



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: 2016133691, 20.02.2015

(24) Дата начала отсчета срока действия патента:
20.02.2015Дата регистрации:
12.09.2017

Приоритет(ы):

(30) Конвенционный приоритет:
20.02.2014 GB 1403038.1

(45) Опубликовано: 12.09.2017 Бюл. № 26

(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: 20.09.2016(86) Заявка РСТ:
EP 2015/025007 (20.02.2015)(87) Публикация заявки РСТ:
WO 2015/124323 (27.08.2015)Адрес для переписки:
191036, Санкт-Петербург, а/я 24, "НЕВИНПАТ"

(72) Автор(ы):

ВИЛЛНЕР Кай (FI),
КАЛЕВО Осси (FI)

(73) Патентообладатель(и):

Гурулоджик Микросистемс Ой (FI)

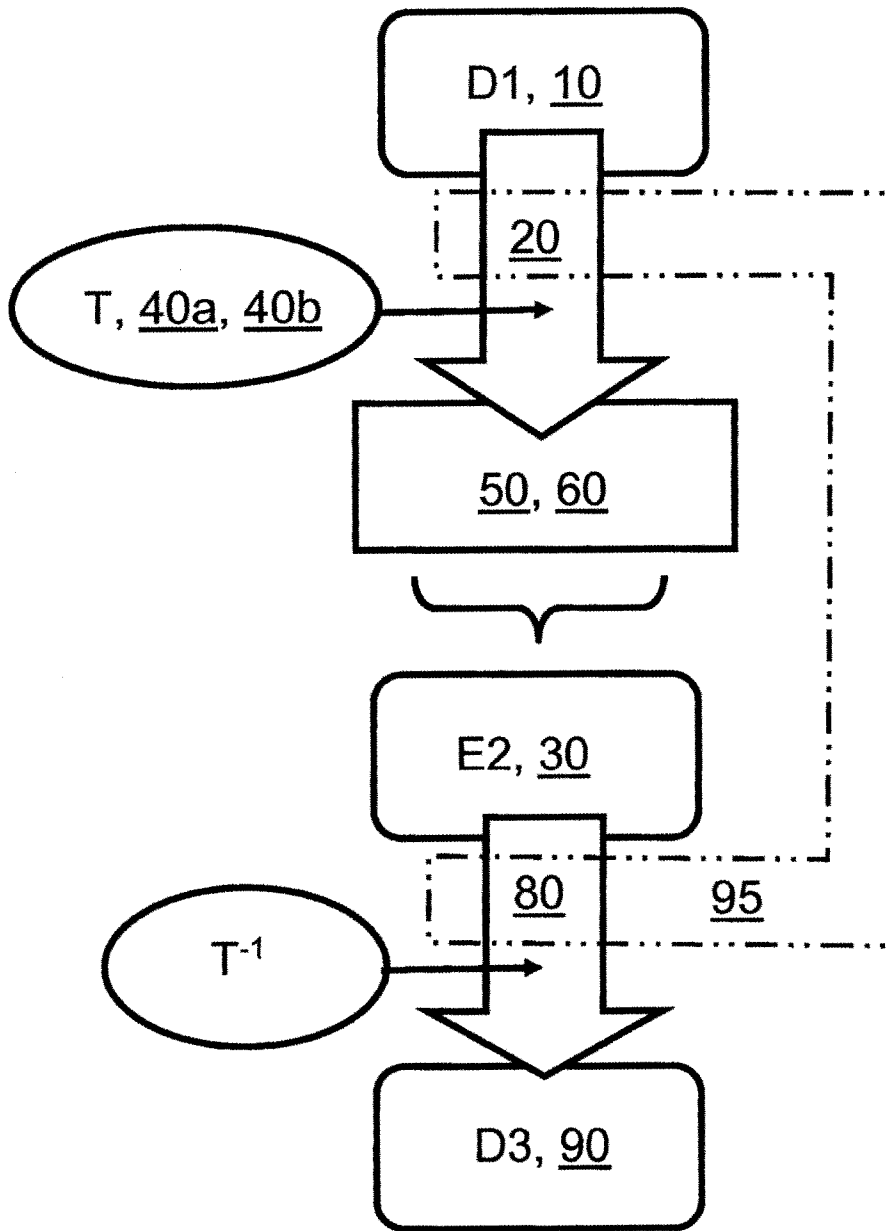
(56) Список документов, цитированных в отчете
о поиске: US 2005/0012648 A1, 20.01.2005. EP
2387243 A2, 16.11.2011. GB 2503295 A,
25.12.2013. RU 2426227 C2, 10.08.2011. WO
2011128268 A1, 20.10.2011.

(54) Устройство и способ для кодирования и декодирования исходных данных

(57) Реферат:

Группа изобретений относится к области кодирования. Техническим результатом является повышение эффективности сжатия кодированных данных. Кодер (20) для кодирования данных (D1, 10) с получением соответствующих кодированных данных (E2, 30), содержащих информацию о частотах, вероятностях или значениях диапазонов различных символов, которые должны быть представлены в кодированных данных (E2, 30), при этом упомянутая информация указывает

символы, к которым относятся упомянутые частоты, вероятности или значения диапазонов, при этом кодер (20) способен включать в кодированные данные (E2, 30) дополнительную информацию, указывающую, включена ли в кодированные данные (E2, 30) информация о частотах, вероятностях или значениях диапазонов для упомянутых различных символов. 7 н. и 14 з.п. ф-лы, 3 ил., 2 табл.



Фиг.1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(51) Int. Cl.
H03M 7/30 (2006.01)
H03M 7/40 (2006.01)
H04N 19/91 (2014.01)

(12) **ABSTRACT OF INVENTION**

(21)(22) Application: **2016133691, 20.02.2015**

(24) Effective date for property rights:
20.02.2015

Registration date:
12.09.2017

Priority:

(30) Convention priority:
20.02.2014 GB 1403038.1

(45) Date of publication: **12.09.2017** Bull. № 26

(85) Commencement of national phase: **20.09.2016**

(86) PCT application:
EP 2015/025007 (20.02.2015)

(87) PCT publication:
WO 2015/124323 (27.08.2015)

Mail address:
191036, Sankt-Peterburg, a/ya 24, "NEVINPAT"

(72) Inventor(s):

**WILLNER Kai (FI),
KALEVO Ossi (FI)**

(73) Proprietor(s):

Gurulogic Microsystems Oy (FI)

(54) **DEVICE AND METHOD FOR ENCODING AND DECODING INITIAL DATA**

(57) Abstract:

FIELD: physics.

SUBSTANCE: encoder (20) for encoding data (D1, 10) to obtain corresponding encoded data (E2, 30) containing information about the frequencies, probabilities, or range values of different symbols to be represented in the encoded data (E2, 30), wherein the said information indicates the symbols, to which the said frequencies, probabilities, or range values

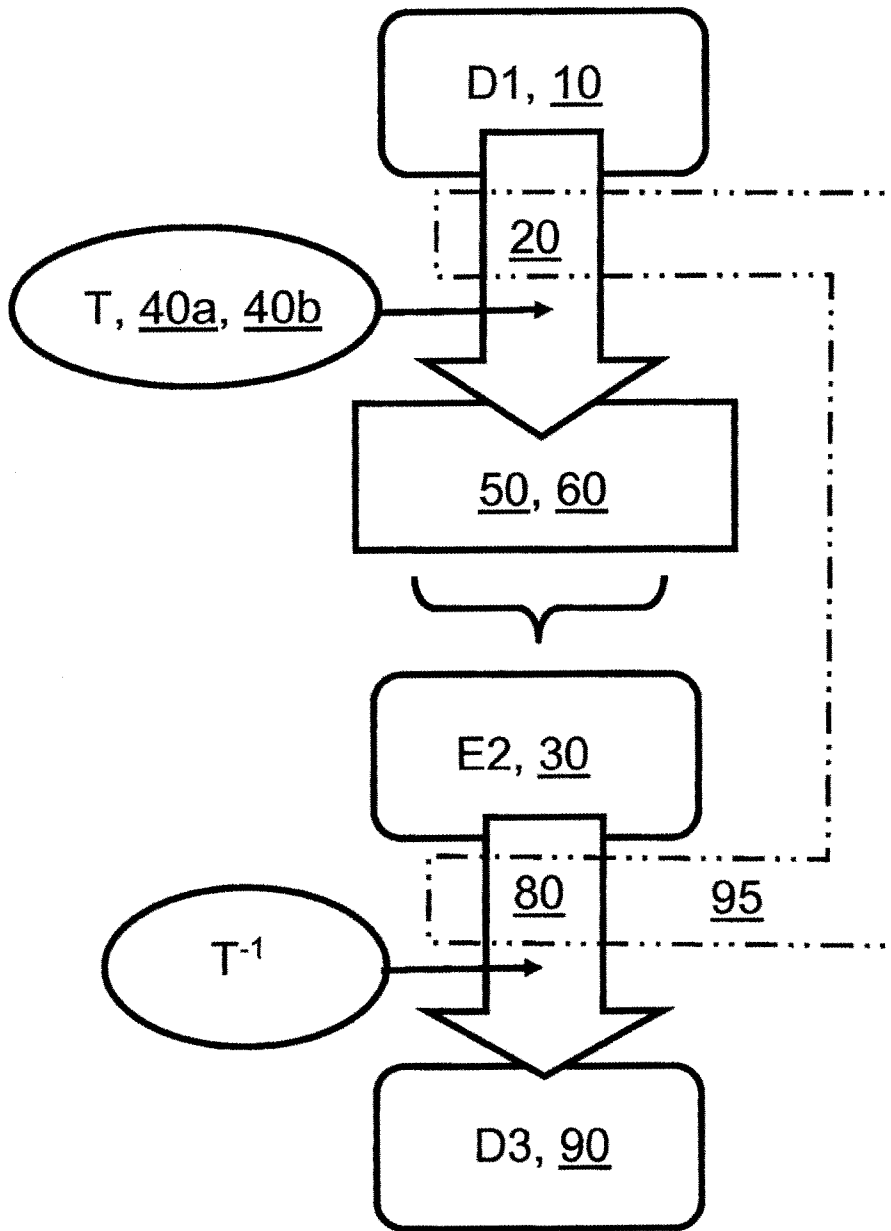
relate, wherein the encoder (20) is able to include additional information into the encoded data (E2, 30), indicating, whether the encoded data (E2, 30) include information about the frequencies, probabilities, or range values for the said different symbols.

