

Сверхнадёжное исправление ошибок на основе МПД алгоритмов для баз данных систем ДЗЗ

В.В. Золотарёв, Р.Р. Назиров

*Институт космических исследований РАН
117997 Москва, ул. Профсоюзная, 84/32
E-mail: zolutasd@yandex.ru*

Рассмотрены алгоритмы мажоритарного типа для декодирования символьных данных с улучшенными на много десятичных порядков характеристиками достоверности. Они относятся к классу многопороговых декодеров (МПД) с минимальной теоретически возможной линейной сложностью реализации. Эти методы обладают свойством сходимости к решению оптимального переборного декодера даже при весьма высоком уровне шума канала. Показано, что во многих случаях корректирующие возможности QМПД вообще недоступны для традиционных алгоритмов декодирования кодов Рида-Соломона сколько угодно большой длины. Предлагается использовать эти алгоритмы в особо надёжных сверхбольших базах данных, в том числе и для систем ДЗЗ.

Введение

В системах дистанционного зондирования Земли (ДЗЗ) важной и особенно сложной является задача быстрой записи и высоконадёжного хранения принятой дискретной информации в базах данных (БД) очень большого объёма. Эта задача наиболее эффективно решается методами помехоустойчивого кодирования [1-4].

Применение специальных многопороговых алгоритмов для недвоичных кодов в БД для ДЗЗ позволяет существенно, иногда на много десятичных порядков повысить достоверность и целостность хранимых, обрабатываемых и восстанавливаемых цифровых данных. Эффективное кодирование создаёт условия для существенно более полной реализации возможностей цифровых сетей и специализированных систем, например, ДЗЗ. Ниже рассмотрены характеристики недвоичных МПД при обработке символьной, байтовой и двухбайтовой информации, закодированной мажоритарным кодом. Выполнено сравнение его возможностей с характеристиками единственно используемых до сих пор недвоичных кодов Рида-Соломона.

Основные свойства QМПД

Результаты работы многопороговых декодеров (МПД) [1, 3, 4, 6, 7] в двоичных гауссовских каналах во многих случаях оказываются совпадающими с характеристиками оптимального декодирования или близкими к ним даже при высоком уровне шума космических, спутниковых и ряда других каналов.

Обобщение многопорогового декодирования (МПД) на недвоичные симметричные каналы рассматривалось в [1, 2, 4, 7]. В [2] отмечалось, что оптимальные декодеры для большого основания кода q , $q > 16$ создать трудно, а помехоустойчивость традиционных декодеров для кодов Рида-Соломона (РС), просто в силу их недостаточно большой длины, мала. Именно поэтому применение длинных недвоичных кодов с возможно меньшей сложностью декодирования представляется крайне желательной. Именно такими свойствами и обладают недвоичные (называемые также q -ичными) коды с мажоритарным декодированием, методом, который далее обозначается как QМПД.

В [1, 5] доказано, что решения QМПД сходятся к решениям оптимальных переборных алгоритмов в том смысле, что в каналах с симметричными случайными ошибками все изменения в декодируемых символах приводят к переходу к более правдоподобным решениям. Проблема состоит только в том, чтобы при возможно более высоком уровне шума были условия для продолжения поиска всё более правдоподобных кодовых слов до тех пор, пока не будет достигнуто наиболее правдоподобное решение, которое и является, как известно, решениями ОД, оптимальным решением.

Как известно, обычно решения ОД являются результатом полного перебора. В то же время QМПД, обеспечивая достижение решения ОД при достаточно высоком уровне шума, остаётся алгоритмом с минимальной теоретически возможной линейной от длины кода сложностью, т.е. числом операций декодера. Именно поэтому применение недвоичных МПД особенно перспективно, поскольку областей применения кодов РС очень много, их эффективность весьма невелика, а потребность в значительном улучшении характеристик помехоустойчивости во всех подобных технических системах огромна. Поэтому рассмотрим сравнительные характеристики QМПД и обычных декодеров РС.

