УДК 004.03

Повышение надежности хранения информации в микрокомпьютерных комплексах персональной идентификации

Е.Г. Андрианова, Д.А. Крюков, А.Б. Петров

Аннотация

Рассматривается подход использования кодов коррекции ошибок в памятимикрокомпьютерных комплексов с ограниченными ресурсами, таких как SMART-карты, RFID-метки и т. п., построенных на электронных интегральных схемах.Проведен анализ эффективности и степень расхода памяти для кода Хемминга. В качестве альтернативы известному коду Хемминга, предложены решения, основанные на помехоустойчивом кодировании Рида-Соломона, а так же продольном контроле избыточности.

Ключевые слова

smart; rfid; код хемминга; код рида-соломона; lrc-код

Введение

Микрокомпьютерные комплексы персональной идентификации (МКПИ), построенные на базе технологий RFID (RadioFrequencyIDentification), радиочастотная идентификация)и SMART, реализуют крайне важную функцию, они является носителями исключительных данных владельца. Яркими представителями этого класса являются, например, банковская карта, электронные ключи доступа, чип-ключи автомобиля, цифровое телевидение, SIM-карта мобильного аппарата. Очевидно, что ценность информации в МКПИ намного превышают их рыночную стоимость. Потеря или искажение информации являются критичными и не допустимы для них. Таким образом, долговременное хранение и своевременное предоставление информации является ключевой функцией каждого МКПИ. В открытой печати практически отсутствуют публикации о повышении надежности МКПИпутем разработки и модификации методов и алгоритмов обработки, хранения и ввода-вывода информации. Данные устройства не используют алгоритмов обеспечения целостности информации и надежности функционирования, что может породить возникновение ошибок, во многом, не зависящих от действий владельца.В статье

предлагается рассмотреть решение о применении алгоритмов помехоустойчивого кодирования для SMART- и RFID-устройств, построенных на электронных интегральных схемах, содержащих, помимо модулей памяти – микропроцессор с арифметико-логическим устройством, устройство управления, устройство ввода-вывода, регистры [1]. МКПИ предполагают наличие специализированной файловой системы (ChipOperationSystem, GSM FileSystem), обеспечивающей большой набор сервисных операций и средств безопасности и предусматривающей разграничение доступа к информации.

Количественные оценки алгоритмов помехоустойчивого кодирования

Для хранения идентификационных данных в картах с микросхемой памяти используется постоянное запоминающее устройство (ПЗУ). Для хранения остальных данных, в том числе изменяемых, в SMART-картах и RFID-устройствах используется электрически программируемое постоянное запоминающее устройство (ЭСППЗУ). Данный вид памяти также как и память ПЗУ является энергонезависимым, но в отличие от неё в ЭССПЗУ можно производить запись в период эксплуатации чипа. Память ЭСППЗУ является странично-ориентированной со средним размером страницы 1-32 байт. В силу технических особенностей ЭСППЗУ, для записи данных в ячейку памяти необходимо предварительное стирание всей страницы. Это означает, что запись происходит всегда для всей страницы целиком. Скорость записи в ЭСППЗУ составляет порядка 3-10 мс на одну страницу памяти. Объем памяти ЭСППЗУ от 64 до нескольких сотен килобайт. Поскольку объем памяти ЭСППЗУ ограничен, возникают определенные условия для использования алгоритмов и методов помехоустойчивого кодирования данных, хранимых на карте. Странично-ориентированный характер памяти ЭСППЗУ также создает определенную специфику использования методов помехоустойчивого кодирования в этом виде памяти. Другой технической особенностью ЭСППЗУ является ограниченность числа процедур записи, которые можно произвести в страницу памяти ЭСППЗУ. Этот предел сильно варьируется в зависимости от конструктивных особенностей чипа и таких условий эксплуатации, как температура и напряжение питания чипа, и находится чаще всего в диапазоне 10000-100000 циклов записи/считывания.Всякая страница памяти ЭСППЗУ при достижении определенного числа циклов записи может выйти из строя. Это означает, что страница памяти будет более непригодна для хранения в ней данных. Данные, находившиеся в ней, будут потеряны.Другой особенностью ЭСППЗУ является то, что данные, хранимые в этом виде памяти, могут храниться в ней только определенное время. Производителями

гарантируется время хранения порядка 5 лет. Когда срок службы ЭСППЗУ приближается к концу, время хранения данных может составлять несколько часов или минут.

Указанные ошибки являются характерными для данного вида памяти. Причиной иных ошибок могут быть, например, ошибки записи данных, вызванные помехами в цепи питания. Другим источником повреждения данных может служить статическое электричество.Таким образом, корректирующие коды, используемые в данном виде памяти, должны быть преследовать цель,прежде всего,восстановить данные целой страницы памяти.

Далее будем рассматривать перезапись *i*-ой страницы данных как событие, имеющее вероятность P_i^d. Величину P_i^d можно определить как отношение числа осуществленных циклов стирания/записи *i*-ой страницы к числу процедур записи за период эксплуатацииМКПИ. Вероятности P_i^d вместе составляют вектор вероятностей $P^d = \left(P_0^d, \ldots, P_{k-1}^d\right)$, где k – число страниц данных.Определим также вектор вероятностей $P^c = \left(P_0^c, \ldots, P_{m-1}^c\right)$ как вектор, элементами которого являются вероятности P_i^c составляющие событие перезаписи хотя бы одной страницы данных во время процедуры записи данных, зависимой от *i*-ой страницы контроля. События перезаписи страниц контроля в отличие от событий перезаписи страниц данных, в общем случае, не могут быть рассмотрены как независимые. Также определим вектор $\tilde{P}^c = \left(\tilde{P}_0^c, \ldots, \tilde{P}_{m-1}^c\right)$, элементами которого являются вероятности $\tilde{P}_i^c$ перезаписи соответствующих страниц контроля.

