

Раздел I

Математические методы синтеза систем

Н.Н. Гринченко, Г.В. Овечкин

ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ ДЛЯ ЦИФРОВЫХ СИСТЕМ СВЯЗИ¹

Введение. При передаче цифровых данных по каналам с шумами всегда существует вероятность того, что принятые данные будут содержать ошибки. Наиболее эффективным способом исправления ошибок является применение методов помехоустойчивого кодирования. На сегодняшний день теории кодирования известно множество кодов и методов их декодирования, различающихся корректирующей способностью, сложностью реализации и рядом других параметров [1]. В статье рассматривается эффективность лучших современных методов коррекции ошибок, позволяющих работать вблизи пропускной способности канала.

Характеристики методов коррекции ошибок. На рис.1 представлены зависимости вероятности ошибки на бит P_b на выходе декодера от уровня шума канала E_b/N_0 для различных методов коррекции ошибок. Данные зависимости получены для случая работы в канале с аддитивным белым гауссовым шумом (АБГШ) при использовании двоичной фазовой модуляции (ФМ2) и квантовании выходного сигнала демодулятора на 16 уровней. Данная модель канала достаточно точно описывает реальные спутниковые и многие другие типы каналов.

На данном рисунке кривые «АВ, $K=7$ », «АВ, $K=9$ » и «АВ, $K=15$ » отражают эффективность применения классического алгоритма Витерби (АВ) [2] для декодирования сверточных кодов с кодовой скоростью $R=1/2$ и длиной кодирующего регистра $K=7, 9$ и 15 соответственно. Данный метод является оптимальным, однако его сложность растет экспоненциально с ростом длины K , и поэтому на практике декодеры с $K > 9$ не используются. А поскольку при использовании кодов с $K \leq 9$ эффективность АВ оказывается невысокой возникает необходимость поиска более простых с точки зрения практической реализации декодеров, обеспечивающих в то же время близкое к оптимальному декодирование даже очень длинных кодов.

Значительно лучшими характеристиками обладают многопороговые декодеры (МПД) [1, 3–6], являющиеся модификацией обычного порогового декодера Мессе [7]. МПД в процессе контроля принятых кодовых последовательностей многократно исправляют информационные символы пришедшего из канала сообщения. Эти декодеры обладают важнейшим строго доказанным свойством – сходимостью к решению оптимального декодера при линейной зависимости сложности реализации от длины кода.

Кривая «МПД, $d=9$ » на рис.1 соответствует работе МПД самоортогонального кода (СОК) [1] с $R=1/2$ и минимальным кодовым расстоянием $d=9$. Возможности

¹ Работа выполнена при поддержке РФФИ, грант №05-07-90024в

МПД для кода с $R=1/2$ и $d=13$ представлены на рис.1 кривой «МПД, $d=13$ ». В обоих случаях длина кода не превосходила $n=10000$ бит, и при декодировании использовалось до $I=15$ итераций. Как следует из графиков, МПД для достаточно длинных кодов оказывается лучше практически реализуемых декодеров Витерби. Основные соотношения между характеристиками АВ и МПД имеют место и при других кодовых скоростях.

