Институт информационных технологий

Кафедра: Математическое и программное обеспечение ЭВМ

Дисциплина: Теория Информации

ЛАБОРАТОРНАЯ РАБОТА № 2-3

Тема: Методы эффективного и помехоустойчивого кодирования

Выполнил:

студент гр. 1ПИб-02-2оп-23

Кринкин Олег Алексеевич

Проверил:

доцент, к.т.н. Ганичева Оксана Георгиевна

ЗАДАНИЕ

Для текста из Лабораторной работы 1 и, опираясь на промежуточные результаты этой работы, выполнить

Часть 1: методы эффективного кодирования

1. Построить код Шеннона- Фано.
2. Построить код Хаффмена.
3. Для каждого полученного кода оценить степень сжатия через среднее число символов на одну букву исходного алфавита ( lср) и энтропию
4. Проверить выполнение теоремы Шеннона ( lср > H)

Часть 2: помехоустойчивое кодирование

А=11000110001110111101

В=10101011110001000011

С= 10000011110011110001

1. Для заданных кодовых комбинаций (А,В,С) выполнить кодирование с контролем четности. Внести ошибку в 2 комбинации. Выполнить проверку и «найти» ошибочные комбинации.
2. Для этих же кодовых комбинаций (А,В,С) выполнить матричные проверки (4х5) по методу с контролем четности. Внести ошибку в 2 комбинации. Выполнить проверку и найти ошибочный разряд.
3. Построить код Хемминга для заданного числа.

Число соответствует номеру в списке группы для всех, чей номер от 20 до 30. Тем, у кого номер в списке группы от 1 до 19 – надо к своему номеру прибавить 30. (Например, в списке N равен 4, тогда к нему надо прибавить 30 и получим число 34. Именного для него строим код).

ХОД РАБОТЫ

Часть 1: методы эффективного кодирования.

Для выполнения лабораторной работы был использован текст из предыдущей работы:

Кто такой Клод Шеннон? В математических кругах Шеннон был популярной фигурой. Он считается "отцом информационного века". В 21 год он опубликовал важную магистерскую работу, которая заложила основу для будущих цифровых компьютеров. И это еще не все. Шеннон является основоположником теории информации, нашедшей применение в современных высокотехнологических системах связи. Он также предложил использовать слово "бит" для обозначения наименьшей единицы информации. Однако Клод Шеннон был не только гениальным теоретиком, но и веселым и изобретательным человеком.

1. Построение кода Шеннона-Фано.
2. Используемый текст размещён в ячейке A2.
3. В ячейки B2-B54 выписаны все символы используемого алфавита (33 буквы, 10 цифр, специальные символы).
4. При помощи Excel формулы (1.1) для каждого символа алфавита найдена вероятность, с которой он встречается в тексте.

где, $A$2 – ячейка с текстом B3 – ячейка с символом

1. При помощи пользовательской сортировки весь алфавит отсортирован по убыванию количества символа в тексте (следовательно, и по убыванию вероятности) (рис. 1).

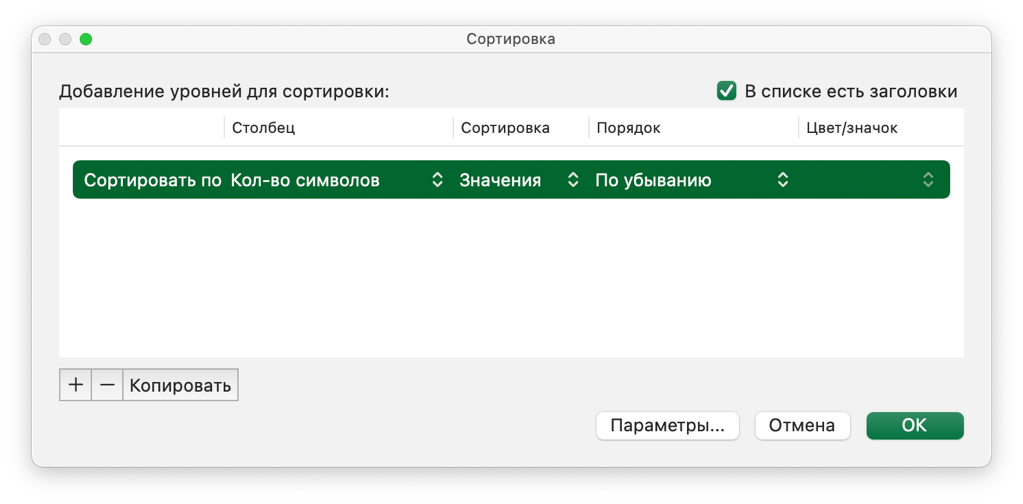


Рис. 1: Настройки пользовательской сортировки

1. Вероятности в столбце D были сгруппированы в две группы таким образом чтобы суммы вероятностей в каждой из них были примерна равны. Верхняя группа отмечалась зелёным цветом, а нижняя – синим (рис. 2). Сумма групп вычислялась путём выделения ячеек, после чего в нижней части окна Excel появлялось значение суммы (рис. ).

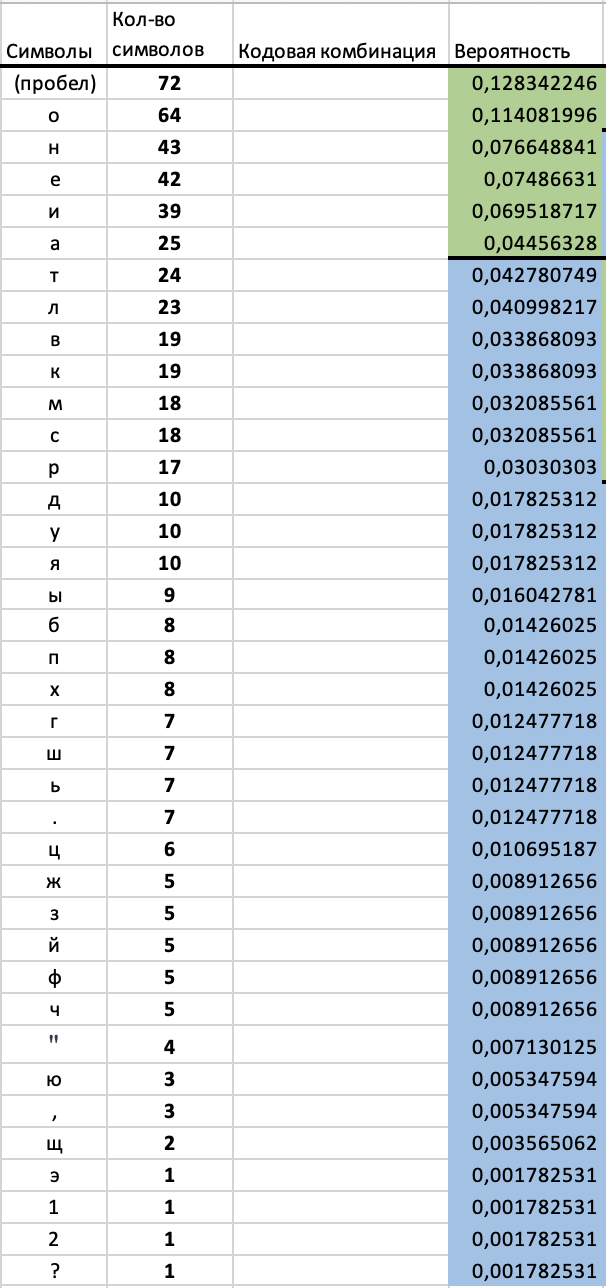


Рис. 2: Разделение вероятностей на группы

1. Полученные группы были разделены на ещё две подгруппы. Данное действие повторялось для каждого следующего набора подгрупп пока в ней не оставался лишь один элемент. Результат разделения подгрупп представлен на рис. 3.

