

Содержание

ВВЕДЕНИЕ	4
1. АЛГОРИТМ КОДИРОВАНИЯ ХАФФМАНА.....	6
1.1. Цель.....	6
1.2. Краткие теоретические сведения.....	6
1.3. Задания для самостоятельной проработки.....	14
1.4. Рекомендации к выполнению лабораторного задания.....	14
Контрольные вопросы.....	26
2. АЛГОРИТМ КОДИРОВАНИЯ ХЕММИНГА	27
2.1. Цель.....	27
2.2. Краткие теоретические сведения.....	27
2.3. Задания для самостоятельной проработки.....	36
2.4. Рекомендации к выполнению лабораторного задания.....	37
Контрольные вопросы.....	43
3. АЛГОРИТМ КОДИРОВАНИЯ РИДА–МАЛЛЕРА.....	45
3.1. Цель.....	45
3.2. Краткие теоретические сведения.....	45
3.3. Задания для самостоятельной проработки.....	50
3.4. Рекомендации к выполнению лабораторного задания.....	51
Контрольные вопросы.....	53
4. СВЕРТОЧНЫЕ КОДЫ	54
4.1. Цель.....	54
4.2. Краткие теоретические сведения.....	54
4.3. Задания для самостоятельной проработки.....	62
4.4. Рекомендации к выполнению лабораторного задания.....	62
Контрольные вопросы.....	67
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	69

ВВЕДЕНИЕ

Предлагаемое учебно-методическое пособие необходимо для усвоения материала, преподаваемого по дисциплинам "Алгоритмы кодирования и шифрования информации" и "Основы кодирования и шифрования информации". Материал пособия составлен в соответствии с требованиями стандарта специальности «Радиоэлектронные системы и комплексы» по данной дисциплине. В основу пособия положены основополагающие принципы построения цифровых систем связи.

Для успешного усвоения разделов дисциплины необходимы знания, полученные студентами при изучении предшествовавших дисциплин, касающихся высшей математики, информатики, теории электросвязи, цифровых устройств в связи. Материал дисциплины будет использоваться для изучения специальных дисциплин, в которых цифровая обработка сигналов на базе сигнальных процессоров имеет первоочередное значение.

Цель преподавания дисциплины:

- обеспечить студентов навыками ориентации в специфических особенностях различных видов кодов;
- сформировать у студентов системы научных понятий и научно упорядоченных базовых представлений об основах построения, параметрах и характеристиках систем кодирования и декодирования в цифровых каналах связи;
- обеспечить студентов профессиональным владением основами построения и расчета характеристик кодеков радиотехнических систем средствами LabVIEW.

В результате изучения дисциплины студенты должны:

- освоить принципы построения кодеров и декодеров и их основные функции;
- обучиться расчету параметров различных кодеров и декодеров средствами LabVIEW;

- освоить навыки проектирования кодеров и декодеров с помощью алгоритмических средств LabVIEW, основанных на работе с процессорами цифровой обработки сигналов.

При решении задач проектирования кодеров и декодеров определяются основные показатели эффективности их функционирования, которые рассмотрены в отдельных разделах. Для выполнения заданий при самостоятельной проработке и проведения лабораторных экспериментов используются персональный компьютер и среда графического программирования LabVIEW фирмы National Instruments.

В первом разделе при изучении соответствующих рекомендаций изложены основы сжимающего кода Хаффмана и основы алгоритмического проектирование кодера и декодера средствами LabVIEW.

Во втором разделе изложены основы помехоустойчивого кода Хемминга и алгоритмическое проектирование его кодера и декодера при создании виртуального прибора LabVIEW.

В третьем разделе изложены основы помехоустойчивого кодирования с высокой корректирующей способностью на примере кода Рида–Маллера и алгоритмическое проектирование его кодера при создании виртуального прибора LabVIEW.

В четвертом разделе рассмотрены основы помехоустойчивого кодирования на примере сверточных кодов и алгоритмическое проектирование кодера и декодера при создании виртуального прибора LabVIEW.

Перед выполнением заданий для самостоятельной проработки полезно ознакомиться с базовой частью основ программирования в LabVIEW.

1. АЛГОРИТМ КОДИРОВАНИЯ ХАФФМАНА

1.1. Цель

Изучение принципа построения кодера Хаффмана и его алгоритмической модели в LabVIEW.

1.2. Краткие теоретические сведения

Алгоритм кодирования Хаффмана – адаптивный алгоритм оптимального префиксного кодирования алфавита с минимальной избыточностью. Префиксный код – код, обеспечивающий запись кодового слова последовательно без разделителей с однозначным восстановлением исходного сообщения, т.е. префиксным кодом является тот, в котором для любого слова a не существует слова ab , где b – любая последовательность символов (никакое более короткое слово не является началом другого более длинного слова кода). Префиксные коды широко используются в различных областях информационных технологий, на них основаны алгоритмы сжатия информации, используемые в различных протоколах передачи данных. К префиксным кодам относятся такие распространённые коды, как Юникод, Код Хаффмана, Код Фибоначчи и др. [1].

Например, код, состоящий из слов 0, 10 и 11, является префиксным, и сообщение 01001101110 можно разбить на слова единственным образом: 0 10 0 11 0 11 10. Код, состоящий из слов 0, 10, 11 и 100, префиксным не является, и то же сообщение 01001101110 можно трактовать несколькими способами: 0 10 0 11 0 11 10 и 0 100 11 0 11 10.

Применение кода Хаффмана. Алгоритмы сжатия нашли широкое применение при передаче информации в системах связи, поскольку они предоставляют возможность передавать значительно меньшие объемы данных, увеличивая пропускную способность каналов связи. Вероятностный метод сжатия по алгоритму Хаффмана основывается на построении кодовой таблицы, в которой каждому символу присваивается код, длина которого обратно пропорциональна частоте появления этого символа. Существуют две

разновидности вероятностных методов, различающихся способом определения вероятности появления каждого символа: статические методы, использующие фиксированную таблицу частоты появления символов, рассчитываемую перед началом процесса сжатия (используется при сжатии изображений); динамические или адаптивные методы, в которых частота появления символов все время меняется и по мере считывания нового блока данных происходит перерасчет начальных значений частот (используется в протоколах связи).

