

УДК 681.327.636

И. И. Коршевер, П. А. Полозков

(Новосибирск)

АРХИВНЫЙ НАКОПИТЕЛЬ ЦИФРОВЫХ ДАННЫХ
НА ВИДЕОМАГНИТОФОНЕ

Рассмотрен комплекс программных и аппаратных средств архивной памяти на основе бытового магнитофона. Предложены способы повышения надежности хранения информации путем жесткой синхронизации данных и двухкаскадного помехозащищенного линейного блокового кодирования. Приведены результаты статистического анализа ошибок, полученные при испытаниях опытного образца архивного накопителя.

Постановка задачи. Для архивирования больших объемов цифровых данных в современных системах используют магнитные, оптические и магнитооптические внешние запоминающие устройства (ВЗУ). Магнитные ВЗУ характеризуются малой плотностью записи и, как следствие, большой удельной стоимостью хранения бита информации. Стоимость хранения для ВЗУ на оптических дисках в 100 раз ниже, чем для магнитных ВЗУ, но из-за невозможности повторной перезаписи на оптических дисках оба эти типа устройств обычно используются совместно. Сравнительные характеристики различных накопителей цифровых данных [1] приведены в табл. 1. Разрабатываемые в настоящее время магнитооптические накопители предоставляют возможность многократной перезаписи, но пока что их производство дорого.

Таблица 1

Характеристика	Оптические ЗУ	Магнитные ЗУ		
	Оптический диск	НМД	НМЛ	НМЛ повышенной плотности
Емкость, Мбайт	1000—4000	2500	180	2000
Плотность записи, Мбит/см ²	100	10	0,2	2
Поток записи/считывания, Мбайт/с	0,1—5	0,1—3	1	3
Исходная частота ошибок на 1 бит	10 ⁻⁴	10 ⁻⁶	10 ⁻⁶ —10 ⁻⁷	10 ⁻⁵ —10 ⁻⁶
Долговечность, лет	10—40	2—3	2—5	1—2
Стоимость хранения, цент/бит	10 ⁻⁸ —10 ⁻⁹	10 ⁻⁴ —10 ⁻⁵	2 · 10 ⁻⁶	5 · 10 ⁻⁶
Среднее время доступа, мс	100—500	10—30	30000	5000

Таблица 2

Производитель	Продукт	Поток, Мбайт/с	Стоимость, доллар/Мбайт	Частота появления ошибок	Примечание
"Amrex"	DST 600	15—20	—	10^{-12}	Одноролочное устройство емкостью 25—165 Гбайт
"E-systems"	EMASS-1300	15—20	0,04	10^{-13}	—
"Metrum"	RSS-600	1,4	0,08	—	600 кассет типа VHS по 10,4 Гбайт на каждой
"Storage Tek"	4400	22—36	0,3	—	16 модулей по 6000 лент по 200 Мбайт на каждой

По всем этим причинам для записи непрерывного потока данных и их долговременного хранения до сих пор по-прежнему, как правило, используют цифровые магнитофоны и архивные накопители, имеющие удельную стоимость хранения порядка $1,25 \cdot 10^{-8}$ долларов/бит. Основные параметры некоторых типов накопителей на магнитной ленте приведены в табл. 2. В последнее время исключительное распространение получают устройства, работающие с магнитной лентой в технологии HELICAN-SCAN, являющейся ныне стандартом для магнитных архивных систем высокой плотности (наиболее известными изделиями данного типа являются накопители фирмы EXABYTE (США)). В соответствии с этой технологией запись узких дорожек на ленту производится по диагонали, в результате чего плотность записи достигает 30—100 бит/дюйм².

Описываемая здесь разработка представляет собой комплекс программных и аппаратных средств, позволяющих использовать в качестве архивной памяти бытовой видеоманитофон (именно в нем и была впервые применена и применяется поныне технология HELICAN-SCAN), работающий в режиме цифрового магнитофона, а в качестве магнитного носителя — видеокассеты в стандарте VHS.

Удельная стоимость хранения информации того же порядка, что и для оптических дисков ($2,5 \cdot 10^{-10}$ долларов/бит). Однако низкое качество носителя информации и низкая надежность бытового устройства препятствуют получению приемлемой достоверности считываемых данных (10^{-9} — 10^{-12} ош./бит).