EFFECT: increasing the efficiency of the encoded data compression.

21 cl, 3 dwg, 2 tbl

RU 2 630 750 C 1

RU 2 630 750 C 1



Фиг.1

ОБЛАСТЬ ТЕХНИКИ

Раскрытие настоящего изобретения относится к способам кодирования данных (D1) для генерации соответствующих кодированных данных (E2). Кроме того, раскрытие настоящего изобретения также относится к способам декодирования указанных выше кодированных данных (E2) для генерации соответствующих декодированных данных (D3) и/или перекодированной версии декодированных данных (D3). Помимо этого, раскрытие настоящего изобретения также относится к кодерам и декодерам, которые способны реализовать указанные выше способы и совместно образуют кодеки. Дополнительно раскрытие настоящего изобретения относится к компьютерным программным изделиям, содержащим машиночитаемый носитель информации, на котором хранятся машиночитаемые инструкции, исполняемые вычислительным устройством для выполнения указанных выше способов.

ПРЕДПОСЫЛКИ СОЗДАНИЯ ИЗОБРЕТЕНИЯ

На обзорной схеме, показанной на фиг. 1, известные кодеру 20 способы кодирования входных данных (D1) 10 для генерации соответствующих кодированных выходных данных (E2) 30 задействуют один или более процессов 40a, 40b преобразования, применяемых к входным данным (D1) 10, для генерации соответствующих кодированных преобразованных данных 50, которые связаны с данными 60 кодовых таблиц, определяющих одну или более кодовых таблиц, задающих один или более используемых процессов преобразования. Кодированные преобразованные данные 50 и данные 60 кодовых таблиц, в комбинации образующие кодированные выходные данные (E2) 30, часто с помощью носителя данных и/или через сеть передачи данных передаются в один или более декодеров 80, которые способны декодировать кодированные выходные данные (E2) 30 с целью генерации соответствующих декодированных данных (D3) 90. Дополнительно декодированные данные (D3) представляют собой восстановленную версию данных (D1), однако не ограничены только этой версией, например, если выполняется перекодирование в одном или более декодеров 80. Во многих случаях желательно, чтобы кодированные выходные данные (E2) 30 сжимались относительно входных данных (D1) 10. Кроме того, желательно, чтобы кодированные выходные данные (E2) 30 сжимались по существу без потерь, так чтобы декодированные данные (D3) 90 как можно точнее воспроизводили входные данные D1 (10). Достижимая степень сжатия кодированных выходных данных (E2) 30 относительно входных данных (D1) 10 потенциально недостаточна, если данные 60 таблиц кодирования занимают значительный объем по отношению к кодированным преобразованным данным 50, то есть данные 60 таблиц кодирования соответствуют значительному объему служебных данных. В определенных известных реализациях кодера 20 в состав данных 60 кодовых таблиц входит информация, указывающая частоты, вероятности или диапазоны кодированных "символов", присутствующих в кодированных преобразованных данных 50, которые являлись исходными символами во входных данных (D1) 10. Данные 60 кодовых таблиц не ограничены только этими перечисленными данными. Дополнительно данные 60 кодовых таблиц также содержат информацию, относящуюся к используемым способам преобразования и их параметрам.

В некоторых известных реализациях для модификации вероятностей, включенных в данные 60 кодовых таблиц, используется контекстная информация. Использование контекстной информации является очень эффективным способом модификации вероятностей, но только для определенных типов данных, например, в случае кодирования значений пикселей в изображении или значений, полученных после преобразования изображения.

Базовая реализация интервального (диапазонного) кодека описывается, например, на сайте Википедии, на странице, посвященной интервальному (диапазонному) кодированию [Интервальное кодирование - Википедия, свободная энциклопедия (по состоянию на 13 декабря 2013 года), URL: http://en.wikipedia.org/wiki/Ranqe_encoding]. В этой базовой реализации используется по меньшей мере один символ <EOM> (End-Of-Message, конец сообщения). Арифметическое кодирование, URL: http://en.wikipedia.org/wiki/Arithmetic_coding, очень напоминает интервальное кодирование, однако в этом варианте кодирования используется одно число с плавающей запятой [0,1] и символ <EOD> (End-Of-Data, конец данных). Кроме того, интервальное кодирование широко используется в тех приложениях, в которых требуется сжатие данных, таких как кодирование видеосигналов и изображений. Оно очень эффективно при кодировании наборов данных большого объема, если известны вероятности различных возможных символов. Однако если характер данных, подлежащих кодированию, таков, что вероятности символов заранее неизвестны в одном или более декодерах 80, то необходимо передавать вероятности символов совместно с кодированными данными (E2) 30. В противном случае эффективность сжатия при интервальном кодировании значительно снижается.

Доставка значений частоты, вероятности или диапазона требует включения дополнительных данных, то есть "расходования данных", что влияет на общую эффективность сжатия, достижимую в кодере 20. Кроме того, повышенная эффективность сжатия при интервальном кодировании достигается, если передаются более точные значения частот, вероятностей или диапазона символов. Однако передача очень точных значений частот, вероятностей или диапазона требует большего объема данных. Другими словами, должен соблюдаться компромисс между передачей точных значений частот, вероятностей или диапазона для символов и достижением максимально возможной эффективности сжатия в кодированных данных (E2).

В современных условиях объем повсеместно передаваемых и сохраняемых данных быстро возрастает. Поскольку ширина полосы пропускания и память для хранения данных являются ограничивающими параметрами, большое значение приобретает потребность в эффективных способах сжатия данных. Таким образом, возникает необходимость в усовершенствованных способах сжатия данных, которые обеспечивают повышенную эффективность сжатия, достигаемую при использовании интервального кодирования, например, в тех случаях, когда размер сжимаемых данных не очень большой. Следовательно, возникает потребность в способе, который позволяет использовать интервальное кодирование в тех случаях, если применение интервального кодирования и подобных ему способов ранее не представлялось возможным и/или было не выгодно.

В ранее опубликованном европейском патентном документе EP 0166607 A2 (King Reginald Alfred) описывается способ, согласно которому сжатие в сеансах передачи выборок (звуковых) данных, в частности символов с временным кодированием (TES, Time Encoded Symbol), потенциально улучшается путем представления битов нелинейным образом (в виде кодов переменной длины) вместо линейного представления (в виде кодов фиксированной длины). Способ использует первичные коды для доставки из кодера в декодер порядка вероятности появления элементов. То есть, согласно этому способу по мере поступления элементов передаются не сами вероятности, а только их порядок.

В более раннем патенте США №6650996 B1 (Garmin Ltd.) описываются системы, устройства и способы, применяющие структуру данных, в которой используются первое

поле данных, определяющее структуру декодирования данных, кодированных с помощью канонического способа Хаффмана, и второе поле данных, определяющее таблицу символов. Кроме того, таблица символов адаптируется для предоставления символа, связанного с низкочастотным индексом.

5 СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Раскрытие настоящего изобретения направлено на реализацию усовершенствованного способа кодирования данных (D1) для генерации соответствующих кодированных данных (E2), при этом в результате выполнения способа достигается более высокая эффективность сжатия кодированных данных (E2), например, в контексте интервального кодирования.

Настоящее изобретение также направлено на реализацию усовершенствованного кодера, выполняющего указанный выше усовершенствованный способ.

В соответствии с первым аспектом предлагается кодер для кодирования данных (D1) с целью генерации соответствующих кодированных данных (E2), содержащих информацию, указывающую частоту, вероятность или диапазон одного или более символов, которые должны быть представлены в кодированных данных (E2), при этом кодер способен включать в кодированные данные дополнительную информацию, указывающую, включена ли в кодированные данные (E2) информация, указывающая частоту, вероятность или диапазон для по меньшей мере одного из одного или более символов.

Настоящее изобретение, преимущество которого состоит в том, что выборочное включение в кодированные данные (E2) информации о частоте, вероятности или диапазоне выполняется как функция от появления символов в кодируемых данных (D1), способно повысить эффективность кодирования, например достижимый коэффициент сжатия, например, таким способом, который в высшей степени полезен при кодировании небольших объемов данных (D1).

Далее более подробно объясняется предшествующее выражение "включена ли информация, указывающая..."

Преимущество кодера, соответствующего первому аспекту, состоит в том, что кодер использует усовершенствованный способ, позволяющий при интервальном кодировании более эффективно доставлять значения частот, вероятностей или диапазона, связанные с кодированными данными, сгенерированными с использованием этого способа.

Кроме того, указанный выше усовершенствованный способ позволяет использовать интервальное кодирование в тех случаях, когда должно кодироваться небольшое количество символов по сравнению с количеством возможных символов. При использовании известных технологий интервального кодирования невозможно и/или непрактично кодировать такие данные из-за количества битов, требуемых для доставки значений частот, вероятностей или диапазона; а именно: в этом случае появляются непрактичные служебные данные, связанные с доставкой информации о частоте, вероятности или диапазоне из кодера в соответствующий декодер с использованием известных технологий интервального кодирования.