Алгоритм кодирования Хаффмана используется в протоколах сжатия данных MNP (Microcom Networking Protocol), используемых в модемах. Принцип работы MNP-модема заключается в использовании при передаче информации блоков переменной длины. Модем принимает от компьютера подлежащие передаче данные и собирает их в пакет, который затем передается по линии другому MNP-модему. При сборке этого пакета информации вычисляется контрольная сумма, которая передается в конце пакета. Размер блока можно изменять от 64 до 256 байт с шагом в 64 байта, причем, на высококачественных телефонных линиях можно использовать блоки большей длины, что увеличивает скорость передачи [2].

MNP-модемы различаются по классам – от 1 до 9. Каждый класс отличается от предыдущего более высокой производительностью и расширенными возможностями. Для более совершенных классов требуется более совершенное аппаратное обеспечение, более мощный процессор для микроконтролера модема. Широко используются MNP5- и MNP7-протоколы, отличающиеся от более совершенного протокола V.42bis и имеющие совместимость.

Протокол MNP5 реализует комбинацию адаптивного кодирования с применением кода Хаффмана и группового кодирования. Реальная скорость их передачи возрастает вдвое по сравнению с номинальной скоростью передачи данных модемом.

Протокол MNP7 использует более эффективный (по сравнению с MNP5) алгоритм сжатия данных и позволяет достичь коэффициента сжатия порядка 3,1. MNP7 использует улучшенную форму кодирования методом Хаффмана в

сочетании с марковским алгоритмом прогнозирования для создания кодовых последовательностей минимально возможной длины.

Протокол V.42bis использует алгоритм словарного типа Лемпеля-Зива-Уэлча (LZW-алгоритме), обеспечивая преимущество достижения коэффициента сжатия 4,1 файлов с оптимальной структурой.

Другое применение алгоритма Хаффмана известно при сжатии 24-битных изображений в JPEG (Joint Photographic Expert Group – подразделение в рамках ISO – Международной организации по стандартизации). Алгоритм JPEG основан на дискретном косинусоидальном преобразовании – ДКП (Discrete-Cosine Transform – DCT), применяемом к матрице изображения для получения некоторой новой матрицы коэффициентов. ДКП раскладывает изображение по амплитудам некоторых частот.

Структура алгоритма JPEG [2] состоит из следующих этапов, последний из которых – это алгоритм Хаффмана:

- изображение из цветового пространства RGB преобразуется в цветовое пространство YCrCb (иногда называют YUV). В нем Y – яркостная составляющая, а Cr, Cb – компоненты, отвечающие за цвет (хроматический красный и хроматический синий);
- исходное изображение разбивается на матрицы 8 x 8;
- применяется ДКП к каждой рабочей матрице;
- для каждой компоненты (Y, U и V) задается своя матрица квантования (на этом шаге осуществляется управление степенью сжатия, и происходят самые большие потери);
- матрица 8 x 8 преобразуется в 64-элементный вектор при помощи "зигзаг"-сканирования;
- полученный вектор свертывается с помощью алгоритма группового кодирования;
- производится кодированием по Хаффману с фиксированной таблицей.

Структура алгоритма Хаффмана при заданном наборе символов, представляющих алфавит передаваемого сообщения ABCDEFGH, выглядит следующим образом [1]:

1. Формируется первый ($k = 1$) столбец, где все буквы алфавита записываются в порядке убывания значений вероятностей их встречаемости в исходном, подлежащем кодированию файле. В соответствие каждой букве ставится значение вероятности. Две буквы, занявшие в столбце предпоследнюю и последнюю позиции, с левой стороны снизу отмечаются двоичными символами соответственно "0" и "1".

2 При известном k -м столбце строится $(k + 1)$ -й столбик по тому же принципу, что и предыдущий, с той лишь разницей, что буквы, отмеченные в предыдущем столбце двоичными символами, в последующем столбце отсутствуют. В новом столбце их представляет одна составная буква со значением вероятности, равным сумме вероятностей слагаемых букв.

3. Этот процесс продолжается до тех пор, пока в очередном столбце под номером $k - n$ (n – число букв в алфавите) не окажется одна-единственная составная буква, представляющая весь алфавит исходного текста со значением вероятности, равным единице. Этот последний столбец выполняет лишь контрольную функцию.

4. Для определения кодового набора, соответствующего интересующей букве, необходимо, начиная с последнего столбца, переходя поочередно к предыдущим столбцам, рассматривая каждый раз только те буквы, которые занимают в очередных столбцах последние два места, записать код символов по Хаффману. Если в указанном участке очередного столбца оказывается интересующая буква или какая-либо составная буква, включающая эту букву, то в очередном разряде ее кодового набора записываем двоичный символ, которым отмечена эта буква или включающая ее составная буква. Этот процесс продолжаем до тех пор, пока в очередном столбце не встретится одна интересующая буква, как показано на рис.1.

Например, для кодового набора, соответствующего букве E, при движении от столбца с номером $k = 8$ к столбцу с номером $k = 7$, в котором составная буква (ACDEFGH), содержащая букву E, дает основание для записи в первом разряде кодового набора буквы E символа "0". В следующем столбце буква E встречается в составной букве (EFGH) в столбце под номером $k = 6$.

Поэтому во второй позиции кодового набора буквы E записываем символ 0. В пятом столбце наличие в составной букве EF позволяет в третьей позиции буквы E записать символ 1. Далее буква E ниже встречается во втором столбце. Исходя из этого четвертую позицию кодового набора заполняем символом 0. В итоге букве E приписываем кодовый набор 0010. Поступая аналогичным образом, получим кодовые наборы, соответствующие остальным буквам алфавита, как приведено в табл.1 [1].

$k = 1$		$k = 2$		$k = 3$	
B	0,44	B	0,44	B	0,44
A	0,08	(GH)	0,16	(GH)	0,16
C	0,08	A	0,08	(EF)	0,16
D	0,08	C	0,08	A	0,08
E	0,08	D	0,08	-----	
F	0,08	-----		₀ C	0,08
-----		₀ E	0,08	₁ D	0,08
₀ G	0,08	₁ F	0,08		
₁ H	0,08				
$k = 4$		$k = 5$		$k = 6$	
B	0,44	B	0,44	B	0,44
(GH)	0,16	(ACD)	0,24	-----	
(EF)	0,16	-----		₀ (EFGH)	0,32
-----		₀ (GH)	0,16	₁ (ACD)	0,24
₀ (CD)	0,16	₁ (EF)	0,16		
₁ A	0,08				
$k = 7$		$k = 8$			
-----		(ABCDEFGH)	1,00		
₀ (ACDEFGH)	0,56				
₁ B	0,44				

Рис. 1. Последовательность формирования кодового дерева по алгоритму Хаффмана

Особенности алгоритма Хаффмана. Из оптимальности алгоритма

Хаффмана не следует единственность варианта достижения этого оптимума. Например, кодирование букв А, В, С, D и E при следующих значениях вероятностей их встречаемости: $P(A) = 0,49$; $P(D) = 0,16$; $P(B) = 0,17$; $P(E) = 0,01$; $P(C) = 0,17$ с помощью алгоритма кодирования Хаффмана дает результат, приведенный в табл. 2.

