Reed-Muller Codes Achieve Capacity on Erasure Channels

Henry D. Pfister
with S. Kudekar, S. Kumar, M. Mondelli, E. Şaşoğlu, R. Urbanke

November 16th, 2015
Reed-Muller Codes (I)

- Codes by Muller, a decoder by Reed, both in 1954
  - Multivariate polynomial-evaluation codes over binary field
  - Minimum distance $\approx \sqrt{N}$ (Not so good!)

- Very popular in theoretical computer science (TCS)
  - locally decodable, locally testable, probabilistic proof systems

- Capacity-Achieving Conjectures
  - By Shu Lin: “RM Codes are Not So Bad” (Tokyo ITW, 1988)
  - By Costello and Forney for Rate-1/2 and BI-AWGN, 2007

- First known conjecture in print by Dumer and Farrell in 1994
  - They show BCH codes achieve capacity on BEC as rate $\to 1$
  - Open problem stated for Reed-Muller codes at constant rates
Closely related to polar codes

From Hadamard matrix, one choice of rows generates Reed-Muller and some other polar codes.

In fact Arikan remarked:

*It is interesting that the possibility of RM codes being capacity-achieving codes under ML decoding seems to have received no attention in the literature*

Under MAP, Reed-Muller observed to be better than polar (Arikan and Mondelli-Hassani-Urbanke)
Reed-Muller Codes (II)

- Closely related to polar codes
  - From Hadamard matrix, one choice of rows generates Reed-Muller and some other polar codes

In fact Arikan remarked:

*It is interesting that the possibility of RM codes being capacity-achieving codes under ML decoding seems to have received no attention in the literature*

- Under MAP, Reed-Muller observed to be better than polar (Arikan and Mondelli-Hassani-Urbanke)
- In 2014, Abbe-Shpilka-Wigderson showed capacity achieving for rates $\to 0, 1$ (erasures) and rates $\to 0$ (errors)
Reed-Muller Codes (II)

- Closely related to polar codes
  - From Hadamard matrix, one choice of rows generates Reed-Muller and some other polar codes

In fact Arikan remarked:

*It is interesting that the possibility of RM codes being capacity-achieving codes under ML decoding seems to have received no attention in the literature*

- Under MAP, Reed-Muller observed to be better than polar (Arikan and Mondelli-Hassani-Urbanke)
- In 2014, Abbe-Shpilka-Wigderson showed capacity achieving for rates $\to 0, 1$ (erasures) and rates $\to 0$ (errors)
- Can they achieve capacity for constant rates?
1. RM codes achieve capacity at all rates (under MAP decoding.)

2. Let \( X^n = (X_1, X_2 \ldots X_n) \) be iid \( \text{Bern}(1/2) \).
Can They?

Let $\{C_n\}$ be a sequence of codes with rates $r_n \to r \in (0, 1)$. Suppose the permutation group of $C_n$ is doubly transitive and then, $\{C_n\}$ achieves capacity on the BEC under bit-MAP.

Important Consequences:

- Reed-Muller codes achieve capacity.
- Primitive narrow-sense BCH codes achieve capacity.
- Affine-invariant codes achieve capacity.
- Extends to block-MAP for Reed-Muller and BCH by Kumar-Pfister and Kudekar-Mondelli-Sasoglu-Urbanke.
Can They?

YES!

Let \( \{C_n\} \) be a sequence of codes with rates \( r_n \to r \in (0, 1) \)

- Blocklengths \( N_n \to \infty \)
- Suppose the permutation group of \( C_n \) is doubly transitive and
- Then, \( \{C_n\} \) achieves capacity on the BEC under bit-MAP
Can They?

YES!

Let \( \{C_n\} \) be a sequence of codes with rates \( r_n \rightarrow r \in (0, 1) \)

- Blocklengths \( N_n \rightarrow \infty \)
- Suppose the permutation group of \( C_n \) is doubly transitive and
- Then, \( \{C_n\} \) achieves capacity on the BEC under bit-MAP

Important Consequences

- Reed-Muller codes achieve capacity
- Primitive narrow-sense BCH codes achieve capacity
- Affine-invariant codes achieve capacity
- Extends to block-MAP for Reed-Muller and BCH
- By Kumar-Pfister and Kudekar-Mondelli-Sasoglu-Urbanke
Few Remarks

