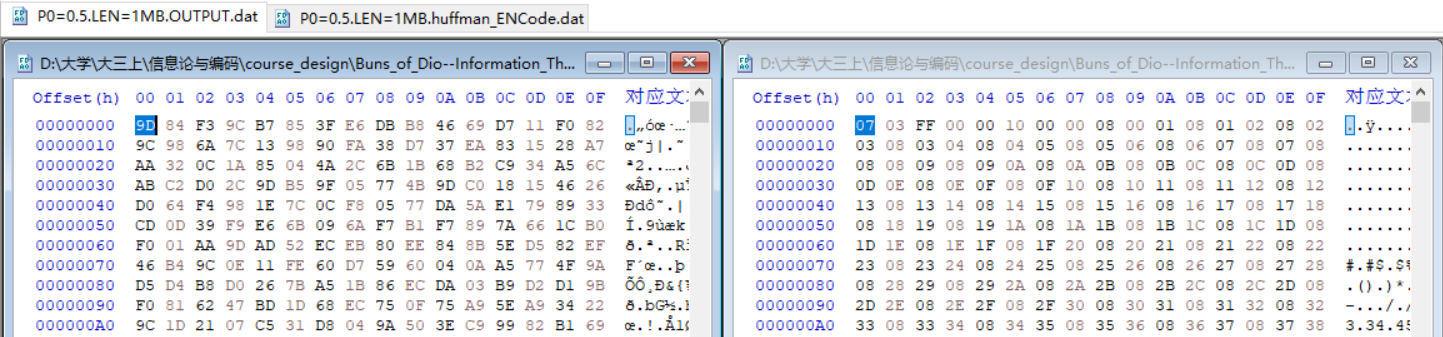
python byteSourceEncoder.py unit-test\pmf.byte.p0=0.5.csv unit-test\P0=0.5.LEN=1MB.OUTPUT.dat unit-test\P0=0.5.LEN=1MB.huffman\_ENCode.dat

程序运行完毕后得到如下文件：



可以看到文件大小与编码前的原文件大小不同，甚至比原文件要大一些，这是因为原文件是由等概率的二元DMS生成，字节概率分别是比较均匀的，而霍夫曼编码基于变长码的思想，因此对信源符号分布较均匀的序列的压缩效果微弱甚至反而会增大文件大小。

使用HxD的文件比较功能对两者的数据进行比较：



可以看到两者数据亦不相同，说明编码器是对原文件进行了处理的。

1. 解码测试：

在cmd中输入如下命令，利用解码器对编码后的文件进行解码：

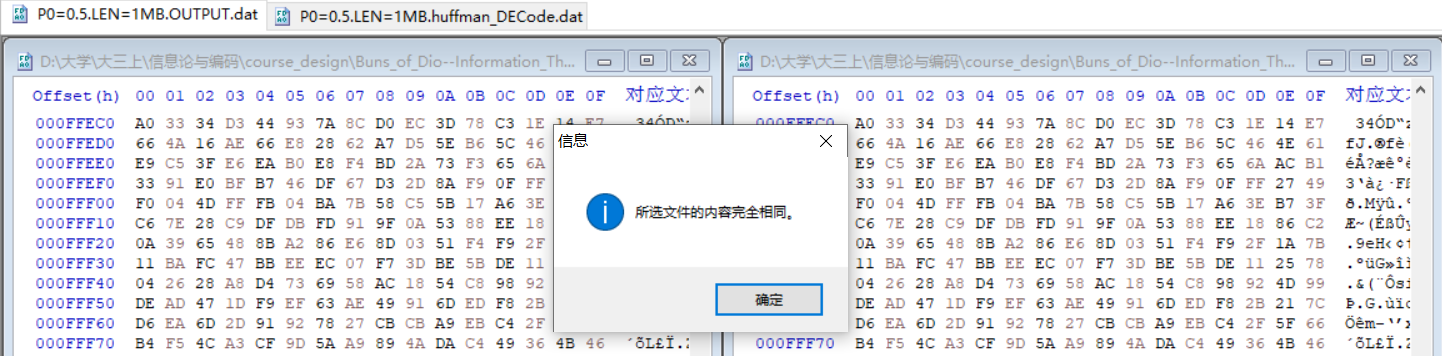
python byteSourceDecoder.py unit-test\P0=0.5.LEN=1MB.huffman\_ENCode.dat unit-test\P0=0.5.LEN=1MB.huffman\_DECode.dat