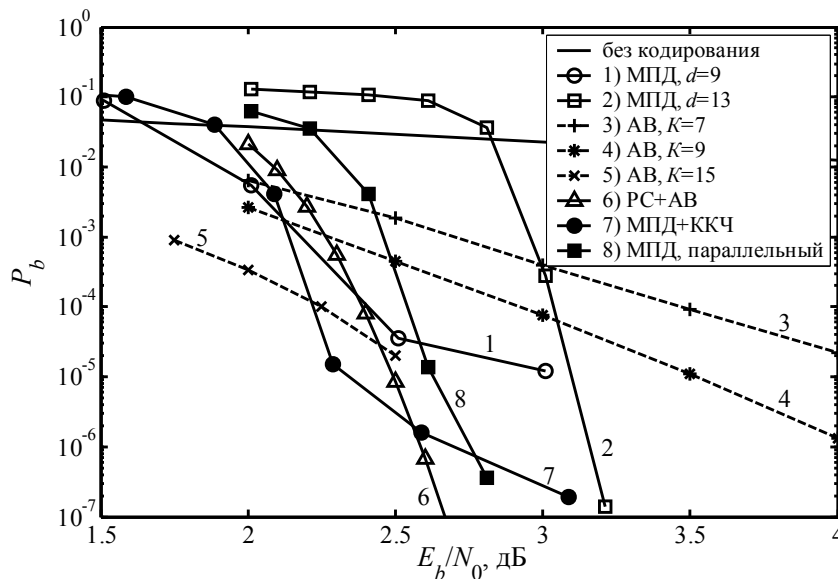


Рис.1. Характеристики методов декодирования в канале с АБГШ

Приведенные выше данные показывают, что доступные для реализации оптимальные АВ уступают МПД в широком диапазоне кодовых скоростей и уровней шума в канале. При этом МПД выполняет только простейшие операции сложения и сравнения небольших целых чисел [1], что обуславливает его крайнюю простоту при всех вариантах программной или аппаратной реализации [8].

Каскадные коды. Важным этапом в развитии теории кодирования является появление каскадных кодов [9], в основе построения которых лежит идея совместного использования нескольких составляющих кодов. Данный подход позволил существенно повысить эффективность применения кодирования по сравнению с базовыми некаскадными методами.

Наиболее широкое распространение в реальных системах связи нашла каскадная схема, в которой внешним кодом является код Рида-Соломона (РС), а внутренним – сверточный код, обычно декодируемый с помощью оптимального АВ [1]. Результаты моделирования такого каскадного кода, состоящего из кода РС длиной $n=255$, с кодовым расстоянием $d=33$ и сверточного кода с $K=7$, $R=1/2$, представлены на рис.1 кривой «РС+АВ». Заметим, что данная схема сопоставима с МПД по эффективности при существенно большей сложности реализации и при несколько меньшей общей кодовой скорости.

Дальнейшим развитием каскадных кодов стали турбокоды (Turbo Convolutional Codes – ТСС), образующиеся путем каскадирования двух или более составляющих кодов. Первые турбокоды, предложенные Беру и соавторами в [10],

состояли из двух параллельно соединенных через перемежитель сверточных кодов. В общем случае турбокоды могут получаться как при параллельном, так и последовательном соединении кодов, разделенных перемежителем.

На рис.2 представлены характеристики турбокода с кодовой скоростью $R=1/2$, конструктивной длиной составляющих рекурсивных систематических сверточных кодов $K=4$, Max-Log-MAP алгоритмом декодирования составляющих кодов, 8 итерациями декодирования и перемежителем, рекомендованным консультативным комитетом по космической связи (кривая «TCC, CCSDS») или псевдослучайным перемежителем (кривая «TCC, random»). Длина турбокода в обоих случаях была равна 10000 бит. Как видно из рисунка, последний турбокод способен работать почти в 1 дБ от пропускной способности канала. Недостатками данных кодов являются все же достаточно высокая сложность декодирования и наличие эффекта «насыщения» вероятности ошибки, заключающегося в уменьшении скорости убывания вероятности ошибки декодирования в области меньших шумов канала.