- **Scope of the work**
  - Linear Codes, Erasure Channels, MAP Decoding

- **Buzzwords**
  - EXIT functions, monotone boolean functions, $k$-transitivity

- **Amalgamation**
  - EXIT functions (from iterative decoding)
  - Automorphism/Permutation groups (from algebraic coding)
  - Monotone boolean functions (from computer science)

- **Why do they achieve capacity?**
Proof
Basic Setup

▶ Binary linear code $\mathcal{C} \subset \{0, 1\}^N$ is a $K$-dim. subspace of $\mathbb{F}_2^N$

▶ Binary Erasure Channel, parametrized by $p$

\[
\begin{align*}
X = \{0, 1\} & \quad \rightarrow \quad \{0, 1, *\} = Y \\
0 & \overset{1-p}{\underset{p}{\rightarrow}} 0 \\
1 & \overset{1-p}{\underset{p}{\rightarrow}} 1
\end{align*}
\]

▶ $\underline{X} = (X_1, \ldots, X_N) \leftrightarrow$ uniform codeword from $\mathcal{C}$

▶ $\underline{Y} = (Y_1, \ldots, Y_N) \leftrightarrow$ received sequence from $\underline{X}$
MAP Decoding on Erasure Channels

Set of Consistent Codewords

\[ C(y) = \{ x \in C \mid x_i = y_i \text{ when } y_i \neq * \} \]
MAP Decoding on Erasure Channels

Set of Consistent Codewords

\[ C(y) = \{ x \in C \mid x_i = y_i \text{ when } y_i \neq * \} \]

MAP Decoding of bit \( X_i \)

\[ |C(y)| = 1 \iff \text{one can recover codeword } X \]
MAP Decoding on Erasure Channels

Set of Consistent Codewords

\[ C(y) = \{ x \in C \mid x_i = y_i \text{ when } y_i \neq * \} \]

MAP Decoding of bit \( X_i \)

- \( |C(y)| = 1 \iff \text{one can recover codeword } X \)
- If \( x_i \) is the same for all \( x \in C(y) \) \( \iff \text{one can recover } X_i \)
  - \( H(X_i | Y = y) = 0 \)
MAP Decoding on Erasure Channels

Set of Consistent Codewords

\[ C(y) = \{ x \in C \mid x_i = y_i \text{ when } y_i \neq \ast \} \]

MAP Decoding of bit \( X_i \)

- \( |C(y)| = 1 \iff \text{one can recover codeword } X \)
- If \( x_i \) is the same for all \( x \in C(y) \iff \text{one can recover } X_i \)
  - \( H(X_i|Y = y) = 0 \)
- Otherwise
  - Half of codewords in \( C(y) \) have \( x_i = 0 \) and half have \( x_i = 1 \)
  - uniform codeword \( \iff \) uniform posterior
  - \( H(X_i|Y = y)) = 1 \)
MAP Decoding on Erasure Channels

Set of Consistent Codewords

\[ C(y) = \{ x \in C \mid x_i = y_i \text{ when } y_i \neq \ast \} \]

MAP Decoding of bit \( X_i \)

- \(|C(y)| = 1 \iff \text{one can recover codeword } X\)
- If \( x_i \) is the same for all \( x \in C(y) \) \( \iff \text{one can recover } X_i\)
  - \( H(X_i | Y = y) = 0 \)
- Otherwise
  - Half of codewords in \( C(y) \) have \( x_i = 0 \) and half have \( x_i = 1 \)
  - uniform codeword \( \iff \text{uniform posterior} \)
  - \( H(X_i | Y = y) \) = 1
- \( H(X_i | Y = y) \) is either 0 or 1 (Boolean)
MAP Decoding on Erasure Channels: Prob. of Bit-Error

Error Prob. of bit $X_i$

- Bit-MAP decoder $D_i: \mathcal{Y}^N \rightarrow \mathcal{X} \cup \{\ast\}$
- Error prob. of bit $i$: $P_{b,i} = \Pr(D_i(Y) = \ast)$
- Average bit error prob. $P_b = \frac{1}{N} \sum_i P_{b,i}$
MAP Decoding on Erasure Channels: Prob. of Bit-Error

Error Prob. of bit $X_i$

- Bit-MAP decoder $D_i : \mathcal{Y}^N \rightarrow \mathcal{X} \cup \{\ast\}$
- Error prob. of bit $i$: $P_{b,i} = \Pr(D_i(Y) = \ast)$
- Average bit error prob. $P_b = \frac{1}{N} \sum_i P_{b,i}$