Определим коэффициент эффективности кода как векторную величину:

$$\rho = \left(\rho_0, \ldots, \rho_{m-1}\right) = \frac{1}{\max P_j^d} \tilde{P}^c. \qquad (1)$$

Эта величина характеризует то, насколько чаще происходит перезапись каждой страницы контроля, чем самой нагруженной страницы данных. Если считать, что при достижении гарантийного числа циклов стирания/записи страница памяти выходит из строя, то в действительности число страниц памяти, необходимых для хранения контрольной информации, равно сумме величин ρ_i, округленных да большего целого. Обозначим эту величину $\tilde{m}$ и приведем формулу для ее вычисления:

$$\tilde{m} = -\sum_{i=0}^{m-1} \left[-\rho_i\right]. \qquad (2)$$

Определим относительный расход памяти ε и действительный относительный расход памяти $\tilde{\varepsilon}$:

$$\varepsilon = \frac{m}{k}, \tag{3}$$

$$\tilde{\varepsilon} = \frac{\tilde{m}}{k}. \tag{4}$$

Организация хранения данных в ЭСППЗУ на основе кода Хэмминга

Рассмотрим самокорректирующийся код Хэмминга. Будем обозначать через q мощность алфавита кода $C(N,m)$. Так как всякий символ кодового слова представляет собой двоичную последовательность фиксированной длины l, очевидно, что $q = 2^l$. А так как доступ к памяти ЭСППЗУ побайтовый, наиболее целесообразным будет выбирать l так, чтобы либо в байте умещалось целое число символов-последовательностей, либо в символе помещалось целое число байт и одновременно в странице помещалось целое число символов. В первом случае l может быть равно 1, 2, 4, 8. Как правило, размер страницы памяти ЭСППЗУ равен 2 в некоторой степени, а потому и во втором случае l будет степенью двойки. Следовательно, длина l двоичной последовательности-символа должна быть равна степени двойки.

Рассмотрим код Хэмминга $C_H\left(\frac{q^m-1}{q-1}, \frac{q^m-1}{q-1}-m\right)$. Число N страниц памяти ЭСППЗУ, в которых будет размещаться код, в общем случае не будет совпадать с $\frac{q^m-1}{q-1}$. Но требуемый код C может быть построен путем применения к коду Хэмминга операции укорочения кода [2], заключающуюся в удалении из кода нескольких информационных символов. Защищенность кода C_H от ошибок при этом нисколько не уменьшается.Для возможности получения кода C из кода Хемминга C_H с помощью операции укорочения, параметр m кода Хемминга должен удовлетворять неравенству

$$\frac{q^m-1}{q-1} \geq N, \tag{5}$$

из которого может быть получена нижняя оценка для числа m страниц контроля:

$$m \geq \log_q\left(N(q-1)+1\right). \tag{6}$$

Чтобы показать, что код $C(N, N-m)$ получен из кода Хемминга $C_H\left(\frac{q^m-1}{q-1}, \frac{q^m-1}{q-1}-m\right)$, будем также обозначать его как $C_H(N, N-m)$.

В качестве примера, выпишем в таблицу возможные параметры кода $C_H\left(N, N-m\right)$ для страниц памяти размером 32 байта с минимальными значениями m, и относительный расход памяти ε для каждого кода. В качестве параметра N будем брать различные степени двойки как один из наиболее часто встречаемых вариантов в ЭСППЗУ:

Таблица 1

Некоторые коды $C_H\left(N, N-m\right)$ для 32-битных страниц памяти ЭСППЗУ и алфавита кодирования из двоичных последовательностей длины l.

Длина кода N	Длина последовательности-символа l (бит)	Мощность q алфавита кода	Число контрольных символов m	Относительный расход памяти ε
128	1	2	8	6.67%
128	2	4	5	4.06 %
128	4	16	3	2.4%
128	8	256	2 (гр. Синглтона)	1.59%
256	1	2	9	3.64%
256	2	4	5	1.99%
256	4	16	3	1.19%
256	8	256	2 (гр. Синглтона)	0.79%
512	1	2	10	1.99%
512	2	4	6	1.19%
512	4	16	4	0.79%
512	8	256	3	0.59%
512	16	65536	2	0.39%
1024	1	2	11	1.09%
1024	2	4	6	0.59%
1024	4	16	4	0.39%
1024	8	256	3	0.29%
1024	16	65536	2	0.2%

Как видно из таблицы 1, относительный расход памяти кодов вида $C_H\left(N, N-m\right)$ не велик, и уменьшается с увеличением длины N кодовых слов и мощности q алфавита кодирования. Более реалистичную оценку расхода памяти дает действительный относительный расход памяти $\tilde{\varepsilon}$, для оценки которого необходимо более детальное исследование работы кода. Найдем матрицу зависимостей кода.

Матрица зависимостей Ω кода Хемминга $C_H\left(\frac{q^m-1}{q-1}, \frac{q^m-1}{q-1}-m\right)$ может быть построена из проверочной матрицы кода Хемминга. При этом получается, что матрица Ω будет состоять из строк, являющихся всевозможными комбинациями единиц и нулей. При этом выполнены условия:

1. В каждой строке матрицы не менее двух единиц;
2. Всякая строка, представляющая из себя упорядоченный набор единиц и нулей, встречается в матрице ровно $(q-1)^{x-1}$ раз, где x число единиц.

Таким образом, все страницы данных для кода Хемминга $C_H\left(\frac{q^m-1}{q-1}, \frac{q^m-1}{q-1}-m\right)$ имеют от 2 до m зависимых страниц контроля. При этом число страниц данных, имеющих x зависимых страниц контроля всего $(q-1)^{x-1}C_m^x$, где C_m^x число сочетаний из m по x.

Каждая страница контроля имеет $q^{m-1}-1=\sum_{x=2}^{m}(q-1)^{x-1}C_{m-1}^{x-1}$ зависимых страниц данных. В случае укороченного кода число γ_i^c зависимых страниц данных для каждой *i*-ой страницы контроля будет меньше. Очевидно, наиболее рациональным при операции укорочения кода будет отбрасывать страницы, имеющие максимальное число зависимых страниц контроля. Также условимся делать выбор удаляемых страниц так, чтобы после их удаления разница между величинами γ_i^c не превышала 1 и страницы с меньшим значением γ_i^c при нумерации страниц данных шли первыми.

Таким образом, операция укорочения кода Хемминга будет однозначно заданной, а запись $C_H(N, N-m)$ будет обозначать вполне определенный код, получаемый из кода $C_H\left(\frac{q^m-1}{q-1}, \frac{q^m-1}{q-1}-m\right)$.При удалении из кода $C_H\left(\frac{q^m-1}{q-1}, \frac{q^m-1}{q-1}-m\right)$ не более $(q-1)^{m-1}$ страниц данных, для каждой страницы контроля число зависимостей уменьшится на число удаленных страниц. При удалении всех страниц данных, имеющих x зависимых страниц контроля величина γ_i^c уменьшится на $\frac{x}{m}(q-1)^{x-1}C_m^x$. Следующее утверждение дает нижнюю границу для значений γ_i^c кода $C_H(N, N-m)$.