Таблица 1

Сформированный код Хаффмана для заданного алфавита сообщения

Буква	Двоичный код
А	011
В	1
С	0100
D	0101
E	0010
F	0011
G	0000
Н	0001

Таблица 2

Сформированный код Хаффмана для заданного алфавита сообщения

Буква	Двоичный код
А	1
В	01
С	000
D	0010
E	0001

Так как в процессе кодирования сообщениям сопоставляются только концевые узлы, полученный код Хаффмана является префиксным и, следовательно, всегда однозначно декодируемым. Из приведенных таблиц видно, что полученный код является неравномерным, причем сообщению с максимальной вероятностью появления соответствует минимальная

длительность кодовой комбинации, а сообщению с минимальной вероятностью – максимальная.

Рассмотрев влияние одиночной ошибки в одном из разрядов принятой кодовой последовательности на результат декодирования, можно заключить, что искажение одного двоичного элемента (0 \rightarrow 1) может привести к появлению ошибок в нескольких кодовых словах (треку ошибок). Это является существенным недостатком рассмотренного метода кодирования.

Другим недостатком является необходимость априорной информации при декодировании сообщения о таблице кодов Хаффмана, которая использовалась в кодере.

Следовательно, длина сжатого сообщения увеличивается на длину таблицы частот, которая должна посылаться впереди данных, что приводит к увеличению размеров выходного сообщения.

Таким образом, необходимость наличия полной частотной статистики перед началом кодирования требует двух итераций: одной – для построения модели сообщения (таблицы частот и дерева), другой – для кодирования [3,4–6].

Если источник дискретных сообщений, алфавит которого K , то при кодировании сообщений данного источника двоичным, равномерным кодом, потребуется $L_{pk} = \log_2 K$ двоичных элементов на кодирование каждого сообщения. Если вероятности $P(a_i)$ появления всех сообщений источника равны, то энтропия источника (или среднее количество информации в одном сообщении) максимальна и равна $H_{max}(x) = \log_2 K$.

В данном случае каждое сообщение источника имеет информационную емкость $\log_2 K = L_{pk}$ бит, и для его кодирования требуется двоичная комбинация не менее L_{pk} элементов. Каждый двоичный элемент, в этом случае, будет переносить 1 бит информации. Если при том же объеме алфавита сообщения не равновероятны, то, как известно, энтропия источника будет меньше

$$H_{\text{всех}}(x) = -\sum_{i=1}^K p(a_i) \log_2 p(a_i) < H_{\text{max}}(x).$$

Если и в этом случае использовать для передачи сообщения L_{pk} -разрядные

кодовые комбинации, то на каждый двоичный элемент кодовой комбинации будет приходиться меньше чем 1 бит. Появляется избыточность, которая может быть определена по следующей формуле:

$$D = \frac{H_{\max}(x) - H_{\text{реал}}(x)}{H_{\max}(x)} = 1 - \frac{H_{\text{реал}}(x)}{H_{\max}(x)} \left[\frac{\text{бит}}{\text{на элемент}} \right].$$

Таким образом, на одно сообщение будет затрачено в среднем меньшее единичных элементов, чем при равномерном:

$$\bar{l} = \sum_{i=1}^K p(a_i) l_i < L_{\text{рк}}.$$

Шеннон показал, что:

1. Нельзя закодировать сообщение двоичным кодом так, что средняя длина кодового слова \bar{l} была численно меньше величины энтропии источника сообщений $H_{\text{реал}}(x) \leq \bar{l}$, где $\bar{l} = \sum_{i=1}^K p(a_i) l_i$.

2. Существует способ кодирования, при котором средняя длина кодового слова немногим отличается от энтропии источника $H_{\text{реал}}(x) \leq \bar{l} < H_{\text{реал}}(x) + \varepsilon$.

При полученных сведениях о среднем количестве информации можно дать оценку эффективности кодирования. Эффективность неравномерных кодов оценивается следующими параметрами:

- коэффициентом статистического сжатия (КСС), который характеризует уменьшение числа двоичных символов на знак при применении методов эффективного кодирования по сравнению с применением равномерного кода:

$$K_{\text{сс}} = \frac{L_{\text{рк}}}{\bar{l}},$$

где $l_{\text{рк}} = \log_2 K = H_{\max}(x)$ – число двоичных элементов на кодирование каждого слова;

$H_{\max}(x) = \log_2 K$ – максимальна энтропия источника.

Пределы изменения КСС лежат в диапазоне от равномерного кода до наилучшего способа кодирования:

$$\frac{H_{\max}(x)}{H_{\text{реал}}(x)} \geq K_{\text{сэ}} \geq 1;$$

• коэффициентом относительной эффективности (КОЭ), который показывает степень близости средней длины кодовой комбинации к теоретически возможному пределу:

$$K_{\text{сэ}} = \frac{H_{\text{реал}}(x)}{\bar{l}}.$$

Данный коэффициент позволяет сравнить эффективность применения различных методов эффективного кодирования.

1.3. Задания для самостоятельной проработки

1. Создать виртуальный прибор для оценки вероятности встречаемости символов в считанном с диска файле.

2. Создать виртуальный прибор для упорядочения массивов символов и вероятностей по убыванию вероятностей встречаемости символов в кодируемом сообщении.

3. Создать виртуальный прибор для выделения двух символов с минимальными вероятностями в каждом наборе и формирования матриц нулей и единиц.

4. Создать виртуальный прибор для формирования кодовых комбинаций Хаффмана, закодированного сообщения и вычисления коэффициента сжатия.

5*. Создать виртуальный прибор для записи кодовой таблицы Хаффмана и закодированного сообщения в файл.

6*. Создать виртуальный прибор для декодирования сообщения, закодированного по коду Хаффмана.

7. Исследовать работоспособность кодера Хаффмана и построить зависимости статистического коэффициента сжатия и реального коэффициента сжатия от числа бит кодируемых символов.