Error Prob. as Entropy

$$H(X_i|Y) = \sum_y \Pr(Y = y)H(X_i|Y = y) = P_{b,i}$$

$$P_{b,i}(p) = H(X_i|Y) \quad P_b(p) = \frac{1}{N} \sum_i H(X_i|Y)$$

Implicit parametrization by channel erasure probability $p$
Capacity-Achieving Codes on Erasure Channels

Suppose \( \{C_n\} \) is a sequence of codes with rates \( r_n \to r \in (0, 1) \)

If \( P_b^{(n)}(p) \to 0 \) for all \( p < 1 - r \),

then \( \{C_n\} \) is Capacity-Achieving

Remarks

▶ \( C \) has length \( N \), \( K \) info bits, and \( N - K \) parity bits

▶ Rate \( r = \frac{K}{N} \) and redundancy \( 1 - r = \frac{N - K}{N} \)

▶ Must correct almost all patterns with fraction \( 1 - r \) erasures

▶ Strong Requirement!
Capacity-Achieving Codes on Erasure Channels

Suppose \( \{C_n\} \) is a sequence of codes with rates \( r_n \to r \in (0, 1) \)

If \( P_b^{(n)}(p) \to 0 \) for all \( p < 1 - r \),

then \( \{C_n\} \) is Capacity-Achieving

Remarks

- \( C \) has length \( N \), \( K \) info bits, and \( N - K \) parity bits
- Rate \( r = K/N \) and redundancy \( 1 - r = (N - K)/N \)
- Must correct almost all patterns with fraction \( 1 - r \) erasures
- Strong Requirement!
EXtrinsic Information Transfer Function

- A popular tool in the iterative decoding community
- In 1999, introduced by ten Brink to visualize iterative decoding
- In 2003, formalized by Ashikhmin, Kramer, ten Brink
**MAP EXIT Functions**

![Diagram showing the process of Map EXIT Functions](image)

**EXtrinsic Information Transfer Function**

- A popular tool in the iterative decoding community
- In 1999, introduced by ten Brink to visualize iterative decoding
- In 2003, formalized by Ashikhmin, Kramer, ten Brink

**Definition**

- **(Bit-i EXIT Function)**
  \[ h_i(p) = H(X_i \mid \underline{Y}_{\sim i}) \]

- **(Average EXIT Function)**
  \[ h(p) = \frac{1}{N} \sum_i h_i(p) \]

- \( \underline{Y}_{\sim i} = (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_N) \)

- Parameterized by channel erasure probability \( p \)
EXIT Functions: Bit-Erasure Probability

\[ h_i(p) = H(X_i \mid Y_{\sim i}) \]

\[ P_{b,i} = H(X_i \mid Y) \]
EXIT Functions: Bit-Erasure Probability

\[ h_i(p) = H(X_i \mid \underline{Y}_i) \]

\[ P_{b,i} = H(X_i \mid \underline{Y}) \]

\[ H(X_i \mid \underline{Y}) = \Pr(Y_i = \ast) H(X_i \mid \underline{Y}_i, Y_i = \ast) \]
\[ + \Pr(Y_i = X_i) H(X_i \mid \underline{Y}_i, Y_i = X_i) \]
\[ = pH(X_i \mid \underline{Y}_i) \]
EXIT Functions: Bit-Erasur Probability

\[ h_i(p) = H(X_i \mid Y_{\sim i}) \]

\[ P_{b,i} = H(X_i \mid Y) \]

\[ H(X_i \mid Y) = \Pr(Y_i = *) H(X_i \mid Y_{\sim i}, Y_i = *) \]
\[ + \Pr(Y_i = X_i) H(X_i \mid Y_{\sim i}, Y_i = X_i) \]
\[ = pH(X_i \mid Y_{\sim i}) \]

\[ P_{b,i}(p) = ph_i(p) \]
\[ P_b(p) = ph(p) \]
EXIT Functions: Area Theorem

\[ h_i(p) = H(X_i | Y_{\sim i}) \quad h(p) = \frac{1}{N} \sum_i h_i(p) \]
EXIT Functions: Area Theorem

Msg. $\rightarrow C(N, K) \rightarrow X \rightarrow \text{BEC}(p) \rightarrow Y$

$h_i(p) = H(X_i|Y_{\sim i})$

$h(p) = \frac{1}{N} \sum_i h_i(p)$

Area Theorem

$\int_0^1 h(p) dp = K / N$

- Conservation Principle
- Not satisfied by $P_b$
Capacity and EXIT Functions

Suppose \( \{C_n\} \) is a sequence of codes with rates \( r_n \to r \)