Напомним, что собственно QМПД декодеры детально описаны в [1, 2, 4, 5]. Отметим наиболее существенный момент, характеризующий предложенный новый алгоритм. Как и в случае двоичных кодов, нельзя утверждать, что улучшение решения при многократных попытках декодирования имеет место до тех пор, пока не будет достигнуто решение ОД. На самом деле и в блоковых, и в сверточных кодах возможны конфигурации ошибок, не исправляемые в QМПД, но которые могут быть исправлены в ОД. Поэтому основной способ повышения эффективности QМПД состоит в поиске кодов, в которых такие неисправляемые конфигурации ошибок довольно редки даже при большом уровне шума. Это задача решается при выборе кодов для QМПД по минимуму уровня проявления эффекта размножения ошибок [1, 2, 4].

Отметим ещё два важных свойства QМПД, которые оказываются полезными при рассмотрении характеристик этого алгоритма. Первым из них состоит в том, что, как и в случае двоичных цифровых потоков в гауссовских каналах, QМПД не покинет решение ОД, если он достиг его [1]. А в случае появления только одного отличия в текущем решении QМПД от единственного оптимального решения при последующих попытках декодирования такого сообщения декодер почти всегда достигнет оптимального решения. Если же декодер устроен так, что сначала оценивается наиболее правдоподобное изменение решения в сообщении, которое затем и реализуется, то декодер вообще обязательно перейдёт к оптимальному решению. Совокупность этих свойств очень важна для устойчивой и очень частой сходимости решений QМПД к решению ОД. Поэтому их удобно так и назвать устойчивостью относительно решения ОД.

Другим важнейшим моментом является то, что по сравнению с традиционным подходом к двоичным мажоритарным схемам, в QМПД для изменения декодируемого символа достаточно наличие не абсолютного, а только относительного строгого большинства проверок, которые анализируются пороговым элементом декодера. Например, в самоортогональном коде с минимальным кодовым расстоянием $d = 9$ ошибка в декодируемом символе будет исправлена даже в том случае, если из 9 его проверок (включая и контролируемый символ разностного регистра) правильными будут только 2, а остальные 7 – ошибочными. Такого невозможно представить для двоичных кодов, а для QМПД данная ситуация типична. Единственным условием для этого частного примера являются разные значения проверок относительно декодируемого информационного символа. А для больших значений основания кода q именно это условие практически всегда и реализуется. Эти свойства QМПД существенно расширяют возможности недвоичного многопорогового алгоритма при работе в больших шумах, сохраняя при этом весьма малую сложность процедур мажоритарного типа и в q -ичных каналах.

Оценки характеристик QМПД

Наиболее частые комбинации ошибок, приводящие к ошибкам даже при использовании оптимальных переборных методов в недвоичных симметричных каналах были рассмотрены в [1-6]. В [1, 2, 4, 5] были предложены нижние оценки вероятностей таких событий. Вероятности ошибки в первом символе кода при использовании QМПД были рассмотрены в [1, 5]. Они важны при оценках вероятностей ошибок канала, при которых удаётся хотя бы немного понизить плотность ошибок в принятом сообщении. Ясно, что если на первой итерации QМПД несколько снижает число оставшихся ошибок по сравнению с их первоначальным количеством, то работа декодера на остальных итерациях существенно облегчается и действительно появляется вполне реальная

возможность достижения им решения ОД. Значит, эти вероятности можно отнести к верхним оценкам характеристик QMPD. Указанные совокупности верхних и нижних оценок вероятностей ошибки QMPD позволяют осуществлять простое предварительное определение возможностей этого алгоритма в двоичных каналах.