运行完毕后得到的文件如下：



可以看到文件大小与编码前的原文件大小相同。

使用HxD的文件比较功能对解码后的文件和原文件进行比较：



HxD弹窗提示所选文件的内容完全相同，说明解码后可以完整无失真地还原原文件，单元测试通过。

* codding\_effect部分：

理论推导：

1. 压缩比

编码前文件长度为l=1048576字节，编码后文件长度为L=1049351字节，故压缩比

1. 平均码长
2. 编码效率
3. 编码前的文件的信息熵

信源消息概率分布，故信息熵：

1. 编码后的文件的信息熵

测试结果：

在cmd中输入如下命令，运行codding\_effect：

运行完毕后得到如下结果：

（解释各数据含义），可以看到实际测试结果与理论推导结果相接近，可以认为是正确的，codding\_effect程序单元测试完成。

## 信道编解码

### 信道编码模块

**要求**

* 模块输入：编码前的文件
* 输入参数：编码方案选择，允许用户自由选择以下6种方案中的一种
  + N长的重复码：N是5、7、9
  + (*n*, *k*)线性分组码：(7,4)、(15,11)、(31,26)
* 模块输出：编码后的文件，包含编码方案相关参数、码字序列

程序API如下：

byteChannelCoder.exe INPUT NUMBER OUTPUT

INPUT path to the encoder input file

NUMBER ： 这六种方案的代号，比如输入参数为1表示第一种方案

OUTPUT path to the encoder output file

### 信道解码模块

**要求**

* 模块输入：解码前的文件（格式与信道编码模块的输出相同）
* 模块输出：解码后的文件

程序API如下：

byteChannelDecoder.exe INPUT OUTPUT

INPUT path to the decoder input file

OUTPUT path to the decoder output file

### 信道编解码指标计算模块

**要求**

* 模块输入：编码前的文件、编码后的文件、解码后的文件
* 模块输出：（输出形式自定）
  + 误码率（汉明失真，错误二元消息数/总二元消息数）
  + 编码前的信源信息传输率（信息比特/二元消息）
  + 编码后的信源信息传输率（信息比特/二元消息）
  + 压缩比（编码前文件长度:编码后文件长度）

程序API如下：

calcByteChannelCoder.exe INPUT1 INPUT2 OUTPUT

– INPUT1：编码前的文件的路径

– INPUT2：编码后的文件的路径

– OUTPUT：信源编码计算结果文件的路径

格式为

"误码率","编码前的信源信息传输率","编码后的信源信息传输率"," 压缩比"

例如：当调用这个API时，输出结果的文件中会追加以下内容：

"0.98743","0.86754","0.9999","1.5345"

### 单元测试

* channelEncoder/channelDecoder部分：

理论推导：

注意测试时避免将未经验证的channelEncoder和channelDecoder程序串联起来自我验证。

测试结果：

待编码的原文件如下：



对6种方案分别进行测试，以5长的重复码为例

1. 编码测试：

在cmd中输入如下命令进行5长的重复码编码：

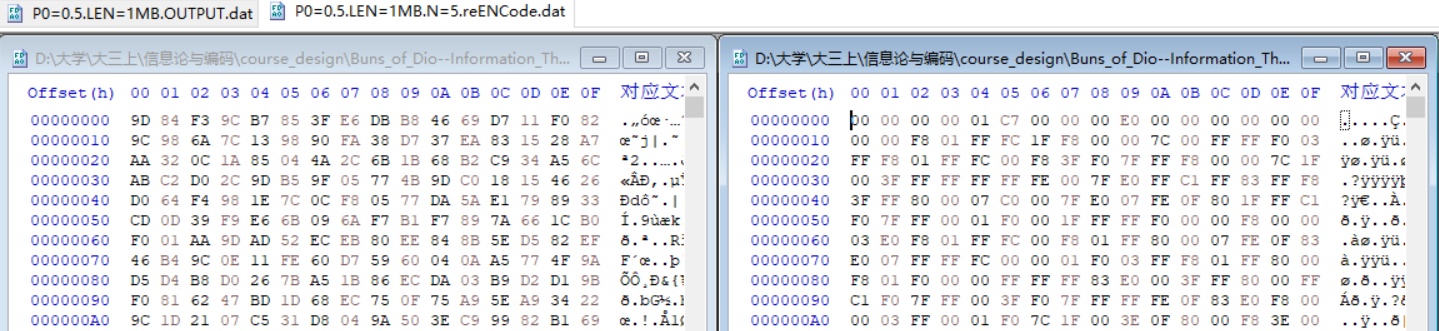
python channelEncoder.py -r unit-test\P0=0.5.LEN=1MB.OUTPUT.dat unit-test\P0=0.5.LEN=1MB.N=5.reENCode.dat 5