В данной статье для увеличения времени хранения и достоверности считываемой информации при минимальной избыточности цифровых данных на видеоленте предлагается ряд аппаратно-программных решений, суть которых — жесткая синхронизация данных и 2-каскадное помехозащитное кодирование кодом Рида — Соломона (R—S-код) с совокупной избыточностью не более 15 %. Приводятся результаты статистического анализа ошибок при считывании цифровых данных и результаты испытания опытного образца накопителя и его технико-экономические показатели.

Синхронизация цифровых данных. Проблема синхронизации регистрируемых данных в накопителях на магнитном носителе рассмотрена в [2, 3]. В данном проекте для самосинхронизации и сведения к нулю постоянной составляющей регистрируемых данных были использованы средства, заложенные в стандартный телевизионный сигнал чересстрочной развертки [4], генерируемый системой управления бытовым видеоманитофоном, в котором предусмотрена большая избыточность для жесткой строчной синхронизации и привязки уровня сигнала.

Более жесткие требования для синхронизации предъявляются к частоте регистрируемых импульсов, размещаемых внутри периода строчной развертки. Для полосы пропускания видеотракта магнитофона 2—4 МГц и частоты строчной развертки 15 кГц количество бит в каждой строке выбираем равным 128. Тогда погрешность подстройки фазы при считывании не должна превышать 0,2 % от периода синхроимпульсов. Для получения такой точности была использована система автоподстройки частоты и фазы (ФАПЧ), где в петле обратной связи авторегулирования выделяется первая производная разности фаз без потери накопленной интегратором постоянной составляющей. При этом сигнал расстройки частоты стробирующего генератора пропорционален величине $A + (A - B)/2$, где A — текущее значение разности фаз входного и подстраиваемого синхроимпульсов, а B — предыдущая разность фаз.

Причины и характер появления испорченных данных. Исследование проводилось с помощью специальной программы, написанной на языке С для универсального 32-разрядного сигнального микропроцессора TMS320C30, использованного в макете накопителя [5]. На рис. 1 изображена структура сегментной наклонно-строчной записи двумя головками на видеоленте [6], а также возможные причины и характер появления испорченных данных по результатам этих исследований.

Ошибки при считывании вызваны тремя основными причинами: 1) локальные дефекты носителя, возникающие при изготовлении ленты, продольные царапины и поперечные замятия, появляющиеся при эксплуатации; 2) загрязнение поверхности барабана магнитной головки видеоманитофона в результате осыпания магнитного слоя ленты; 3) возможный срыв синхронизации данных вследствие первых двух причин или в результате колебаний скорости движения ленты, возникающих вследствие длительного засорения магнитных головок и ненормальной реакции автоматической подстройки скорости движения ленты в видеоманитоне, а также при торможении ленты в видеокассете.

Запись и считывание псевдослучайных последовательностей проводились на различных видеолентах и с помощью двух видеоманитонов марок "Akai" и "Sanyo". Обозначенные на рис. 1 всплески испорченных данных, возникающие в результате всех указанных причин, кроме засорения головок и длительной потери их контакта с лентой, достаточно равномерно распределяются между сегментами видеозаписи, совпадающими с полукадрами телевизионного поля и простирающимися от края к краю видеоленты под острым углом. Для исправления пакетных ошибок указанного происхождения вполне достаточно организовать коды, блоки которых будут совпадать с полукадрами видеосигнала и которые будут способны исправить несколько пакетов ошибок, составляющих малый процент всего кода (так называемые внутриблочные коды). Размер блока в стандартном полукадре, содержащем 312 строк (на

