

А.Н. Кокоулин, Е.Л. Кон

Пермский национальный исследовательский
политехнический университет

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ ДВУХСТУПЕННЫХ АЛГЕБРАИЧЕСКИХ КОДОВ В СИСТЕМАХ НАДЕЖНОГО ХРАНЕНИЯ ИНФОРМАЦИИ

Обосновывается возможность усовершенствования алгоритма исправления стираний при двухступенных алгоритмах декодирования итеративными и каскадными кодами. Показано, что при синдромном декодировании кодовых слов внешнего кода Хемминга существуют комбинации стираний кратности $e > d_{\min} - 1$, которые могут быть исправлены. Количество подобных комбинаций увеличивается для укороченных внешних кодов. В результате реализации эффективного алгоритма декодирования ГСК могут быть улучшены основные характеристики системы надежного хранения информации – исправляющая способность и время декодирования.

Системы повышения надежности хранения и передачи данных в информационных системах, как правило, для парирования дефектов носителей информации используют многоступенные алгебраические коды. Важнейшими задачами, решаемыми при разработке подобных систем, являются выбор алгебраического кода, выбор параметров кодовых групп и поиск эффективных алгоритмов кодирования и декодирования [1]. Некоторые рекомендации по выбору типа кода, выбору длин кодовых групп и методики оценки эффективности алгоритмов кодирования и исправления ошибок в системах повышения надежного хранения информации на магнитных дисках (СНХИ) были приведены в предыдущих публикациях авторов [1, 2, 3, 5, 6]. В данной статье рассматриваются методы повышения эффективности алгоритма исправления ошибок (стирания) кодов Хемминга, входящих в состав двумерных итеративных кодов СНХИ, за счет целенаправленного подбора проверочной матрицы и предлагается методика оценки числа потенциально исправимых ошибок стирания вблизи границы $e = d_{\min} - 1$. Ранее

нами было доказано, что, несмотря на существование границы Хемминга [4], определяющей исправляющую способность кода по известному минимальному кодовому расстоянию d_{\min} , существуют некоторые комбинации ошибок стирания кратностью, большей $e = d_{\min} - 1$, которые могут быть исправлены при реализации переборного алгоритма декодирования [3, 5, 6]. Были вычислены оценки вероятности трансформации ошибок при реализации переборного алгоритма и показана эффективность подобных алгоритмов в сравнении с базовыми алгоритмами исправления стираний.

Структура итеративного кода СНХИ. Итеративный код СНХИ реализован в виде совокупности двух кодов: внутренним кодом (кодом строки) является циклический код, кодек внутреннего кода реализуется аппаратно в каждом накопителе на магнитных дисках в виде посекторного контроля CRC/ECC (рис. 1), а внешним кодом (кодом столбца) является код Хемминга с реализуемым программно кодеком. Кодек внутреннего кода работает в режиме обнаружения и исправления ошибок, кодек внешнего кода работает в режиме исправления стираний. Внешний код может быть выбран из множества циклических (ЦСК) или групповых (ГСК) кодов.

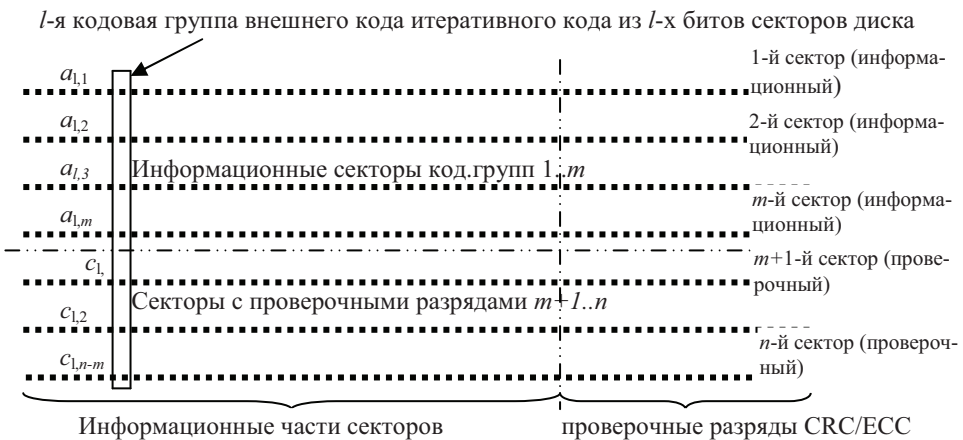


Рис. 1. Структура помехоустойчивого кода СНХИ

На рис. 1 представлена структура кодовой группы итеративного кода СНХИ на магнитном диске с внешним кодом Хемминга (кодом столбца), и внутренним кодом (кодом строки) является аппаратно реализуемый. Подобная схема реализована и в отказоустойчивых дисковых массивах технологии RAID-2.