Утверждение 1. *Если код* $C_H(N, N-m)$ *получается из кода Хемминга* $C_H\left(\frac{q^m-1}{q-1}, \frac{q^m-1}{q-1}-m\right)$ *операцией укорочения по страницам данных, имеющим максимальные значения* γ_i^d, *тодля величин* γ_j^c *выполнены неравенства:*

$$\gamma_j^c \geq \left\lceil \frac{2k}{m} \right\rceil, \tag{7}$$

$$\gamma_j^c \le q^{m-1} - 1. \qquad (8)$$

Доказательство.Общее число зависимостей между страницами данных и контроля в силу $\gamma_i^d \ge d - 1$ не меньше $2k$. Так как укорочение кода осуществляется так, чтобы разница между величинами γ_j^c не превышала 1, всякая величина γ_j^c будет не меньше $\left[\frac{2k}{m}\right]$.

Неравенство (8) получается из неравенства для кода Хемминга. При укорочении кода число зависимостей только уменьшится, а потому неравенство останется справедливым.

Неравенство (7) может быть переписано как $\gamma_j^c \ge \left[\frac{2}{\varepsilon}\right]$. Величина ε не должна превышать 20%. Следовательно, можно заключить, что для любого кода $C_H\left(N, N-m\right)$ должно быть выполнено неравенство: $\gamma_j^c \ge 10$.Можно показать, что для кода $C_H\left(N, N-m\right)$ величины γ_j^c могут быть сколь угодно уменьшены путем увеличения параметра m у исходного кода Хэмминга. При этом значение ε будет увеличиваться. Более интересным является тот факт, что во многих случаях при увеличении параметра m величины γ_j^c уменьшаются «быстрее», чем увеличивается значение ε.

Утверждение 2.*Обозначим средние величины* γ_j^c *кодов* $C_H\left(N, N-m-1\right)$ *и* $C_H\left(N, N-m\right)$ *как* $\gamma^c\left(m\right)$ *и* $\gamma^c\left(m+1\right)$, *соответственно. Если у кода* $C_H\left(N, N-m\right)$ *есть страницы данных, имеющие число зависимостей* $\gamma_i^d > 2$, *товыполнено неравенство:*

$$\gamma^c\left(m+1\right) < \frac{m}{m+1}\frac{N-m-1}{N-m}\gamma^c\left(m\right), \qquad (9)$$

иначе выполнено равенство:

$$\gamma^c\left(m+1\right) = \frac{m}{m+1}\frac{N-m-1}{N-m}\gamma^c\left(m\right). \qquad (10)$$

Доказательство.Сначала рассмотрим случай, когда для кода $C_H\left(N, N-m\right)$ величины $\gamma_i^d = 2$ для всех i. Число страниц данных, имеющих только 2 зависимых страницы контроля, для кода $C_H\left(\frac{q^m-1}{q-1}, \frac{q^m-1}{q-1} - m\right)$ равно $\left(q-1\right)C_m^2$. Следовательно, для производного кода $C_H\left(N, N-m\right)$ должно быть выполнено неравенство $N - m \le \left(q-1\right)C_m^2$. Но тогда тем более выполнено неравенство $N - m - 1 \le \left(q-1\right)C_{m+1}^2$, означающее то, что в коде $C_H\left(N, N-m-1\right)$

для всех страниц данных $\gamma_i^d = 2$. Значения $\gamma^c(m)$ и $\gamma^c(m+1)$ равны $\frac{2(N-m)}{m}$ и $\frac{2(N-m-1)}{m+1}$, соответственно. Подстановкой этих значений в равенство (10) легко убедиться в его справедливости.

В случае если для кода $C_H(N, N-m)$ есть страницы данных с $\gamma_i^d > 2$, общее число страниц данных может быть представлено как сумма $N-m = \sum_{x=2}^{t}(q-1)^{x-1} C_m^x + f$, где t - некоторое число меньше m, а f - положительный остаток не больше чем $(q-1)^t C_m^{t+1}$. Для кода $C_H(N, N-m-1)$ возможно аналогичное представление $N-m-1 = \sum_{x=2}^{s}(q-1)^{x-1} C_{m+1}^x + g$. Так как для любых $m > x > 2$ справедливо $C_m^x < C_{m+1}^x$, получаем, что либо $s < t$, либо $s = t$ и $g < f$. Сопоставляя первые $N-m-1$ страниц данных кода $C_H(N, N-m)$ страницам данных кода $C_H(N, N-m-1)$, получаем, что число зависимостей для страниц данных второго кода меньше числа зависимостей соответствующих страниц данных первого кода. Кроме того, в коде $C_H(N, N-m)$ последняя страница данных, имеющая $a > 2$ зависимостей, оказывается «лишней». Поэтому общее число зависимостей кода $C_H(N, N-m-1)$ как минимум на 3 меньше числа зависимостей кода $C_H(N, N-m)$. Получаем: $\sum_{j=0}^{m}\gamma_j^c(m+1) \le \sum_{j=0}^{m-1}\gamma_j^c(m) - a$, где a - максимальное число зависимостей для страниц данных кода $C_H(N, N-m)$. Осуществим преобразования с неравенством:

$$(m+1)\gamma^c(m+1) \le m\,\gamma^c(m) - a,$$

$$\gamma^c(m+1) \le \frac{m}{m+1}\gamma^c(m) - \frac{a}{m+1}.$$

Преобразуем неравенство (9): $\gamma^c(m+1) < \frac{m}{m+1}\gamma^c(m) - \frac{m}{m+1}\frac{1}{N-m}\gamma^c(m)$.

Для доказательства достаточно показать, что $\frac{m\,\gamma^c(m)}{(m+1)(N-m)} < \frac{a}{m+1}$. Заметим, что $\frac{m\,\gamma^c(m)}{N-m}$ не что иное, как среднее число зависимостей страниц данных кода $C_H(N, N-m)$. Это число обязательно меньше a. Справедливость неравенства (9) доказана.

Для оценки эффективности кода Хемминга нам нужно произвести оценку значений вектора P^c при различных значениях P^d. Будем считать что нумерация страниц данных строится методом исчерпывания: в качестве *i*-ой страницы данных выбирается страница с максимальным значением вероятности перезаписи среди еще непронумерованных страниц данных. При такой нумерации страницы с наибольшей вероятностью перезаписи будут иметь наименьшее число зависимостей.