Кроме того, усовершенствованный способ, соответствующий первому аспекту, может использоваться с применением или без применения одного или более символов <ЕОМ>, которые служат для указания "конца сообщения". Во всех примерах вариантов осуществления настоящего изобретения, описываемых далее, один или более символов <ЕОМ> не используются, однако один или более символов <ЕОМ> могут дополнительно включаться во все примеры осуществления путем применения для них, например, значения 1 вероятности с последующим вычитанием 1 из любого другого значения

вероятности. Такое решение не требует доставки нового значения вероятности, поскольку значение вероятности уже известно. Вероятность символа <EOM> также может дополнительно вычисляться на основе количества символов данных. Если вероятность вычисляется, символ <EOM> добавляется в таблицу вероятностей и ему присваивается значение 1 частоты. Сумма частот также увеличивается на 1, и это влияет на вероятности всех символов. Для этого также требуется, чтобы дополнительное значение вероятности доставлялось для символа <EOM>.

Усовершенствованный способ, соответствующий первому аспекту, может использоваться для кодирования данных любого типа, поэтому он характеризуется широкой областью потенциального применения; к таким "данным любого типа", подлежащим кодированию, дополнительно относятся, помимо прочего, данные одного или более следующих типов: данные генетических последовательностей, данные последовательностей ДНК, данные последовательностей РНК, захваченные звуковые сигналы, захваченные видеосигналы, захваченные изображения, текстовые данные, сейсмографические данные, сигналы датчиков, данные, преобразованные из аналоговой в цифровую форму (ADC, Analog-To-Digital), данные биомедицинского сигнала, календарные данные, экономические данные, математические данные, двоичные данные. Согласно усовершенствованному способу другой способ модификации вероятностей заключается в использовании адаптивных вероятностей и диапазонов при кодировании данных для генерации соответствующих кодированных данных. Адаптивные вероятности удобно применять для кодирования больших по объему потоков данных, в особенности, если вероятности символов изменяются переменным образом, но медленно, в различных частях потока данных. Однако адаптивные вероятности менее удобны для меньших потоков данных, поскольку для адаптации вероятностей с целью согласования с корректными вероятностями символов в потоке данных требуются дополнительные вычислительные ресурсы и время.

Усовершенствованный способ, согласно первому аспекту, используется в кодере, и соответствующая инверсия усовершенствованного способа используется в соответствующем декодере, и, как указано выше, кодер и декодер совместно образуют кодек.

Дополнительно в декодере выполняется перекодирование для переформатирования декодированных в нем данных, например, с целью предоставления совместимых данных устройству визуализации декодера или устройству хранения данных декодера. Согласно усовершенствованному способу используется битовая информация в виде 0 и 1 для описания, требуется ли доставка значения диапазона, значения вероятности или значения частоты для различных символов, которые могут присутствовать среди символов потока данных. Этот усовершенствованный способ дополнительно также использует одно или более пороговых значений вероятности для определения, какие символы требуют своего собственного доставляемого значения вероятности, то есть бит 1, и для каких символов вероятность может быть пропущена, например, с использованием бита 0. Дополнительно для всех символов данных с пропускаемой вероятностью может использоваться значение 1 вероятности в кодере и в декодере. В альтернативном варианте все те символы данных, для которых не требуется свое собственное значение частоты или вероятности, то есть диапазона, дополнительно поставляются с управляющим символом (escape symbol), то есть диапазоном, в схеме интервального кодирования. Кроме того, все значения символов с пропускаемой вероятностью дополнительно доставляются, например, в виде исходных несжатых значений.

Далее более подробно разъясняются следующие концепции (i)-(iii):

(i) "кодируемый символ известен", или

(ii) в кодере: "включается ли информация, указывающая частоту, вероятность или диапазон для по меньшей мере одного символа из одного или более символов, включенных в кодированные данные", или

5 (iii) в декодере: "какие символы кодированные в кодированных данных должны декодироваться с использованием информации о частоте, вероятности или диапазоне".

Согласно вариантам осуществления, соответствующим раскрытию настоящего изобретения, не существенно, должны ли всегда быть известны сами символы, но существенно, к каким символам относятся вероятности. Таким образом, в вариантах
10 осуществления, соответствующих раскрытию настоящего изобретения, понятно, к каким символам, например к очереди данных, относится вероятность, но при этом не обязательно на основе контекста понятно, какой фактический символ используется на практике. Таким образом, информация о вероятности должна доставляться, например, в декодер, и она должна быть доступна по меньшей мере для одного символа. Однако
15 не столь важно знать, какой фактический символ используется на практике.

Дополнительно также существует вероятность того, что значение символа фактически известно, и в таком случае эта информация, безусловно, используется.

Далее описывается следующий пример. Если доставляется таблица частоты, вероятности или диапазона, то при доставке важными параметрами являются
20 максимальный индекс таблицы и информация о том, какие символы кодированы с использованием доставленных вероятностей. Максимальный индекс таблицы определяет количество различных символов или количество возможных различных символов, доступных в доставленной таблице, а именно: в предоставленных входных данных D1. Рассмотрим следующий пример:

25 4, 3, 0, 1, 0, 4, 3, 4

Фактический максимальный индекс равен 4 (а минимальный индекс равен 0), это означает, что в данных может содержаться 5 (максимальное значение - минимальное значение + 1 = 4 - 0 + 1) различных символов (0, 1, 2, 3, 4). С учетом того, что в данных D1 в действительности существуют только четыре различных символа (0, 1, 3, 4), таблица
30 также может доставляться с использованием значения '3' в качестве максимального индекса, а именно: количества доступных различных символов, вместо 4 возможных различных символов. Если значение '3' использовано в качестве максимального значения индекса таблицы, то некоторый другой механизм дополнительно применяется для доставки информации о том, какие символы используются для каждого из индексов
35 таблицы.

Эти способы более подробно разъясняются в других частях раскрытия настоящего изобретения. Если символы упорядочены, то одной из возможностей является доставка фактического максимального индекса (4) и битов доступности, (например, 11011 в данном случае). В результате такого вида доставки выполняется преобразование вида
40 "индекс 0 таблицы = символ 0", который называется индексом, и связанная пара символов равна (0, 0). Таким же образом, оставшиеся пары индексов и символов представляют собой следующую последовательность: (1, 1), (2, 3) и (3, 4). Кроме того, можно использовать пары индексов и символов непосредственно с целью определения используемых индексов для различных индексов таблицы и затем доставлять
45 максимальный индекс таблицы, равный 3, или количество различных пар индексов и символов, равное 4. При использовании такого способа индексы кодовой таблицы обычно упорядочены, и необходимо доставлять только символы. Такой способ в значительной степени предпочтительно применять, если символы в доставляемой

таблице отсортированы на основе своих частот.

Например, в этом случае пары индексов и символов могут представлять собой следующую последовательность: (0, 4), (1, 0), (2, 3) и (3, 1), и в определенном порядке доставляются только значения 4, 0, 3, 1. В некоторых ситуациях:

- 5 (а) эти используемые пары индексов и символов определены предварительно;
- (б) иногда доставляется индекс таблицы пар используемых символов и индексов, в других случаях доставляются пары индексов и символов совместно с кодированными данными;
- (с) в некоторых других случаях декодер может извлечь пары используемых символов
- 10 и индексов из известного местоположения, и
- (д) в иных ситуациях декодер может извлечь пары используемых индексов и символов из местоположения, информация о котором доставляется с кодированными данными.

Частоты частей данных, подлежащих кодированию в кодере, применяющем указанный выше усовершенствованный способ, соответствующий первому аспекту,

15 часто взаимно различны, и часто их относительная энтропия данных также взаимно различна, по этой причине при кодировании полезно разделять, то есть подразделять, данные на множество частей для генерации кодированных данных (E2).

Преимущественно для различных частей используются различные кодовые таблицы. Усовершенствованный способ позволяет, например, один большой фрагмент данных

20 более эффективно разделить на меньшие части, то есть на меньшие фрагменты данных, поскольку при этом оптимизируется доставка кодовой или таблицы частот. Такое разделение данных, подлежащих кодированию, позволяет получить значительные преимущества в том, что касается энтропии данных (D1), подлежащих кодированию, и, таким образом, при разделении возможно обеспечить более высокую степень сжатия

25 кодированных данных (E2) по отношению к соответствующим некодированным данным (D1).

Дополнительно, в соответствии с указанным первым аспектом кодер способен включать в кодированные данные (E2) дополнительную информацию о том, включена ли в кодированные данные (E2) выраженная в виде одного бита доступности

30 информация, указывающая частоту, вероятность или диапазон для по меньшей мере одного символа из одного или более символов. Кроме того, дополнительно кодер способен сообщать о включении информации, указывающей на наличие в кодированных данных (E2) информации о частоте, вероятности или диапазоне, путем использования одного битового значения "1" доступности и сообщать о том, что информация о частоте, вероятности или диапазоне в кодированных данных (E2) отсутствует, путем

35 использования одного битового значения "0" доступности. Такие битовые значения доступности также можно использовать и наоборот. Кроме того, такая информация о доступности также может доставляться в виде количества доступных индексов и затем - присутствующих индексов информации. Дополнительно доставляются фактические

40 индексы или при доставке используется дельта-кодирование. Дополнительно различные способы кодирования используются для уменьшения объема данных, требуемых для доставки информации о доступности. В выбранном способе кодирования дополнительно включается количество доступных индексов, или способ кодирования требует доставки количества индексов, или способ требует доставки битовых значений доступности для

45 всех символов, если доставляется кодовая таблица.

Дополнительно в кодере информация, указывающая частоту, вероятность или диапазон одного или более символов, представленных в кодированных данных (E2), динамически изменяется кодером как функция от характера, то есть характеристик

данных (D1), подлежащих кодированию; к таким характеристикам дополнительно относятся следующие характеристики: структура данных (D1), размер данных (D1), диапазон значений, представленных в данных (D1), размер элементов, представленных в данных (D1), метаданные, связанные с данными (D1). Кроме того, такие
5 характеристики относятся, например, к типу данных (D1), содержимому данных (D1) и/или к статистическим показателям данных (D1). Кроме того, дополнительно кодер способен подразделять эти подлежащие кодированию данные (D1) на множество частей и генерировать индивидуально для каждой части соответствующую дополнительную
10 информацию о том, включена ли в кодированные данные (E2) информация, указывающая частоту, вероятность или диапазон для по меньшей мере одного из одного или более символов.