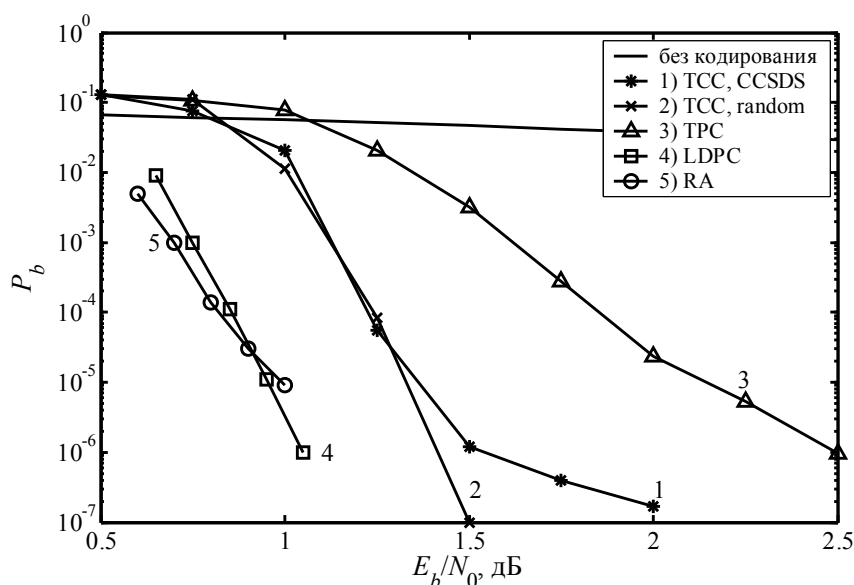


Рис.2. Характеристики турбоподобных методов декодирования в канале с АБГШ

Развитием ТСС стали турбокоды произведения (Turbo Product Codes – TPC) [11, 12], при построении которых используется последовательное каскадирование двух или нескольких линейных блочных кодов, таких как коды с контролем четности, коды Хэмминга, коды Боуза-Чоудхури-Хоквингема. Среди них наибольший интерес представляют ТРС коды, построенные на основе кодов Хэмминга.

На рис.2 представлены характеристики ТРС, построенного на основе трех блочных (n, k) кодов с параметрами (16, 11), (8, 7) и (32, 26) (кривая «ТРС»). Общая кодовая скорость таких ТРС кодов равна 0.489, а длина – 4096 бит. При декодировании данных кодов использовалось до 8 итераций. Заметим, что данные коды обладают меньшей сложностью декодирования, чем турбокоды и у них практически отсутствует область «насыщения» вероятности ошибки. Но ТРС коды показывают хорошие характеристики только при высоких кодовых скоростях.

Прекрасные характеристики, получаемые при итеративном декодировании турбокодов, стимулировали активные исследования по применению данного метода к декодированию других кодов. В результате оказалось возможным получить еще лучшие характеристики при декодировании низкоплотностных (Low Density Parity Check – LDPC) кодов [13, 14], предложенных Галлагером еще в 1962 г. Для декодирования LDPC кодов обычно используется разновидность итеративного декодирования, называемая алгоритмом распространения надежности (Belief Propagation), работающего с графом кода.

На рис.2 представлены характеристики нерегулярных низкоплотностных кодов с кодовой скоростью $1/2$ и длиной 10000 бит (кривая «LDPC»). Как следует из данного рисунка, эффективность LDPC кодов оказывается лучше эффективности турбокодов, что особенно заметно при больших длинах кодов и при малых вероятностях ошибки декодирования. При этом сложность практической реализации декодера LDPC кодов существенно ниже сложности реализации декодера турбокодов. К недостаткам данных кодов можно отнести достаточно высокую сложность кодирования.

Коды повторения-накопления (Repeat-Accumulate – RA) [15] сочетают в себе достоинства турбокодов (линейная сложность кодирования) и низкоплотностных кодов (линейная сложность декодирования). Данные коды являются разновидностью кодов с последовательным каскадированием. Кодер состоит из повторителя, перемежителя и аккумулятора. В роли последнего обычно выступает рекурсивный сверточный кодер с кодовой скоростью 1.

Кривая «RA» на рис.2 отражает характеристики нерегулярных RA кодов с длиной 10000 бит и кодовой скоростью $R=1/2$. Видно, что эффективность RA кодов оказывается лучше эффективности турбокодов и сопоставима с эффективностью LDPC кодов.