В целом ряде случаев оказывается также очень полезным оценить предельные возможности QMPD при малом уровне шума канала, когда практически все его конечные решения с большой долей вероятности оказываются оптимальными. В этом случае нижняя оценка для вероятности ошибки ОД, совпадающая с вероятностью ошибки декодирования алгоритма QMPD для двоичного самоортогонального кода [1, 3, 4] с минимальным кодовым расстоянием d оказывается просто равной вероятности того, что декодируемый символ и все относящиеся к нему проверочные символы приняты ошибочно. Иначе говоря, вероятность ошибки декодера в отдельном символе равна $P_s(e) = p_0^d$, где p_0 – вероятность ошибки в рассматриваемом канале. Во многих случаях такая простая оценка позволяет на первом этапе определения возможностей QMPD найти его минимальную достижимую вероятность ошибки со вполне приемлемой точностью. В случае необходимости точность этой простейшей ошибки можно повысить, если рассмотреть также вероятность таких событий, когда один из проверочных символов принят правильно, а также вариант, при котором информационный символ принят правильно, но есть 2 проверочных символа, ошибки в которых имеют одинаковые значения. В этих двух группах событий решения QMPD и ОД также будут ошибочны. Отметим, что если основание кода q будет небольшим, одного порядка с кодовым расстоянием d , то необходимо оценивать гораздо более сложные комбинации ошибок в принятом сообщении, рассмотренные, например, в [1, 4, 5].

Характеристики помехоустойчивости QMPD

Ниже рассмотрены результаты моделирования работы QMPD в двоичном симметричном канале qCK, а также возможности обычных декодеров кодов РС [1, 7, 8].

На рис. 1 представлены вероятности ошибки на символ для кодов Рида-Соломона, которые обозначены как RSn , где n – длина кода, выраженная в числе символов. Отметим, что реально сейчас используются только коды РС длины 256 или менее. Коды длины 4096 символов и, тем более, длины $n=65536$ (каждый символ – размером 16 битов), в обозримом будущем реализации не подлежат.

Здесь же пунктирными линиями показаны возможности кодов с мажоритарным декодированием при $R=1/2$ для случая $q=256$ (символ – один байт) для разных длин кодов, поскольку при любом $q>2$, как и в двоичном случае, для QMPD можно строить сколько угодно длинные коды с различными значениями кодового расстояния d и кодовой скорости R . Эти коды отмечены как M400 и M4000 с числами, обозначающими длины кодов, выраженные числом символов. Далее,

**Характеристики двоичных блочных МПД
и кодов РС с $R=1/2$**

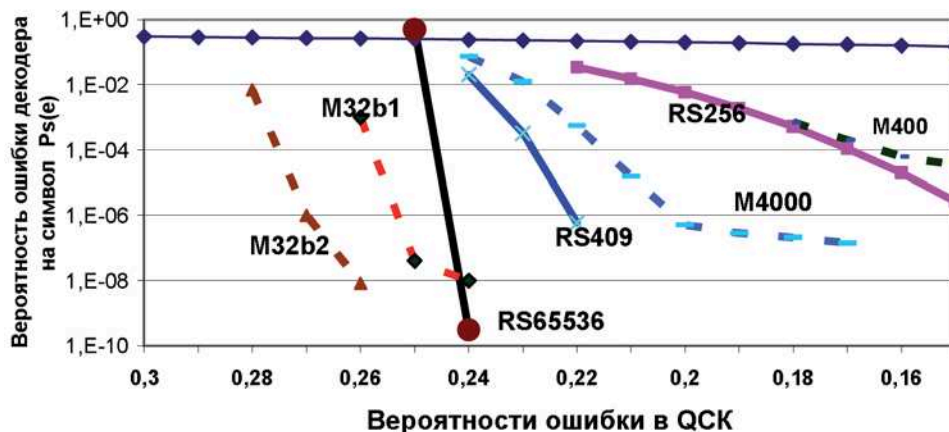


Рис. 1

обозначение M32b1 соответствует QМПД для кода длины 32000 символов. Как видно из рис. 1, возможности QМПД во всех случаях сопоставимы или лучше, чем у весьма сложных декодеров кодов РС. Более того, очень простой для реализации последний декодер для кода длины 32000 оказывается способным обеспечить простейшими мажоритарными методами помехоустойчивость, принципиально недостижимую даже для кода РС длины 65536 двухбитовых символов, декодер для которого не будет создан никогда. А если перейти к двухбайтовым недвоичным кодам с мажоритарным декодированием, то его характеристики для длины кода 32000 символов будут соответствовать графику M32b2, ещё более показательному по уровню помехоустойчивости в области, где коды РС уже не работают. При этом QМПД для двухбайтовых символов практически ни в чём не сложнее однобайтового, так как даже обычные микропроцессоры одинаково просто и быстро работают и с однобайтовыми символами, и с 2-х и даже иногда с 8-байтовыми словами со скоростью, обработки, выраженной в битах, во столько же раз большей.