Методика оценки числа потенциально исправимых ошибок стирания. Определим зависимость числа и кратности исправимых комбинаций векторов стирания от вида проверочной матрицы для двухступенного итеративного кода, используемого в СНХИ. Выявленные и обоснованные закономерности подтвердим на конкретном примере кода, использованного нами при реализации СНХИ на гибких магнитных дисках – внешнего группового (ГСК) кода (18,12,4).

Как известно, используемый код гарантированно исправляет ошибки стирания кратностью до $e = d_{\min} - 1 = 3$, но способен исправлять и некоторые комбинации кратностью 4 и 5, что было подтверждено в ходе практических испытаний СНХИ с переборным алгоритмом декодирования. Это обусловлено тем, что при исправлении стираний нам известны возможные позиции с ошибками, и из всего множества пар (синдром:вектор ошибки) мы можем выбирать только те пары, которые соответствуют определенному правилу. Рассмотрим подробнее процесс декодирования.

Допустим, при считывании секторов с диска возникли ошибки считывания в i секторах, относящихся к одной кодовой группе итеративного кода (см. рис. 1). То есть после декодирования внешнего кода несколько строк помечены как недостоверные (или заменены символами стирания). Тогда в каждом кодовом слове внутреннего кода будет находиться i символов стирания в позициях, соответствующих поврежденным секторам. Обозначим этот вектор ошибок стирания как e_i , в дальнейшем мы будем называть его маской. Необходимо учесть, что для контроля целостности современных магнитных дисков используются такие коды CRC/ECC, которые обеспечивают вероятность обнаружения ошибок не выше $10^{-12} \dots 10^{-14}$. Поэтому вероятностью появления ошибок перехода в бинарном алфавите, возникших помимо ошибок стирания во внутреннем кодовом слове, можно пренебречь. Тогда можно рассматривать весь процесс декодирования кода столбца как исправление ошибок в канале со стираниями и считать, что ошибки сосредоточены в позициях, соответствующих вектору e_i , а все остальные символы кодового слова являются истинными.

Покажем, что в этом случае мы можем использовать принцип синдромного декодирования для исправления ошибок стирания. Для кода с числом информационных разрядов m и длиной проверочной части $k = m - n$ каждому значению синдрома соответствует в среднем $2^m/n$ векторов ошибок. При работе в канале с ошибками появление

в кодовом слове ошибок кратности $s > \lceil d_{\min} - 1 \rceil / 2$ гарантированно приводит к неправильному распознаванию синдрома, так как из всего множества комбинаций векторов ошибок выбирается наиболее вероятный вектор, имеющий наименьший вес. В то же время, если декодеру известны позиции с заведомо истинными символами и позиции, помеченные как недостоверные, то из всего множества пар (синдром:вектор ошибки) можно выбрать только те векторы, которые соответствуют маске e_i . Это правило позволяет уменьшить размерность пространства решений, особенно для векторов ошибок стирания невысокой кратности, не превышающей $d_{\min} + 2$. Тогда для реализации алгоритма исправления стираний достаточно иметь либо статическую двухмерную таблицу с синдромами, в которой одним измерением будут комбинации масок, вторым измерением – значения синдромов, а в ячейках будут искомые комбинации векторов стирания, либо просчитывать синдромы динамически для каждого значения маски.

В качестве альтернативы можно предложить разработанный авторами переборный алгоритм, перебирающий все комбинации векторов стираний внутри маски до тех пор, пока синдром не станет равным нулю [3, 5, 6]. Можно показать также, что необходимым и достаточным признаком исправимости ошибок при конкретном значении маски является отсутствие повторений значений синдрома в соответствующей строке таблицы.