程序运行完毕后得到如下文件：



可以看到文件大小与编码前的原文件大小不同，N=5长的重复编码后的文件约比原文件大5倍。

使用HxD的文件比较功能对两者的数据进行比较：



可以看到两者数据亦不相同，说明编码器是对原文件进行了处理的。

1. 解码测试：

在cmd中输入如下命令，利用解码器对编码后的文件进行解码：

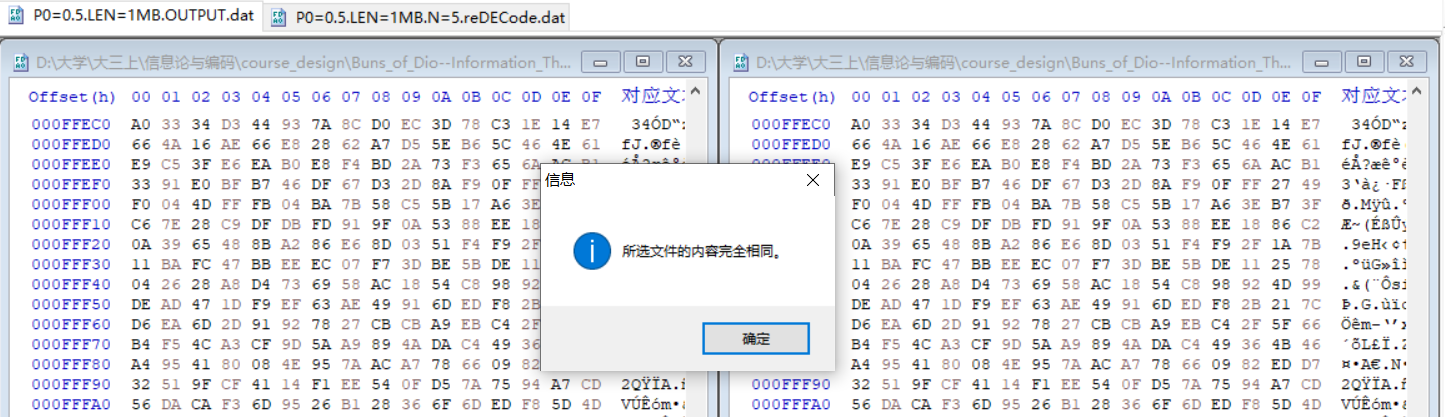
python channelDecoder.py unit-test\P0=0.5.LEN=1MB.N=5.reENCode.dat unit-test\P0=0.5.LEN=1MB.N=5.reDECode.dat