Каскадные коды, декодируемые с использованием многопорогового декодера. Высокие характеристики, показываемые обычным некаскадным МПД, способствуют его широкому применению в составе различных кодовых конструкций, поскольку эффективность последних непосредственно связана с эффективностью их составных элементов.

Особое место среди кодовых схем на базе МПД занимает каскадирование с кодами контроля по четности (ККЧ) [4], использование которых также позволяет повысить эффективность применения кодирования. Особенность данной схемы заключается в том, что такое каскадирование практически не требует дополнительных затрат на оборудование (в схему кодирования требуется добавить лишь один сумматор по модулю 2), тогда как использование в каскадном коде, например, кодов Рида-Соломона, несравненно труднее.

На рис.1 показаны результаты моделирования каскадной схемы, состоящей из СОК с $R=1/2$ и $d=9$, декодируемого с помощью МПД, и ККЧ длиной 50 бит (кривая «МПД+ККЧ»). Заметим, что использование простейшего ККЧ совместно с СОК позволяет получить дополнительный энергетический выигрыш около 1 дБ (по сравнению с кривой «МПД, $d=9$ ») при вероятности ошибки на выходе декодера $P_b=10^{-5}$. Также отметим, что каскадный код, состоящий из кода Рида-Соломона (255, 223, 33) и декодера Витерби для кода с $R=1/2$ и $K=7$, даже при меньшей общей кодовой скорости, уступает каскадной схеме на базе МПД при $P_b>10^{-6}$. К сожалению, рассмотренный способ каскадирования позволяет улучшить характеристики МПД только в области его эффективной работы.

Для приближения границы эффективной работы МПД к пропускной способности канала возможно его применение в схемах параллельного

кодирования [5], в основе построения которых лежит выделение в некотором базовом СОК, например с кодовой скоростью $R=1/2$, двух составляющих СОК с большими кодовыми скоростями. В такой схеме на первых итерациях декодируется первый составляющий СОК, после чего в процесс декодирования включается второй СОК. Заметим, что в результате на последних итерациях МПД просто декодирует базовый СОК с $R=1/2$, в то время как в обычных каскадных кодах кодовая скорость внешнего кода близка к единице. Данное свойство обеспечивает существенное преимущество МПД перед другими каскадными конструкциями.

Для примера кривой «МПД, параллельный» на рис.1 представлены результаты моделирования схемы с параллельным кодированием в канале с АБГШ для СОК с $R=1/2$, $d=13$. Как следует из анализа представленных графиков, применение параллельного кодирования позволяет приблизить границу эффективной работы МПД к пропускной способности канала примерно на 0,5 дБ.

Сложность МПД при параллельном кодировании оказывается даже меньше сложности обычного МПД, поскольку в данном случае на первых итерациях декодирования некоторые элементы синдромного регистра просто не участвуют в процессе вычисления суммы на пороговом элементе.

Заметим, что приведенные для МПД результаты относились только к двоичным кодам и системам с двоичной фазовой модуляцией. Но данный метод уже реализован и для недвоичных кодов, обеспечивая характеристики близкие или несколько более высокие, чем у кодов Рида - Соломона. Кроме того, МПД показывает высокие характеристики в каналах со стираниями и в системах с многопозиционной модуляцией, что делает его универсальным средством простого достижения высокого уровня помехозащищенности сообщений в системах широкого спектра применения. Дополнительную информацию об алгоритмах класса МПД можно найти на тематическом веб-сайте ИКИ РАН www.mtdbest.iki.rssi.ru.