Далее на рис. 2 показаны возможности QМПД и кодов РС при кодовой скорости $R=7/8$. Сплошными линиями в тех же обозначениях, что и на рис.1, представлены вероятности ошибки на символ для кодов РС.

Пунктирными линиями помечены коды с QМПД декодированием и длинами от 2560 до 32000 байтов (символ - 8 битов): M2560, M8000 и M32k. Аналогично случаю $R=1/2$, возможность создания кодов РС длины 4096 при $R=7/8$ в ближайшее время останется очень проблематичной, в то время как даже для кодов длины 32000 байтов рассматриваемые недвоичные мажоритарные декодеры останутся очень простыми.

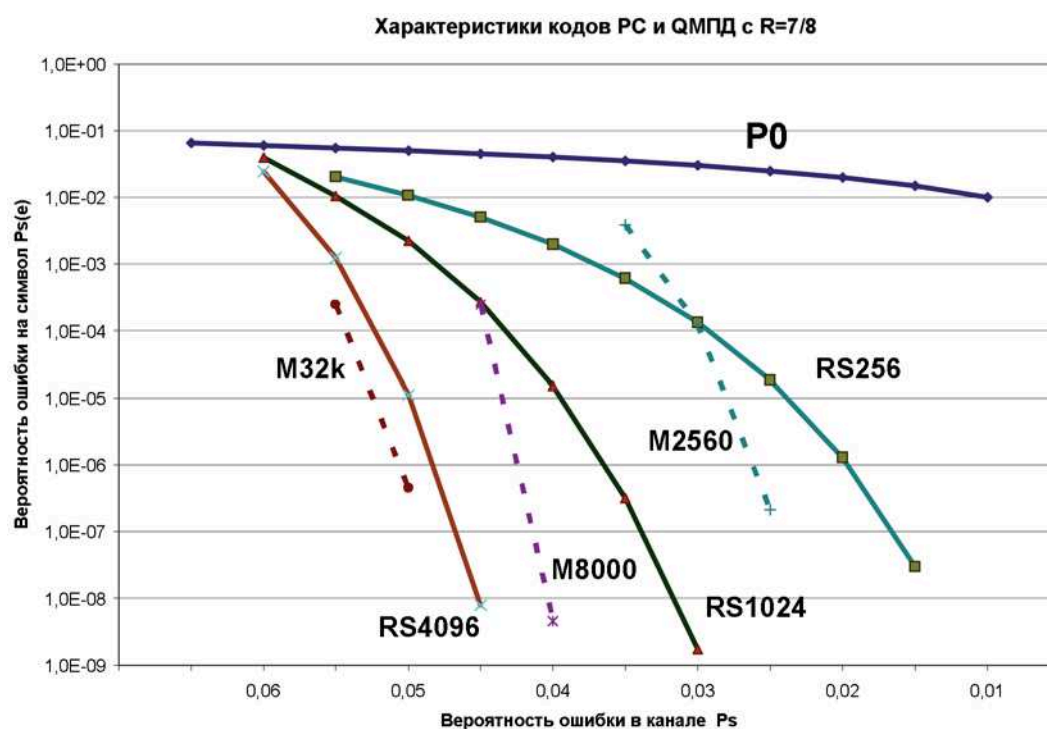


Рис. 2

Наконец, на рис. 3 для малоизбыточных кодов с $R=0,95$ представлены аналогичные характеристики для QМПД и кодов РС. Для удобного сопоставления на рис. 3 приведён также график для кода РС с $n=256$ и $R=7/8$ с рис.2. Пунктирами M80Kb1 и M80Kb2 указаны возможности двух QМПД для кодов длины $n=80000$ и с символами размером 1 и 2 байта.

Из сопоставления кодов РС длины $n=256$ при $R=7/8$ и при $R=19/20$ видно, насколько последний менее эффективен первого и насколько труднее обеспечивать хорошую эффективность при уменьшении избыточности. Тем не менее характеристики малоизбыточных кодов с мажоритар-

ным декодированием на основе QMPД оказываются весьма высокими и могут достигать очень высоких уровней помехоустойчивости, если выбранные коды имеют достаточно большие длины.