1.4. Рекомендации к выполнению лабораторного задания

Для создания диаграммной панели виртуального прибора необходимо

использовать библиотеку Programming и соответствующие разделы работы с файлами File I/O, со строками String и матрицами Array, как показано на рис. 2–4. Фрагмент лицевой панели виртуального прибора для анализа вероятностей встречаемости символов в считанном из текста файле приведен на рис. 5, а на рис. 6 представлена диаграммная панель виртуального прибора.

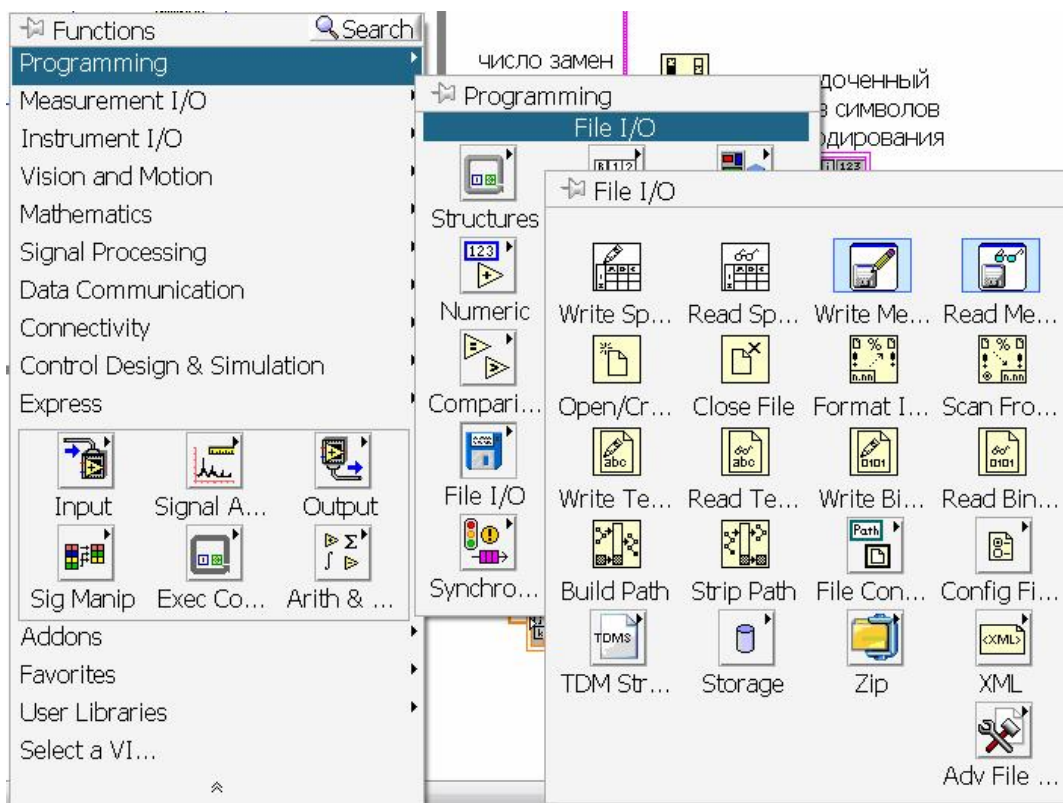


Рис. 2. Разделы библиотеки работы с файлами

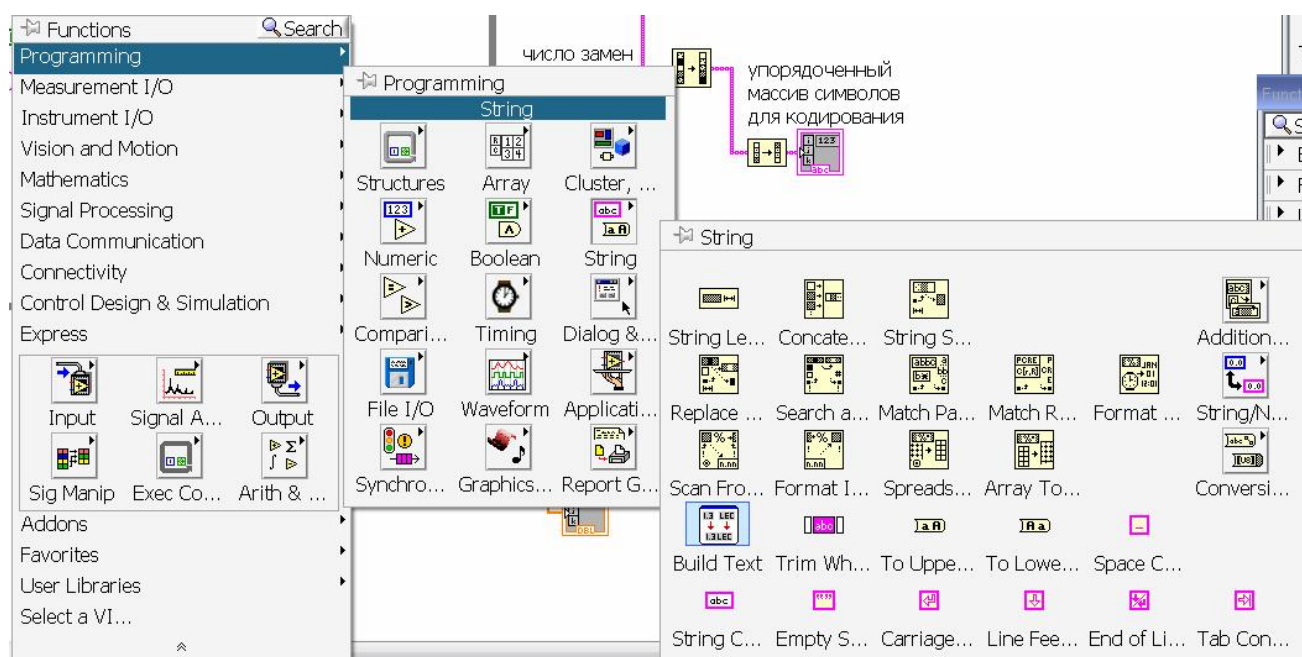


Рис. 3. Разделы библиотеки работы со строками

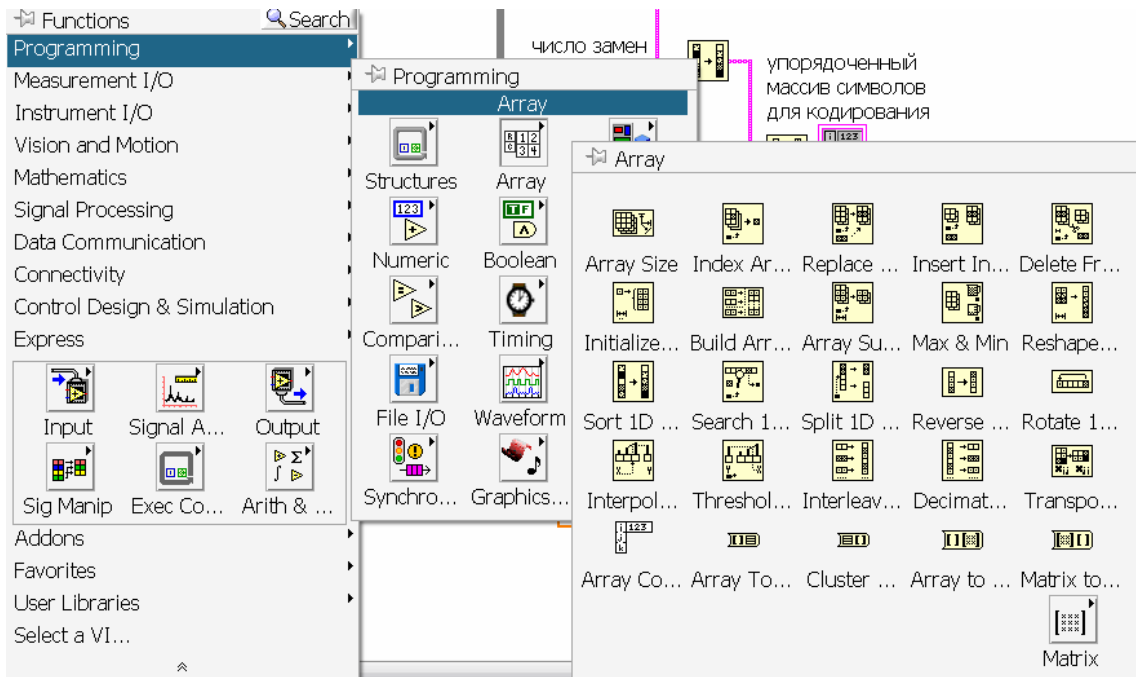


Рис. 4. Разделы библиотеки работы с матрицами

Путь к файлу
 F:\кодирование Хемминга\mnp.txt

Считанный из файла текст
 Протокол MNP7 использует более эффективный (по сравнению с MNP5) алгоритм сжатия данных и позволяет достичь коэффициента сжатия порядка 3:1. MNP7 использует улучшенную форму кодирования методом Хаффмена в сочетании

длина строки символов
 324

число символов для кодирования
 44

массив символов для кодирования	число замен	вероятность встречаемости символа в тексте
0	1	0,0030
П	10	0,0308
р	35	0,1080
о	15	0,0462
т	8	0,0246
к	13	0,0401
л	39	0,1203
М	3	0,0092
Н	3	0,0092
Р	3	0,0092
7	2	0,0061
и	23	0,0709
с	13	0,0401
п	7	0,0216
ь	5	0,0154
з	6	0,0185
	6	0,0185