Заключение. Последнее десятилетие создало условия для достижения в реальной каналообразующей аппаратуре систем связи уровня энергетики, весьма близкого к теоретическим пределам для каналов космической и спутниковой связи. Высокие характеристики в настоящее время могут обеспечить всего несколько методов коррекции ошибок. Но учет проблемы сложности при сопоставимой корректирующей способности показывает, что в этом случае алгоритмы на базе МПД сохраняют близость по своей сложности реализации к простейшему декодеру порогового типа и поэтому в большинстве случаев применения кодирования являются наиболее предпочтительными методами декодирования в приемлемых по стоимости высокоскоростных системах связи.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Золотарев В.В., Овечкин Г.В. Помехоустойчивое кодирование. Методы и алгоритмы. Справочник.-М.: Горячая линия – Телеком, 2004.-126 с.
2. Витерби А. Границы ошибок для сверточных кодов и асимптотически оптимальный алгоритм декодирования // Некоторые вопросы теории кодирования.-М.: Мир, 1970.- С.142–165.
3. Золотарев В.В., Овечкин Г.В. Многопороговые декодеры для каналов с предельно высоким уровнем шума // Телекоммуникации. 2005. №9. С.29–34.
4. Золотарев В.В., Овечкин Г.В. Использование многопорогового декодера в каскадных схемах // Вестник РГРТА. 2003. Вып.11. С.112–115.

5. Золотарев В.В. Параллельное кодирование в каналах СПД // Вопросы кибернетики. 1986. Вып.120. С.56–58.
6. Веб-сайт www.mtdbest.iki.rssi.ru.
7. Мессу Дж. Пороговое декодирование / Пер. с англ.; под ред. Э.Л. Блоха. М.: Мир, 1966.-208 с.
8. Золотарев В.В., Овечкин Г.В. Сложность реализации эффективных методов декодирования помехоустойчивых кодов // 6-я Межд. конф. и выст. «Цифровая обработка сигналов и ее применение». М.: 2004. Т.1. - С.220-221.
9. Форни Д. Каскадные коды // Пер. с англ.; под ред. Самойленко С.И. М.: Мир, 1970.-208 с.
10. Berrou C., Glavieux A., Thitimajshima P. Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes // Proc. of the Intern. Conf. on Commun. 1993. May. P.1064–1070.
11. Press Release, AHA announces Turbo Product Code Forward Error Correction Technology. 1998.
12. Williams D. Turbo Product Code FEC Contribution // IEEE 802.16.1pc-00/35. 2000.
13. MacKay D.J.C., Neal R.M. Near Shannon limit performance of low density parity check codes // IEEE Electronics Letters. Aug. 1996. V.32. №18. P.1645–1646.
14. Richardson T., Shokrollahi M., Urbanke R. Design of capacity-approaching irregular low-density parity-check codes // IEEE Trans. Inform. Theory. 2001. V.47.-P.638–656.
15. Jin H., Khandekar A., McEliece R. Irregular repeat-accumulate codes // Proc. 2nd Int. Symp. on Turbo Codes and Related Topics. 2000, Sept.-P.1–8.

А.В. Пруцков

ОПРЕДЕЛЕНИЕ И ГЕНЕРАЦИЯ СЛОЖНЫХ ФОРМ СЛОВ ЕСТЕСТВЕННЫХ ЯЗЫКОВ ПРИ МОРФОЛОГИЧЕСКОМ АНАЛИЗЕ И СИНТЕЗЕ

За последние десятилетия вычислительная техника совершила большой скачок и заняла важное место в жизни современного общества. Одной из причин такого скачка является развитие средств взаимопонимания человека и машины, что сделало компьютерные системы более удобными для использования. Однако взаимодействие человека и машины осуществляется пока с помощью команд, подаваемых с клавиатуры или манипулятора типа «мышь» через графический интерфейс пользователя (graphic user interface, GUI). Системы общения с человеком на естественном языке позволяют ускорить и упростить работу с компьютерными системами. В настоящее время такие системы получают все большее развитие не только в компьютерной, но и в бытовой технике.

Понимание речи включает два этапа:

- 1) распознавание речи: перевод речи в текстовый вид;
- 2) понимание текста: перевод текста в смысловое содержание.

В процессе понимания текста в свою очередь можно выделить последовательность следующих этапов:

- 1) лексический – выделение в тексте слов;
- 2) морфологический – определение грамматических характеристик (падеж, число, время) выделенных слов;