Дополнительно кодер способен кодировать данные соответствующие по меньшей мере одному из следующих типов: информация генетических последовательностей, информация последовательностей ДНК, информация последовательностей РНК,
15 захваченные звуковые сигналы, захваченные видеосигналы, захваченные изображения, текстовые данные, сейсмографические данные, сигналы датчиков, данные, преобразованные из аналоговой в цифровую форму (ADC), данные биомедицинского сигнала, календарные данные, экономические данные, математические данные, двоичные данные.

20 В соответствии со вторым аспектом предлагается способ кодирования данных (D1) в кодере для генерации соответствующих кодированных данных (E2), содержащих информацию, указывающую частоту, вероятность или диапазон одного или более символов, представленных в кодированных данных (E2), при этом способ включает:

(а) использование кодера для включения в кодированные данные (E2) дополнительной
25 информации о том, включена ли в кодированные данные (E2) информация, указывающая частоту, вероятность или диапазон для по меньшей мере одного символа из одного или более символов.

Дополнительно способ также включает:

(b) использование кодера для включения в кодированные данные (E2) дополнительной
30 информации о том, включена ли в кодированные данные (E2) выраженная в виде одного бита доступности информация, указывающая частоту, вероятность или диапазон для по меньшей мере одного символа из одного или более символов.

Кроме того, способ дополнительно также включает:

(с) использование кодера для сообщения о включении информации, указывающей
35 на наличие в кодированных данных (E2) информации о частоте, вероятности или диапазоне, путем использования одного битового значения "1" доступности и сообщения о том, что информация о частоте, вероятности или диапазоне в кодированных данных (E2) отсутствует, путем использования одного битового значения "0" доступности. Дополнительно для указания такой информации о доступности или недоступности
40 значения "0" и "1" могут использоваться наоборот.

Дополнительно способ включает использование кодера для динамического изменения информации, указывающей частоту, вероятность или диапазон одного или более
символов, представленных в кодированных данных (E2), в зависимости от характера, то есть одной или более характеристик данных (D1), подлежащих кодированию.
45 Предшествующая ссылка в данном случае относится к определению "характера" и "характеристик" как технических терминов.

Дополнительно способ включает использование кодера для разделения подлежащих кодированию данных (D1) на множество частей и генерации индивидуально для каждой

части соответствующей дополнительной информации о том, включена ли в кодированные данные (E2) информация, указывающая частоту, вероятность или диапазон для по меньшей мере одного из одного или более символов.

5 Дополнительно способ включает кодирование данных (D1), соответствующих по меньшей мере одному из следующих типов: данные генетических последовательностей, захваченные звуковые сигналы, захваченные видеосигналы, захваченные изображения, текстовые данные, сигналы датчиков, без ограничения приведенными примерами. Данные генетических последовательностей при генерации прибором для электрофореза из биологических образцов часто занимают значительный объем, достигающий порядка 10 1 терабайта данных, определяющих гетероциклические основания нуклеиновой кислоты и/или аминокислоты. Обработка и передача таких данных представляет собой в настоящее время существенную техническую проблему.

15 В соответствии с третьим аспектом предлагается декодер для декодирования кодированных данных (E2), генерированных кодером, соответствующим первому аспекту.

В соответствии с четвертым аспектом предлагается способ декодирования кодированных данных (E2) в декодере, включающий выполнение инверсии шагов способа, соответствующего второму аспекту.

Дополнительно способ включает:

20 (a) прием сигнала доступности;

(b) прием переданных частот, вероятностей или таблицы диапазонов в кодированных данных (E2);

(c) построение полной таблицы частот, таблицы вероятностей или таблицы диапазонов на основе данных, полученных на шагах (a) и (b); и

25 (d) использование полной таблицы частот, таблицы вероятностей или таблицы диапазонов для интервального декодирования кодированных данных (E2) с целью генерации декодированных выходных данных (D3).

Дополнительно согласно способу сигнал доступности содержит по меньшей мере один бит, указывающий на доступность.

30 Дополнительно согласно способу сигнал доступности содержит по меньшей мере информацию об одном указываемом способе кодирования.

Дополнительно согласно способу сигнал доступности содержит по меньшей мере один индекс, указывающий на доступность, и/или количество индексов.

35 Согласно пятому аспекту предлагается кодек, содержащий по меньшей мере один кодер, соответствующий первому аспекту, и по меньшей мере один декодер, соответствующий третьему аспекту.

40 Согласно sixthому аспекту предлагается компьютерное программное изделие, содержащее машиночитаемый носитель информации, на котором хранятся машиночитаемые инструкции, исполняемые вычислительным устройством кодера, соответствующего первому аспекту, для реализации способа, соответствующего второму аспекту.

45 Согласно seventhому аспекту предлагается компьютерное программное изделие, содержащее машиночитаемый носитель информации, на котором хранятся машиночитаемые инструкции, исполняемые вычислительным устройством декодера, соответствующего третьему аспекту, для реализации способа, соответствующего четвертому аспекту.

Следует принимать во внимание, что признаки настоящего изобретения допускается комбинировать в различных сочетаниях без нарушения объема настоящего изобретения,

определенного прилагаемой формулой изобретения.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

Варианты осуществления, приведенные в настоящем раскрытии изобретения, описываются ниже только в качестве примеров со ссылкой на прилагаемые чертежи, на которых:

на фиг. 1 показана схема кодера, способного кодировать входные данные (D1) для генерации соответствующих кодированных выходных данных (E2), и соответствующего декодера, способного декодировать кодированные данные (E2) для генерации соответствующих декодированных данных (D3);

на фиг. 2 показан алгоритм способа кодирования данных, соответствующий раскрытию настоящего изобретения; и

на фиг. 3 показан алгоритм способа декодирования данных, соответствующий раскрытию настоящего изобретения.

На прилагаемых чертежах подчеркнутые числа используются для обозначения элементов, над которыми расположены эти числа, или элементов, рядом с которыми находятся эти числа. Если число не подчеркнуто, но рядом с ним расположена стрелка, то неподчеркнутое число используется для идентификации общего элемента, на который указывает стрелка.

ОПИСАНИЕ ВАРИАНТОВ ОСУЩЕСТВЛЕНИЯ ИЗОБРЕТЕНИЯ

В целом, при выполнении интервального кодирования данных (D1) должны быть известны вероятности символов, содержащихся в данных (D1), подлежащих кодированию, в противном случае достижимая эффективность сжатия при таком интервальном кодировании значительно снижается. Вероятность каждого символа определяет "диапазон" в процессе кодирования, причем чем больше вероятность, тем больше диапазон. При кодировании символов соответствующие диапазоны обновляются согласно вероятности кодированного символа. Если диапазон становится достаточно небольшим, интервальный кодер выделяет, например, байт и увеличивает диапазон. Таким образом, для более распространенных символов диапазоны уменьшаются медленно, и большее количество символов может быть кодировано перед выделением байта.

Если в кодере реализовано интервальное кодирование, то индекс таблицы частот, индекс таблицы вероятностей или индекс таблицы диапазонов, или значения частот, вероятностей или диапазонов передаются в кодированных данных (E2), генерированных кодером, с использованием некоторой заданной точности, если эти значения не известны в соответствующем кодере заранее; дополнительно кодер способен динамически переключаться между использованием индекса таблицы частот, индекса таблицы вероятности, индекса таблицы диапазонов, значениями частот, значениями вероятностей и значениями диапазона в зависимости от взаимно различных частей данных (D1), подлежащих кодированию. Например, если в интервальном кодеке кодируются восьмибитовые данные, то потенциально формируются 2^8 символов, то есть 256 символов. Если для передачи каждого значения частоты, вероятности или диапазона используется восемь бит, то есть один байт, то для передачи всех значений частот, вероятностей или диапазонов потребуется 256 байт. В некоторых случаях значение 256 байт потенциально превышает объем сжатых данных, полученных из интервального кодера, следовательно, существует ярко выраженная необходимость использования меньшего количества битов для доставки значений частот, вероятностей или диапазона; другими словами, при доставке из кодера в декодер значений частот, вероятностей или диапазона задействованы избыточные данные, объем которых потенциально больше

размера данных, подлежащих кодированию, что соответствует низкому уровню коэффициента сжатия.

Способы кодирования данных (D1), соответствующие раскрытию настоящего изобретения, отличаются от известных способов тем, что помимо значений частот, вероятностей или диапазона символов также передается сигнал доступности, например один бит доступности для каждого символа. Преимущественно этому биту доступности присваивается значение "один" то есть "1" для всех символов, представленных в кодированных данных (E2) и имеющих достаточно высокую вероятность, и присваивается значение "ноль" то есть "0" для всех символов, не представленных в кодированных данных (E2), а также для всех символов, представленных в кодированных данных (E2), но имеющих относительно низкую вероятность.