The following are equivalent

- \( \{C_n\} \) achieves capacity

- \( h^{(n)} \to \begin{cases} 0, & \text{if } p < 1 - r, \\ 1, & \text{if } p > 1 - r. \end{cases} \)

- For all \( \varepsilon > 0 \), \( p_{1-\varepsilon}^{(n)} - p_{\varepsilon}^{(n)} \to 0 \)
Rate-1/2 Reed-Muller Codes

Average EXIT Function $h$

Erasure Probability

$N = 2^3$
Rate-1/2 Reed-Muller Codes

![Graph showing Average EXIT Function](image)

- **Average EXIT Function** $h$
- **Erasure Probability**
- **$N = 2^3$** (black curve)
- **$N = 2^5$** (blue curve)
Rate-1/2 Reed-Muller Codes

Average EXIT Function $h$

Erasure Probability

- Black: $N = 2^3$
- Blue: $N = 2^5$
- Red: $N = 2^7$
Rate-1/2 Reed-Muller Codes

![Graph showing Average EXIT Function $h$ vs. Erasure Probability with curves for different $N$ values: $N = 2^3$, $N = 2^5$, $N = 2^7$, $N = 2^9$.](image-url)
When do EXIT Functions Exhibit 0 – 1 Transition?
EXIT Function as Measure of a Set $\Omega_i$

Set of bad erasures that prevent recovery of $X_i$ from $Y_{\sim i}$

$$\Omega_i \triangleq \{ z_{\sim i} \in \{0, 1\}^{N-1} \mid \exists x \in C, x_i = 1, x_{\sim i} \leq z_{\sim i} \}$$
EXIT Function as Measure of a Set $\Omega_i$

Set of bad erasures that prevent recovery of $X_i$ from $Y_{\sim i}$

$$\Omega_i \triangleq \{ z_{\sim i} \in \{0, 1\}^{N-1} | \exists x \in C, x_i = 1, x_{\sim i} \leq z_{\sim i} \}$$

$$h_i(p) = H(X_i \mid Y_{\sim i}) = \sum_{Y_{\sim i}} \Pr(Y_{\sim i} = y_{\sim i})H(X_i \mid Y_{\sim i} = y_{\sim i})$$

$$= \sum_{z_{\sim i} \in \Omega_i} p^{|z_{\sim i}|} (1 - p)^{N-1-|z_{\sim i}|}$$
EXIT Function as Measure of a Set $\Omega_i$

Set of bad erasures that prevent recovery of $X_i$ from $Y_{\sim i}$

$$\Omega_i \triangleq \{ z_{\sim i} \in \{0, 1\}^{N-1} \mid \exists x \in C, x_i = 1, x_{\sim i} \leq z_{\sim i}\}$$

$$h_i(p) = H(X_i \mid Y_{\sim i}) = \sum_{Y_{\sim i}} \Pr(Y_{\sim i} = y_{\sim i})H(X_i \mid Y_{\sim i} = y_{\sim i})$$

$$= \sum_{z_{\sim i} \in \Omega_i} p^{|z_{\sim i}|}(1 - p)^{N-1-|z_{\sim i}|}$$

$$= \mu_p(\Omega_i)$$

$$\mu_p(\Omega) \triangleq \sum_{a \in \Omega} p^{|a|}(1 - p)^{N-1-|a|}$$
EXIT Function and Monotone Boolean Functions

\[ h_i(p) = \mu_p(\Omega_i) \]

\[ \Omega_i \leftrightarrow \text{Set of bad erasures} \]

Adding erasures only worsens recoverability
EXIT Function and Monotone Boolean Functions

\[ h_i(p) = \mu_p(\Omega_i) \]

\( \Omega_i \leftrightarrow \text{Set of bad erasures} \)

Adding erasures only worsens recoverability

\( \Omega_i \) is Monotone

If \( a \in \Omega_i \) and \( a \leq b \), then \( b \in \Omega_i \)
EXIT Function and Monotone Boolean Functions

\[ h_i(p) = \mu_p(\Omega_i) \]

\( \Omega_i \leftrightarrow \) Set of bad erasures

Adding erasures only worsens recoverability

\( \Omega_i \) is Monotone

If \( a \in \Omega_i \) and \( a \leq b \), then \( b \in \Omega_i \)

\( h_i \) is Monotone Boolean
When do EXIT Functions Exhibit $0 - 1$ Transition?
When do EXIT Functions Exhibit 0 – 1 Transition?