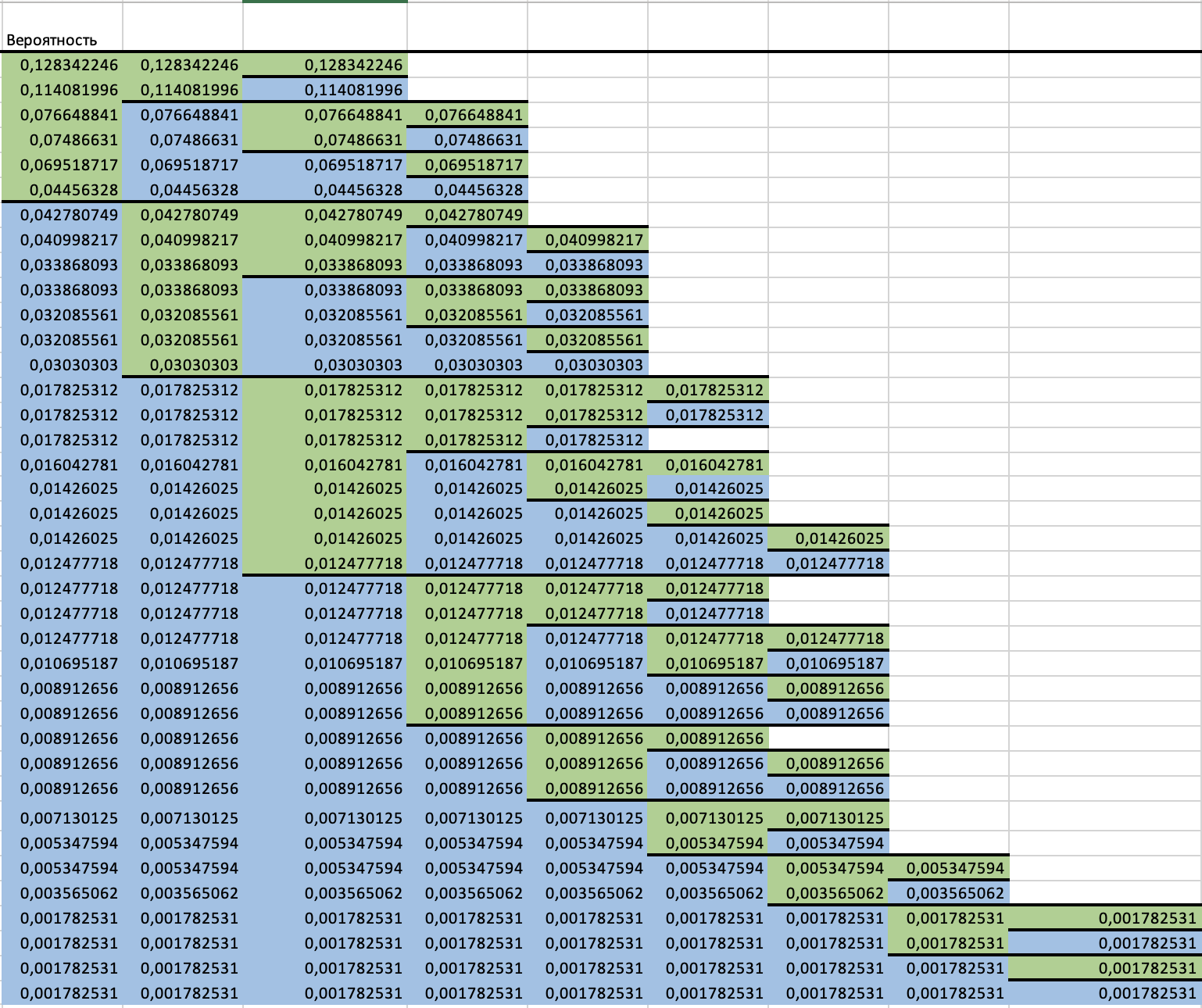


Рис. 3: Разделение подгрупп

1. Исходя из полученного разделения, для каждого символа составлена кодовая комбинация: каждая зелёная ячейка в строке с символом – 1, каждая синяя – 0 (рис. 4).

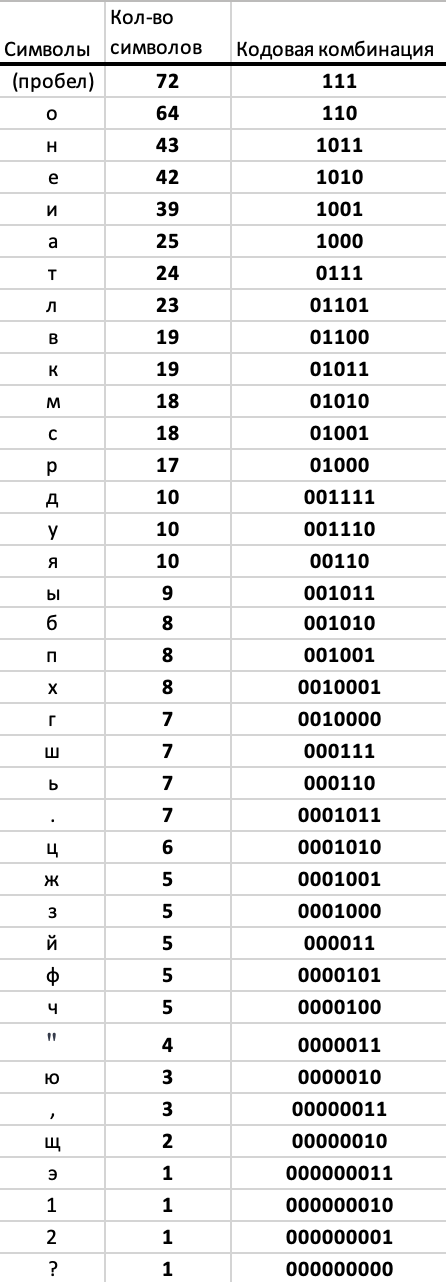


Рис. 4: Составленные кодовые комбинации

1. По формуле Шеннона (1.2) вычислено степень энтропии сообщения.

где, – вероятность появления символа в сообщении.

В Excel это действие было реализовано в 2 этапа: для каждой из вероятностей найден член суммы при помощи формулы (1.3):

где, – вероятность появления символа ()

Затем при помощи формулы (1.4) найдена сумма по найденным членам ряда:

где, – ранее вычисленные члены суммы.

1. По формуле (1.5) найдено среднее число символов на букву сообщения.

где, pi – вероятность появления символа, ki – длина кодовой комбинации символа.

В Excel расчёт среднего числа символов на букву сообщения производился следующим образом: для каждого символа найдено произведение при помощи Excel формулы (1.6);

где, D2 – кодовая комбинация, E2 – вероятность появления символа.

Затем все вычисленные произведения были сложены при помощи формулы (1.7):

где, O2:O39 – вычисленные ранее произведения.

1. Среднее число символов на букву сообщения также показывает и степень сжатия сообщения для последующей передачи.
2. По теореме Шеннона lср не должно быть меньше H(pi) (1.8).

Для проверки условия выполнения теоремы Шеннона в Excel ячейку внесена формула (1.9):

где, E56 – H(pi), F56 – lср.

В ответ в ячейке с формулой было получено сообщение «Выполняется», что означает что код Шеннона построен правильно (рис. 5).