Пусть ${}_1P^d = (p_1,\ldots,p_1)$ и ${}_2P^d = (p_2,\ldots,p_2)$ векторы вероятностей с вероятностями p_1 и p_2 такими, что $p_1 \le p_2$. Если ${}_1P^d$ и ${}_2P^d$ два вектора вероятностей для страниц данных такие, что ${}_1P^d \le {}_2P^d$. Тогда для соответствующих векторов вероятности ${}_1P^c$ и ${}_2P^c$ для страниц контроля справедливо неравенство ${}_1P^c \le {}_2P^c$. Следовательно, если ${}_1P^d \le P^d \le {}_2P^d$, то для соответствующих векторов вероятностей для страниц контроля выполнено неравенство ${}_1P^c \le P^c \le {}_2P^c$. Это означает, что зависимость P^c от P^d в значительной степени может быть изучена при исследовании более простого частного случая, когда все вероятности P_i^d равны некоторому значению p.Элементы вектора $1^k - P^d$ являются вероятностями того, что соответствующие страницы данных не изменяются при процедуре записи. Аналогично, элементы вектора $1^m - P^c$ являются вероятностями того, что не изменяется ни одна из страниц данных, зависимых от соответствующей страницы контроля. Так как мы считаем события перезаписи страниц данных независимыми, вероятность того, что некоторыйнабор страниц останется неизменным, равна произведению соответствующих вероятностей $1-P_i^d$. Иными словами вероятность $1-P_i^c$ равна произведению вероятностей неизменности зависимых с ней страниц данных. То есть, если код имеет матрицу зависимостей Ω,тогда для всякого вектора вероятностей P^d справедливо неравенство: $\ln\left(1^m - P^c\right) = \ln\left(1^k - P^d\right)\Omega$. В соответствии с равенством, все вероятности P_i^c будут иметь значения:

$$1-(1-p)^{\gamma_i^c}. \qquad (11)$$

Сумму элементов вектора вероятностей $\tilde{P}^c$ можно интерпретировать как математическое ожидание E_c числа перезаписываемых страниц контроля. Таким же образом определим величину математического ожидания E_d числа записываемых страниц данных:

$$E_c = \sum_{i=0}^{m-1} \tilde{P}_i^c = \delta \sum_{i=0}^{m-1} P_i^c, \qquad (12)$$

$$E_d = \sum_{i=0}^{k-1} P_i^d \,. \qquad (13)$$

Заметим, что величина E_d всегда должна быть больше или равна 1, что накладывает определенные ограничения на вектор вероятностей P^d. В силу замечаний выше, для любого кода, исправляющего не менее одной ошибки, величина E_c при этом будет не меньше 2.

Определим для помехоустойчивого кода, задаваемого на страницах памяти, функцию $R\left(P^d\right)$ от вектора вероятности P^d:

$$R\left(P^d\right) \stackrel{def}{=} \frac{E_c\left(P^d\right)}{E_d\left(P^d\right)}. \qquad (14)$$

Вычислим по формулам (12), (13), (14) математические ожидания числа перезаписываемых страниц данных и контроля, и функцию $R(p)$:

$$E_d = (N-m)\,p\,,$$

$$E_c = \delta\left(m - \sum_{i=0}^{m-1}(1-p)^{\gamma_i^c}\right) = \left(1-\frac{1}{2^L}\right)\left(m - \sum_{i=0}^{m-1}(1-p)^{\gamma_i^c}\right),$$

$$R(p) = \left(1-\frac{1}{2^L}\right)\frac{m - \sum_{i=0}^{m-1}(1-p)^{\gamma_i^c}}{(N-m)\,p},$$

где L - размер страницы памяти в битах.

Построим графики функции $R(p)$ при различных параметрах, рассмотренных в таблице 1. Для этого предварительно найдем недостающие параметры γ_i^c для кодов таблицы с $N=128$ и $N=256$.

В качестве примера вычислим γ_i^c для первого кода. Двоичный код Хемминга при числе контрольных символов $m=8$ имеет длину кодовых слов $2^8-1=255$ символов. Таким образом, для получения нужного кода должна быть применена операция укорочения и удалено $255-128=127$ символов, отвечающих страницам данных.

Так как $(2-1)^8 C_8^8 + (2-1)^7 C_8^7 + (2-1)^6 C_8^6 + (2-1)^5 C_8^5 = 93 < 127$, могут быть удалены все страницы данных, имеющие более 4 зависимых страниц контроля. Число страниц данных, имеющих четыре зависимых страницы контроля, равно $(2-1)^4 C_8^4 = 70$, и из них необходимо удалить 34 страницы данных. Удалять такие страниц будем попарно, объединяя в пары страницы, имеющих вместе в качестве зависимых все 8 страниц контроля. Такое

разбиение однозначно, и при удалении каждой пары все параметры γ_i^c уменьшается на единицу. Отсюда получаем, что для любой страницы контроля

$$\gamma_i^c = 2^7 - 1 - \left(\frac{8}{8}(2-1)^8 C_8^8 + \frac{7}{8}(2-1)^7 C_8^7 + \frac{6}{8}(2-1)^6 C_8^6 + \frac{5}{8}(2-1)^5 C_8^5 \right) - 17 = 46.$$

В таблице 2 приведены значения величин γ_i^c для кодов таблицы 1 с $N = 128$ и $N = 256$:

Таблица 2.

Число зависимостей γ_i^c страниц контроля от страниц данных для различных кодов Хэмминга.

Длина кода N	Мощность q алфавита кода	Число контрольных символов m	Зависимости страниц контроля γ_i^c
128	2	8	$\gamma^c = 46$
128	4	5	$\gamma_0^c = \gamma_1^c = \gamma_2^c = 69$, $\gamma_3^c = \gamma_4^c = 70$
128	16	3	$\gamma_0^c = 116$, $\gamma_1^c = \gamma_2^c = 117$
128	256	2	$\gamma^c = 126$
256	2	9	$\gamma_0^c = \gamma_1^c = \gamma_2^c = \gamma_3^c = 93$, $\gamma_4^c = \gamma_5^c = \gamma_6^c = \gamma_7^c = \gamma_8^c = 94$
256	4	5	$\gamma_0^c = 170$, $\gamma_1^c = \gamma_2^c = \gamma_3^c = \gamma_4^c = 171$
256	16	3	$\gamma^c = 238$
256	256	2	$\gamma^c = 254$

Введем условные обозначения функции $R(p)$ для кодов таблицы 1. Будем записывать функции как $R1(p), R2(p), R3(p), R4(p)$ для четырех кодов с $N = 128$, и $R5(p), R6(p), R7(p), R8(p)$ для четырех кодов с $N = 256$. На рисунках 1 и 2 представлены графики названных функций:

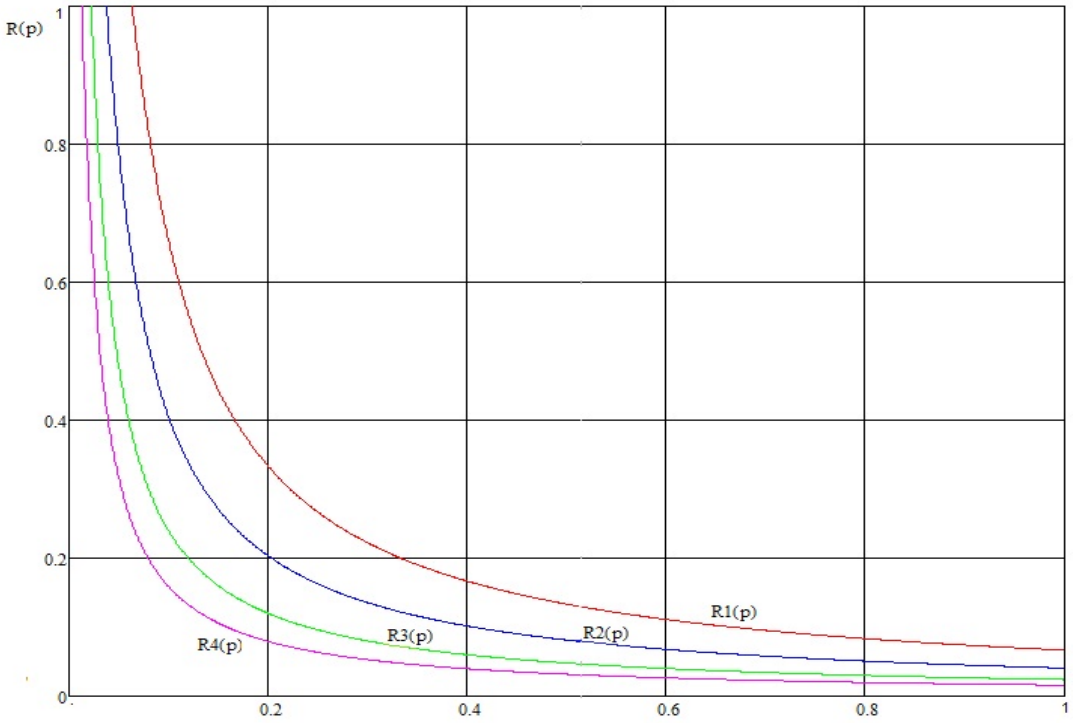

Рис. 1. Графики функций $R1(p), R2(p), R3(p), R4(p)$

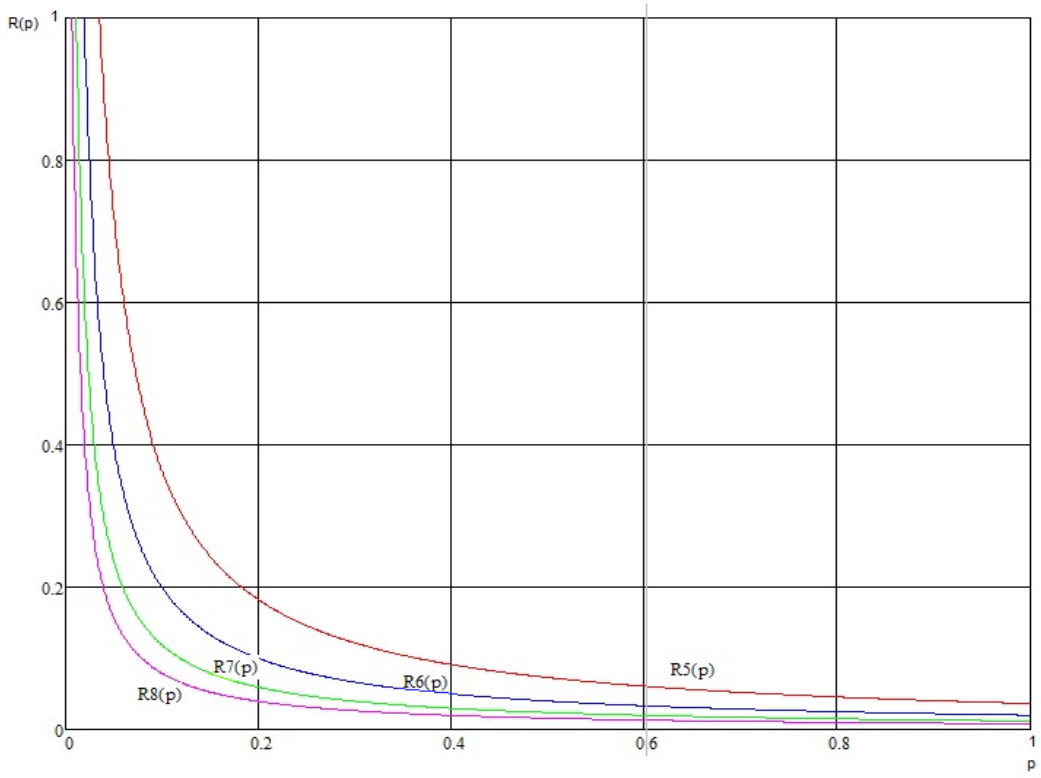

Рис. 2.Графики функций $R5(p), R6(p), R7(p), R8(p)$

Как видно на представленных графиках, повышение мощности алфавита кодирования приводит к уменьшению отношения числа перезаписываемых страниц контроля к числу

перезаписываемых страниц данных. Отсюда уже можно сделать вывод о повышении эффективности кода за счет повышения мощности алфавита кодирования.

Более точную информацию об эффективности кода можно получить путем изучения векторной величины ρ (1) и действительного числа страниц контроля $\tilde{m}$ (2), необходимого для обеспечения минимальной целостности страниц контроля. На рисунке 3 приведены графики функций $\tilde{m}(p)$ для кодов Хемминга из таблицы 1 с длиной кодовых слов $N = 128$.