1	код e_i	результат	кратность	с	и	н	д	р	о	м	ы	.
2				1	2	3	4	5	6	7	8	9
3	*0000001		1	*0000001								
4	*0000010		1		*0000010							
5	*0000011		2	*0000001	*0000011	*0000010						
6	*0000100		1					*0000100				
7	*0000101		2	*0000001			*0000101	*0000100				
8	*0000110		2		*0000010			*0000100	*0000110			
9	*0000111		3	*0000001	*0000011	*0000010	*0000101	*0000100	*0000110	*0000111		
10	*0001000		1									*0001000
31	*0011101	BAD	4	*0000001			*0000101	*0000100			*0001001	*0001000
32	*0011110		4	*0011100	*0011110	*0000010	*0011000	*0000100	*0000110	*0011010	*0010100	*0001000
33	*0011111	BAD	5	*0000001	*0000011	*0000010	*0000101	*0000100	*0000110	*0000111	*0001001	*0001000
34	*0100000		1									
35	*0100001		2	*0000001								
36	*0100010		2		*0000010						*0100010	

Рис. 2. Фрагмент таблицы соответствия синдромов векторам стираний

Для наглядности продемонстрируем использование описанной методики оценки числа потенциально исправимых ошибок стирания вблизи границы $e = d_{\min} - 1$ для кода (7,3,4). Для этого необходимо составить таблицу, фрагмент которой представлен ниже (рис. 2). Таблица

масок ошибок стираний и соответствующих синдромов составляется при помощи специально написанной программы.

В представленной таблице можно увидеть построчное распределение синдромов для каждой из комбинаций e_i , указанной в первом столбце. Второй столбец указывает на наличие коллизий – совпадений значений синдромов для разных комбинаций ошибок внутри одного e_i . Можно заметить, что некоторые из комбинаций стираний кратности 4 (при $d_{\min} = 4$) могут быть исправлены. Третий столбец показывает вес e_i (кратность ошибки стирания), а начиная с четвертого столбца, отсчитываются синдромы: в 4-м столбце синдром равен 1 (0001), в пятом – 2 (0010) и т.д. В ячейках записаны векторы ошибок, соответствующие этим синдромам. Если для одного значения синдрома существует больше одного вектора ошибок, возникает коллизия, и необходимо отказаться от декодирования для данного e_i . Приведем итоговые расчеты количества неисправимых ошибок стирания разной кратности для исследуемого кода (табл. 1).

Таблица 1

Количество неисправимых ошибок стирания для кода (7,3,4)

Кратность	7	6	5	4	3	2	1
Число ошибок кратности $e > d_{\min} - 1$	1	7	21	35	0	0	0
Число неисправимых ошибок при использовании переборного алгоритма декодирования	1	7	21	7	0	0	0

Полученные результаты позволяют сделать вывод о целесообразности использования переборного алгоритма декодирования для кода (7,3,4).

Исследование эффективных алгоритмов декодирования для укороченных внешних кодов. Рассмотрим ситуацию, когда внешний код Хемминга является укороченным. Это означает, что количество пар (синдром: вектор ошибки) сокращается по сравнению с полным кодом. Можно *утверждать*, что при этом вероятность коллизий в одной строке таблицы синдромов также уменьшается. Доказать это утверждение можно, подсчитав число возможных комбинаций векторов ошибок, удовлетворяющих маске e_i , при одинаковом числе ком-

бинаций синдромов. Так, число синдромов равно 2^K , число комбинаций векторов ошибок для маски весом W равно $\sum_{i=1}^W C_N^i$. Нетрудно заметить, что при уменьшении N эта сумма так же пропорционально уменьшается, тем самым уменьшая вероятность коллизий.

Приведем результаты исследования укороченного внешнего кода ГСК (18,12,4) при помощи методики оценки числа потенциально исправимых ошибок стирания вблизи границы $e=d_{\min}-1$, опуская расчеты для маловероятных ошибок высокой кратности (табл. 2). Исходя из результатов расчета, можно убедиться в существенном улучшении способности декодера исправлять ошибки стирания в весовом диапазоне от d_{\min} до $d_{\min}+2$. Ранее нами были выведены и проанализированы формулы, позволяющие получить граничные оценки вероятности исправления ошибок [5], которые полностью подтверждаются результатами проведенных расчетов для различных длин кодов. Так, для кода (18,12,4) вероятность необнаружения ошибки (стирания) снижается с 2×10^{-13} до $1,5 \times 10^{-14}$, т.е. на порядок.

Таблица 2

Количество неисправимых ошибок стирания для кода (18,12,4)

Кратность	10	9	8	7	6	5	4	3	2	1
Число ошибок кратности $e > d_{\min}-1$	43758	48620	43758	31824	18564	8568	3060	0	0	0
Число неисправимых ошибок при использовании переборного алгоритма декодирования	43758	48620	43758	31824	12292	2072	148	0	0	0

Исследование зависимости числа потенциально исправимых ошибок стирания от вида проверочной матрицы. Было замечено, что вероятность исправления ошибок стираний высокой кратности (до $d_{\min}+2$) в некоторой степени зависит от выбора проверочной матрицы укороченного кода. Поскольку в различных источниках по теории кодирования подобного исследования не проводилось, и отсутствуют рекомендации по выбору проверочной матрицы для укороченных кодов ГСК, работающих в канале со стираниями, было проведено экспериментальное исследование. Для поиска наиболее эффективной структуры проверочной матрицы была написана программа, перебирающая все варианты матриц и выполняющая подсчет числа

неисправляемых ошибок стирания. Результаты для укороченного кода представлены в табл. 3.

