Harvard University Department of Mathematics

Expository Pa per

Intro duction to Reed-Solomon Co des

Advisor: Nathan Kaplan

Author: Yo Sup (Joseph) Moon

August 8, 2011

1 Intro duction

1.1 History

Reed-Solomon Codes are error-correcting codes with applications ranging from data retrieval from bar codes and QR codes in our daily lives to sending transmissions to and from spacecrafts launched in deep-space missions. The Reed-Solomon (RS) Code was discovered by Irving Reed and Gus Solomon and was subsequently presented to the world in their paper \Polynomial Codes over Certain Finite Fields" in the Journal of the Society for Industrial and Applied Mathematics (1959). Since then, RS Codes have been an integral contributor to the telecommunications revolution that took place in the last half of the twentieth century. In particular, Reed-Solomon codes are the most frequently used digital error control codes in the world, due their usage in computer memory and non-volaitile memory applications [1]. A hurried list of signi cant applications include the Digital Audio Disk, Deep Space Telecommunication Systems, Error Control for Systems with Feedback, Spread-Spectrum Systems, and Computer Memory.

1.2 Mathematical Background

De nition. The Hamming Distance d(x;y) between two strings x;yfrom a particular alphabet is de ned as the metric measuring the number of di ering coordinates.
De nition. A code of length nis a set containing codewords
$c_i = (c_{i;1}; c_{i;2}; ::::; c_{i;n}); c_{i;j}$ 2F_q:
De nition. A code C $F_{n\,q}$ $_{n1}$ $_{1}$;c ;::::;c ;c) 2C:
(c0;c1
is linear if it is a linear subspace. The dimension kof a code is the dimension of the subspace. Further, the minimum distance dis de ned as the smallest Hamming distance between any two distinct codewords. A linear code of length n, dimension kand minimum distance dis denoted [n;k;d].
De nition. A generator matrix for an [n;k;d] code Cis any k nmatrix Gwhose rows form a basis for C. For any set of kindependent columns of a generator matrix G, the corresponding set of coordinates form an information set.
) 2C)(c 2 n1 0
De nition. A linear code Cis cyclic if
;::::;c

Proposition. Singleton Bound. Let d(x;y) be the Hamming distance between xand y. Let $A_q(n;d)$ represent the maximum number of possible codewords in a q-ary block code of length nand minimum distance d.
Then, $A_q(n;d)$ q_{nd+1q}: If an [n;k;d] over Fexists, then k nd+ 1. De nition. A code Cis Maximum Distance Separable (MDS) if equality holds in the Singleton Bound. Theorem. If Cis an [n;k;d] code, then every nd+1 coordinate position contains an information set. Inaddition, dis the largest number with this property. Proof. Consult [3] De nition. Let $_{1}$;::::; $_{s}$be elements in F . The s smatrix V = [$v_{i;j}$] where $v_{i;j}$= $_{i1\,j}$is called a Vandermonde matrix.
are distinct.
Lemma. A matrix V is nonsingular if the elements $_{1}$;::::; $_{s}$
1

□

2 Approaches to the RS Co de

2.1 Original Conception

Reed and Solomon's original approach in constructing and decoding their code was not done via generator polynomials, but by using nite eld arithmetic.

De nition. A Galois Field is a nite eld of qelements. It is denoted by GF(q). Proposition. GF(q) contains an element such that $_{q1}$m= 1. This element is called primitive. In particular, there is no m<q1 such that $_{2}$= 1. Since is primitive, the elements f1; ; ;:::; $_{q2}$gmust be distinct. Then it follows that GF(q) = f0;1; ; $_{2}$;:::; $_{q2}$g: Multiplication in the eld, then, is well-de ned and obeys the \obvious" rules of exponentiation:

$$x \quad {}^{y} = {}^{x+y} \quad {}^{k1}:$$

$$1$$

$_{0}$;m$_{1}$;m$_{2}$;:::;m$_{k1}$

P(X) := m$_{0}$

:

Addition works in a similar manner.

De nition. A primitive polynomial P(X) over GF(q) is an irreducible polynomial with coe cients in GF(q) and a primitive root in GF(q).