Согласно способам кодирования данных (D1), соответствующим раскрытию настоящего изобретения, осуществляется проверка, достаточно ли мала вероятность символа, например, путем сравнения вероятности символа с пороговым значением или, дополнительно, с несколькими пороговыми значениями. При кодировании символам, отсутствующим в кодированных данных (E2), или символам с вероятностью, меньшей порогового значения, назначается вероятность 'ноль' (используется управляющий код, и для символа диапазон не предусматривается) или 'один' (то есть, маловероятные символы доставляются с наименьшим возможным диапазоном), то есть '0' или '1' независимо от размера всего диапазона, используемого для кодирования входных данных (D1) 10, предоставляемых кодеру 20 для генерации кодированных данных (E2) 30; то же относится к ситуации, когда кодер 20 и один или более декодеров 80 применяют усовершенствованный способ (T) и обратный ему способ ("M), соответственно. Объем A (в битах) кодированных данных, требуемый для передачи частот, вероятностей или диапазонов и сигнала доступности (например, битов доступности в этом примере), составляет количество N_s символов, присутствующих или доступных в кодированных данных (E2), с достаточно высокой вероятностью, умноженное на количество N_B битов, используемое для каждого значения частоты, вероятности или диапазона, плюс один бит доступности для всех символов O_s; другими словами:

$$A = N_s \times N_B + O_s \quad \text{Формула 1}$$

Следующий пример усовершенствованного способа, соответствующего раскрытию настоящего изобретения, применяется в кодере 20 для кодирования значений набора данных, представленных в виде входных данных (D1) 10, для генерации соответствующих кодированных данных (E2) 30. В этом примере поток кодируемых данных (D1) содержит 150 значений данных, которые могут содержать 20 символов (= значений), а именно: от 0 до 19, но только 8 из этих 20 возможных значений данных, а именно: минимальное значение =2 и максимальное значение =19, фактически представлены в текущем потоке данных. Эти минимальное и максимальное значения дополнительно также доставляются в соответствии, например, со способом ODelta-кодирования, раскрытым в заявке на патент GB1303661.1, поданной Gurulogic Microsystems Oy 1-го марта 2013 года; "ODelta-кодирование" относится к способу дельта-кодирования с использованием вычислений со смещением и циклическим переходом для обеспечения в особенности эффективного кодирования данных. Этот способ позволяет сократить некоторое количество битов, требуемых для доставки сигнала доступности. Кроме того, в этом случае также уменьшается диапазон данных значений возможно доступных символов при использовании некоторого способа модификации энтропии, такого, например, как ODelta-кодирование, перед применением интервального кодирования. Основное

преимущество, связанное с доставкой минимального и максимального значений, достигается, если все способы кодирования с модификацией энтропии и интервального кодирования используют одинаковую информацию, которая доставляется из кодера в соответствующий декодер только один раз. Ниже в Табл. 1 показан пример частот и вероятностей символов. Вероятность вычисляется путем деления частоты на общее количество значений данных.

Таблица 1: Пример интервального кодирования, соответствующий раскрытию настоящего изобретения

Значение (символов в данных)	Частота	Вероятность	7-битная вероятность	7-битная вероятность (с упр. символами)	7-битная вероятность (модифицированная)	6-битная вероятность (модифицированная)
2	7	0.047	6	6	6	2
4	2	0.013	2	(0)	(1)	(1)
7	81	0.54	69	69	63	28
9	1	0.007	1	(0)	(1)	(1)
12	39	0.26	33	33	30	13
13	9	0.06	8	8	7	3
14	3	0.02	2	2	2	(1)
19	8	0.053	7	7	6	3
Упр. символ	0	0.0	(0)	3	x 12	x 12
Сумма	150	1.000	128	128	114+14=128	49+15=64

В Табл. 1 существует пара символов с очень низкой вероятностью появления. В случае использования 7 бит для доставки вероятностей, если бы применялось одинаковое количество битов для описания каждой вероятности, то потребовалось бы $20 \cdot 7 = 140$ бит для представления всех вероятностей. Если наличие символов в потоке, то есть в кодированных данных (E2) 30, выражается путем передачи, например, одного бита доступности для символа, то требуется еще $8 \cdot 7 + 20 \cdot 1 = 76$ бит для представления всех вероятностей.

Этот начальный вариант осуществления для реализации усовершенствованного способа, соответствующего раскрытию настоящего изобретения, уже отличается от известных решений вследствие использования отдельных битов доступности для уменьшения количества доставляемых вероятностей с 20 до 8. С учетом других, описываемых ниже вариантов раскрытия настоящего изобретения дополнительно полезно устанавливать пороговое значение, которое определяет, передается ли в декодер заданная вероятность, например, пороговое значение дополнительно устанавливается равным 0,018 при 7-битной вероятности. Модификация порогового значения изменяет количество битов в кодированных данных (E2) 30, требуемых для выражения и доставки вероятностей в один или более декодеров 80. Если в качестве порогового значения для данных, описанных в Табл. 1, выбирается 0,10, то остается только два символа, и $2 \cdot x + 20 \cdot 1$ бит требуется для представления вероятностей в кодированных данных (E2) 30, где x - количество битов, используемых для представления вероятности.

Нахождение оптимального порогового значения, используемого при реализации усовершенствованного способа, соответствующего раскрытию настоящего изобретения,

зависит от объема кодированных данных (E2) 30, подлежащих передаче из кодера 20 в один или более декодеров 80, и от частот символов, представленных в кодированных данных (E2) 30. Далее более подробно разъясняется способ, с помощью которого определяется указанное выше пороговое значение. Например, пусть

5 T = пороговое значение, и

B = количество битов, используемое для представления вероятности, например, 6 или 7 битов, как в примере, приведенном в Табл. 1. Пороговое значение преимущественно располагается в заданных рекомендуемых пределах и должно быть не меньше, чем нижнее пороговое значение T_{Low} , как определяется формулой 2:

$$10 \quad T_{Low} = \frac{1}{2^B}$$

$$\frac{2}{2^B} \leq T_{Recommended} \leq \frac{8}{2^B}$$

Формула 2

15 В последних четырех столбцах Табл. 1 приведены фактические вероятности, передаваемые в один или более декодеров 80. В столбце "7-битная вероятность", приведена вероятность, передаваемая, если 7 бит используется для всех 8 или 20 значений данных с использованием или без использования сигнала доступности, при этом под сигналом доступности понимается сигнал, выражающий наличие символов в потоке. Столбец "7-битная вероятность (с упр. символами)" соответствует вероятности, передаваемой с использованием первого варианта осуществления усовершенствованного способа, соответствующего раскрытию настоящего изобретения, столбец "7(6)-битная вероятность (модифицированная)" соответствует вероятности, передаваемой с использованием второго варианта осуществления усовершенствованного способа, соответствующего раскрытию настоящего изобретения.

25 Вероятности, указанные в скобках в трех последних столбцах Табл. 1, фактически не передаются в один или более декодеров 80, поскольку они выражены с использованием нулевого бита доступности. В этом примере вероятности, передаваемые в один или более декодеров 80, близки к вероятностям, передаваемым с помощью известных способов кодирования, поэтому в данном случае использование усовершенствованного способа, соответствующего раскрытию настоящего изобретения, не влияет в значительной степени на эффективность кодирования фактического диапазона, однако может значительно повысить общую эффективность сжатия путем уменьшения количества битов, требуемых для передачи значений частот, вероятностей или диапазона. Кроме того, следует принимать во внимание, что сумма вероятностей, передаваемая известным способом или с использованием управляющих символов, составляет 128, в то время как в усовершенствованном способе, соответствующем второму варианту осуществления, сумма вероятностей передается, как показано в таблице, то есть составляет либо 114, либо 49. Это возникает из-за резервирования значения 'один' вероятности для каждого значения данных, выраженного нулевым битом доступности, при этом существует 14 или 15 этих значений данных, и 12 значений данных не представлены в потоке, а 2 или 3 значения имеют вероятность ниже порогового значения.

45 В этом примере, в котором используются 7 бит, а также пороговое значение, символы 4 и 9 выражены с помощью нулевого бита доступности совместно со всеми 12 символами, не представленными в потоке данных. Изначально диапазон, называемый управляющим символом и зарезервированный для отсутствующих символов в потоке данных, отсутствует. Теперь необходимо реализовать механизм доставки тех символов

4 и 9, которым не присвоена собственная вероятность. Это задача может эффективно решаться с использованием двух различных вариантов осуществления, описываемых ниже.

Первый вариант осуществления, показанный в приведенной выше Табл. 1 и отмеченный строкой "(с упр. символами)", состоит в создании отдельного управляющего символа и связанного с ним диапазона. Если собственный диапазон для символа отсутствует, то доставка осуществляется посредством управляющего символа (диапазона). Кроме того, значение символа также доставляется в его собственном потоке данных. В этом примере значение символа может доставляться с использованием 5 бит, поскольку диапазон значений символов составляет от 0 до 19. Дополнительно, если отношение доступных символов ко всем возможным символам велико, то есть больше 0,25, то значение символа преимущественно доставляется с использованием немного меньшего количества битов путем доставки значения символа с помощью его соответствующей 0-битной позиции. Например, в рассматриваемом случае значение 4 символа соответствует нулевому биту 3 доступности, а значение 9 символа соответствует нулевому биту 7 доступности битового потока "00100001000011100001". Таким образом, в потоке существует только 14 нулевых битов доступности, и только 4 бита требуется для их представления в кодированных данных (E2). Дополнительно устраняется доставка отдельного потока значений символов, то есть вероятности интервального кодирования переключаются от обычных вероятностей к равным вероятностям всех символов или к равным вероятностям всех 0-битных символов доступности, что возможно после доставки управляющего символа (диапазона) для доставки фактического значения данных, а затем, после доставки фактического значения символа, восстанавливаются обычные вероятности.

Согласно второму варианту осуществления определяется, что всем 0-битным символам доступности выделяется вероятность 1, а оставшийся диапазон резервируется для 1-битных символов доступности из полного диапазона вероятностей. Далее описываются такие решения, показанные в Табл. 1 с пометкой "(модифицированная)". Теперь в описанном выше примере существует 12+2 управляющих символов в 7 битах, и они распределены 14 значениям из 128 доступных значений. Оставшиеся 114 значений используются для тех 6 символов, которые распределяются с использованием 1-битной доступности. Подобный пример также представлен в последнем столбце Табл. 1 при использовании 6 бит для интервальных вероятностей, при этом в данном примере выбирается пороговое значение, равное 0,4. Как указывалось выше, в этот момент эти 0-битные символы доступности в битовом потоке можно рассматривать в качестве управляющих символов, и им всем выделяется собственное значение '1' вероятностей, как показано в скобках в Табл. 1.