Подчеркнем также, что, согласно общим принципам теории кодирования, переход к каскадным принципам кодирования ещё более улучшит характеристики QMPД. При этом сложность декодирования увеличится по сравнению с исходным алгоритмом очень незначительно. Детально сложность QMPД рассматривалась в [1]. Реальная очень незначительная сложность QMPД декодера, выполняющего только операции сложения и сравнения, подтверждается скоростью работы этого алгоритма на ПК средней производительности для ряда типичных параметров кода и канала, которая может составлять миллиард (!) и более декодированных символов за один час полного моделирования функционирования кодера, имитатора шума канала и собственно декодера рассматриваемого класса, что иногда соответствует его программной производительности вплоть до 10^{11} бит/с.

Характеристики QMPД и кодов РС при R=19/20

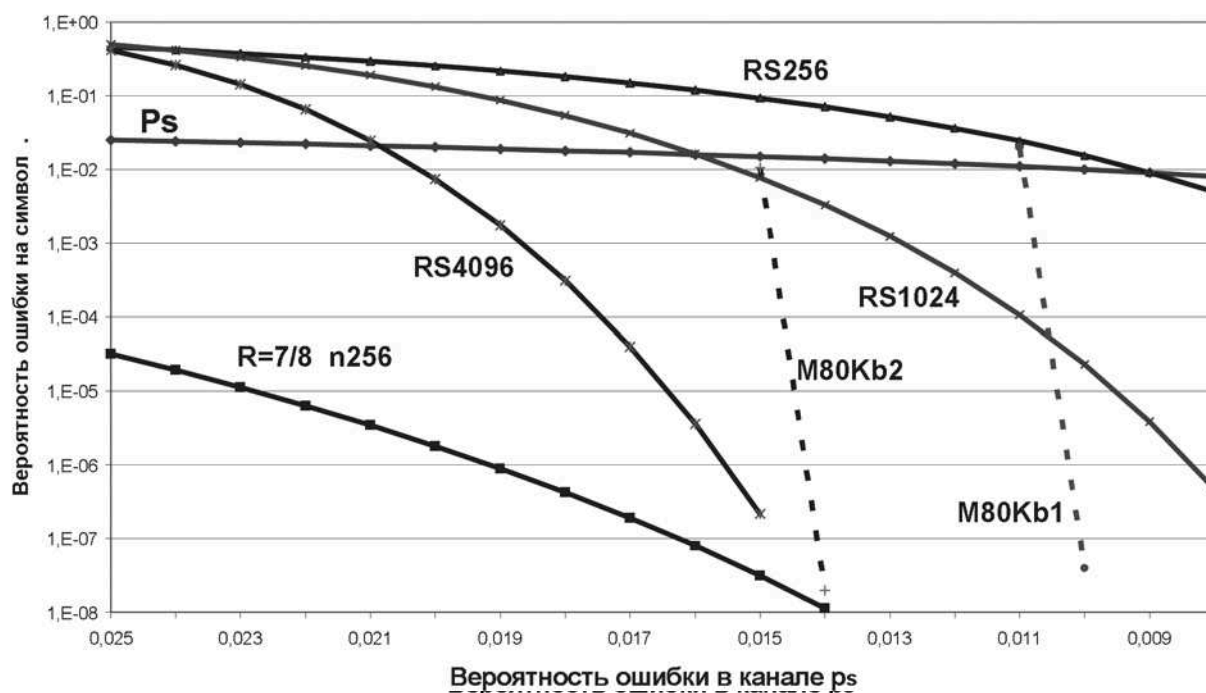


Рис. 3

Дальнейшее улучшение эффективности декодирования возможно при переходе к свёрточным кодам, методам параллельного кодирования, применения кодов с выделенными ветвями и другим мерам, часть из которых описана в [1]. Целый ряд демопрограмм различных декодеров представлен на информационно-методическом и справочном сетевом ресурсе ИКИ РАН www.mtdbest.iki.rssi.ru.