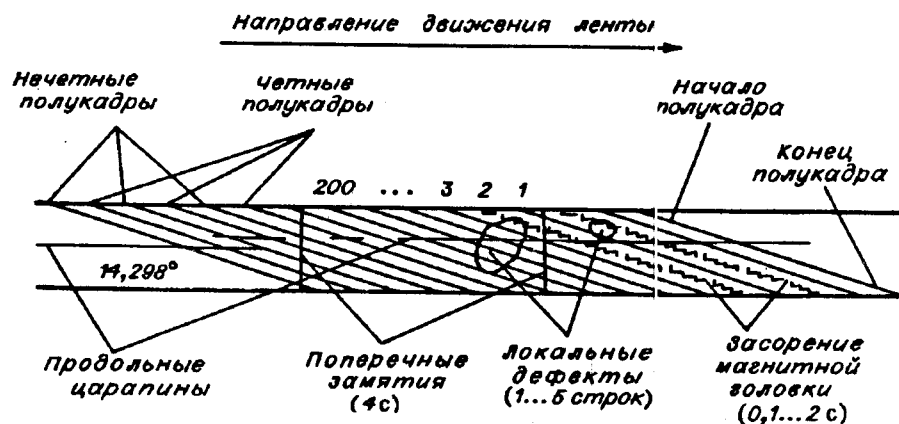


Рис. 1. Сегментная наклонно-строчная запись двумя головками

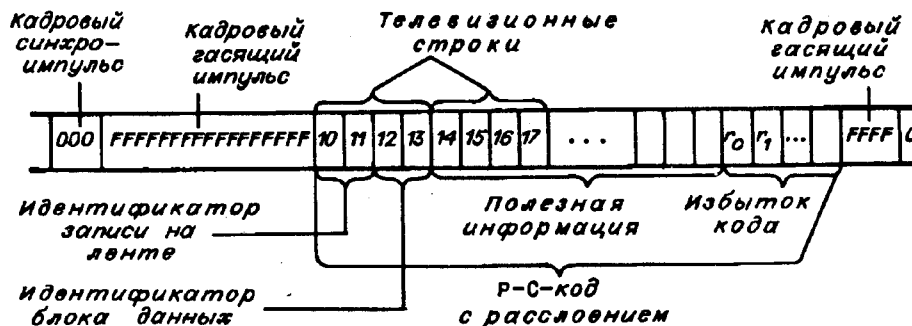


Рис. 2. Структура блока данных на видеоленте

информативную часть изображения остается меньше строк), не превышает $312 \cdot 128 = 4992$ байта цифровых данных при выбранном объеме информации в строке 128 бит.

Организация блоков данных и блоковых кодов, исправляющих ошибки. Для исправляющих пакеты ошибок кодов Рида — Соломона мультипликативный порядок алфавита символов в кодовом слове кратен длине кода [1]. Наиболее удобным для построения кодера/декодера на универсальном процессоре является ряд алфавитов порядка 2^m , где m должно делить нацело длину кода, а также для упрощения процесса упаковки/распаковки данных желательно, чтобы $m = 2^p$, где $p = 1, \dots, 5$ для микропроцессора с разрядностью 32 бита. Однако только для $p = 1, 2$ и 3 длина кода меньше предполагаемой длины блока — 4800 байт. Для $p = 3$ строим P—C-код с расслоением на 16 неукороченных кодовых слов, каждое из которых состоит из 255- и 8-битных символов ($4096 < 4992$). Эти ограничения, между прочим, исключают возможность построения кода в кольце целых чисел.

В качестве пилот-сигналов, вводимых обычно в начале каждого блока при записи на магнитный носитель и образующих дополнительную избыточность информации, целесообразно использовать полукадровые уравнивающие импульсы полного телевизионного кадра. Благодаря этому блоки данных на ленте, защищенные кодами и разделенные полукадровыми импульсами, совпадают с сегментами наклонно-строчной записи на видеоленте. Структура блока данных приведена на рис. 2. Внутри каждого блока первые несколько слов отводятся для идентификации записи на ленте и абсолютного номера блока внутри записи. Такие идентификаторы существенно упрощают процесс поиска начала записи на ленте.

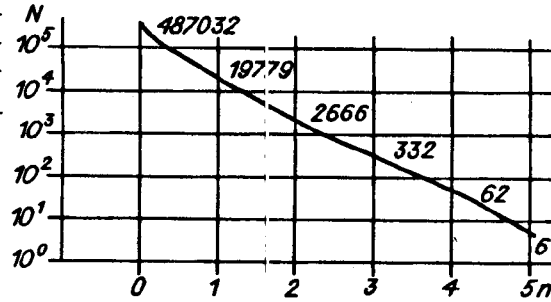
Статистический анализ возникновения ошибок при записи/считывании на видеоленте. Для определения требуемой избыточности внутриблочных кодов Рида — Соломона авторы провели исследование характера ошибок при записи/считывании на видеоленту AGFA с помощью видеоманитофона "Akai". Из графика распределения ошибок на рис. 3 следует, что 500 тыс. блоков данных (полукадров), непрерывно считываемых с видеоленты в течение 3 ч, 19 тыс. содержали по одной испорченной строке, 2,6 тыс. — по две строки, 332 полукадра — по три, 62 — по четыре и 6 полукадров — по пять испорченных строк. Вероятность длительного засорения головок и замедления движения ленты в данном эксперименте была исключена путем предварительной перемотки всей ленты вперед/назад.