Далее описывается, какое количество битов требуется для доставки значений вероятностей. В комбинированном решении (с использованием управляющих символов) используется $7*7+20*1=69$ бит. В этом примере данное решение не является предпочтительным, поскольку оно позволяет сэкономить только семь (76-69) бит при доставке таблицы вероятностей по сравнению с первым вариантом осуществления, соответствующим раскрытию настоящего изобретения, но при этом используется $3*4=12$ бит для доставки дополнительных значений управляющих символов. Таким образом, более высокая вероятность одного или более управляющих символов по сравнению с исходными символами не достаточна для компенсации возрастания объема данных. В этот момент времени вероятность других символов одинакова, и преимущества путем доставки этих символов также не достигаются.

Этот усовершенствованный способ также полезен в другом случае, например, если множество символов являются управляющими, но их вероятность не очень велика. Иногда также возможно использовать меньшее количество битов для управляющих символов по сравнению с исходными символами, поскольку наименьшие вероятности доставляются в любом случае совместно с управляющим кодом. Во втором варианте осуществления с модифицированными вероятностями используется $6*7+20*1=62$ бита (7-битная вероятность) или $5*6+20*1=50$ битов (6-битная вероятность). Дополнительно доставка вероятностей в еще большей степени оптимизируется путем сообщения или доставки информации, определяющей вероятность наиболее часто встречающегося символа в кодированных данных (E2) 30.

В данном случае, если значение наибольшей вероятности меньше 64 и меньше 32, то значения вероятностей дополнительно доставляются с помощью 6 бит (и 5 бит) в этих примерах, поэтому для доставки общей таблицы вероятностей требуется только $6*6+20*1=56$ бит (для примера с 7 битами) и $5*5+20*1=45$ бит (для примера с 6 битами). Кроме того, при доставке этот поток битов доступности, объемом $20*1$, дополнительно сжимается; помимо сжатия битов доступности также возможно реализовать варианты осуществления, соответствующие раскрытию настоящего изобретения, путем сжатия частотных таблиц или значений вероятностей; такое сжатие дополнительно реализуется с использованием одного или более способов сжатия. Это особенно важно, если существует множество возможных доступных символов и различные вероятности 0-битной и 1-битной доступности, то есть множество нулей и множество единиц, или все они расположены в различных частях битового потока в кодированных данных (E2) 30. Дополнительно, если диапазон доступных символов отдельно ограничен, например, путем доставки минимального и/или максимального значений, первый и/или последний символы, о которых известно, что они доступны в потоке, передавать не требуется с использованием сигнала доступности, например, биты доступности для этого или этих значений доставлять не требуется. Требуется только доставлять сигнал доступности, например биты доступности, для диапазона от минимального значения +1 до максимального значения -1. В примере, приведенном в Табл. 1, экономится 31 (76-45) бит по сравнению со случаем, когда наличие символов выражено, то есть по сравнению с первым усовершенствованным способом. Экономия составляет 95 (140-45) бит по сравнению с известными способами. Такая экономия сохраняется и при меньшей точности вероятностей символов, и поэтому данные усовершенствованные способы представляют собой несомненно преимущественные способы доставки данных в этом примере.

Дополнительно информация о вероятностях доставляется в кодированных данных (E2) 30 даже с менее точными значениями вероятностей или с использованием более высокого порогового значения, благодаря чему достигается даже большая экономия размера данных при доставке таблицы вероятностей с помощью кодированных данных (E2) 30.

Экономия при доставке таблицы вероятностей обычно приводит к некоторым потерям в процессе энтропийного кодирования. Такие потери необходимо принимать во внимание при выборе наилучшего возможного способа доставки таблицы для реализации вариантов осуществления, соответствующих раскрытию настоящего изобретения.

Усовершенствованные способы согласно раскрытию настоящего изобретения также в некоторых случаях дополнительно используются для исходных данных для экономии размера данных, например, при сохранении в базах данных, на носителях данных и

т.п. Следует принимать во внимание, что в приведенных выше примерах в исходных данных содержится только 150 символов и 8 различных значений, поэтому исходная информация, объемом $150 \cdot 5 = 750$ бит, может также доставляться с использованием $150 \cdot 3 + 20 \cdot 1 = 470$ бит с помощью информации о битах доступности непосредственно для исходных данных. С помощью интервального кодирования с доставкой таблиц и энтропийным кодированием можно получить меньший размер данных по сравнению с исходными данными. Следует принимать во внимание, что перед интервальным кодированием могут использоваться различные способы модификации энтропии данных. Таким же образом, после интервального декодирования требуется выполнить обратные операции. К таким способам модификации энтропии относятся, например, дельта-кодирование, ODelta-кодирование, то есть способ, раскрытый в заявке на патент GB1303661.1, поданной Gurulogic Microsystems Oy 1-го марта 2013 года (включенной в настоящую заявку посредством ссылки, при этом ODelta-кодирование представляет собой дельта-кодирование с использованием вычислений со смещением и циклическим переходом), модификатор энтропии (EM), способ, раскрытый в заявке на патент GB1303658.7, поданной Gurulogic Microsystems Oy 1 марта 2013 года (включенной в настоящую заявку посредством ссылки), кодирование с переменной длиной строки (RLE, Run Length Encoding) и отдельное кодирование с переменной длиной строки (SRLE, Split Run Length Encoding), способ, раскрытый в заявке на патент GB 1303660.3, поданной Gurulogic Microsystems Oy 1 марта 2013 года (включенной в настоящую заявку посредством ссылки).

Как описывалось ранее, усовершенствованные способы, соответствующие раскрытию настоящего изобретения, преимущественно не используют один или более символов <EOM> при реализации интервального кодирования, поскольку вместо этого может доставляться объем данных, подлежащих кодированию (= объем данных, подлежащих декодированию), или объем кодированных данных (которые требуется декодировать). На основе этого объем данных, подлежащих кодированию, или объем кодированных данных, присутствующих в кодированных данных (E2) 30 после интервального кодирования, может использоваться при декодировании данных (D3) 90 с помощью интервального декодирования. Кроме того, согласно раскрытию настоящего изобретения не требуется доставлять дополнительный символ <EOM> после данных во время интервального кодирования для генерации кодированных данных (E2) 30.

Дополнительно, если доставляются значения вероятностей, например, с помощью кодированных данных (E2) 30 или другим способом, то не требуется доставлять последнее значение вероятности, если известна сумма значений вероятностей, например, эта сумма может предварительно устанавливаться, или она уже может быть доставлена в один или более декодеров 80. Последнее значение вероятности для символа, управляющего символом, <EOM> или действительного символа, присутствующего в данных, дополнительно рассчитывается путем вычитания всех ранее доставленных значений из суммы вероятностей. В любом случае требуется знать сумму, для того чтобы корректно выполнить декодирование значений вероятностей в одном или более декодерах 80. Некоторые биты объема данных, таким образом, могут заранее дополнительно экономиться, поскольку последнее значение вероятности не доставляется.

Хотя примеры раскрытия настоящего изобретения приводятся в отношении интервального кодирования, они также, с соответствующими изменениями, допускают использование арифметического кодирования и двоичного арифметического кодирования, реализуемых в кодере 20. Если такое арифметическое кодирование и двоичное арифметическое кодирование реализовано в кодере 20, управляющий код не

используется в процессе фактического кодирования, вместо этого он используется для определения способа передачи диапазона вероятностей для аналогичных последующих элементов. Однако такой способ менее эффективен, чем описанные выше усовершенствованные способы, соответствующие раскрытию настоящего изобретения, согласно которым используется сигнал доступности и значение вероятности для 5 передачи информации о том, что заданный символ существует, и индикация связанной вероятности заданного символа; в данном случае можно сослаться на известный факт: "Коэффициент сжатия обычно меньше, чем при статическом кодировании (<http://compressions.sourceforge.net/Arithmetic.html#static>). Хотя для коротких входных 10 последовательностей преимущества, получаемые в отсутствие передачи статистических данных, могут быть более существенными". Таким образом, усовершенствованные способы, соответствующие раскрытию настоящего изобретения, позволяют получить значительное техническое преимущество при кодировании данных.

Усовершенствованные способы кодирования, соответствующие раскрытию 15 настоящего изобретения, подходят для использования в устройствах обработки и генерации данных, таких как персональные компьютеры, портативные компьютеры с поддержкой беспроводной связи, смартфоны, планшеты, интеллектуальные терминалы, мобильные (сотовые) телефоны, аудиовизуальные устройства, цифровые камеры, видеокamеры, научное оборудование, устройства генетических последовательностей, 20 телевизионные устройства, устройства наблюдения, а также другие подобные устройства. Усовершенствованные способы дополнительно реализуются в цифровой аппаратуре, например в специализированных интегральных схемах (ASIC, Application Specific Integrated Circuit). Альтернативно или дополнительно усовершенствованные способы 25 допускают реализацию с использованием одного или более программных изделий, записанных на машиночитаемом носителе информации и выполняемых вычислительной аппаратурой. Дополнительно одно или более программных изделий загружаются по сети связи в виде "программных приложений" например, в портативные устройства связи, поддерживающие беспроводную связь, например в мобильные телефоны. Кодер 20 и соответствующий декодер 80 допускают реализацию в виде кодека 95, например, 30 для использования в портативном устройстве захвата сигнала, таком как цифровые камеры, мультимедийное устройство с возможностью доступа в Интернет, игровое устройство, камеры наблюдения, устройство контроля за состоянием окружающей среды и т.п.

На фиг. 2 показаны шаги способа кодирования данных (D1) 10 в кодере 20 для 35 генерации соответствующих кодированных данных (E2) 30. Дополнительно шаги выполняются в показанной последовательности, хотя определенные шаги являются необязательными, и порядок их исполнения может изменяться.

40

45

Таблица 2: Шаги способа кодирования данных

Шаг	Операция	Подробности	
5	100	Прием данных (D1) 10 и анализ характеристик этих данных для определения информации, описывающей вероятности появления символов в этих данных	Обязательный шаг
10	110	Определение, разделены ли данные (D1) 10 на множество частей, каждая из которых кодируется отдельно со связанной вероятностью появления в ней символов	Необязательный шаг
15	120	Генерация данных, указывающих присутствует ли заданный символ в кодированных данных (E2) 30 путем информации, описывающей вероятности появления символов с помощью сигнала доступности, например, выраженного в виде однобитовой информации о доступности, например в виде "1" или "0", или способа кодирования, или количества индексов и их значений.	Обязательный шаг
20	130	Дополнительно применение одного или более пороговых значений для определения, передана ли информация о вероятностях, относящаяся к одному или более символам, присутствующим в данных (D1) 10	Необязательный шаг
25	140	Кодирование символов данных (D1) 10 совместно с информацией, описывающей вероятности появления символов и определенной на шаге 100, а также с идентификационными данными, полученными на шаге 120, для генерации кодированных данных (E2) 30	Обязательный шаг
30			
35			

В одном или более декодерах 80 по существу выполняются шаги, обратные тем, что приведены в Табл. 2, то есть извлекается информация о частоте, вероятности или диапазоне из кодированных данных (E2) 30, полученных на шаге 100, и данные, полученные на шаге 120, применяются для определения, какие символы в кодированных данных (E2) 30 должны декодироваться с использованием информации о частоте, вероятности или диапазоне, и какие символы не должны декодироваться таким способом, при этом результаты таких вычислений объединяются для предоставления декодированных данных (D3) 90. Дополнительно один или более декодеров 80 способны выполнять перекодирование таким образом, чтобы данные D1 и D3 взаимно различались. Такое перекодирование, например, полезно, если один или более декодеров 80 взаимно различны и данные D3 должны визуализироваться на различном презентационном оборудовании, для которого требуется перекодирование с целью настройки, например, изменения разрешения изображения в пикселях и формата

изображения.

На фиг. 3, показаны шаги 200-240 способа декодирования кодированных данных (E2) 30 для генерации декодированных данных (D3) 90; при этом обычно декодированные данные (D3) 90 представляют восстановленную версию данных (D1) 10, хотя способ не ограничен таким вариантом. В процессе функционирования на шаге 200 декодер 80 принимает сигнал доступности, то есть биты доступности, или информацию о способе кодирования, или количество индексов и непосредственно значения индексов. Кроме того, на шаге 210 декодер 80 принимает переданные таблицы частот, вероятностей или диапазонов в данных 60 (которые включены в данные E2, 30), из которых декодер 80 формирует полные таблицы частот, вероятностей или диапазонов. Для битового диапазона "0" преимущественно устанавливается значение "0" или "1", при необходимости, а для битового диапазона "1" устанавливается принятое значение частоты, вероятности или диапазона. Если принята частотная таблица, она преобразуется в таблицу вероятностей, для того чтобы ее можно было использовать при интервальном кодировании. Следует принимать во внимание, что кодеру 20 также всегда требуется преобразовывать частотную таблицу в таблицу вероятностей перед тем, как интервальное кодирование может быть эффективно выполнено. После формирования на шагах 220-240 таблицы вероятностей в декодере 80, в которой общая сумма значений равна степени двух, затем возможно выполнить интервальное декодирование принятых преобразованных данных 50 (которые вставляются в данные E2, 30), а затем сгенерировать декодированные выходные данные (D3) 90. Шаги 200-240 по существу соответствуют инверсии шагов 100-140, показанных на фиг. 2.

Кодер 20 и декодер 80 могут использоваться широким спектром оборудования обработки данных, такого, например, как беспроводные устройства связи, звуковое устройство, аудиовизуальное устройство, вычислительное устройство, устройство видеоконференций, устройство наблюдения, устройство научных измерений, устройство анализа генетических последовательностей, аппаратная и/или программная система связи, аппаратное и/или программное обеспечения сети связи и т.п.

Различные изменения в варианты осуществления настоящего изобретения, описанные выше, могут вноситься при условии сохранения объема настоящего изобретения, определенного прилагаемой формулой изобретения. Такие термины, как "включающий", "содержащий", "встроенный", "состоящий из", "имеющий", "представляет собой", используемые в описании и формуле настоящего изобретения, не должны толковаться как исключаящие использование блоков, компонентов или элементов, явно не описанных выше. Ссылка на компонент, указанный в единственном числе, также должна допускать толкование, связанное со множеством компонентов. Числовые значения, заключенные в скобки в прилагающейся формуле изобретения, помогают разобраться в пунктах формулы изобретения и не должны трактоваться в качестве ограничения предмета изобретения, заявленного посредством этой формулы.

(57) Формула изобретения

1. Кодер (20) для кодирования данных (D1, 10) с получением соответствующих кодированных данных (E2, 30), содержащих информацию о частотах, вероятностях или значениях диапазонов различных символов, которые должны быть представлены в кодированных данных (E2, 30), при этом упомянутая информация указывает символы, к которым относятся упомянутые частоты, вероятности или значения диапазонов, при этом кодер (20) способен включать в кодированные данные (E2, 30) дополнительную информацию, указывающую, включена ли в кодированные данные (E2, 30) информация

о частотах, вероятностях или значениях диапазонов для упомянутых различных символов.

2. Кодер (20) по п. 1, отличающийся тем, что он способен включать в кодированные данные (E2, 30) дополнительную информацию о том, включена ли в кодированные данные (E2, 30) информация о частотах, вероятностях или значениях диапазонов для символов, в виде однобитовой информации о доступности.

3. Кодер (20) по п. 2, отличающийся тем, что он способен сообщать о включении информации о наличии в кодированных данных (E2, 30) информации о частотах, вероятностях или значениях диапазонов для символов путем использования одного битового значения "1" доступности и сообщать о том, что информация о частотах, вероятностях или значениях диапазонов для символов в кодированных данных (E2, 30) отсутствует, путем использования одного битового значения "0" доступности.

4. Кодер (20) по п. 1, отличающийся тем, что информация о частотах, вероятностях или значениях диапазонов символов, представленных в кодированных данных (E2, 30), динамически изменяется кодером (20) как функция от одной или более характеристик данных (D1, 10), подлежащих кодированию.

5. Кодер (20) по п. 4, отличающийся тем, что к одной или более характеристик относится по меньшей мере одна из следующих характеристик: тип данных (D1, 10), содержимое данных (D1, 10) и/или статистические показатели данных (D1, 10).

6. Кодер (20) по п. 4 или 5, отличающийся тем, что он способен разделять подлежащие кодированию данные (D1, 10) на множество частей и генерировать индивидуально для каждой части соответствующую дополнительную информацию о том, включена ли в кодированные данные (E2, 30) информация о частотах, вероятностях или значениях диапазонов для символов.

7. Кодер (20) по любому из пп. 1-5, отличающийся тем, что он способен анализировать данные (D1, 10) для определения появляющихся в них минимального и максимального значений, и информация, описывающая минимальное и максимальное значения, включается кодером (20) в кодированные данные (E2, 30), при этом минимальное и максимальное значения ограничивают диапазон возможных значений доступных символов при использовании интервального кодирования.

8. Кодер (20) по п. 7, отличающийся тем, что он способен представлять в кодированных данных первый символ и/или последний символ, присутствующий в данных (D1, 10), без использования сигнала доступности.

9. Кодер (20) по любому из пп. 1-5, отличающийся тем, что он способен модифицировать энтропию данных (D1, 10), подлежащих кодированию, перед применением к ним интервального кодирования для генерации кодированных данных (E2, 30).

10. Кодер (20) по любому из пп. 1-5, отличающийся тем, что он способен кодировать данные (D1, 10), соответствующие по меньшей мере одному из следующих типов: захваченные звуковые сигналы, захваченные видеосигналы, захваченные изображения, текстовые данные, сейсмографические данные, сигналы датчиков, данные, преобразованные из аналоговой в цифровую форму, данные биомедицинского сигнала, календарные данные, экономические данные, математические данные, двоичные данные.

11. Способ кодирования данных (D1, 10) в кодере (20) для генерации соответствующих кодированных данных (E2, 30), содержащих информацию о частотах, вероятностях или значениях диапазонов различных символов, представленных в кодированных данных (E2), при этом упомянутая информация указывает символы, к которым относятся упомянутые частоты, вероятности или значения диапазонов, и способ включает:

прием данных (D1, 10) и анализ их характеристик для определения информации о частотах, вероятностях или значениях диапазонов символов в этих данных;

генерирование дополнительной информации, указывающей, включена ли в кодированные данные (E2, 30) информация о частотах, вероятностях или значениях диапазонов для упомянутых символов; и

использование кодера (20) для включения в кодированные данные (E2, 30) дополнительной информации о том, включена ли в кодированные данные (E2, 30) информация о частотах, вероятностях или значениях диапазонов для упомянутых различных символов.

12. Способ по п. 11, отличающийся тем, что он также включает:

использование кодера (20) для включения в кодированные данные (E2, 30) дополнительной информации о том, включена ли в кодированные данные (E2, 30) информация о частотах, вероятностях или значениях диапазонов для символов, в виде однобитовой информации о доступности.

13. Способ по п. 12, отличающийся тем, что он также включает:

использование кодера (20) для сообщения о включении в кодированные данные (E2, 30) информации о частотах, вероятностях или значениях диапазонов для символов путем использования одного битового значения "1" доступности и для сообщения об отсутствии информации о частотах, вероятностях или значениях диапазонов для символов в кодированных данных (E2, 30) путем использования одного битового значения "0" доступности.

14. Способ по пп. 11, 12 или 13, отличающийся тем, что он включает использование кодера (20) для динамического изменения информации о частотах, вероятностях или значениях диапазонов символов, представленных в кодированных данных (E2, 30), в зависимости от характера данных (D1), подлежащих кодированию.

15. Способ по п. 14, отличающийся тем, что он включает использование кодера (20) для разделения подлежащих кодированию данных (D1, 10) на множество частей и генерации индивидуально для каждой части соответствующей дополнительной информации о том, включена ли в кодированные данные (E2, 30) информация о частотах, вероятностях или значениях диапазонов для упомянутых символов.

16. Способ по любому из пп. 11-13, отличающийся тем, что он включает кодирование данных (D1, 10), соответствующих по меньшей мере одному из следующих типов: захваченные звуковые сигналы, захваченные видеосигналы, захваченные изображения, текстовые данные, сейсмографические данные, сигналы датчиков, данные, преобразованные из аналоговой в цифровую форму, данные биомедицинского сигнала, календарные данные, экономические данные, математические данные, двоичные данные.

17. Декодер (80) для декодирования кодированных данных (E2, 30), генерированных кодером (20), с получением декодированных данных (D3, 90), отличающийся тем, что он содержит оборудование обработки данных, которое способно:

a) принимать сигнал доступности;

b) принимать переданные частоты, вероятности или значения диапазонов в кодированных данных (E2, 30);

c) формировать полную таблицу частот, таблицу вероятностей или таблицу диапазонов из данных, полученных на шагах (a) и (b); и

d) использовать полную таблицу частот, таблицу вероятностей или таблицу диапазонов для интервального декодирования кодированных данных (E2, 30) с получением декодированных выходных данных (D3, 90).

18. Способ декодирования кодированных данных (E2, 30) в декодере (80),

отличающийся тем, что он включает:

(a) прием сигнала доступности;

(b) прием переданных частот, вероятностей или значений диапазонов в кодированных данных (E2, 30);

5 (c) построение полной таблицы частот, таблицы вероятностей или таблицы диапазонов из данных, полученных на шагах (a) и (b); и

(d) использование полной таблицы частот, таблицы вероятностей или таблицы диапазонов для интервального декодирования кодированных данных (E2, 30) с получением декодированных выходных данных (D3, 90).

10 19. Кодек (20, 80; 95), содержащий по меньшей мере один кодер (20) по любому из пп. 1-10 и декодер (80) по п. 17.

20. Машиночитаемый носитель информации, на котором хранятся машиночитаемые инструкции, исполняемые вычислительным устройством кодера (20) по любому из пп. 1-10 для реализации способа по любому из пп. 11-16.

15 21. Машиночитаемый носитель информации, на котором хранятся машиночитаемые инструкции, исполняемые вычислительным устройством декодера (80) по п. 17 для реализации способа по п. 18.

20

25

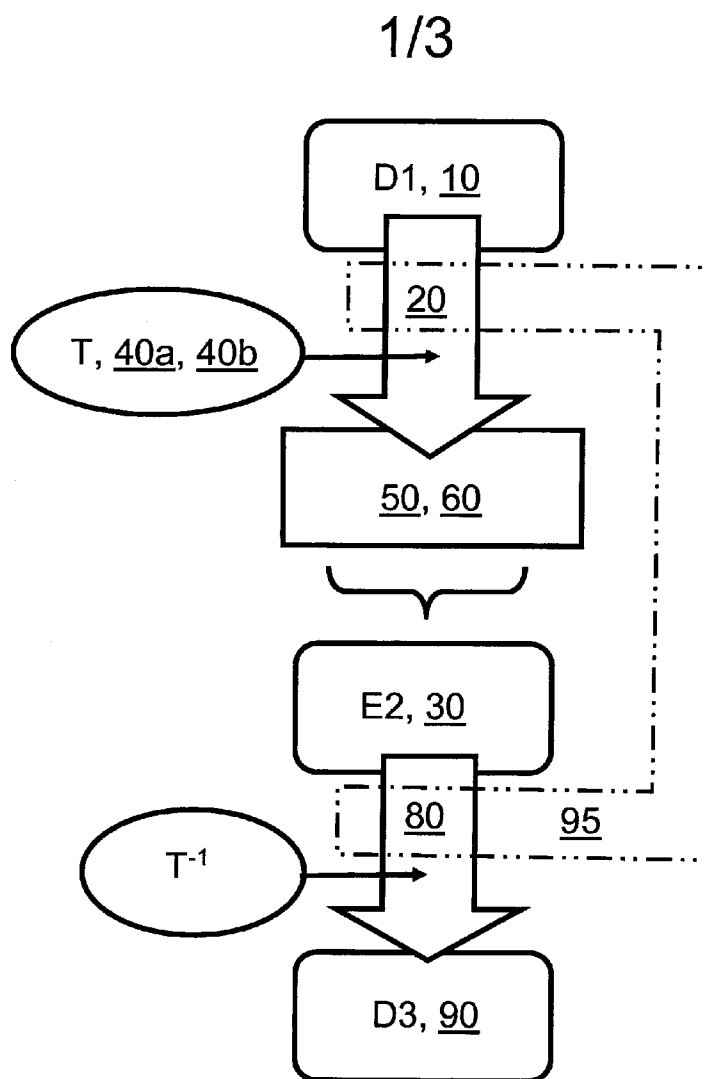
30

35

40

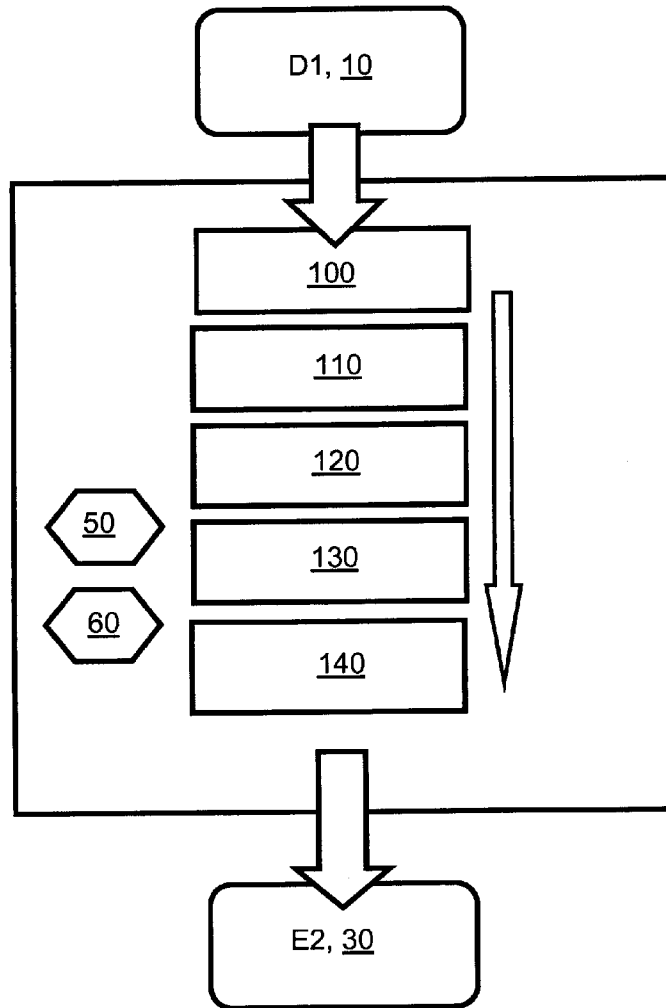
45

Устройство и способ для кодирования и декодирования исходных данных



Фиг. 1

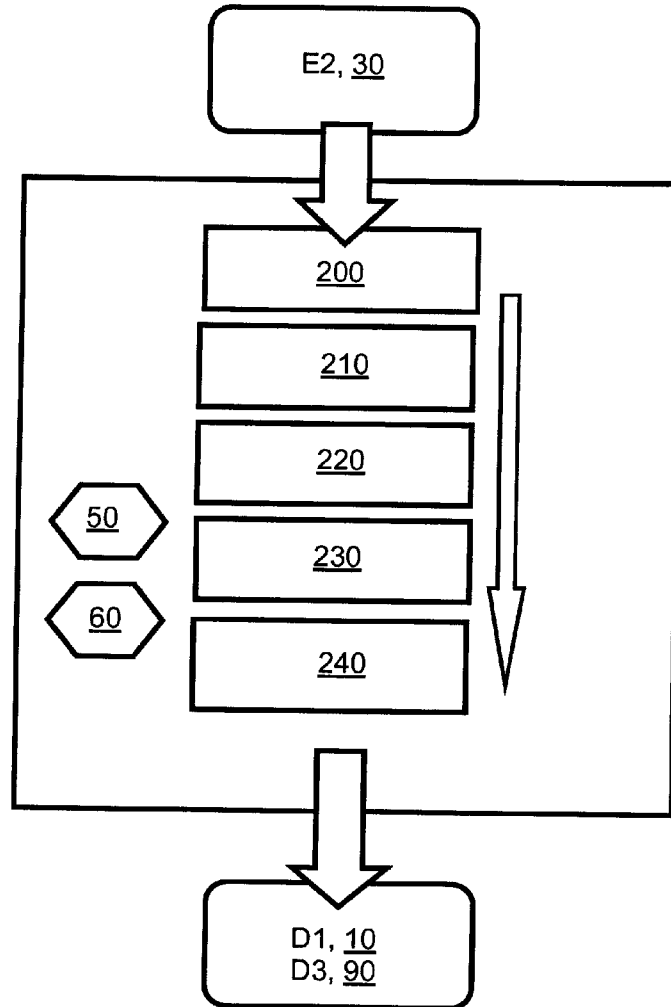
2/3



Фиг.2

Устройство и способ для кодирования и декодирования исходных данных

3/3



Фиг.3