The RS Code is constructed in the following manner: Suppose m:= (m) is a list of information symbols (i.e. the information one wants to encode), with each coordinate taken from GF(q). De ne

+ m x+ + m x

Each code word cis generated by evaluating the polynomial by each member of GF(q):

c= (c$_{0}$;c$_{1}$;:::;c$_{q1}$) = P(0);P( );:::;P($_{q1}$) : (1)$_{k}$The complete code Cis constructed by choosing mover all possible values. Thus it follows that there are qcodewords in C. And since any two polynomials of degree (k1) is another polynomial of degree less than or equal to (k1), we have that Cis linear. We say that this code is of length n= qand dimension k and denote it by RS(n;k).

Now, consider a particular codeword cin C. From (1), we get a system of qlinear equations with k variables:

$$\begin{aligned} & + m_1 + m_2\,{}_2 + + m_{k1}\,{}_{k1} \\ P(0) &= m_0 \\ P(\ ) &= m_0 \\ P(\ {}_2) &= m_0 + m_2\,{}_4 + m_1\,{}_2 \qquad {}^{2(k1)} \qquad (2) \\ . & \quad + + m_{k1} \\ . & \\ . & \\ P(\ {}_{q1}) &= m_0 + m_1\ {}^{q1} + m_2\,{}_{2(q1)} + + m_{k1}\,{}_{(k1)(q1)}: \end{aligned}$$

Proposition. There is a unique solution m= (m$_{0}$;m$_{1}$;:::;m$_{k1}$) for any kchoices of the qlinear equations. This follows from the proposition in the Mathematical Background section, but Reed and Solomon's original proof o ers some insight in their thinking:

2.1 Original Conception 2 APPROACHES TO THE RS CODE Proof. Vandermonde proof
Without loss of generality, suppose we take the rst kequations to solve. Then we wish to solve the set of equations:

We can show that the system of equations have a uni
determinant of the coe cient matrix

0 B1 0 0 0 1 B B
.k1 24. . .2(k1) . .
CC C C CA :

But the determinant of the matrix reduces to that o
determinant of a Vandermonde matrix is nonsingular.
k qtThis result gives the RS code very signi cant er
coordinates of the qcoordinates of the received codew
are choices for the system of equations to solve, with
(becausean incorrect solution will arise when at least 1
When the probability of an error occurring during trans
with the majority count will be the correct solution. Tha

Now, □

0 B1 0 0 0 1 B B B B1 @. . .1 2. . 1
.k1 24. . .2(k1) . . . k12(k1). . .(k1)(k1)
CC C C CA

0 m0 1 0 2 1
BBm1
B Bm2 C B C
B@ C B C
C B C
C B C
C B C
C A @ A

. =

.
.

k2;m
q k

P
(
0
)
P
(
)
P
(
)
:
..
.

t+k1 k qt k

k1)!(t1)!k! <

t+ k1 k

(
)
(
t
+

Suppose that in the decoding process, a certain threshold c is implemented such that signals below the threshold are decoded into zeros and signals above the threshold are decoded into ones. This is called hard decoding method. In certain cases, however, it may be better to simply omit a codeword coordinate rather than to decode it to a value with significant probability of being incorrect. This is called soft decoding.

Suppose then, we

$$\frac{(q-1)^t}{t!}\cdot\frac{(q-t)^k}{k!}$$

$$\frac{(t+1)(t+2)\cdots(t+k-1)}{(k-1)!}\cdot\frac{(t+1)(t+2)\cdots(t+k-k)}{1}<\frac{(qt)((q-1)t)\cdots(qk-t)}{t!}$$

$$(t+1)(t+2)\cdots(t+k-1)<(qt)((q-1)t)\cdots(qk-t)$$

$$(t+1)^{k-1}<(qt)^{t}$$

$$(2t)^{t}<(qk+1)^{t}$$

$$t<\frac{qk+1}{2}.$$

So the code can correct up to $t=\left\lfloor\frac{qk+1}{2}\right\rfloor$ errors. By the Singleton Bound, this is the best possible error correction capability for any code of the same length and dimension. It follows that RS Codes are MDS codes.

The true merit of the RS Code is in its additional ability to deal with erasures.