运行完毕后得到的文件如下：



可以看到文件大小与编码前的原文件大小相同。

使用HxD的文件比较功能对解码后的文件和原文件进行比较：



HxD弹窗提示所选文件的内容完全相同，说明解码后可以完整无失真地还原原文件，5长重复码单元测试通过。

7/9长的重复码和(n, k)线性分组码的测试步骤与5长重复码的步骤相同，不再赘述，经测试均通过。

* channelCoder\_calc部分：

理论推导：

1. 误码率

由于原文件编码后未经信号传输而直接解码，没有噪声干扰，因此相当于无失真编解码，误码率

1. 编码前的信源信息传输率

信源为等概率的二元DMS，消息个数，因此编码前的信源信息传输率

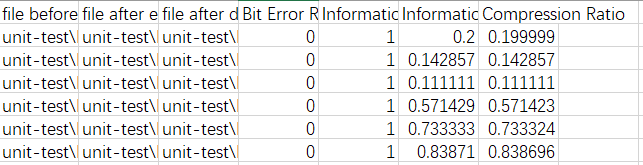
1. 编码后的信源信息传输率

设编码是将每n位二元消息编成N位码长的码字，则

1. 对于5长的重复码，n=1，N=5，故编码后的信源信息传输率
2. 对于7长的重复码，n=1，N=7，故编码后的信源信息传输率
3. 对于9长的重复码，n=1，N=9，故编码后的信源信息传输率
4. 对于(7, 4)线性分组码，n=4，N=7，故编码后的信源信息传输率
5. 对于(15, 11)线性分组码，n=11，N=15，故编码后的信源信息传输率
6. 对于(31, 26)线性分组码，n=26，N=31，故编码后的信源信息传输率
7. 压缩比
8. 对于5长的重复码，编码前文件长度为l=1048576字节，编码后文件长度为L=5242898字节，故压缩比
9. 对于7长的重复码，编码前文件长度为l=1048576字节，编码后文件长度为L=7340050字节，故压缩比
10. 对于9长的重复码，编码前文件长度为l=1048576字节，编码后文件长度为L=9437202字节，故压缩比
11. 对于(7, 4)线性分组码，编码前文件长度为l=1048576字节，编码后文件长度为L=1835026字节，故压缩比
12. 对于(15, 11)线性分组码，编码前文件长度为l=1048576字节，编码后文件长度为L=1429895字节，故压缩比
13. 对于(31, 26)线性分组码，编码前文件长度为l=1048576字节，编码后文件长度为L=1250245字节，故压缩比

测试结果：

根据不同编码类型输入相应参数，运行channelCoder\_calc，运行完毕后得到如下结果：



前三列分别为：编码前文件路径、编码后文件路径、解码后文件路径；后四列分别为：误码率、编码前的信源信息传输率、编码后的信源信息传输率、压缩比。可以看到实际测试结果与理论推导结果相接近，可以认为是正确的，channelCoder\_calc程序单元测试完成。

# 系统仿真与分析

要求

* 在本节推导以下待测指标的计算公式
  + 信源的信息率*RS*（信息比特/秒）
  + 信道的数据率*rc*（数据比特/秒）
  + 信道的输入信息率*Rci*（信息比特/秒）
  + 信道的输出信息率*Rco*（信息比特/秒）
  + 信宿关于信源的信息率*RI*（信息比特/秒）
  + 信宿的误码率*er*
* 在以下每种情况下
  + 在“条件设定”小节中，详细描述固定条件、可变条件的具体数值
    - 所有情况下的“信源的数据率”固定为*rs*=1（数据比特/秒）
  + 在“理论推导”小节中，计算所有待测指标的理论数值
  + 在“仿真”小节中，描述仿真的操作过程和待测指标的仿真结果
  + 在“分析”小节中，对比待测指标的理论数值和仿真结果，分析异同及原因

## 理想情况下

### 条件设定

**要求**

* 固定条件
  + 信源的数据率*rs*（数据比特/秒）
  + 信源为等概率二元DMS
  + 信道为无噪无损的离散无记忆信道
  + 不使用信源编码
  + 不使用信道编码
* 可变条件
  + 无

### 理论推导

### 仿真

### 分析

## 信源非理想情况下

### 条件设定

**要求**

* 固定条件
  + 信源的数据率*rs*（数据比特/秒）
  + 信道为无噪无损的离散无记忆信道
  + 不使用信道编码
* 可变条件
  + 2种信源概率分布：0.1, 0.3
  + 无信源编码、有信源编码

### 理论推导

### 仿真

### 分析

## 信道非理想情况下

### 条件设定

**要求**

* 固定条件
  + 信源的数据率*rs*（数据比特/秒）
  + 信源为等概率二元DMS
  + 不使用信源编码
* 可变条件
  + 3种信道错误传递概率：0.01, 0.2, 0.5
  + 无信道编码、重复码、(*n*,*k*)线性分组码

### 理论推导

### 仿真

### 分析

## 一般非理想情况下

### 条件设定

**要求**

* 固定条件
  + 信源的数据率*rs*（数据比特/秒）
* 可变条件
  + 2种信源概率分布：0.1, 0.3
  + 3种信道错误传递概率：0.01, 0.2, 0.5
  + 无信源编码、有信源编码
  + 无信道编码、重复码、(*n*,*k*)线性分组码

### 理论推导

### 仿真

### 分析

# 系统改进【可选】

**要求**

* 针对仿真所出现的现象，提出对系统设计、实现、仿真等方面的改进设想。
* 分析说明改进的目的、原理和可行性依据。
* 实施对系统的改进，并通过仿真验证改进的效果。

# 总结

## 设计完成情况

**要求**

* 列出已完成的内容
* 列出未完成的内容，及导致未能完成的原因

## 经验与收获

**要求**：在本次设计过程中，所得到的经验与收获

# 附录——相关技术简介

**要求：**简要介绍本设计中所使用的

* 计算机语言
  + 官方名称、版本号、官网URL
* 主要函数包/模组
  + 只需包含非标准函数包/模组等（不是计算机语言自带的）
  + 官方名称、版本号、官网URL
  + 在本设计中的用途
* 工具软件
  + 编程开发、报告撰写、作图、其他辅助工具等
  + 官方名称、版本号、官网URL
  + 在本设计中的用途