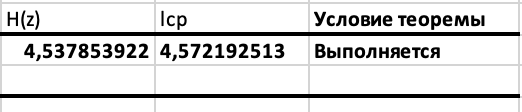


Рис. 5: Проверка по теореме Шеннона

1. Построение кода Хаффмана.
2. Столбцы с алфавитом и вероятностями скопированы из предыдущей таблицы (рис. ).

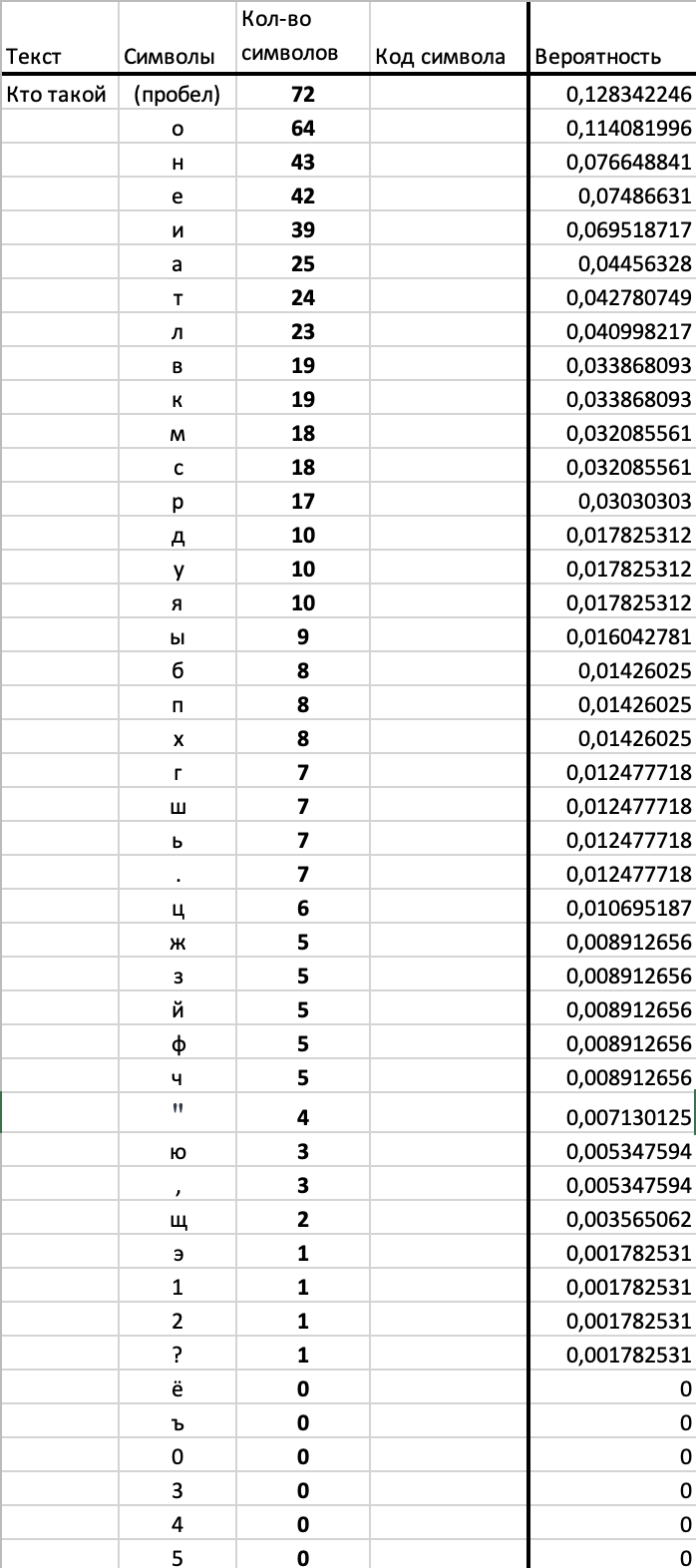


Рис. 6: Таблица вероятности по алфавиту

1. В предыдущей таблице вероятности уже были отсортированы в порядке убывания, поэтому этого более делать не нужно.
2. Сбоку от столбца с вероятностями были суммированы две его последние ячейки, после чего ячейке с суммой присвоен цвет (рис. 7).

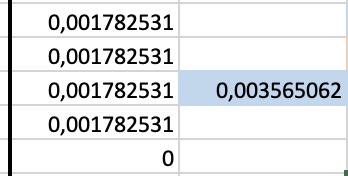


Рис. 7: Сумма двух последних ячеек

1. Выше ячейки с суммой были вставлены значения вероятностей из столбца слева, за исключением тех, что участвовали в сумме. Вероятности в новом столбце также были отсортированы по убыванию.
2. Шаги 3-4 были повторялись до тех пор, пока сумма двух последних ячеек в столбце не оказалась равной единице. Часть результата повторения шагов представлена на рис. 8.

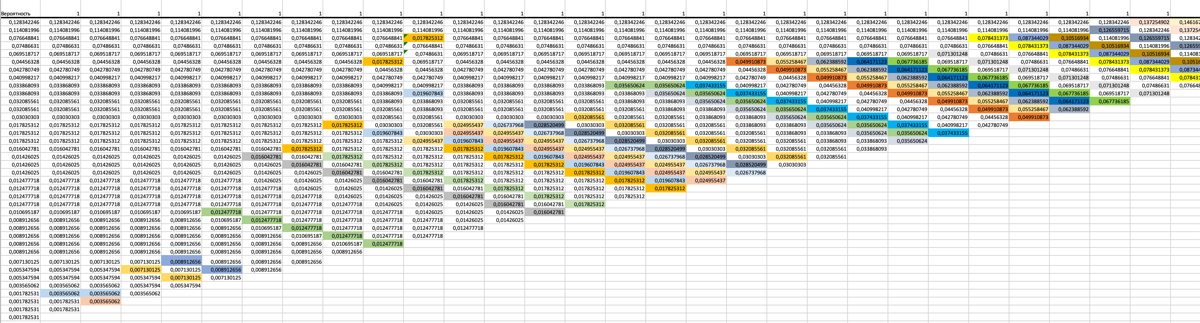


Рис. 8: Построение кода Хаффмана

1. При помощи онлайн-сервиса было построено дерево Хаффмана: для каждой цветной ячейки найдено два элемента дерева. Полученное дерево представлено на рис. 9.

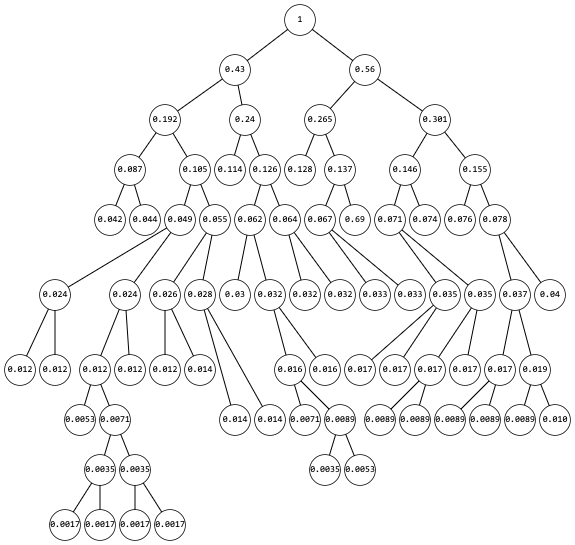


Рис. 9: Построенное дерево Хаффмана

1. В построенном дереве все элементы, которые не состоят из двух других, являются вероятностью появления какого-либо символа. Для каждой такой вероятности найден код: двигаясь по дереву сверху вниз записывались цифры 0 или 1. Если составной элемент находится справа – записывается 1, а если слева – 0. Построенные по вероятностям коды – это то, как будет кодироваться тот или иной символ с данной вероятностью. Полученные коды для каждого символа представлены на рис. 10.