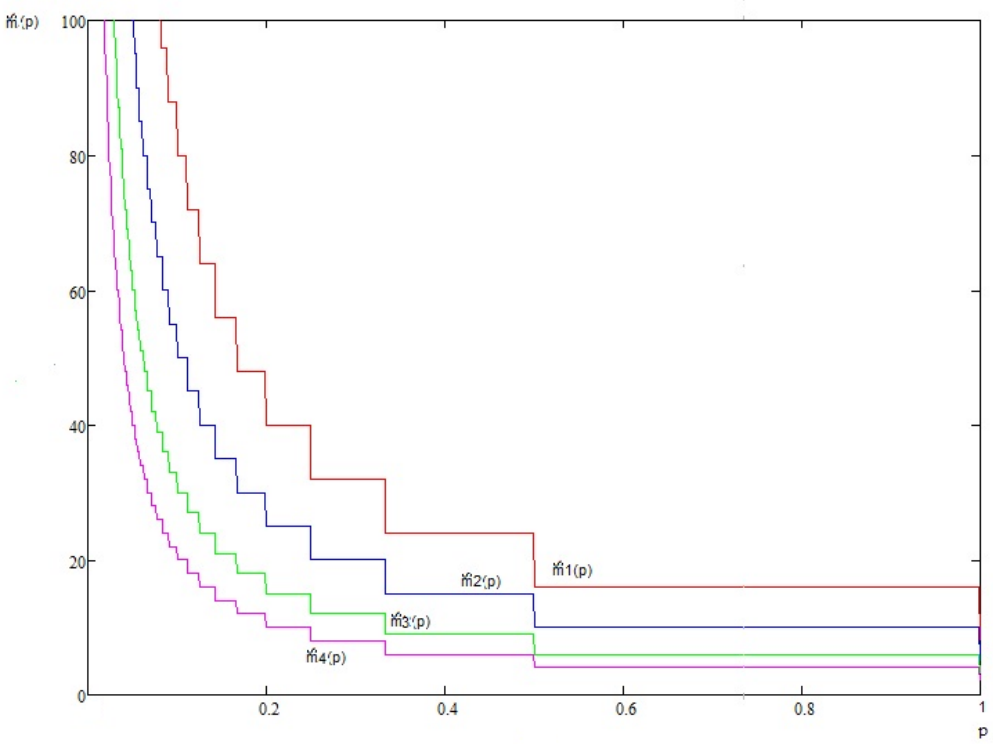

Рис. 3. Графики функций $\tilde{m}(p)$ для кодов Хемминга таблицы 1 с $N = 128$

На рисунке 4 приведены графики функций $\tilde{m}(p)$ для кодов Хемминга из таблицы t1 с длиной кодовых слов $N = 256$:

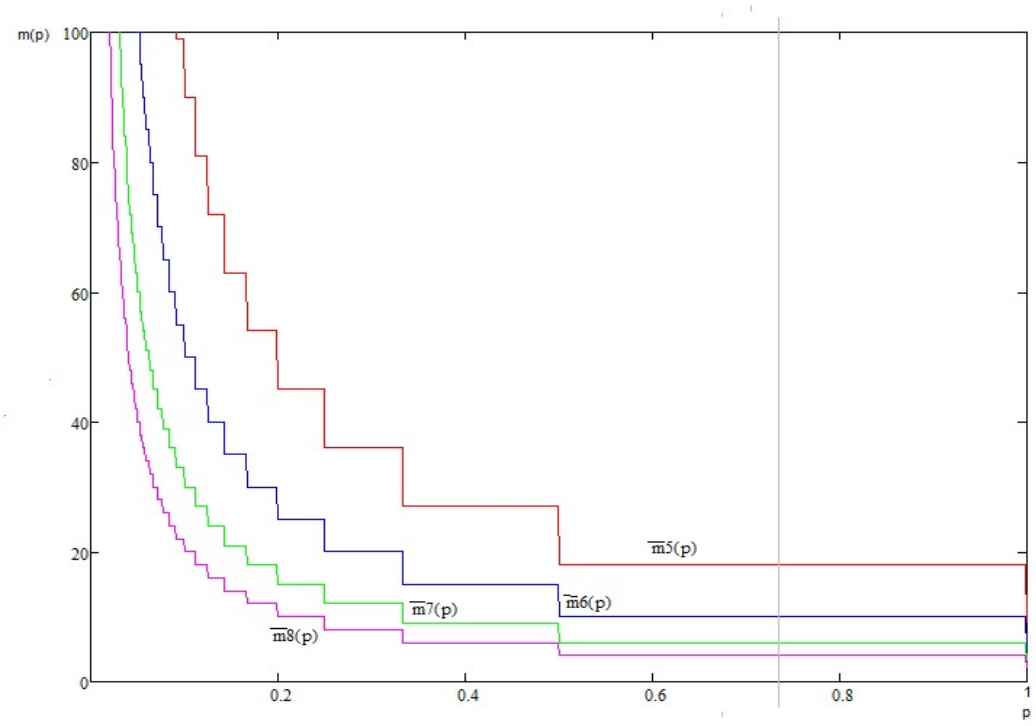

Рис. 4. Графики функций $\tilde{m}(p)$ для кодов Хемминга таблицы 1 с $N = 256$

На представленных графиках можно видеть, что с понижением вероятности перезаписи страниц данных сильно повышается требуемое число страниц контроля. Это вызвано тем, что при уменьшении вероятности p математическое ожидание числа перезаписываемых страниц данных может сильно снизиться, в то время как страниц контроля такое снижение если и будет, то не значительным.

Из графиков можно сделать вывод, что для рассматриваемых кодов приемлемый расход памяти на кодирование начинается при вероятностях не менее 5-20%. Более наглядно это можно увидеть на графиках относительного расхода памяти $\tilde{\varepsilon}$, приведенных на рисунке 5:

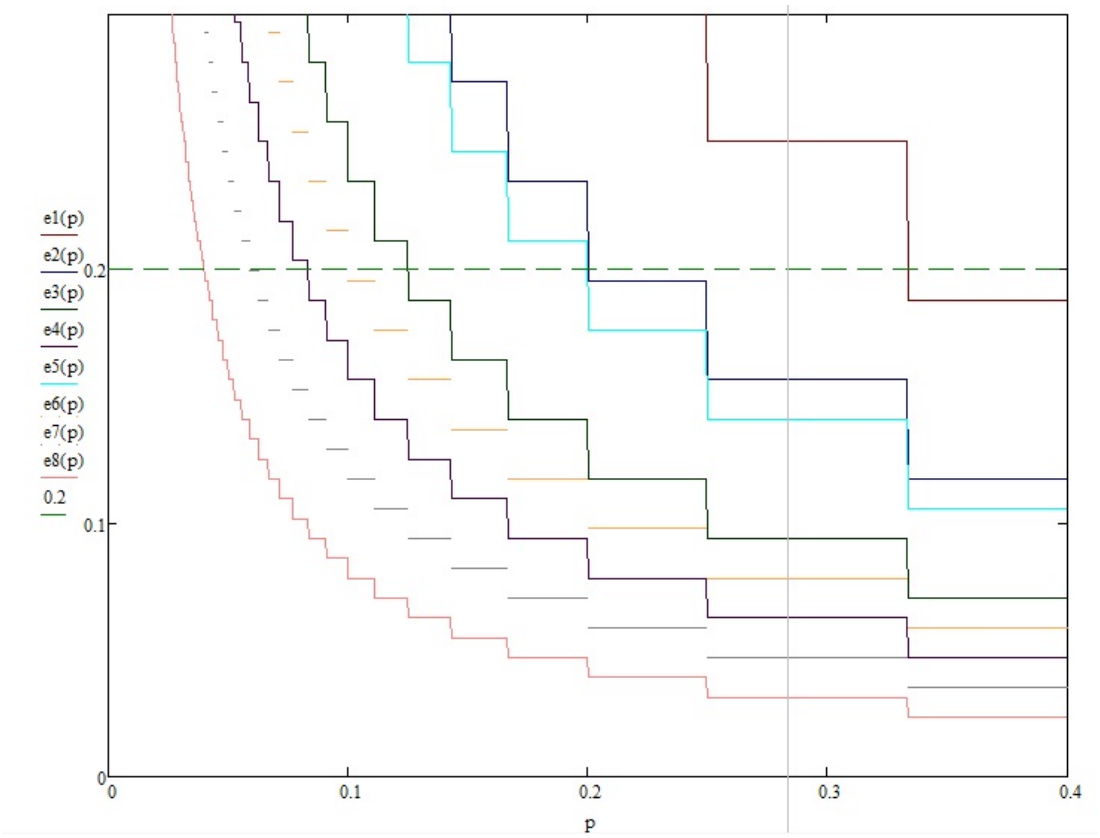

Рис. 5. Действительный относительный расход памяти для кодов Хемминга таблицы t1 с $N = 128$ и $N = 256$

Из графиков видно, что коды, имеющие большую длину кодовых слов, более эффективно расходуют память. Пунктирной линией на графике обозначен предельный уровень 20% расхода памяти, установленный нами в предыдущих разделах. Даже самый эффективный из кодов таблицы 1 имеет приемлемый расход памяти, только начиная с вероятности перезаписи страницы данных около 5%, т.е. при условии, что при каждой процедуре записи в среднем будет перезаписываться не менее 6 страниц данных.

Рассмотренные коды $C_H\left(N, N-m\right)$ являются кодами с минимальным значением параметра m в соответствии с формулой (6). В соответствии с утверждением 2 величины зависимостей γ_j^c уменьшаются при увеличении параметра m. Данное обстоятельство может служить в пользу большей эффективности кода с большим значением параметра m. В качестве примера исследуем зависимость эффективности кода от m для двоичных кодов вида $C_H\left(128,128-m\right)$. В таблице 3 приведены значения величин γ_j^c для таких кодов.

Таблица 3.

Число зависимостей γ_i^c страниц контроля от страниц данных для некоторых двоичных кодов $C_H\left(128,128-m\right)$

Длина кода N	Число контрольных символов m	Зависимости страниц контроля γ_i^c	Относительный расход памяти ε
128	8	$\gamma^c = 46$	6.25%
128	9	$\gamma_0^c = \gamma_1^c = \gamma_2^c = 35$, $\gamma_3^c = \ldots = \gamma_8^c = 36$	7.03%
128	10	$\gamma_0^c = 30$, $\gamma_1^c = \ldots = \gamma_9^c = 31$	7.81%
128	11	$\gamma_0^c = 26$, $\gamma_1^c = \ldots = \gamma_{10}^c = 27$	8.59%
128	12	$\gamma_0^c = \ldots = \gamma_5^c = 23$, $\gamma_6^c = \ldots = \gamma_{11}^c = 24$	9.38%
128	13	$\gamma_0^c = \ldots = \gamma_5^c = 20$, $\gamma_6^c = \ldots = \gamma_{12}^c = 21$	10.16%
128	14	$\gamma_0^c = 17$, $\gamma_1^c = \ldots = \gamma_{13}^c = 18$	10.93%
128	15	$\gamma_0^c = \ldots = \gamma_5^c = 15$, $\gamma_6^c = \ldots = \gamma_{14}^c = 16$	11.71%
128	16	$\gamma_0^c = \ldots = \gamma_{11}^c = 14$, $\gamma_2^c = \ldots = \gamma_{15}^c = 15$	12.5%

Обозначим функции $R(p)$ для кодов $C_H(128,128-m)$ из таблицы 2.3 как $R_1(p), R_2(p), \ldots, R_9(p)$.

На рисунке 6 изображены графики указанных функций в области, где их поведение представляет наибольший интерес.

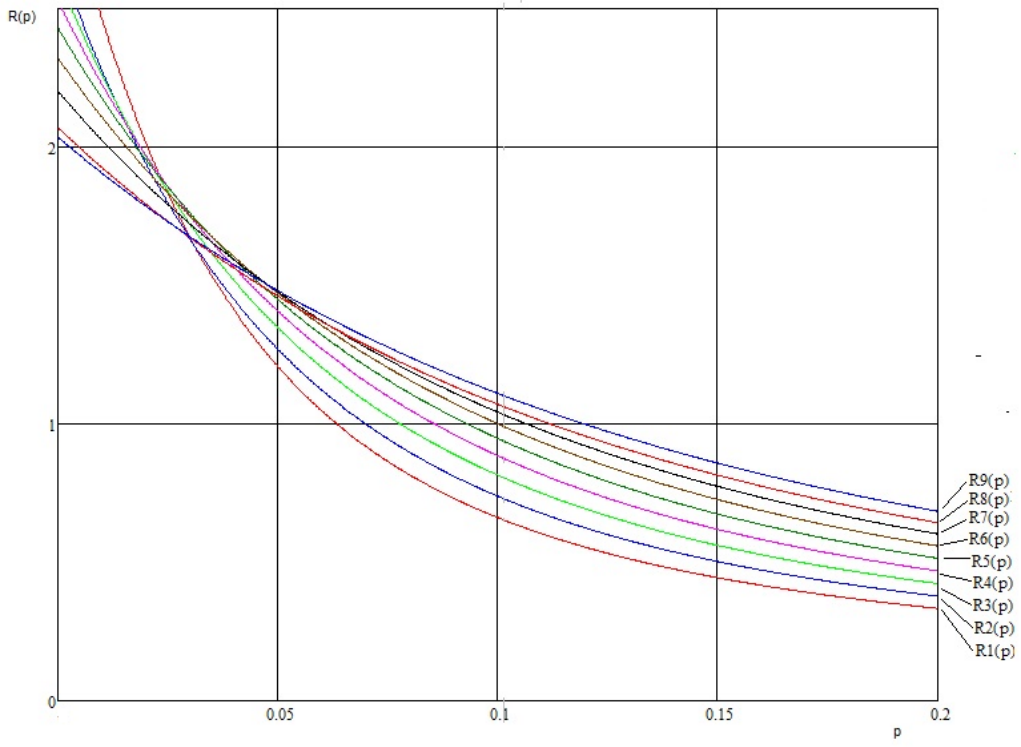

Рис. 6. Графики функций $R_1(p), R_2(p), \ldots, R_9(p)$

Как можно заключить из представленных графиков функций, при малых значениях вероятности p коды $C_H\left(128,128-m\right)$ с большим значением параметра m оказываются более эффективными. Но при вероятностях больших некоторого значения ситуация кардинально меняется, более эффективными оказываются коды с меньшим значением параметра m.