Важные области применения новых методов кодирования

Кроме естественных областей применения простых высокоэффективных методов кодирования в сетях связи, как это следует из рассмотренных выше характеристик, очень широки возможности использования QMPД для кодирования информации на DVD дисках и других носителях больших объёмов информации, в сверхбольших базах аудио- и видеоданных с намного более высоким уровнем достоверности, чем это было доступно до недавнего времени, а также при обновлении, восстановлении и использовании хранимых данных. Наиболее известными системами такого типа оказываются БД терабайтной ёмкости, в частности, для ДЗЗ. При этом легко обеспечить

и оперативный постоянный контроль за качеством хранимой информации, а также корректировку данных вследствие старения и возникающих дефектов носителя. Производительность кодеров для QМПД, которые оказываются предельно простыми одноканальными узлами, может быть также легко доведена до уровня $2,5 \div 20$ Гбайт/с. Их использование в процессе записи данных в БД никогда не может быть ограничивающим фактором для высокоскоростных систем ДЗЗ.

Принципиально новый уровень помехоустойчивости, достигаемый с помощью QМПД, позволяет решать перечисленные сложные задачи высокодостоверного хранения без дополнительной доработки этих алгоритмов или всего лишь при незначительной их адаптации к возможным дополнительным требованиям, возникающим в подобных масштабных цифровых системах.

Выводы

Возможность очень простого исправления ошибок в длинных двоичных кодах при эффективности, близкой к уровню, доступному только оптимальным переборным алгоритмам, открывает принципиально новые возможности для кодирования символьной информации, т.е. основным видам данных, практически непосредственно используемым современным информационным обществом при создании аудио-, видео- и других видов информации. Кодирование обеспечивает высокое контролируемое качество хранимой, передаваемой и обрабатываемой информации. Применение простых и одновременно высокоэффективных методов кодирования может создать новые стандарты высоко-надёжного информационного обеспечения всех аспектов развития нашей цивилизации.

Разработки алгоритмов МПД поддерживались Научным советом по комплексной проблеме "Кибернетика" АН СССР и ИКИ РАН.

Дополнительная информация – на специализированном тематическом двуязычном веб-сайте ИКИ РАН www.mtdbest.iki.rssi.ru.

Исследования велись с финансовой поддержкой РФФИ по гранту №08-07-00078а.

Литература

1. *Золотарёв В.В.* Теория и алгоритмы многопорогового декодирования // М.: Радио и связь, Горячая линия - Телеком, 2006. 270 с.
2. *Золотарёв В.В.* Многопороговое декодирование для информационных потоков с байтовой структурой // Мобильные системы. М., 2006. №3. С.25-27.
3. *Золотарёв В.В.* Характеристики каскадирования многопороговых декодеров для спутниковых каналов связи // 5-я Международная конференция и выставка «Цифровая обработка сигналов и её применение», Доклады-2, Москва, 2003. С.353-355.
4. *Золотарёв В.В., Овечкин Г.В.* Помехоустойчивое кодирование. Методы и алгоритмы. Справочник // "Горячая линия - Телеком", Москва, 2004. 128 с.
5. *Золотарёв В.В.* Алгоритмы кодирования символьных данных в вычислительных сетях // В сб.: "Вопросы кибернетики", Научный совет по комплексной проблеме «Кибернетика» АН СССР, ВК-106, М., 1985. С.54-62.
6. *Золотарёв В.В., Назиров Р.Р., Чулков И.В.* Оптимальное декодирование в цифровых спутниковых каналах при дистанционном зондировании Земли // Современные проблемы дистанционного зондирования Земли из космоса: Физические основы, методы и технологии мониторинга окружающей среды, потенциально опасных объектов и явлений. Сборник научных статей. М.: ООО «Азбука-2000», 2007. Выпуск 4. Т. 1. С. 229-235.
7. *Zolotarev V.V., Nazirov R.R., Chulkov I.V.* The Quick Almost Optimal Multithreshold Decoders for Noisy Gaussian Channels // RCSGSO International Conference ESA in Moscow, Russia, June, 2007.