Рис. 5. Лицевая панель виртуального прибора

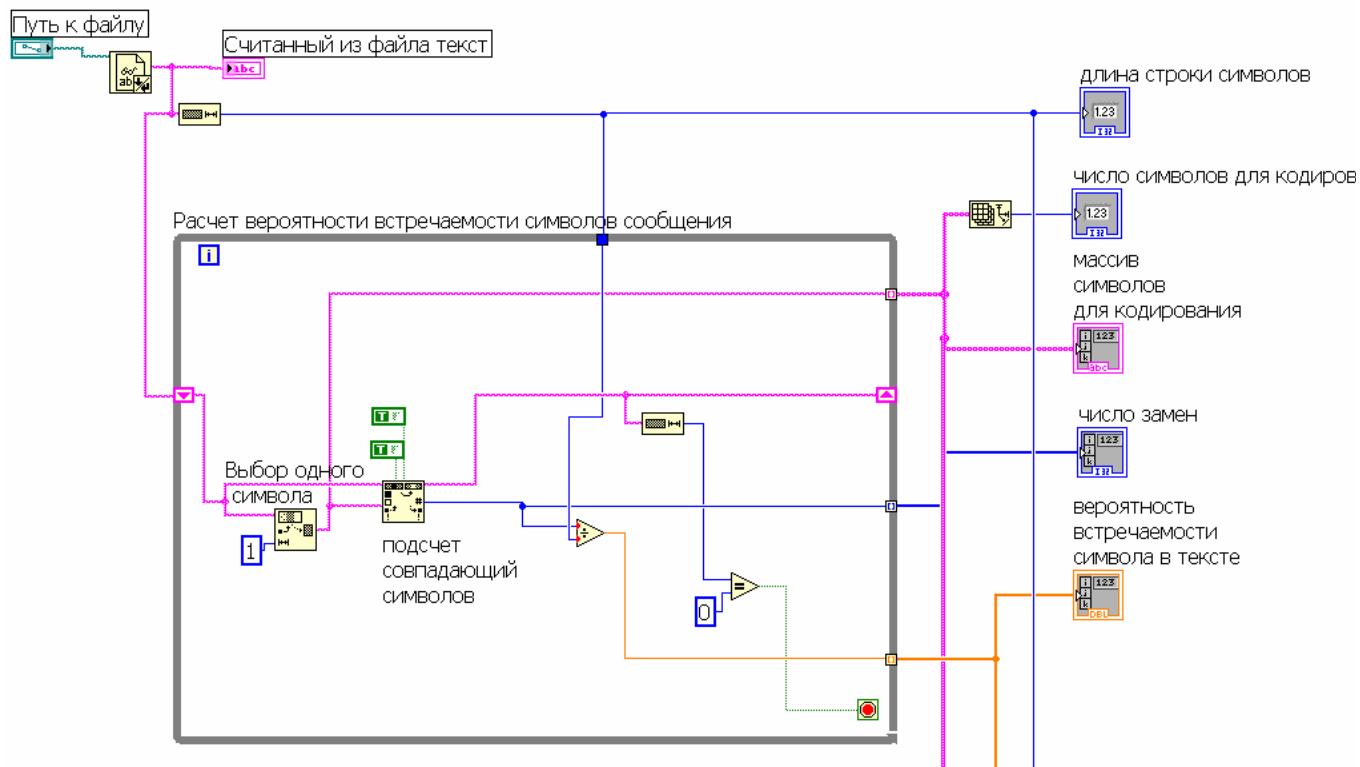


Рис. 6. Диаграммная панель виртуального прибора

Для упорядочения массива вероятностей встречаемости символов можно использовать упрощенный алгоритм, позволяющий упорядочить символы и их вероятности таким образом, чтобы выделить два символа с наименьшими вероятностями в каждом наборе и сформировать кодовые слова Хаффмана для соответствующих кодируемых символов, причем в матрицах логических пробников для логических нулей и единиц строки соответствуют кодируемым символам, а столбцы соответствуют строкам последних двух элементов наборов кодового дерева Хаффмана. Фрагменты лицевой панели виртуального прибора для этого случая приведены на рис. 7, а на рис. 8–10 представлены соответствующие фрагменты диаграммных панелей виртуального прибора. Приведенный алгоритм упорядочения массива вероятностей встречаемости символов является упрощенным представлением «пузырьковой» сортировки методом прямого обмена.

Для осуществления сжатия входного сообщения необходимо поэлементно сравнить входное сообщение с набором символов алфавита и при совпадении по соответствующему индексу выбирать кодовое слово Хаффмана

из кодовой таблицы и добавлять его в закодированное сообщение.

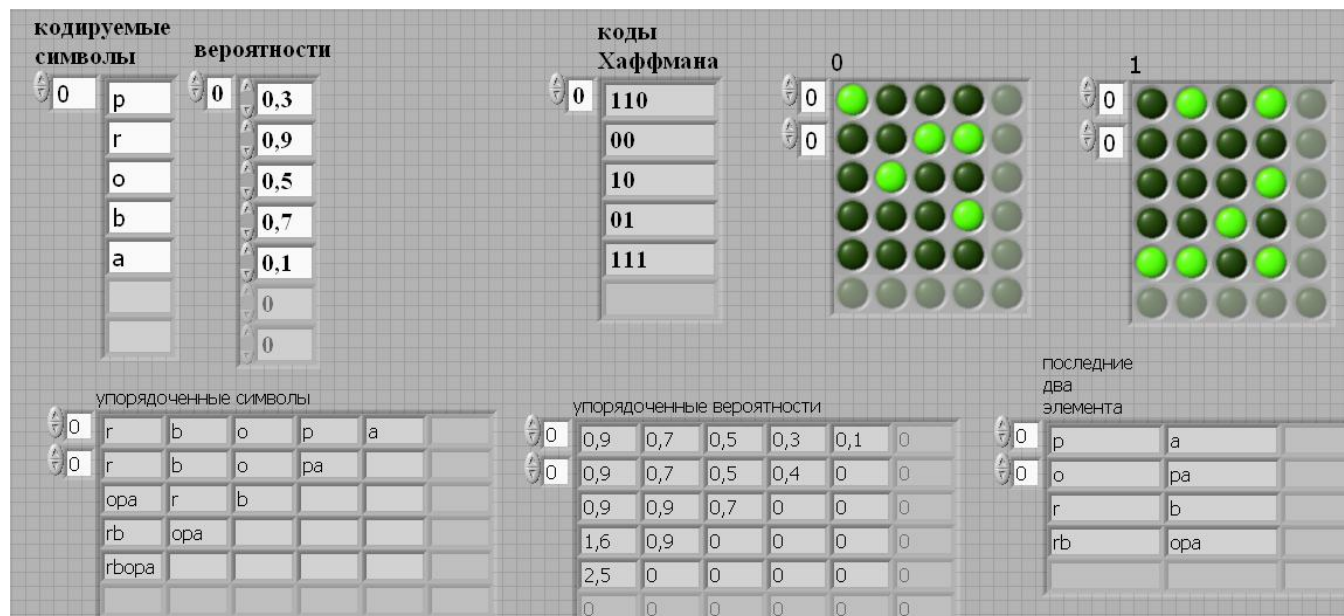


Рис. 7. Фрагмент лицевой панели виртуального прибора для упорядочения массивов символов по вероятностям их встречаемости и формирования кодовой комбинации Хаффмана

Данный процесс показан на рис. 11 во фрагменте диаграммной панели виртуального прибора. Расчет коэффициентов сжатия показан на фрагменте диаграммной панели виртуального прибора, изображенном на рис. 12, на котором статистический коэффициент сжатия вычисляется как отношение размерности входного двоичного кода к средней длине кодового сообщения Хаффмана (сумме произведений вероятности встречаемости символов на длину соответствующего кодового слова Хаффмана), оптимальный коэффициент сжатия – как отношение длины исходного сообщения в битах к длине закодированного по Хаффману сообщения и реальный коэффициент сжатия – как отношение длины исходного сообщения в битах к длине закодированного по Хаффману сообщения вместе с размерностью кодовой таблицы. Фрагмент лицевой панели виртуального прибора приведен на рис. 13. Организовать запись закодированных по Хаффману битов сообщения вместе с кодовой таблицей, содержащей двоичные коды символов алфавита и соответствующие коды Хаффмана, уже не представит большого труда.