Организация хранения данных в ЭСППЗУ на основе кодаРида-Соломона

Коды Рида-Соломона $C_{RS}\left(q-1,q-1-2t\right)$ в отличие от кодов Хэмминга в общем случае способны исправлять более одной ошибки и всегда удовлетворяют границе Синглтона. В случае кодов, исправляющих $t=1$ ошибок, также можно строить кодовые слова, беря по одному символу из каждой страницы данных.

В случае произвольного значения параметра t возможно построение кодовых слов таким образом, чтобы из каждой страницы данных бралось t символов, а $2t$ контрольных символов таким же образом размещались по двум страницам контроля. Так как код исправляет t ошибок, любой сбой одной страницы памяти будет скорректирован. Подобное использование кодов, исправляющих более одной ошибки, имеет неоспоримое преимущество в плане исправления ошибок, несвязанных со сбоем страниц памяти. В случае возникновения более чем одной ошибки в пределах одного кодового слова, такой код справится с ней в отличие от кода, исправляющего одну ошибку.

При любом значении параметра t код будет иметь две страницы контроля. Причем, при изменении любого информационного символа будет следовать обязательное изменение всех $2t$ контрольных символов. Следовательно, каждая страница контроля зависит от всех страниц данных, - величины $\gamma_i^c=q-1-2t$ для всех страниц контроля.

Также как и в случае кодов Хемминга, длина кодовых слов может не совпадать с длиной кода Рида-Соломона, и требуемый код может быть построен путем операции укорочения. Полученный код будем обозначать $C_{RS}\left(N,N-m\right)$, где $m=2t$. Производный код $C_{RS}\left(q-1,q-1-2t\right)$ очевидно должен удовлетворять условию:

$$q-1\le Nt. \qquad (15)$$

Заметим, что также как и для кодов Хемминга, $q=2^l$, а параметр t должен быть таким, чтобы число q-символов одной страницы памяти делилось на t. Так как размер страницы памяти равен обычно степени двойки, параметр t сам должен быть степенью двойки.

Как и в предыдущем разделе для кодов Рида-Соломона возможно изучение расхода памяти. Но в отличие от кодов Хемминга, здесь ожидаемый расход памяти для всех кодов будет одним и тем же, так как число страниц контроля всегда равно двум, и они зависят от всех страниц данных. Более того коды Рида-Соломона будут расходовать память также как код $C_H\left(N, N-M\right)$ из предыдущего раздела.

Возможности применения продольного контроля избыточности

В предыдущих разделах рассмотрены возможности исправления сбоя страниц памяти с помощью кодов, обнаруживающих не менее одной ошибки. Все эти коды имеют минимальный вес кодовых слов $d \geq 3$. Коды с минимальным весом $d = 2$ способны только обнаруживать единичную ошибку, но не исправлять ее. Рассмотрим гипотетическую ситуацию, когда аппаратно-программные возможности SMART-карты или RFID-метки позволяют обнаруживать сбой, а также номер вышедшей из строя страницы. Наличие такой дополнительной информации позволяет использовать код с $d = 2$ как код, исправляющий одну ошибку.

Простейшим кодом с минимальным весом $d = 2$ является код с битом четности. Применительно к организации кодовых слов, используемой нами, данный код оказывается ничем иным как кодом продольного контроля избыточности LRC. Данный код будет иметь одну страницу контроля равную сумме по модулю два всех страниц данных. Следовательно, текущий расход памяти на контрольную информацию для этого кода меньше в два и более раз, чем для кодов с весом $d \geq 3$.

По сравнению с кодами с весом $d = 3$ число циклов перезаписи в страницах контроля также снизится примерно в два раза. Соответственно в такое же число раз понизится число страниц контроля, необходимых для обеспечения сохранности данных до первого выхода из строя страницы данных.

Алгоритм кодирования для данного кода является достаточно простым и компактным, и его машинный код легко может быть размещен в электронных интегральных схемах МКПИ.

Алгоритм восстановления потерянных данных может быть реализован следующим образом:

1. аппаратно-программными средствами МКПИобнаруживается сбой *i*-ой страницы данных;
2. потерянная страница данных вычисляется как сумма остальных страниц данных и страницы контроля.

Так как сбой страницы памяти может быть обнаружен без применения процедуры декодирования, она осуществляется только после обнаружения ошибки. Сама процедура декодирования по вычислительной сложности совпадает с процедурой кодирования. Код ее алгоритма также является достаточно компактным и может быть размещен в памяти SMART-карты либо RFID-метки.

Заключение

В статье рассмотрены реализации помехоустойчивого кодирования, основанные на кодах Хемминга, Рида-Соломона, продольном контроле избыточности. Анализ эффективности повышения надежности путем использования кода Хемминга для данных, хранящихся в ЭСППЗУ, показал, что повышение мощности алфавита кодирования приводит к уменьшению отношения числа перезаписываемых страниц контроля к числу перезаписываемых страниц данных, в свою очередь, повышение вероятности перезаписи страниц данных приводит к сильному росту требуемых число страниц контроля. Таким образом, коды Хемминга, имеющие большую длину кодовых слов, более эффективно расходуют память. Окончательный выбор алгоритма кодирования зависит от производительности МКПИ, степенью защищенности хранимых данных и степенью избыточности, определяемой объемом памяти.

Библиографический список

[1]. WolfgangR., WolfgangE.SmartCardHandbook. WILEY, 2011. – 1025 стр.

[2] Блейхут Р. Теория и практика кодов, контролирующих ошибки. М.: Мир, 1986. – 576 стр.

Сведения об авторах

Андрианова Елена Гельевна, доцент Московского государственного технического университета радиотехники, электроники и автоматики (МГТУ МИРЭА), к.т.н.;
тел.: (903)7748505, (495)4338533, e-mail: andrianova@mirea.ru

Крюков Дмитрий Алексеевич, аспирант Московского государственного технического университета радиотехники, электроники и автоматики (МГТУ МИРЭА),
тел.: (985) 4444085; e-mail: dm.bk@bk.ru

Петров Андрей Борисович, декан Московского государственного технического университета радиотехники, электроники и автоматики (МГТУ МИРЭА), д.т.н., профессор, тел.: (901)5602900, (495)4338533, e-mail: petrov@mirea.ru