Path Ahead: Symmetric Monotone Boolean Functions Exhibit Sharp 0 – 1 Transitions
Avg. EXIT Function $h$, not $h_i$, satisfies Area Theorem

Bit-i EXIT Function $h_i$, not $h$, is Monotone Boolean
Avg. EXIT Function $h$, not $h_i$, satisfies Area Theorem

Bit-i EXIT Function $h_i$, not $h$, is Monotone Boolean

What about symmetry?
Group Symmetry

The **Permutation Group** $\mathcal{G}$ of code $\mathcal{C}$ is defined as

$$\mathcal{G} = \{ \pi \in S_N \mid \pi(x) \in \mathcal{C} \quad \forall \ x \in \mathcal{C} \}$$
Group Symmetry

The Permutation Group $\mathcal{G}$ of code $\mathcal{C}$ is defined as

$$\mathcal{G} = \{\pi \in S_N | \pi(x) \in \mathcal{C} \ \forall \ x \in \mathcal{C}\}$$

Transitive Permutation Groups

$\mathcal{G}$ is transitive if for all $i, j$, $\exists \pi \in \mathcal{G}$ such that $\pi(i) = j$  

$\mathcal{G}$ is doubly transitive if for all distinct $i, j, k$, $\exists \pi \in \mathcal{G}$ such that $\pi(i) = i$ and $\pi(j) = k$
EXIT Functions Under Group Symmetry

Proposition

- If $\mathcal{G}$ is transitive

  \[ h_i(p) = h_j(p) = h(p) \quad \text{for all } 0 \leq p \leq 1 \]

- If $\mathcal{G}$ is doubly transitive

  $\Omega_i$ is invariant under a transitive permutation group
Under double transitivity: $h_i = h$ and $\Omega_i$ is transitive
EXIT Functions Under Double Transitivity

Under double transitivity: \( h_i = h \) and \( \Omega_i \) is transitive

Symmetric Monotone Boolean Functions Exhibit Sharp 0 – 1 Transitions
EXIT Functions Under Double Transitivity

Under double transitivity: $h_i = h$ and $\Omega_i$ is transitive

Symmetric Monotone Boolean Functions Exhibit Sharp $0 - 1$ Transitions

Avg. EXIT Function $h$ must exhibit a sharp $0 - 1$ transition!
(Symmetric) Monotone Boolean Functions invariant under Transitive Permutation Group

Bernoulli($p$) Product Measure $\mu_p$ on $\{0, 1\}^N$

$$f : \{0, 1\}^N \rightarrow \{0, 1\}, \quad h(p) = \mathbb{E}_{\mu_p} [f].$$
(Symmetric) Monotone Boolean Functions invariant under Transitive Permutation Group

Bernoulli($p$) Product Measure $\mu_p$ on $\{0, 1\}^N$

$$f : \{0, 1\}^N \rightarrow \{0, 1\}, \quad h(p) = \mathbb{E}_{\mu_p} [f].$$

**Popular Theorem in TCS**

- Shown in early 1990s
- By Friedgut-Kalai, Talagrand, Bourgain-Kahn-Kalai-Linial
- Below, $p_t = h^{-1}(t)$

$$p_{1-\varepsilon} - p_\varepsilon \leq 2C \frac{\log \frac{1}{\varepsilon}}{\log N}, \quad p_{1-\varepsilon} - p_\varepsilon \rightarrow 0.$$
1. RM codes achieve capacity at all rates (under MAP decoding.)

2. Let $X^n = (X_1, X_2 \ldots X_n)$ be iid $\text{Bern}(\frac{1}{2})$. 

[Checkmark]
Other Symmetric Monotone Boolean Functions

Monotone graph properties
(i) arguments to function indicate which edges present
(ii) invariance under relabeling of vertices gives symmetry

Hamming weight greater than $r$
Clearly symmetric and monotone
Capacity via Symmetry

Generality

- How general is this phenomenon?
- Proof heavily exploits MAP decoding on erasure channels
- Abbe et al. have shown for BSC when rate $\to 0$

Open Questions

- Extension to general BMS channels
- Practical decoders that achieve capacity for non-trivial rates
- Extension to rates converging to 0 or 1 (ala Friedgut)
- Capacity-achieving schemes for other systems?
  - Quantum codes, Rate-Distortion, Compressed Sensing