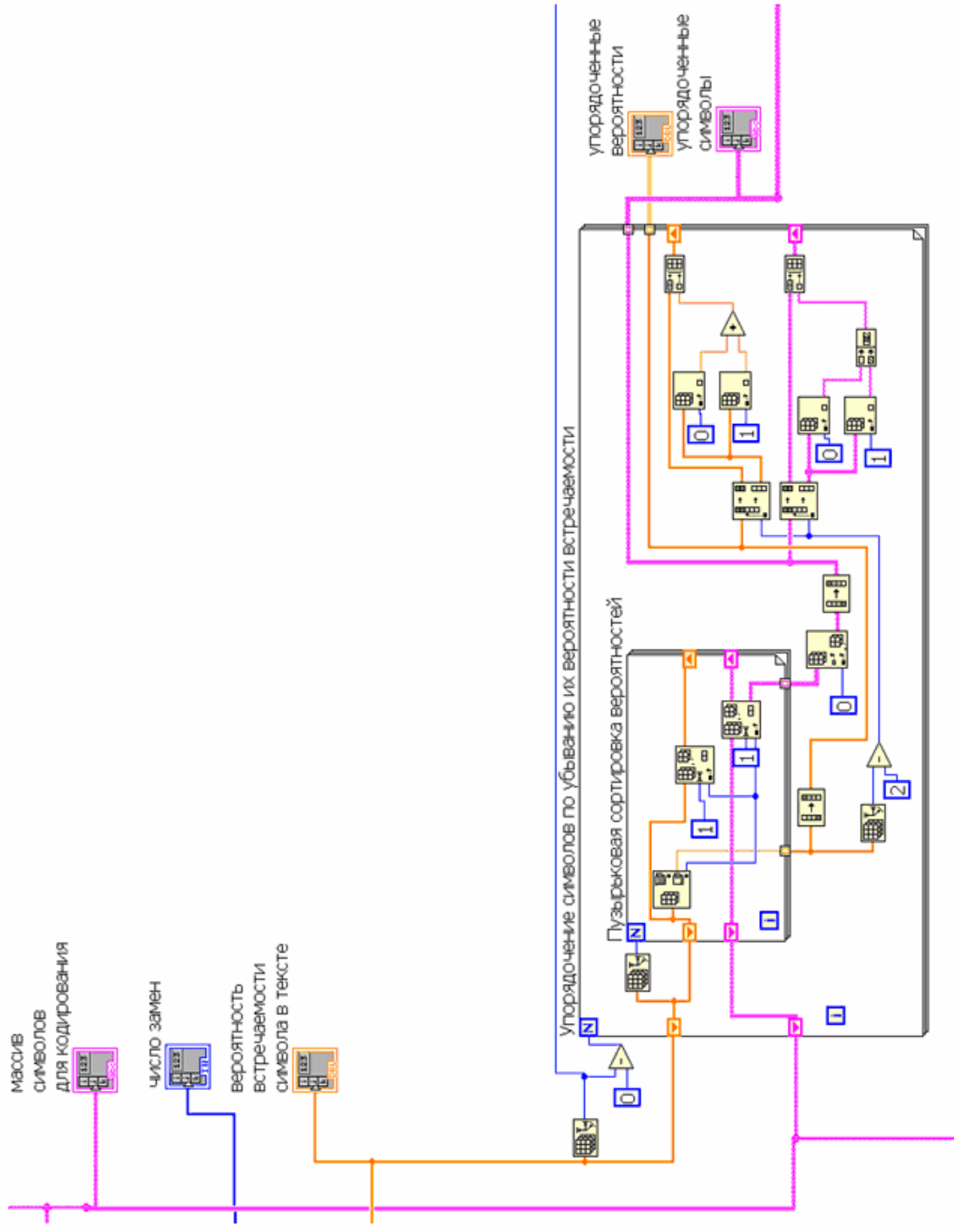


Рис. 8. Фрагмент диаграммной панели виртуального прибора для упорядочения символов по убыванию вероятности их встречаемости

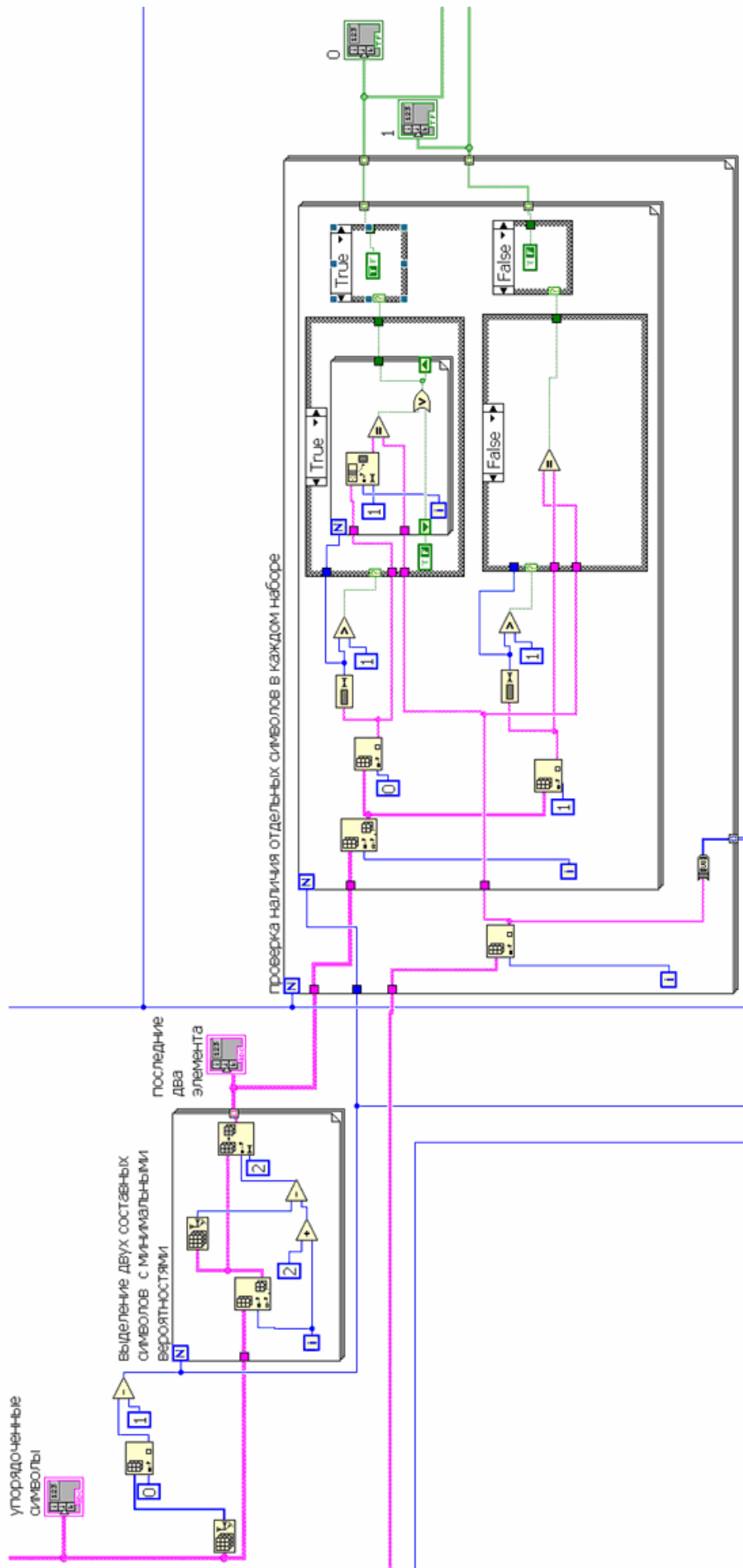


Рис. 9. Фрагмент диаграммной панели виртуального прибора для выделения двух символов с минимальными вероятностями в каждом наборе и формирования матриц нулей и единиц