Эмпирическая формула вероятности появления неисправляемой конфигурации ошибок в блоке данных при исправлении не меньше t пакетов ошибок, длина каждого из которых не больше одной телевизионной строки, имеет вид

$$P(t) \text{ ош./блок} = 4,1 \cdot 10^{-2} \cdot 7,5^{-t},$$

где $t > 0$.

Рис. 3. Ломаная распределения испорченных строк в блоках данных на видеоленте (N — число полукадров, n — число ошибок на полукадр)



Исходная вероятность ошибок, полученная как сумма ошибок, составляет $4,6 \cdot 10^{-2}$ ош./блок ($1,4 \cdot 10^{-6}$ ош./бит при размере блока $256 \cdot 128$ бит). Чтобы получить $10^{-9} - 10^{-12}$ ош./бит с учетом описанного выше характера исправляемых ошибок, необходимо добиться, по меньшей мере, $10^{-9} - 10^{-12}$ ош./строку или $2,5 \cdot 10^{-7} - 2,5 \cdot 10^{-10}$ ош./блок данных. Задав $t = 8$, получим вероятность появления неисправляемой конфигурации ошибок в блоке $P = 6 \cdot 10^{-10}$ ош./блок. Другими словами, число исправляемых блоковым кодом пакетов по $4 \cdot 32$ бит испорченной информации в каждом блоке должно быть не меньше восьми.

Пользуясь методикой построения исправляющего ошибки кода, описанной в [1], определим требуемое конструктивное межблочное расстояние для заданного числа исправляемых ошибок $t = 8$:

$$d = 2t + 1 = 17.$$

Число избыточных символов $d - 1 = 16$ составляет 6,3 % от длины неукороченного P—C-кода над GF(256).

Исправление блочных ошибок. Восстановление информации, потерянной целыми блоками при длительном засорении головок или в результате сильных колебаний скорости движения ленты, может быть реализовано двумя способами. Первый, более простой, но менее удобный в пользовании способ — это перезапись с верификацией для исключения возникновения так называемых «тяжелых ошибок» при записи и введение повторного считывания данных. Второй способ — дополнительная защита информации от ошибок с помощью таких избыточных кодов, которые, с одной стороны, были бы ортогональны к внутриблочным кодам и, с другой — к пространственному размещению тяжелых ошибок на ленте таким образом, чтобы, как и в случае с «легкими» ошибками, всплески потерянной информации располагались как можно более равномерно между множеством кодов небольшой длины и с малым процентом избыточности. Такой метод борьбы с ошибками особенно необходим во многих применениях, где поток информации непрерывен и нет возможности верификации и повторной ее перезаписи.

Единственно возможным способом распределения тяжелых ошибок является межблочное расслоение информации и образование межблочных кодов вдоль магнитной ленты. При этом чем длиннее простирается элементарный код на ленте, тем меньшая избыточность кода требуется для исправления длительного сбоя. Длина межблочного кода вдоль ленты ограничена в конечном счете только размером одной непрерывной записи на видеоленте.

На рис. 4 изображена структура записи на видеоленте с защитой от ошибок с помощью двойного P—C-кода. Кратность расслоения межблочных кодов может быть установлена программно в зависимости от размера записи и при этом будет наилучшей.

Параметры и программы кодирования/декодирования для межблочных кодов могут быть теми же, что и для внутриблочных. Тяжелые ошибки возникают крайне редко, поэтому их исправление может быть выполнено как «на лету», так и в режиме останова видеоленты в зависимости от предоставленного аппаратного ресурса.

Технико-экономические показатели накопителя. В рамках данного проекта создан контроллер для видеомагнитофона, представляющий собой плату

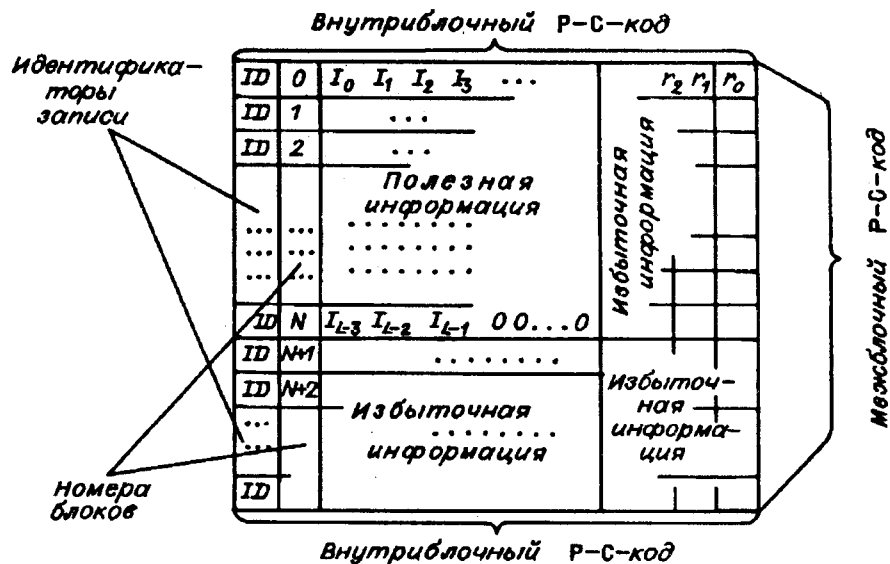


Рис. 4. Структура записи на видеоленте с защитой от ошибок двойным Р—С-кодом

в стандарте ISA и использующий систему синхронизации данных. В качестве блочного кодера/декодера использован процессор СП-30, построенный на базе процессора TMS320C30.

Специально разработанные процедуры кодера/декодера позволяли программно задавать кратность расчленения блока данных на элементарные кодовые слова, оптимально настраивая корректирующую способность (255,236,19) Р—С-кода на размер блока. Поэтому одни и те же процедуры использовались как для внутриблочного, так и для межблочного кодирования/декодирования данных записи на видеоленту.

При использовании одного лишь внутриблочного кодирования данных «на лету» и при преобладающем пакетировании ошибок по 128 бит расчетная вероятность ошибок составила $2 \cdot 10^{-13}$ ош./бит. Объем оперативной памяти микропроцессора при этом — 32 Кбайта.

Для 180-минутной кассеты по цене 5 долларов за штуку (при потоке 180 Кбайт/с информационная емкость кассеты составляет 1944 Гбайта) стоимость хранения 1 Мбайта информации на 600 кассетах (как, например, это имеет место в накопителе фирмы "Metrum" (см. табл. 2)) составляет 0,3 цента/Мбайт. При расчете стоимость базового оборудования (видеомагнитофон, ПЭВМ IBM PC/AT, контроллер и арифметический процессор) принималась равной 2 тыс. долларов.

При использовании двухкаскадного Р—С-кода кодирование данных также удалось выполнить в реальном времени при потоке 180 Кбайт/с. Вероятность появления ошибок при этом измерить не удалось. Для успешного восстановления блока данных при кратности перемежения межблочных кодов 255, записанного на катастрофически испорченном (нарисованный шариковой авторучкой крест, занимающий всю ширину ленты) участке, потребовалось от 4 до 20 с (до 5 итераций продольно-поперечного декодирования) работы процессора TMS320C30, снабженного дополнительно 1 Мбайтом динамической памяти. Стоимость хранения 1 Мбайта при этом увеличивается до 0,5 цента.

В дальнейшем планируется разработать два варианта дешевых накопителей на видеомагнитофоне или видеоплейере: 1-й вариант с совмещением системы контроллера, микропроцессора серии TMS320 и 32 Кбайта статической памяти на одной плате в стандарте ISA; 2-й вариант беспроцессорный, в котором кодирование и восстановление данных будет выполняться «на лету» на управляющей ЭВМ при условии некоторого снижения избытка кодирования и

значительном ограничении вычисляемого числа синдромов ошибок для обнаружения безошибочных блоков при восстановлении данных.

СПИСОК ЛИТЕРАТУРЫ

1. Типикин А. П., Петров В. В., Бабанин А. Г. Коррекция ошибок в оптических накопителях информации.—Киев: Наук. думка, 1990.
2. Балашов Е. П., Атанасов Д. Х. Накопители информации с подвижным магнитным носителем.—Л.: Энергоиздат, 1982.
3. Щербина В. И. Цифровая звукозапись.—М.: Радио и связь, 1989.
4. ГОСТ 7845-79. Система вещательного телевидения. Основные параметры. Методы измерений.—Взамен ГОСТ 7845-72, ГОСТ 19432-74; Введен 01.01.80.—М.: Изд-во стандартов, 1979.
5. Texas Instruments: Third-Generation TMS320 User's Guide, 1988.
6. Гончаров А. В., Харитонов М. И. Канал изображения видеомagneтофона.—М.: Радио и связь, 1983.

Поступила в редакцию 27 апреля 1994 г.