2.2 Generator Polynomials 2 APPROACHES TO THE RS CODE allow for \erasures" of the coordinates of codewords. Since only kuncorrupted linear equations are needed
to retrieve the information bits, up to qkcoordinates may be erased. Let vdenote the number of erasures. As long as$_{t+k1\ k}$ <$_{qvt\ k}$, errors can be corrected. Similar to before, we have: t+ k1k < qvt k
() (t+ l1)!k!(t1)! < (qvt)! k!(qkvt)!
()2t+ v<qk+ 1: Thus, a Reed-Solomon code can correct up to terrors and verasures, as long as the above inequality is satis ed.

2.2 Generator Polynomials

The generator polynomial method to the construction of the RS Code is the most commonly used approach today [1].
For a cyclic code C, we interpret c2Cas such:
$_{nk}$ ^{nk} (c$_0$;c$_1$;:::;c$_{n2}$;c$_{n1}$) 7!c$_0$ ^{2}x + x$_{n1}$ + + 0 g g

g(x) =

length dimension 2

: Proposition. A cyclic code can always be de ned using a generator polynomial g(x) = g 1 2x x + x$_{n1}$

0 c 1 x + 1

+

+

c n 1

^{2} Y i=

1
k1
X
i
:
4
+ +
g. A vector cis a codeword in the code generated by g(x) if and only if the corresponding code polynomial c(x) = cis a multiple of g(x). From the proposition, we have that for kinformation symbols m= (m) and its corresponding polynomial m(x), any codeword c(x) is generated by the following relation:

c(x) = m(x)g(x): The cyclic Reed-Solomon code with symbols from GF(q) have length q1, with one fewer coordinatethan the codewords generated via Reed and Solomon's original construction. Note that the eld GF(256) is a popular choice of the nite eld because each of the 256 eld elements can be represented as a binary byte (8 bits).

Construction We wish to construct a t-error correcting RS code of length q1 with entries in GF(q). Take a primitiveelement 2GF(q). We construct the generator polynomial g(x) such that ghas 2tconsecutive powers of as its roots:

It follows that a code polynomial can have degree 2t d q2. The dimension of the code is then k= q2t1. The Maximum Distance Separation relation applies here as well:Error correction capacity = bc This method of construction provides a very convenient way for determining the correctness of the received word. One must simply make sure that the code polynomial has the necessary roots as dictated by the construction.

2.3 Fourier Transf orms

De nition. The Fourier Transform is a map that takes a complex-valued function of a real variable to another. The domain of the map is called the time domain and the image of the map is called the frequency domain.

The notion of time domain and frequency domain is ambiguous in a nite eld setting, but the operation itself is well-de ned and leads to useful consequences [7].

De nition. Take 2GF(q) a primitive element in the nite eld. The Galois eld Fourier transform of a vector c= ($c_0;c_1;:::;c_{n1}$) is the de ned as follows:
Ff($c_0;c_1;:::;c_{n1}$)g= ($C_0;C_1;:::;C_{n1}$);where
$C_j = \sum_{0}^{n1} c_i \ ij$:
Theorem. Over GF(q), a eld of characteristic p, a vector and its spectrum are related by
$C_j = \sum_{0}^{n1} \ ijc_i$
c_i = 1n mod p $\sum_{0}^{n1} \ ijC_j$:

Proof. In any eld, we have the relation ^{n1}1 = (x1)(x + x^{n2}r+ + x+ 1): By de nition, is a root of the polynomial on th ^{n}x that ^{r}is a root of the last term for all r6= 0 modulo n.This is equi
$\sum_{n1j=0}^{} X^{r} = 0$ for r6= 0 mod n:
j
and when r= 0, $\sum_{0}^{n1} \ ^{rj} = n \bmod p$
which is not zero if nis not a multiple of p. = $\sum_{k=0}^{n1} X^{ck} \ ^{(ki)j} = (n \bmod p)c_i$:
We thus obtain:
$\sum_{n1j=0}^{} X^{ij} \sum_{k=0}^{n1} X^{ck}$
In addition, if the eld has characteristic p, there exists an integer ssuch that q1 = p^{s}1 0 mod n. Consequently, nis not a multiple of p, and n mod p6 0.This proves the claim. Theorem. 1. The polynomial c(x) := c_0+ c_1x+ + $c_{n1}x^{n1}$has a root at $_j$i the jth spectral component C_j= 0.