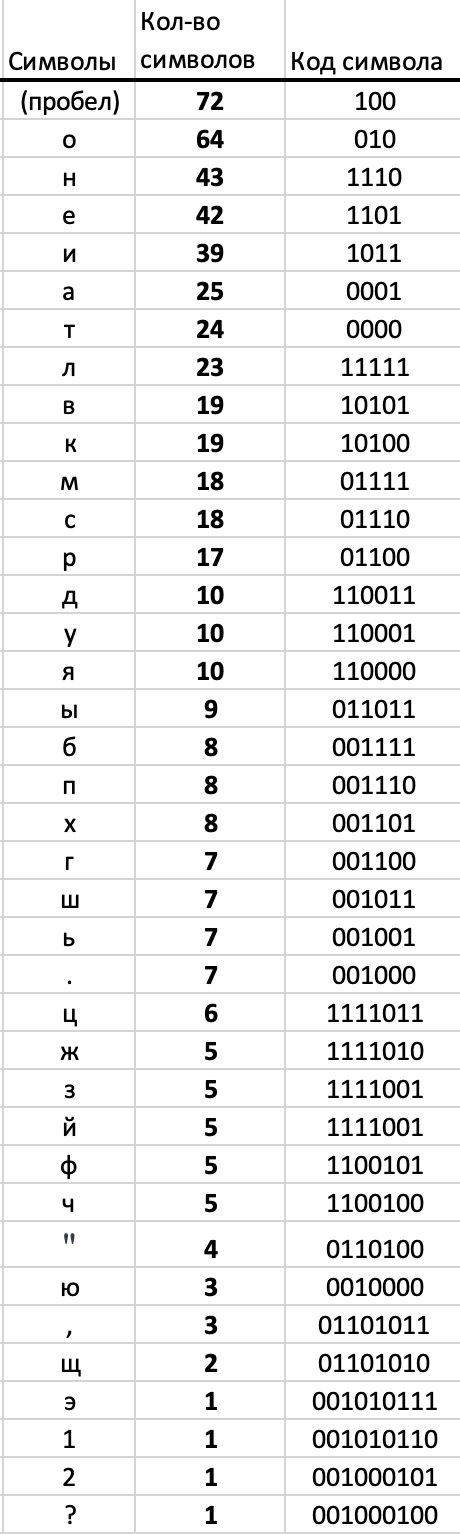


Рис. 10: Коды для каждого их символов

1. Также как и для кода Шеннона-Фано для построенного кода Хемминга были найдены значения среднего числа символов на букву сообщения, а также энтропия. Проверено выполнение теоремы Шеннона. Полученные результаты представлены на рис. 11.

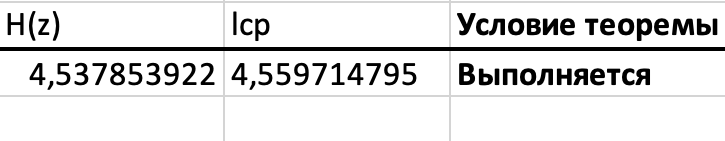


Рис. 11: Результаты построения кода Хемминга

Часть 2: помехоустойчивое кодирование.

1. Кодирование с контролем чётности.

Даны следующие три числа:

А=11000110001110111101

В=10101011110001000011

С= 10000011110011110001

1. Для каждого числа путём сложения по модулю два рассчитан контрольный разряд. Если количество единиц числа чётно – контрольный разряд равен нулю, если же нет, то единице. Во всех заданных числах контрольный разряд равен нулю (рис. 12).

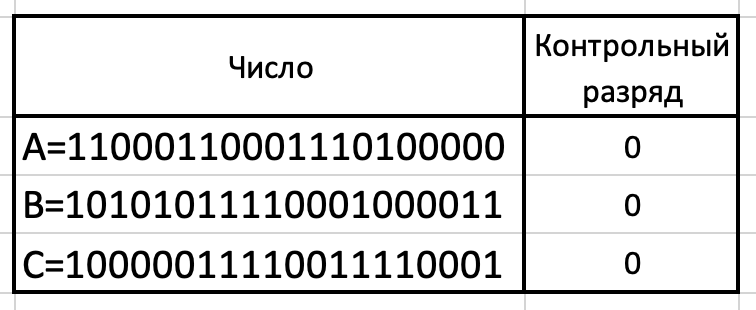


Рис. 12: Вычисление контрольных разрядов

1. В два числа внесено по одной ошибке. Для проверки в какие числа она была внесена был заново вычислен контрольный разряд (рис. 13).

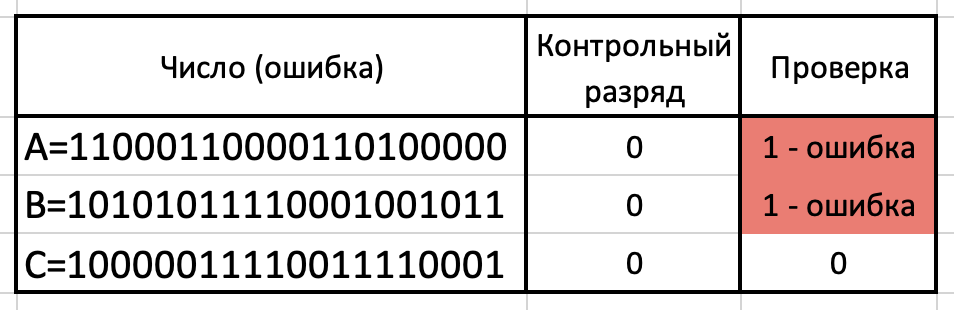


Рис. 13: Обнаружение значений с ошибкой

1. Матричное кодирование.
2. Каждое из чисел выписано в матрицу размером 5x4. Для каждых столбца и строки матрицы найден контрольный разряд (рис. ).

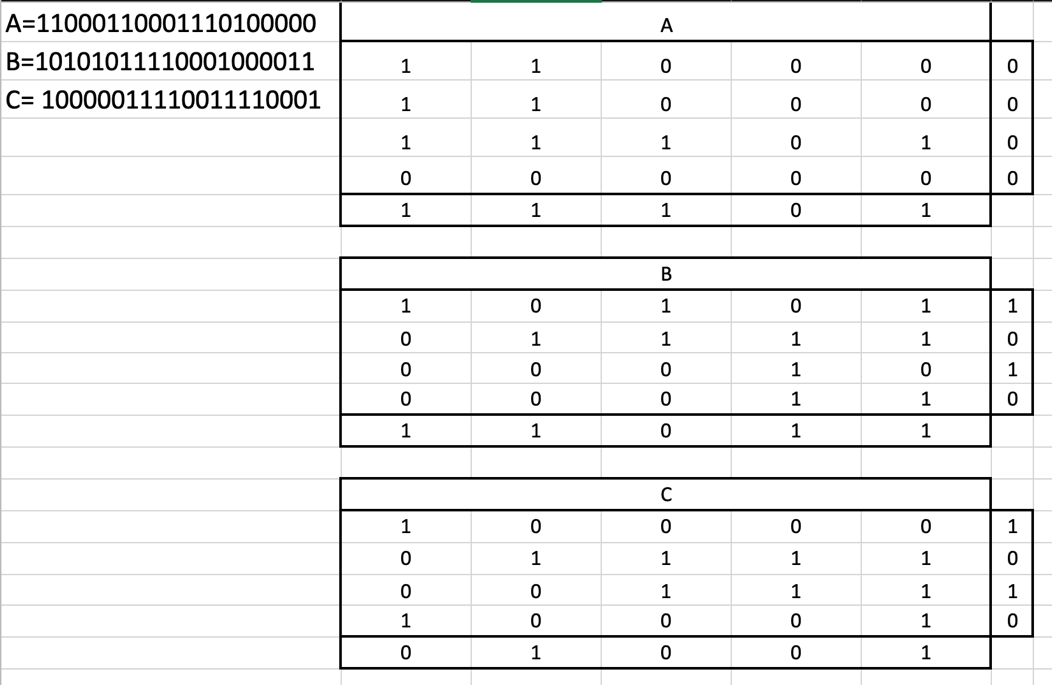


Рис. 14: Выполнение матричного кодирования

1. В каждое число в матрице внесена ошибка. Для нахождения ошибочного бита были заново вычислены контрольные разряды. Ошибка находится на пересечении отличающихся контрольных разрядов (рис. 15).