Таблица 3

Количество неисправимых ошибок для нескольких вариантов проверочных матриц кода ГСК (18,12,4)

Кратность	10	9	8	7	6	5	4	3	2	1
Число ошибок кратности $e > d_{\min} - 1$	43758	48620	43758	31824	18564	8568	3060	0	0	0
Число неисправимых ошибок при использовании переборного алгоритма декодирования для проверочной матрицы, созданной удалением подряд идущих столбцов слева	43758	48620	43758	31824	12292	2072	148	0	0	0
Число неисправимых ошибок при использовании переборного алгоритма декодирования для наилучшего варианта матрицы	43758	48620	43758	31824	9356	1456	104	0	0	0

Можно сделать предположение, требующее дальнейшего исследования и доказательства, что вероятность исправления ошибок для укороченных кодов тем выше, чем больше среднее кодовое расстояние между столбцами проверочной матрицы. Иными словами, при выборе вида проверочной матрицы целесообразно учитывать данный факт и проводить сравнительный анализ всех вариантов матриц. Наилучший и неоптимизированный вид матриц приведен на рис. 3.

0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	0	0	1	1	1	1	1	1	1	0
0	0	0	1	0	0	0	1	1	1	0	0	0	1	1	1	1	0
0	0	1	0	0	0	1	0	1	1	0	1	1	0	0	1	1	0
0	1	0	0	0	0	1	1	0	1	1	0	1	0	1	0	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

a

0	0	0	0	0	1	0	1	0	1	0	0	1	0	1	0	1	1
0	0	0	0	1	0	0	1	0	1	1	1	1	1	0	1	0	0
0	0	0	1	0	0	0	1	1	1	0	0	0	1	1	1	1	0
0	0	1	0	0	0	1	0	1	1	0	1	0	0	1	1	0	0
0	1	0	0	0	0	1	1	0	0	1	0	1	0	0	0	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

б

Рис. 3. Неоптимизированный (*a*) и наилучший (*б*) вид матриц

Выводы:

1. Проведенное теоретическое исследование укороченных внешних кодов Хемминга, работающих в канале со стираниями, показало возможность исправления некоторых комбинаций ошибок кратности, превышающей $e=d_{\min}-1$.

2. Эти результаты согласуются с полученными ранее граничными оценками исправляющей способности переборного алгоритма системы СНХИ. По результатам эксперимента вероятность неисправления ошибки снижается на порядок при использовании переборного алгоритма.

3. Предложена методика оценки числа потенциально исправимых ошибок стирания вблизи границы $e=d_{\min}-1$ и написана программа, составляющая таблицу масок ошибок стираний и соответствующих синдромов.

4. Дальнейшее улучшение исправляющей способности алгоритма декодирования может быть достигнуто при использовании оптимизированных проверочных матриц.

Библиографический список

1. Кокоулин А.Н., Кон Е.Л. Выбор критериев и методики оценки вычислительной сложности алгоритмов кодирования блоковыми алгебраическими кодами // Информационные управляющие системы: сб. науч. тр. – Пермь, 2002. – С. 32–39.

2. Кокоулин А.Н. Система надежного хранения информации на магнитных дисках // Информационные управляющие системы: сб. науч. тр. – Пермь, 1999. – С. 64–70.

3. Кокоулин А.Н., Кон Е.Л. Особенности реализации кодека помехоустойчивого кода в составе системы обеспечения надежного хранения информации на магнитных дисках (СНХИ) // Информационные управляющие системы: сб. науч. тр. – Пермь, 2000. – С. 23–29.

4. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки. – Мир, 1976.

5. Кокоулин А.Н. Методы и алгоритмы повышения надежности хранения информации на магнитных дисках в системах RAID // Вестник Казанского государственного технического университета им. А.Н. Туполева. – Казань: Изд-во КГТУ, 2007. – № 2. – С. 72–76.

6. Кокоулин А.Н. Эффективные алгоритмы кодирования и исправления ошибок в отказоустойчивых системах // Информационные технологии моделирования и управления. – Пермь: Изд-во Перм. гос. техн. ун-та, 2007. – №3(37). – С. 348–354.

Получено 19.09.2011