□

3 ENCODING AND DECODING OF RS CODE 2. The polynomial C(X) := C_0+ C_1x+ + $C_{n1}$$x_{n1}$has a root at $_i$i the ith time component c_i= 0. Proof. 1. This follows from the equation
c($_j$) = $^{n1i=}_{0}$X c_i $_{ij}$ = C_j:

= 1n mod p $^{n1j=}_{0}$X $_{ij}C_j$

2. Similarly, from c_i
the claim follows.

Corollary. A word polynomial has 2tconsecutive powers of as roots i the spectrum of the corresponding word has 2tconsecutive zero coordinates.

From the corollary we see that the GFFT method is a dual to the generator polynomial method. The GFFT method thus leads to a series of e cient encoders and decoders not available by generator polynomials alone, using the powerful vocabulary of the Fourier Transform.

□

3 Enco ding and Deco ding of RS Co de

The current view of the RS codes considers the RS codes to be a special case of the BCH codes$_1$of length n= q1. This perspective coupled with the perspectives previously introduced lead to a series of e cient encoding and decoding algorithms, which the author will discuss here.

+ d_1x+ + d_{k1} x_{k1} be a data polynomial with each d_i

Enco ding

Let the polynomial d(x) = d_0

c= (c_0 ;c_1;:::;c_{n1}) = (p ;p $^{nk1}_{:}$ $_0$;d ;:::;d

1

2GF(q) (each codeword c2Cis constructed from each choice of d, and the entire code is constructed by varying over all d). One way of encoding RS codes is to compute c(x) = d(x)g(x), where gis the generator polynomial; however, this is not a systematic method since the kdata symbols are not explicitly present in the codeword. Hence this leads to an extra step being required in the decoding process to extract information from the codeword. A way to improve this inadequacy is to encode by using the Cauchy Matrix.
Let d(x) be as above. A codeword can be systematically encoded by adding nkparity check symbols to the data symbols:

$_1$;:::;p ;d 1 k1
$_0$ 6

$A_{i;j}$
i

see [2]

)

De ne the systematic generator matrix G:

G= I Awhere Iis the k kidentity matrix and Ais a k (nk) matrix, known as a Cauchy Matrix. Ais constrained by these de nitions:

$= u x_i v_j + \quad ;0 \quad i \; k1; 0 \quad j \; nk1$

y_j

= n1i
n1 l) ;0 i k1
= n1kj
= 1Q0 l k1= Y0 l k1 (n1i
0nk1) is p= dA

j

i

v
j

p = k1 i dᵢAᵢ;ⱼ ;0 j nk1:
j i=
0
X
iu
+ yj

Deco ding

xj ;0 i k1 y;0 j nk1 u

The parity check vector ;:::;p
p= (p
j d i
k1i=0X
x

= v
A co m m on ly us ed de co di ng pr oc es s for po ly no mi al- ge ne rat ed R S co de

n1kj0 n1l ;0 j nk1:

sis Syndrome-Based Decoding. Dene an error polynomial as e(x) = e_0 + e_1x+ + $e_{n-1}x^{n-1}$. An input polynomial for the decoder is then:

$$v(x) = c(x) + e(x) = v_0 + v_1 x + \quad + v_{n1} \quad x_{n1})$$

i

2t:

$$v(_j) = c(_j) + e(_j = e(_j) = \quad {}^{n1}{}_{i=} \; e_{i\,ij}; j = 1;2;:::;2t:$$

$_0X$

;

where the v's are the components of the received word. We evaluate the polynomial at the roots of the generator

or
po
ly
no
mi
al
;
::::;

W
e
de
n
e
th
e
jth
sy
nd
ro
m
e
of
va
s
th
e
fo
uri
er
tra
ns
for
m
ati
on
:

$S_j := v(_j) = \sum_{0}^{n1i=} v_{i\ ij}; j = 1;2;::::2t:$

Suppose that 0 terrors occur in unknown locations $i_1;i_2;::::;i$. Then, $e(x) = e_{i_1}$'s. $_iX_1 + + e_i\ _iX$

De ne $Y_I := e_{iI}$for $I = 1;2;::::$; and $X_I = _{iI}$

$S = Y_1X_1 + Y_2X_2 + Y X$

1

.

We
wish
to n
d
the
i's
and
the
e
for
l=
1;
2;:
::;
.
Fr
o
m
ev

al
ua
tin
g
th
e
sy
nd
ro
m
es
w
e
ob
tai
n
a
sy
st
e
m
of
eq
ua
tio
ns
wi
th

un
kn
o
w
ns
:

$= Y_1 X_{j1} . X^{j2} + + Y X_j$

X_{2t} :

. . .

$+ Y_2 .$

$X_{2t\,2}$

$^{2}_{t}$

S_j

$= Y_1 X_{1} + + Y$

$+ Y_2$

7

S_{2t}

 By the construction of the syndromes, at least 1 solution exists. The decoding process is completed by nding a solution with the least amount of errors.

4 RS Co des in Standard Usage

The author will conclude the paper with an extended overview of a single application of the Reed-Solomon Codes.

Compact Discs

RS Codes are extensively used in the error-correcting used in reading and writing Compact Discs (CDs). When sound is recorded, sound-waves are converted from an analog source to a digital medium by a process called \sampling". The most commonly used sampling rate is 44.1 Khz, a rate chosen to be compatible with an already existing standard for video recording. The amplitude of a waveform is sampled at an instance in time and is is assigned to a binary string of length 16, one for the left and one for the right channel of sound. Each vector of length 16 is cut in half, thereby producing 4 bytes of data per second. When the sound is encoded, a combination of two RS codes called the Cross-Intervealed Reed Solomon Code (CIRC) is used, with the design to correct long burst errors. Now we are ready to describe the process of encoding.

4.0.1 Encoding Note that the encoding method was designed to combat burst errors. Six samples of each of the four bytes
are grouped into a frame consisting of 24 bytes. Denote the ith pair of bytes in the left channel as Land its right channel counterpart as R_i. Then a frame consists of the following pattern:$_i$
$L_1R_1L_2R_2 \quad L_6R_6$

^{3} $_4$^ R_5^ ^
R_4L L R_6:

1. The Spread The odd-numbered samples
L_1R_1;L_3R_3;L_5R_5
L_2^ R_2; ^ L_4^ R_4; ^
L_6^ R_6
L_1R_1^ L_2^ R_2L_3
: The frame goes through a series
of permutations before it is encoded to a master CD.
are grouped with the even-numbered samples ^from two frames ahead. Thus the new arrangment consists of the following pattern:
R ^ ^{5} ^{6}
L

2. Internal Rearrangment The new frame is then arranged so that the even and odd
bytes are separated:

$$L_1L_3L_5R_1R_3R_5\text{^ }L_2\text{^ }L_4\text{^ }L_6\text{^ }R_2\text{^ }R_4\text{^ }R_6:$$

$_1$The motivation behind this is to separate samples as far apart as possible inside the new frame. 3. Adding redundancy via CThis 24-byte message, an element in F_{256}^{24}. The encoder adds four bytes of redundancy, namely Pis then encoded by an encoder using a [28,24,5] Reed-Solomon Code, which we will henceforth call Cand P_2such that the new frame consists of the following pattern:$_1$

$$L_1L_3L_5R_1R_3R_5P_1P_2\text{^ }L_2\text{^ }L_4\text{^ }L_6\text{^ }R_2\text{^ }R_4\text{^ }R_6::$$

We see that this further separates the odd-numbered samples from the even-numbered samples.

4 RS CODES IN STANDARD USAGE 4. 4-frame delay interleaving
Each code $2C_1$
c_i
is interleaved into a matrix with 28 rows and a large but determined number of columns in the following manner:
is placed in column 1, the rst byte of c_2
. ^{1} $^{1}_{2}$is placed in column 5, second byte of c_2

.