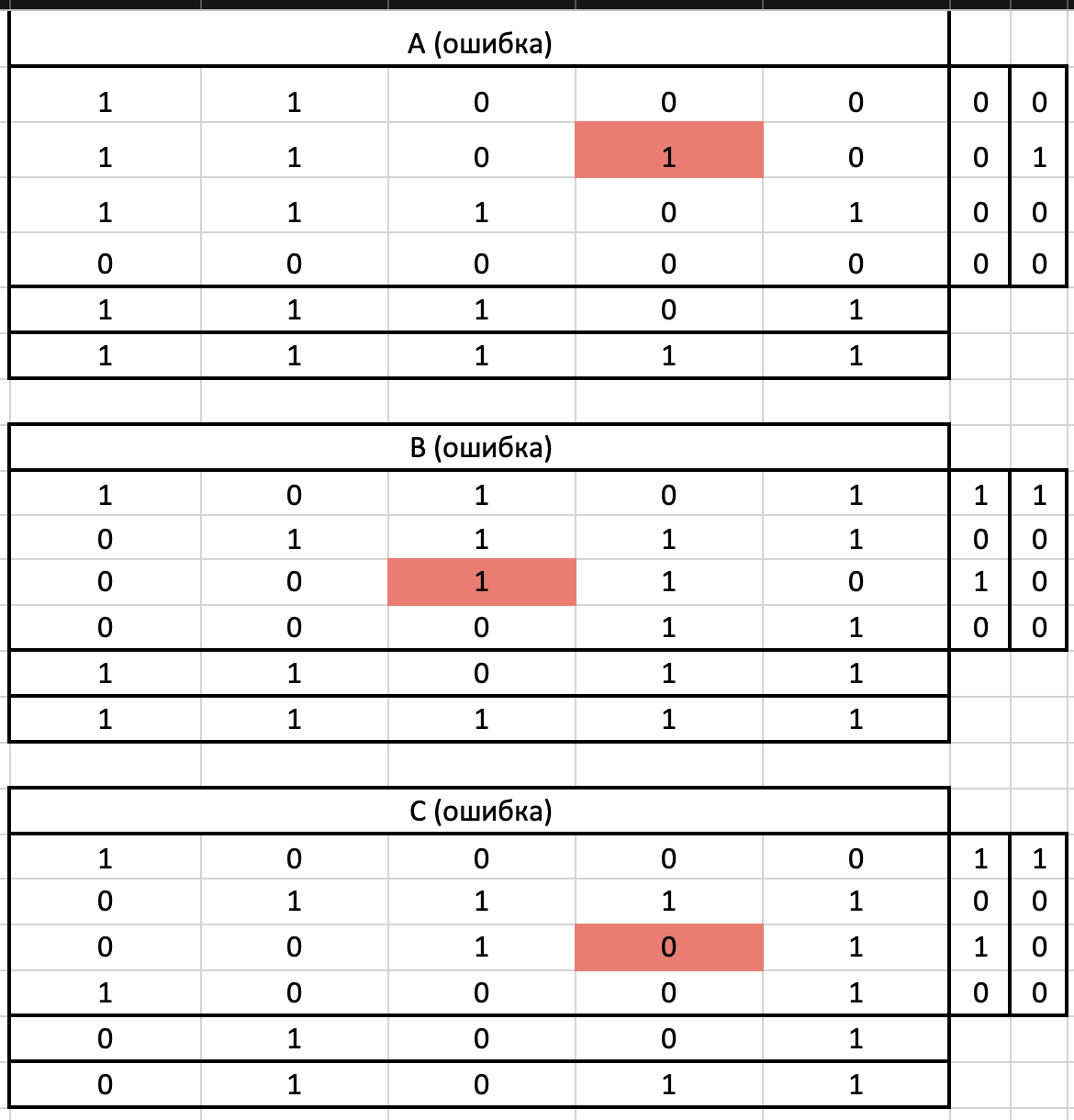


Рис. 15: Нахождение ошибочных битов в матрицах

1. Код Хемминга.

Число для кодирования – 45 (30 + 15 (порядковый номер в списке группы)).

1. Число 45 было переведено в двоичную систему счисления (101101).
2. Для двоичного числа найдено количество информационных разрядов (единиц в числе) () и контрольных разрядов () по формуле (2.1).
3. Создана таблица из 9 ячеек (6 разрядов числа + 3 контрольных). В ячейки, начиная с крайней правой, в соответствующие степеням двойки ячейки (1, 2, 4) записаны нули как контрольные разряды. В оставшиеся ячейки слева направо записаны разряды числа. Полученная таблица представлена на рис. 16.

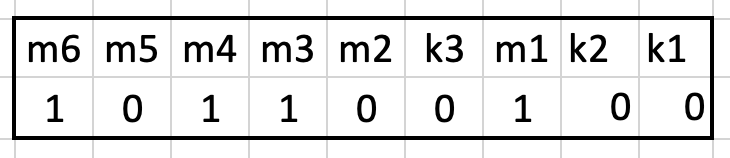


Рис. 16: Вспомогательная таблица разрядов

1. Вычислены контрольные разряды:

* Первый – сумма по модулю два по битам через один (2.2).
* Второй – сумма по модулю два по битам через два (2.3).
* Третий – сумма по модулю два по битам через четыре (2.4).

1. Вычисленные контрольные разряды записаны в таблицу вместо нулей (рис. 17).

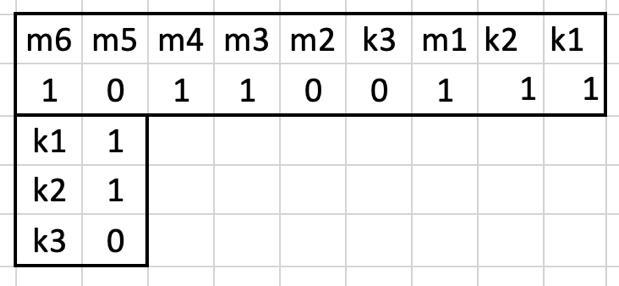


Рис. 17: Построенная таблица разрядов

1. В случайное место таблицы внесена ошибка. Для её обнаружения по идентичному алгоритму были пересчитаны контрольные разряды. Вычисленные контрольные разряды записаны справа налево, что дало на выходе положение ошибочного бита начиная справой стороны. Результат нахождения ошибки представлен на рисунке ниже (рис. 18):