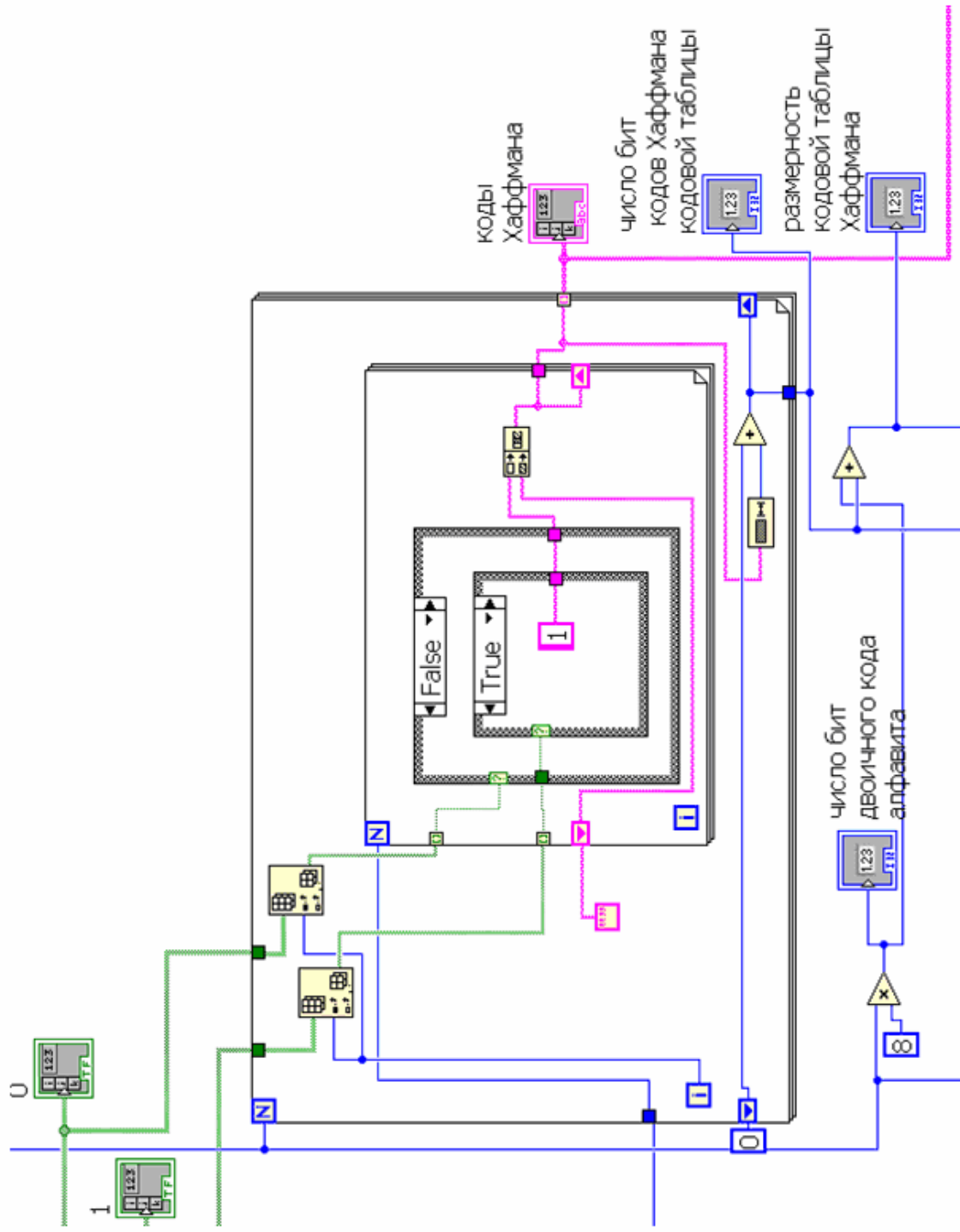


Рис. 10. Фрагмент диаграммы панели виртуального прибора для формирования кодовой таблицы Хаффмана

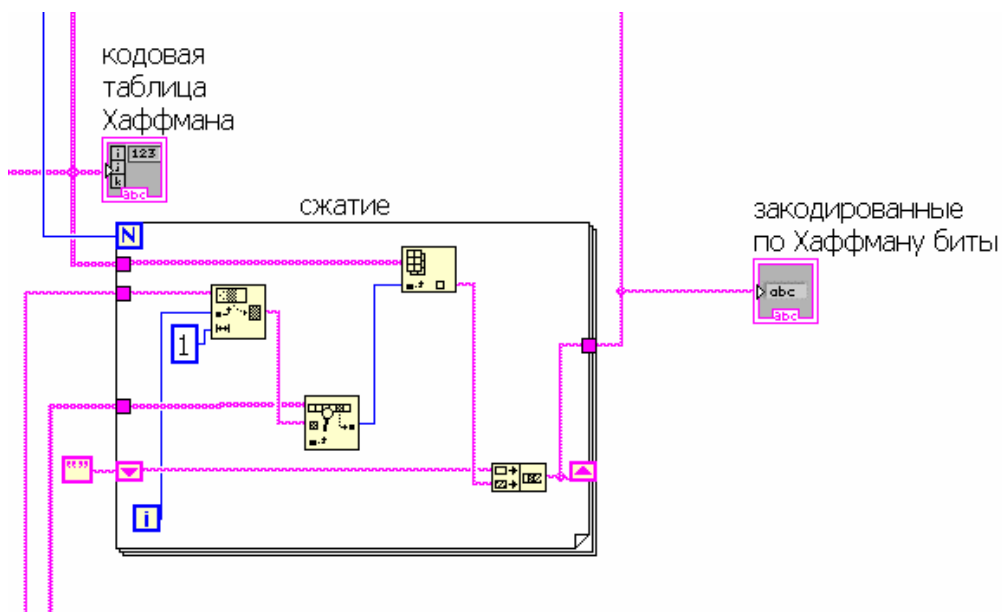


Рис. 11. Фрагмент диаграммной панели виртуального прибора для формирования закодированного сообщения по Хаффману

В итоге, пользуясь возможностями LabVIEW, можно создать библиотечный модуль – вложенный виртуальный прибор кодера Хаффмана (см. рис. 14), который можно использовать при создании других виртуальных приборов для анализа протоколов сжатия информации [7].

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru