This document contains supplemental results on the structural properties of the RM code not included in the journal paper.

For notational reference, $\mathbf{G_m}$ is

$$\mathbf{G_m} = \begin{bmatrix} \underline{1} \\ \mathbf{M_m} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & \dots & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 1 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & \dots & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & \dots & 1 & 0 & 1 & 0 \end{bmatrix},$$

were $\underline{1}$ denotes the binary string of length 2^m with all entries equal to 1, and the m by 2^m submatrix $\mathbf{M_m}$ consists of lexicographically decreasing binary columns of length m. Observe that the i^{th} row of $\mathbf{G_m}$, for $1 < i \leq m+1$, consists of 2^{i-1} alternating runs of ones and zeros, and that each run is of length 2^{m-i+1}. The code $C(m)$ is generated by $\mathbf{G_m}$.

0.1 Relationship between the input message and the run-lengths of its codeword

Let $\mathbf{a_m} = (a_0, a_m, a_{m-1}, ..., a_2, a_1)$ be a binary string of length $m+1$ and let $\mathbf{c}$ be a codeword in $C(m)$ such that $\mathbf{c} = \mathbf{a_m}\mathbf{G_m}$. The bit a_0 multiplies the all-ones row of $\mathbf{G_m}$ and therefore does not affect the number of runs of the resulting codeword, i.e. $\mathbf{a_m} = (a_0, a_m, a_{m-1}, ..., a_2, a_1)$ and $\mathbf{a_m}' = (\overline{a_0}, a_m, a_{m-1}, ..., a_2, a_1)$ result in complement codewords (with the same number of runs). In the following we replace a_0 by x to indicate that the value of a_0 does not matter.

We denote by $R_m(a_0, a_1, ..., a_{m-1}, a_m)$ the total number of runs in $\mathbf{c}$. The following result provides a closed-form expression for $R_m(a_0, a_1, ..., a_{m-1}, a_m)$ in terms of $\mathbf{a_m}$.

Lemma 1 *The number of runs in the codeword* $\mathbf{c}$ *given by* $\mathbf{c} = \mathbf{a_m G_m}$ *where* $\mathbf{a_m}$ = $(a_0, a_m, a_{m-1}, ..., a_2, a_1)$ *is* $R_m(a_0, a_1, ..., a_{m-1}, a_m) = 2^{m-1}a_1 + 2^{m-2} + 1/2 - \sum_{k=2}^{m} 2^{m-k-1}(-1)^{\sum_{i=1}^{k} a_i}$.

Proof: By construction the bottom $m-1$ rows in $\mathbf{G_m}$ when viewed as a $m-1$ by 2^m matrix, are the same as the matrix obtained by concatenating the matrix consisting of the bottom $m-1$ rows in $\mathbf{G_{m-1}}$ with itself. In constructing $\mathbf{c}$ from codewords in $C(m-1)$, if the runs at the point of concatenation are the same, the concatenation results in the merging of two runs, otherwise no runs are altered.

Therefore, a linear combination of the bottom $m-1$ rows in $\mathbf{G_m}$ produces a codeword in $C(m)$ which has either $2R$ or $2R-1$ runs, where R denotes the number of runs of the codeword produced by the same linear combination of rows in $\mathbf{G_{m-1}}$. In particular, the number of runs is $2R$ if the auxiliary codeword in $C(m-1)$ (the one constructed from the same linear combination) had different outermost bits, and the number of runs is $2R-1$ if the outermost bits are the same. The former (latter) case occurs when the linear combination consists of an odd (even) number of participating rows.

Then, when $a_m = 0$ we have the following:

$$R_m(x, a_1, a_2, ..., a_{m-1}, 0) = \begin{cases} 2R_{m-1}(x, a_1, a_2, ..., a_{m-1}), & \text{if } \sum_{i=1}^{m-1} a_i \bmod 2 \equiv 1, \\ 2R_{m-1}(x, a_1, a_2, ..., a_{m-1}) - 1, & \text{if } \sum_{i=1}^{m-1} a_i \bmod 2 \equiv 0 \, . \end{cases}$$

The value $a_m = 1$ has the effect of complementing the left half of the codeword obtained from a linear combination of the bottom $m-1$ rows of $\mathbf{G_m}$, and leaving the right half intact.

Therefore,

$$R_m(x, a_1, a_2, ..., a_{m-1}, 1) = \begin{cases} 2R_{m-1}(x, a_1, a_2, ..., a_{m-1}), & \text{if } \sum_{i=1}^{m-1} a_i \bmod 2 \equiv 0, \\ 2R_{m-1}(x, a_1, a_2, ..., a_{m-1}) - 1, & \text{if } \sum_{i=1}^{m-1} a_i \bmod 2 \equiv 1 \end{cases}$$

We can jointly write these two expressions as

$$R_m(x, a_1, a_2, ..., a_{m-1}, a_m) = 2R_{m-1}(x, a_1, a_2, ..., a_{m-1}) - 1/2(-1)^{\sum_{i=1}^{m} a_i} - 1/2.$$

To obtain the formula for $R_m(x, a_1, ..., a_{m-1}, a_m)$, we expand as follows,

$$
\begin{aligned}
&R_m(x, a_1, ..., a_m) \\
&\quad = 2R_{m-1}(x, a_1, a_2, ..., a_{m-1}) - 1/2(-1)^{\sum_{i=1}^{m} a_i} - 1/2 \\
&\quad = 2\left[2R_{m-2}(x, a_1, a_2, ..., a_m) - 1/2(-1)^{\sum_{i=1}^{m-1} a_i} - 1/2\right] - 1/2(-1)^{\sum_{i=1}^{m} a_i} - 1/2 \\
&\quad = 4R_{m-2}(x, a_1, a_2, ..., a_{m-2}) - (-1)^{\sum_{i=1}^{m-1} a_i} - 1 - 1/2(-1)^{\sum_{i=1}^{m} a_i} - 1/2 \\
&\quad = 4\left[2R_{m-3}(x, a_1, a_2, ..., a_{m-3}) - 1/2(-1)^{\sum_{i=1}^{m-2} a_i} - 1/2\right] - (-1)^{\sum_{i=1}^{m-1} a_i} - 1 - 1/2(-1)^{\sum_{i=1}^{m} a_i} - 1/2 \\
&\quad = 8R_{m-3}(x, a_1, a_2, ..., a_{m-3}) - 2(-1)^{\sum_{i=1}^{m-2} a_i} - 2 - (-1)^{\sum_{i=1}^{m-1} a_i} - 1 - 1/2(-1)^{\sum_{i=1}^{m} a_i} - 1/2 \\
&\quad \vdots \\
&\quad = 2^{m-2}R_2(x, a_1, a_2) - 2^{m-3-1}(-1)^{\sum_{i=1}^{3} a_i} - 2^{m-3-1} - 2^{m-4-1}(-1)^{\sum_{i=1}^{4} a_i} - 2^{m-4-1} - ... \\
&\qquad -2^{m-(m-1)-1}(-1)^{\sum_{i=1}^{m-1} a_i} - 2^{m-(m-1)-1} - 2^{m-m-1}(-1)^{\sum_{i=1}^{m} a_i} - 2^{m-m-1} \\
&\quad = 2^{m-2}2R_1(x, a_1) - 2^{m-2}1/2(-1)^{\sum_{i=1}^{2} a_i} - 2^{m-2}1/2 - 2^{m-4}(-1)^{\sum_{i=1}^{3} a_i} - 2^{m-4} - ... \\
&\qquad -2^{0}(-1)^{\sum_{i=1}^{m-1} a_i} - 2^{0} - 2^{-1}(-1)^{\sum_{i=1}^{m} a_i} - 2^{-1} \\
&\quad = 2^{m-1}(1 + a_1) - \\
&\quad \left[2^{m-3}(-1)^{\sum_{i=1}^{2} a_i} + 2^{m-4}(-1)^{\sum_{i=1}^{3} a_i} + ... + 2^{-1}(-1)^{\sum_{i=1}^{m} a_i} + 2^{m-3} + 2^{m-4} + ... + 1 + 1/2\right] \\
&\quad = 2^{m-1}a_1 - \sum_{k=2}^{m} 2^{m-k-1}(-1)^{\sum_{i=1}^{k} a_i} + 2^{m-1} - \left[2^{m-2} - 1 + 1/2\right] \\
&\quad = 2^{m-1}a_1 + 2^{m-2} + 1/2 - \sum_{k=2}^{m} 2^{m-k-1}(-1)^{\sum_{i=1}^{k} a_i},
\end{aligned}
$$

which completes the proof. ∎

It is also useful to know how to quickly determine the input message based on the number of runs in the codeword it generates. Let $N_{1,m}$ be the integer denoting the number of runs of a codeword in $C(m)$, and let $\mathbf{a_m}(N_{1,m}) = (a_0, a_m, ..., a_2, a_1)$ be the input message whose codeword has $N_{1,m}$ runs.

First observe that $|\sum_{k=2}^{m} 2^{m-k-1}(-1)^{\sum_{i=1}^{k} a_i}| \leq 2^{m-2} - 1/2$. Thus, for $a_1 = 1$, $R_m(x, 1, ..., a_{m-1}, a_m)$ is in the interval $[2^{m-1} + 1, 2^m]$ and for $a_1 = 0$, $R_m(x, 0, ..., a_{m-1}, a_m)$ is in the interval $[1, 2^{m-1}]$. Thus, for the given m, if $N_{1,m} \geq 2^{m-1} + 1$, a_1 must be 1, otherwise it must be zero. Moreover, note that $R_m(x, 1, ..., a_{m-1}, a_m)$ + $R_m(x, 0, ..., a_{m-1}, a_m)$ evaluates to $2^m + 1$, since

$$\left(2^{m-1} + 2^{m-2} + \tfrac{1}{2} - \sum_{k=2}^{m} 2^{m-k-1}(-1)^{1+\sum_{i=2}^{k} a_i}\right) + \left(2^{m-2} + \tfrac{1}{2} - \sum_{k=2}^{m} 2^{m-k-1}(-1)^{\sum_{i=2}^{k} a_i}\right)$$
$$= 2^{m-1} + 2^{m-1} + 1 - \sum_{k=2}^{m} 2^{m-k-1}\left((-1)^{1+\sum_{i=2}^{k} a_i} + (-1)^{\sum_{i=2}^{k} a_i}\right)$$
$$= 2^m + 1 + 0.$$

To evaluate the remaining a_2 through a_m, we determine the contribution of a_2 through a_m to $N_{1,m}$. This contribution $N_{2,m}$ is $N_{1,m}$ for $a_1 = 0$ and is $2^m + 1 - N_{1,m}$ for $a_1 = 1$.

Having determined a_1, observe that $R_m(x, 0, a_2, ..., a_{m-1}, a_m) = R_{m-1}(x, a_2, ..., a_{m-1}, a_m)$, since the i^{th} row of $\mathbf{G_m}$ for $1 \leq i \leq m$ is constructed from the i^{th} row of $\mathbf{G_{m-1}}$ by repeating each entry twice. Thus, a codeword constructed from the linear combination of a subset of these particular rows of $\mathbf{G_m}$ has the same number of runs as the codeword in $C(m-1)$ constructed from the counterpart rows of $\mathbf{G_{m-1}}$.

We now view a_2 as the value that multiplies the last row of $\mathbf{G_{m-1}}$, just like a_1 did for $\mathbf{G_m}$. By using the same line of arguments as for a_1, conclude that if $N_{2,m} \geq 2^{(m-1)-1} + 1$, a_2 is 1, otherwise it is 0. The contribution $N_{3,m}$ of a_3 through a_m is $N_{2,m}$ if $a_2 = 0$ and is $2^{m-1} + 1 - N_{2,m}$ for $a_2 = 1$. Compare $N_{3,m}$ to $2^{(m-2)-1} + 1$, and if below, set $a_3 = 0$, else $a_3 = 1$. Repeat evaluating $N_{i,m}$ and a_i until a_m is determined.

Recall that input messages $(1, a_m, ..., a_2, a_1)$ and $(0, a_m, ..., a_2, a_1)$ result in complement codewords which thus have the same number of runs.

The steps for determining the input message $\mathbf{a_m}(N_{1,m}) = (x, a_m, ...a_2, a_1)$ for the given integer $N_{1,m}$ can be outlined as follows:

1. Set $i = 1$.
2. Set $a_i = 1(N_{i,m} \geq 2^{m-i} + 1)$.
3. Set $N_{i+1,m} = (2^{m-i+1} + 1 - N_{i,m})1(a_i = 1) + N_{i,m}1(a_i = 0)$.
4. If $i = m$ return strings $(1, a_m, ..., a_2, a_1)$ and $(0, a_m, ..., a_2, a_1)$, else go back to Step 2 with $i \to i + 1$.