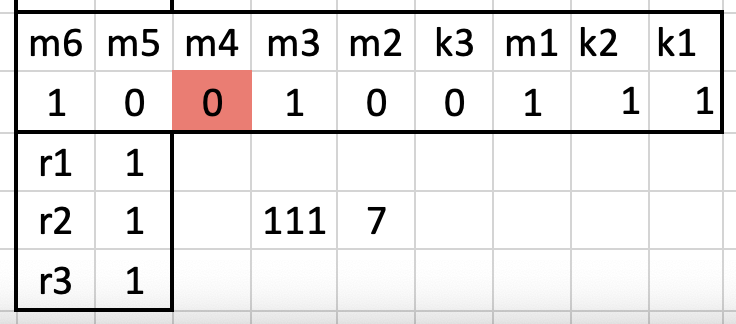


Рис. 18: Нахождение ошибки

РЕЗУЛЬТАТ ВЫПОЛНЕНИЯ РАБОТЫ

Ниже представлены все построенные в ходе выполнения лабораторной работы таблицы (табл. 1-5):

Таблица 1

Кодирование методом Шеннона-Фано

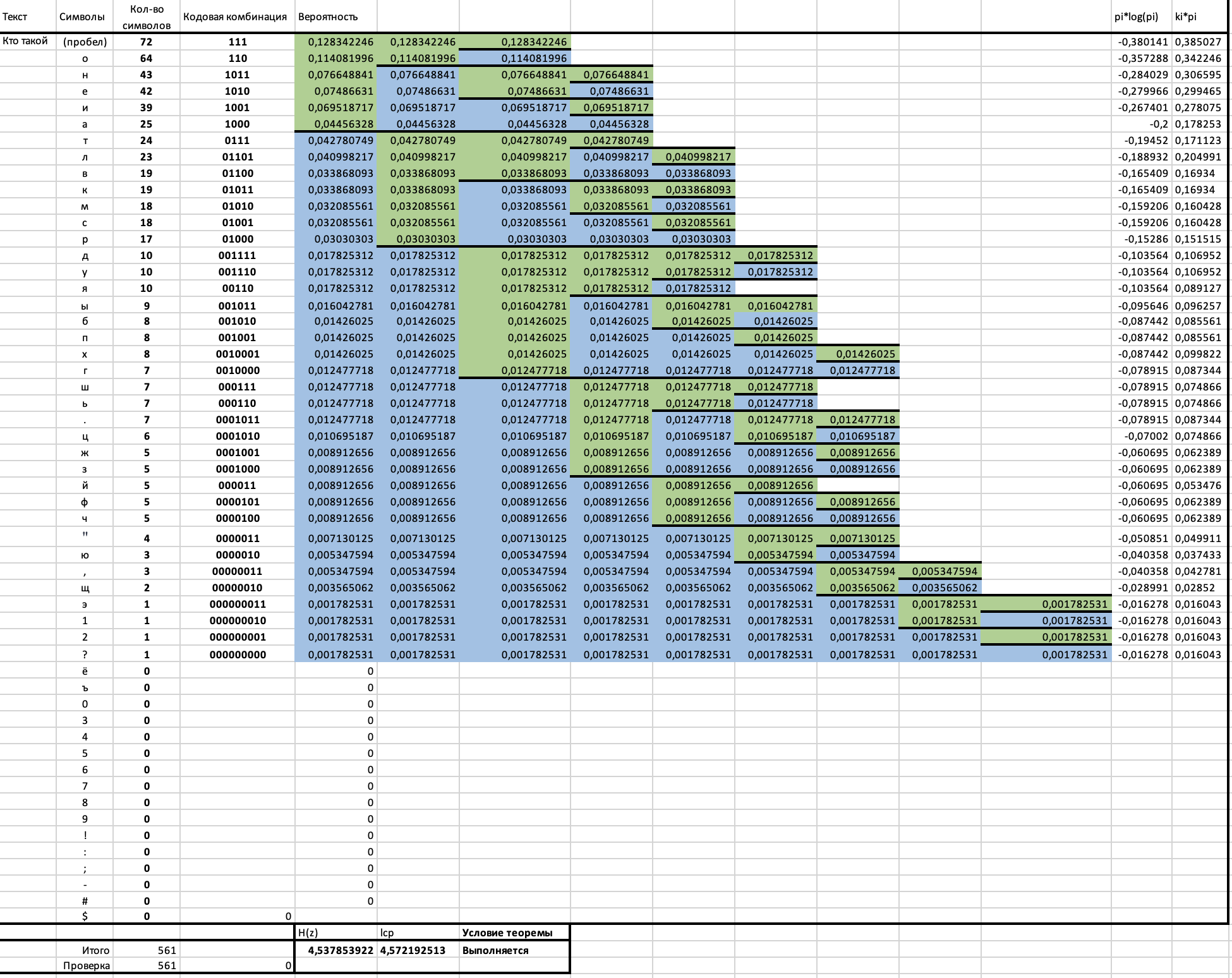


Таблица 2

Кодирование методом Хаффмана

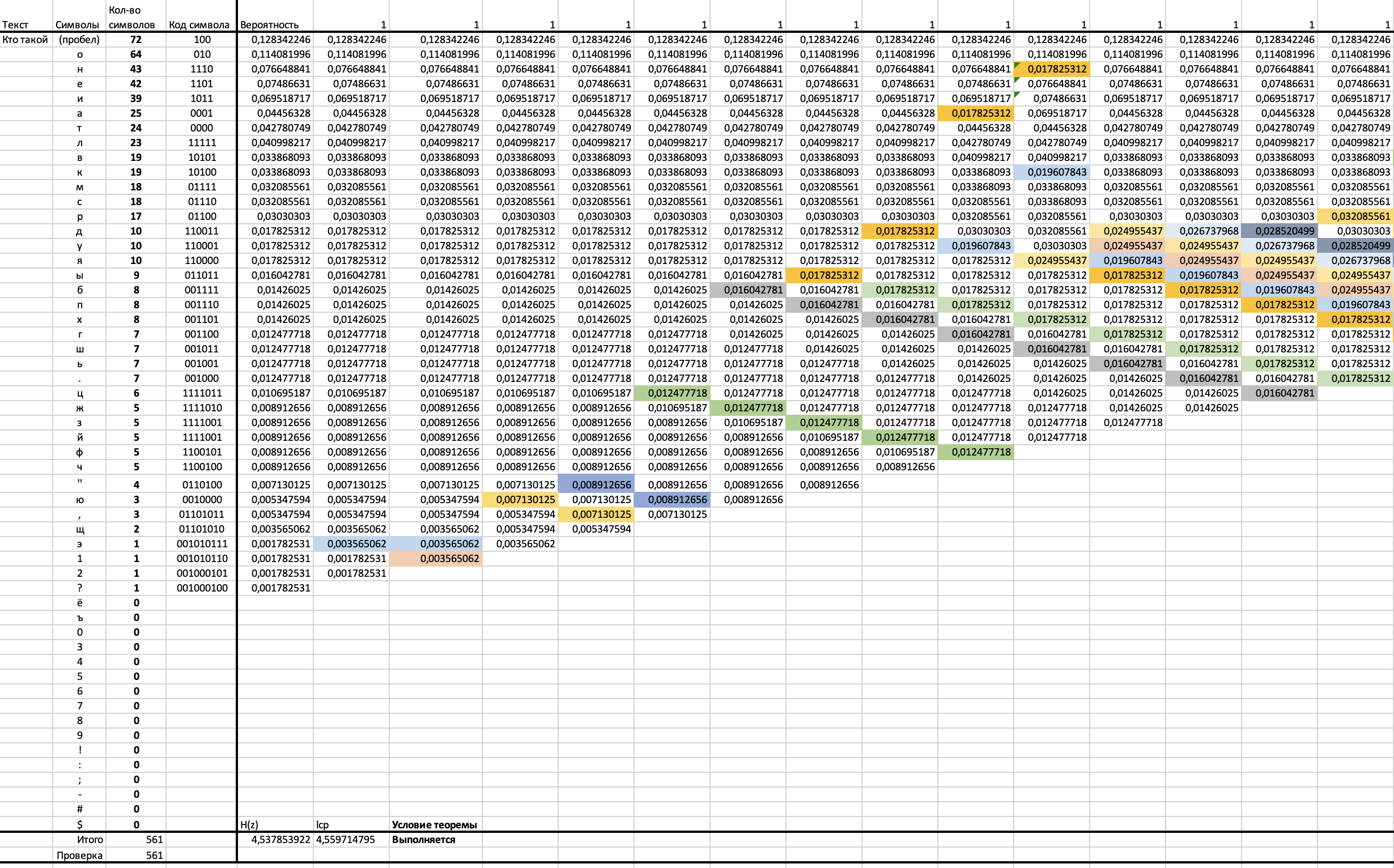


Таблица 2. Продолжение 1

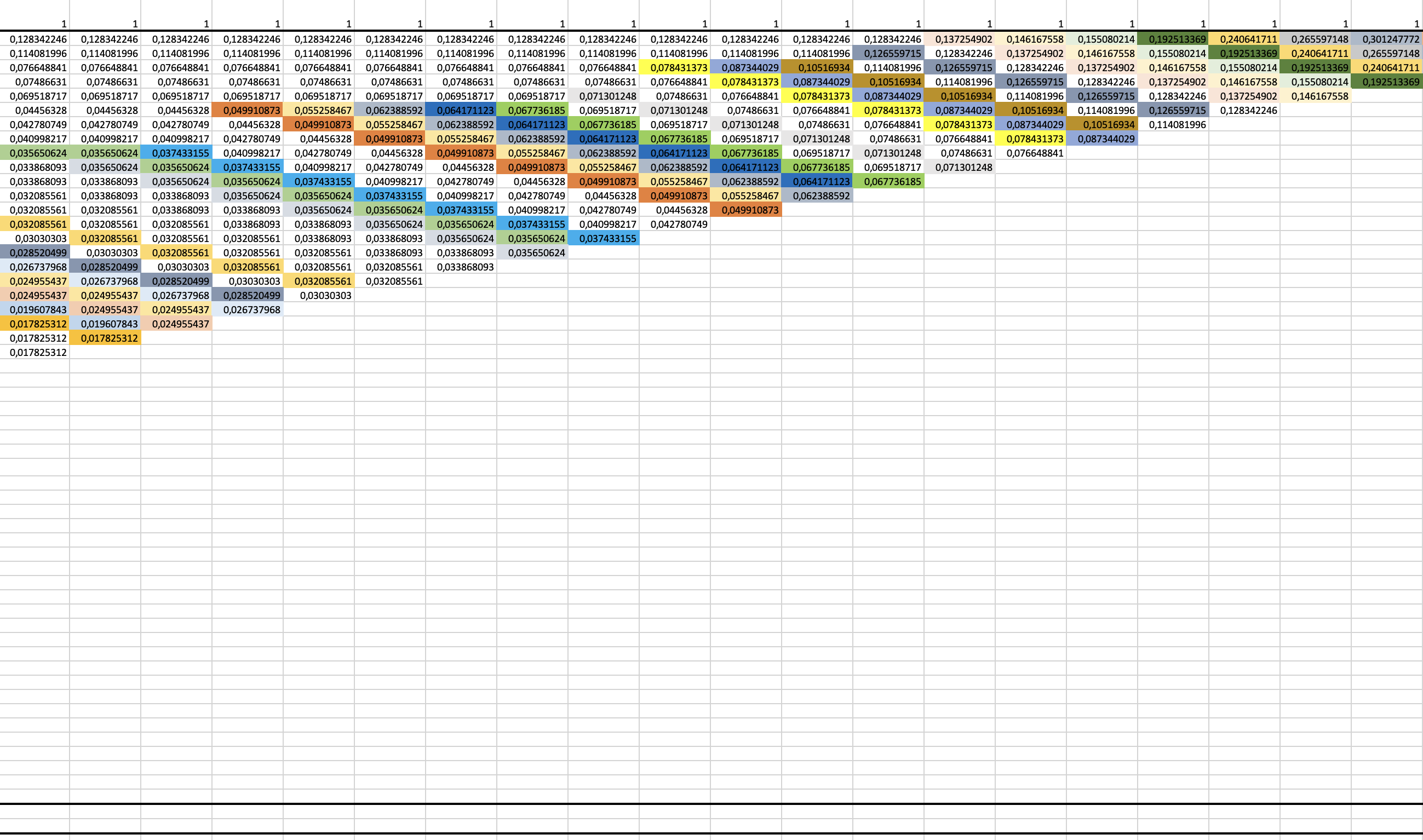


Таблица 2. Продолжение 2

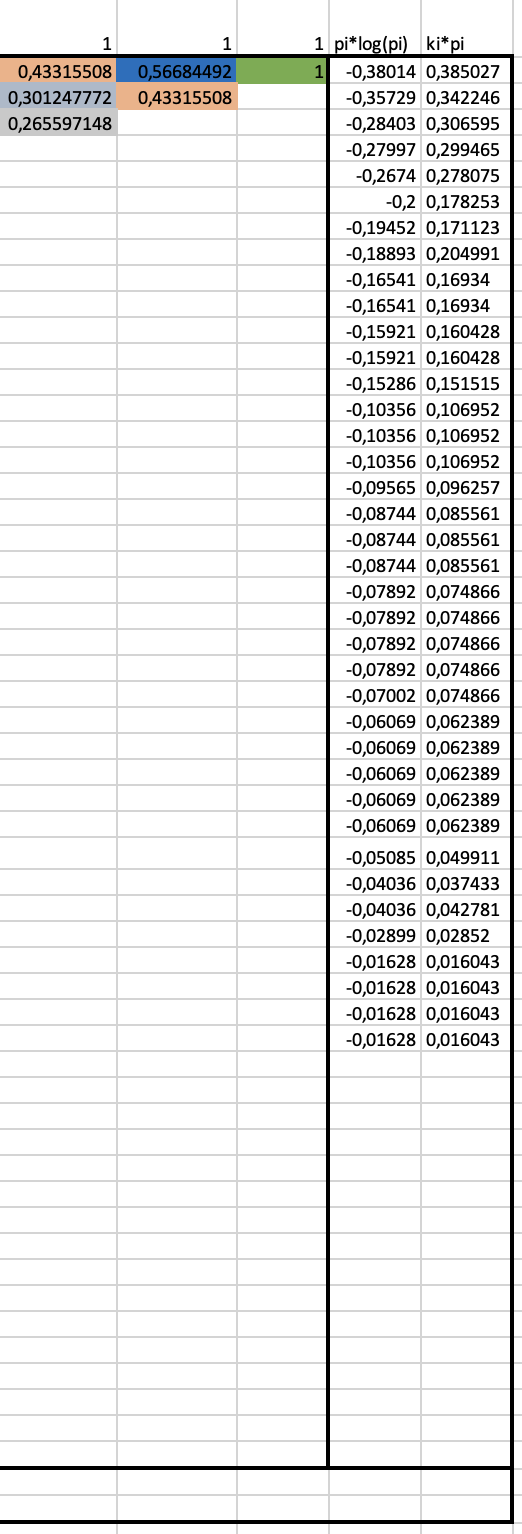


Таблица 3

Кодирование с контролем чётности

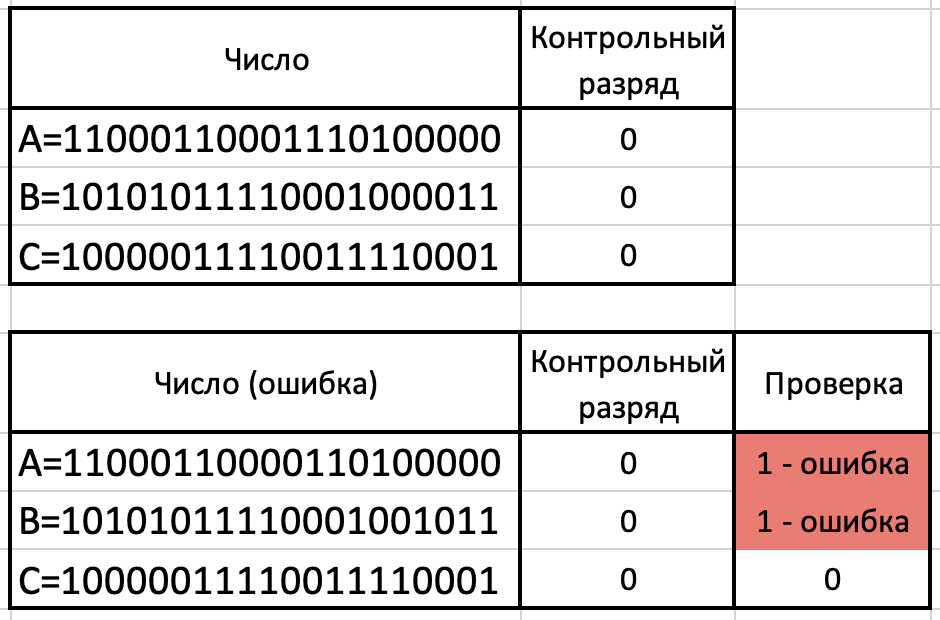


Таблица 4

Матричное кодирование

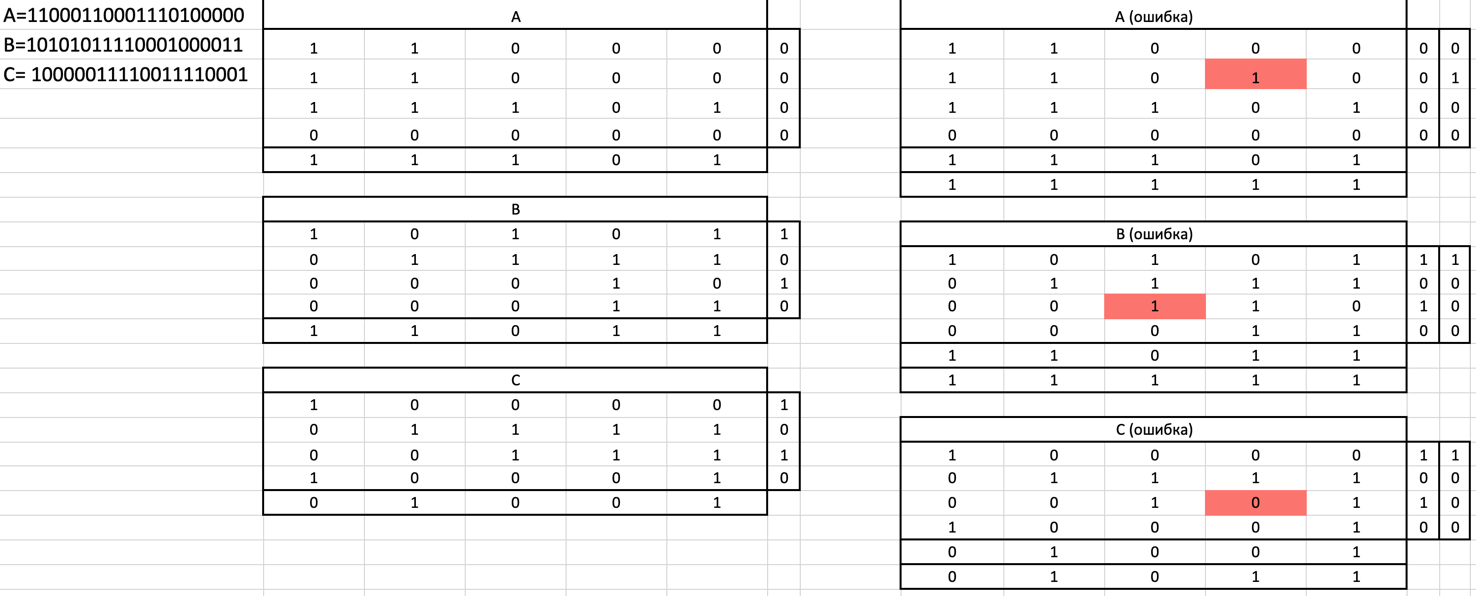
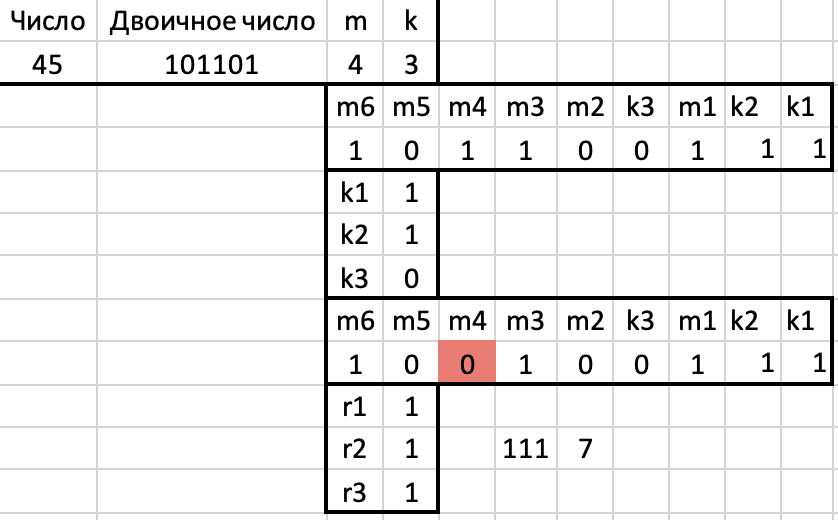


Таблица 5

Построение кода Хемминга



ЗАКЛЮЧЕНИЕ

В ходе лабораторной работы было выполнено эффективное и помехоустойчивое кодирование сообщений различными способами.

Выполняя эффективное кодирование, были использованы два метода: метод Шеннона-Фано и метод Хаффмана. Степень сжатия информации для первого метода составила . Для второго метода степень сжатия информации составил Энтропия для сообщения равна . Из этого получаем что

Данное неравенство показывает, что, используя метод Хаффмана можно более эффективно закодировать сообщение, т.к. при использовании этого метода степень сжатия более близка к значению энтропии, чем при использовании метода кодирования Шеннона-Фано.

Выполняя помехоустойчивое кодирование, были использованы следующие методы: кодирование с контролем чётности, матричное кодирование, код Хемминга.

Используя кодирование с контролем чётности в сообщение был добавлен один контрольный бит, что помогло обнаружить ошибочное сообщение.

Используя матричное кодирование в сообщение, были добавлены контрольные биты для каждых строк и столбцов матрицы, которые позволили не только найти ошибку, но и исправить её.

Используя код Хемминга в исходное сообщение, было добавлено некоторое количество контрольных разрядов, по которым удалось найти и исправить ошибку.

Каждый метод помехоустойчивого кодирования отличался разной степенью избыточности: для первого способа добавлялся 1 бит информации, для второго – 9, а для третьего 3 (5 для сообщения из 20 бит). Исходя из этого, очевидно, что код Хемминга позволяет избегать ошибок при передаче сообщений при добавлении меньшего количества избыточности.