.
Row 1 The rst byte of c in column 2, etc. Row 2 Begin with four bytes of 0. Second byte of cis placed in column 6, etc. Row 28 Begin with 4 (281) = 112 bytes of 0. 28th byte of c is placed in column 113, 28th byte of c

$c_{2;1}$ $c_{3;1}$ $c_{4;1}$ $c_{5;1}$ $c_{6;1}$ $c_{7;1}$ $c_{8;1}$ $c_{9;1}$ $c_{10;1}$ $c_{11;1}$ $c_{12;1}$ $c_{13;1}$
$c_{5;2}$ $c_{6;2}$ $c_{7;2}$ $c_{8;2}$ $c_{9;2}$
consistent:
Example. $^{1;3}$ $c_{5;3}$
$^{1;4}$

$^{1;2}$ $c_{2;2}$ $c_{3;2}$ $c_{4;2}$
$c_{2;3}$ $c_{3;3}$ $c_{4;3}$
$c_{1;1}$ . . —
is placed in column 114, etc In addition, we pad all the rows except for the last with 0's so that the number of columns are
0 0 0 0 c 0 0 0 0 0 0 0 0 c 0 0 0 0 0 0 0 0 0 0 0 0 c .
$^{1\ 4}_{2}$

This will leave the original codeword diagonally on the matrix with slope
$_2$and interleaving Each column in the
matrix $c2F_{28\ 256}$

9

. 5. C
is then encoded by a [32;28;5] RS code, denoted by C. This thusly generates a group of codewords, each consisting of 32 bytes.

The codewords are then rearranged so that the odd-numbered symbols are matched up with the evennumbered symbols from the next codeword, and so on. The motivation for this interleaving again comes from the desire to break up other short bursts that may be present even after the 4-delay interleaving. After the rearrangement, the codewords are written consecutively in one long stream, which is then redivided into a segment of 32 bytes; at this point a 33rd byte is added that includes control and display information that the CD player requires to display the playing time, composer, and the title of the piece.

6. Imprinting A pit is an indentation in the CD that the laser reads and interprets as data. A land is the space between the pits. Each bit is of length 0.3 m when it is imprinted on the CD. A land-to-pit or pit-to-land transition is marked with a 1, and the pit and land itself is marked with a string of 0's. There are physical limitations to CD writing, so each land or pit must be between 0.9 and 3.3 m in length. Hence a pair of 1's must be separated by at least two 0's and at most ten 0's.
For this reason, a look-up map from bytes to 14-length strings called the eight-to-fourteen modulation (EFM) map is implemented for the actual imprinting process.

Example. A part of the EFM Table

Data	Code	Data	Code
100	01000100100010	114	10010010000010
101	00000000100010	115	00100000100010
102	01000000100010	116	01000010000010
103	00100100100010	117	00000010000010
104	01001001000010	118	00010001000010
105	10000001000010	119	00100001000010
106	10010001000010	120	01001000000010
107	10001001000010	121	00001001001000
108	01000001000010	122	10010000000010
109	00000001000010	123	10001000000010
110	00010010000010	124	01000000000010
111	00100001000010	125	00001000000010
112	10000000100010	126	00010000100010
113	10000010000010	127	00100000000010

The strings are imprinted on the CD in succession; therefore there may be instances when two strings satisfying the criterion de ned by the proposition do not in conjunction satisfy the criterion. For instance, 10010000000100 and 00000000010001 are allowable strings, but 10010000000100000000000010001 is not. Therefore, merge bits de ned as 001 must be appended to the end of the 14-length strings, resulting in a string of length 17.

At the end of the 33 17 = 561 bits, 24 synchronization and 3 merge bits are appended. The result is that for each frame of six samples of information 588 bits are written onto the CD.

4.0.2 Decoding Decoding takes place as to exactly reverse the process of encoding. It is implemented in two main steps:

1. Decoding with C_2

$_2$The rst step entails reading the disc, removing synchronization, control and display bits, and demodulating the 17-bit codes into their respective byte formats. The data is presented as a stream of bytes. The stream is divided up into a segment of 32-bytes. Recall that by construction each 32-byte segment contains 16 bytes of the odd-indexed bytes from a codeword (possibly with errors) and 16 bytes of the even-indexed bytes from the next codeword. Because $C_2$is a [32;28;5] code, it is able to correct up to 2 errors. But in this case, correction is only made when a single error exists; otherwise Cis used to only detect multiple errors. The reasoning behind this can be summarized by evaluating the probability of error.

can detect 2 or 3 errors. The probability that $_6$.

C_2

$$= \quad ^2$$

$$\begin{matrix} 2 \\ 5 \\ 6 \end{matrix} \qquad \begin{matrix} 8 \\ 1 \\ 6 \end{matrix}$$

$$^{32} \quad ^{2}_{8} \qquad 1$$

10

will not detect greater or equal to four errors when using it to correct for single errors

laim. When correcting for 1 error, C_2 1:9 10

Proof. Assume that all codewords are equally likely. The probability that Cwill not de approximately the number of codewords in spheres of radius 1 about each codewor total number of codewords. Hence we have that the probability is :

[1 + 32(256 1)] 256 256_4

1:9 10_6:

4 RS CODES IN STANDARD USAGE Claim.

When correcting for 2 errors, the probability that three or more errors will be undetected by Cis approximately 7:5 $10^{3.2}$

Proof. Assume again that all codewords are equally likely to occur. The probability that C_2will not detect three or more errors is approximately the number of codewords in spheres of radius 2 about each codeword divided by the total number of codewords. Hence we have that the probability is :

$$\frac{256^{28}[1 + 32\ 1\ (256\ 1) +]\ 256^{32\ 2}\ (256\ 1)^{2}\ 3:}{256^{32}} \quad 7:5\ 10$$

2. Decoding with C_1

There are three possibilities that can result from step 1 of the decoding process, each one resulting from the determination of the decoder.

(a) No Errors If the decoder determines that there are no errors, the 28-byte vector is extracted from the 32-byte

vector and is passed on to step 2. (b)

Single Error

Error is corrected, and the 28-byte vector is passed on to step 2. (c) Two or more errors

It passes on a 28-byte vector with every component marked as erasures. From the array as described in (4), the diagonal elements along the line of slope 1=4 are extracted.There are two schemas that the decoding algorithm uses:

is used only to correct erasures. C_1

Schema 1: In schema 1, $C_1$11

1

can correct up to four erasures. Schema 2: In schema 2, Cis used to correct one error and two erasures. Decoding methods for this step has not been standardized and the choice of method is largely

dependant on the manufacturer of the decoding device. The choice may depend upon many considerations such as power consumption and price.

3. Interpolation and Silencing Even after the error-detection and correction algorithms implemented in the decoding process, errors

may still remain in the code. One additional step better suits the data for output. This step is conditional on the existence of burst errors.

Case 0: If there are no burst errors in a neighborhood of a sample, then linear interpolation is used to \conceal" the error.

Case 1: If there are burst errors in a neighborhood of a sample, the neighbors of a sample are unreliable for linear interpolation. In this case, a procedure called \muting" is used. This procedure entails gradually weakening reliable samples 32 samples before the burst and gradually strengthening the next 32 reliable samples after the burst, with the burst altered to have zero values. This muting process happens over a few milliseconds and it is essentially inaudible [1].

References

[1] S.B. Wicker and V.K. Bhargava, \An Introduction to Reed-Solomon Codes" in Reed-Solomon Codes and Their Applications, eds. S.B. Wicker and V.K. Bhargava. New York: IEEE Press, 1994, pp.1-15

[2] J.H. van Lint, Introduction to Coding Theory. Springer, New York 3rd Edition, 1999. [3] W. Cary Hu man; Vera Pless, Fundamentals of Error-Correcting Codes. Cambridge University Press;

Cambridge, UK 1st Edition, 2010. [4] J.L. Massey, \Shift-register synthesis and BCH decoding," IEEE Trans. Inform. Theory IT-15 (1969) [5] K.A.S. Immink, \Reed Solomon codes and the compact disc," in Reed-Solomon Codes and Their Appli-

cations, eds. S.B. Wicker and V.K. Bhargava. New York: IEEE Press, 1994, pp.41-59 [6] T. Yaghoobian and I.F. Blake \Reed Solomon and Algebraic Geometry Codes," in Reed-Solomon Codes

and Their Applications, eds. S.B. Wicker and V.K. Bhargava. New York: IEEE Press, 1994, pp.292-314 [7] R.E. Blahut, \Transform Techniques for Error Control Codes," IBM Journal of Research and Develop-

ment, Volume 23 pp.299-315, 1979.