Coding on Dual-Parameter Barrier Channels beyond Worst-Case Correction

Yuval Ben-Hur and Yuval Cassuto
The Andrew and Erna Viterbi Department of Electrical and Computer Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 3200003, Israel
Email: {yuvalbh@campus, ycassuto@ee}.technion.ac.il

***Abstract*—This paper studies coding on channels with the barrier property: only errors to and from a special barrier state are possible. This model is motivated by storage media that have heterogeneous state structure, not admitting the usual multi-bit scaling of the representation states. Our contributions include derivation of the channel capacity, efficient maximum-likelihood and list decoding algorithms, and finite-block-length analysis using random codes. This work is the first that addresses a barrier channel with separate parameters for the transitions into and out of the barrier state. Earlier work addressed special-case single-parameter models, and focused primarily on the worst-case coding performance.**

I. Introduction

A key property of next-generation solid-state data-storage devices, sought by researchers, manufacturers and customers, is extreme information density. The race toward denser storage runs two main scaling trajectories: (1) packing more memory cells per unit area, and (2) increasing the number of representation levels per memory cell. However, both scaling mechanisms suffer from reliability challenges: higher cell density increases read/write noise+interference, and more representation levels shrink the noise margins and slow down read and write. Industry commonly combines these two mechanisms through the *multi-bit cell* principle, in which an integer power of 2 number of levels are uniformly spaced across the cell's read/write dynamic range.

Some emerging storage media may be able to span multiple representation levels/states, but without the regularity needed for the multi-bit scaling principle. Hence, in this paper we study a different representation-scaling principle we call *barrier multi-level cell*. This principle requires a media technology in which one of the Q levels is designated as the *barrier level*, and the dominant error transitions are from non-barrier levels to the barrier, and from the barrier to non-barrier levels. A sample motivation for the barrier model for the ternary case ($Q = 3$) is shown in Fig. 1, where the barrier level is added in between two well-separated levels. The barrier model may also be useful for magnetic memories with $Q > 3$ (see, for example [1]), if one magnetic state is more "stable", and thus attracts most of the errors from the other states. Another motivating example is next-generation electrically erasable programmable read-only memories (EEPROMs), where information can be stored in three levels, but transitions between the highest level and the lowest level are physically not possible [3].

Toward reliable storage over such media, this paper advances the study of coding over *barrier channels*. In the sequel, a barrier channel is defined by two parameters: p and q, specifying the transition probabilities *into* and *out of* the barrier state, respectively (see Fig. 2). We call this channel the *dual-parameter barrier channel*. Two special cases of the dual-parameter barrier channel have already been studied in prior work: [2] addresses the one-directional case ($q = 0$), and [3] studies another single-parameter barrier channel in which $p = q/2$. Since practical channels may not fall into one of these special cases, it is motivated to study the general p,q case. For example, the channel described in Fig. 1 has p and q values that fit neither of the prior models. Another contribution of the present work is extending error-correction capabilities beyond the guaranteed-correction regime, which was the focus of the prior work.

The formal definition of the dual-parameter barrier channel is now given:

Definition 1: Let $\mathbb{Z}_Q \triangleq \{0, \ldots, Q-1\}$. For any input $X \in \mathbb{Z}_Q$ and output $Y \in \mathbb{Z}_Q$ and parameters $0 \leq p, q \leq 1$, the Q-ary dual-parameter barrier channel $W_Q(p, q)$ has the transition

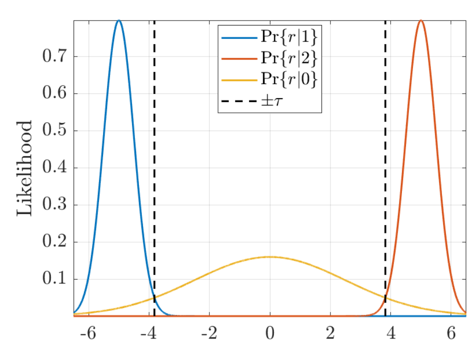

Fig. 1: Sample cell-level distribution of a ternary media motivating the barrier model. The corresponding barrier parameters are $p = 9 \cdot 10^{-3}$, $q = 1.3 \cdot 10^{-1}$. The transition probability between the extreme levels is negligible.

probabilities

$$P(Y|X) = \begin{cases} 1-p, & Y = X, \quad X \in \{1, Q-1\} \\ p, & Y = 0, \quad X \in \{1, Q-1\} \\ 1-q, & Y = X, \quad X = 0 \\ q/(Q-1), & Y \neq X, \quad X = 0 \\ 0, & otherwise \end{cases} \tag{1}$$

The important special case of $Q = 3$ is denoted $W(p,q) \triangleq W_3(p,q)$, and is called the *ternary* dual-parameter barrier channel. A diagram describing $W(p,q)$ is given in Fig. 2.

The rest of the paper is organized as follows. In Section II, we derive the capacity of $W(p,q)$. In Section III, we derive a maximum-likelihood decoder (MLD) for codes constructed by the state-of-the-art method of [3]. The new MLD is more efficient than the prior MLD [3] thanks to a reduction to MLD of lower-alphabet classical codes in the Hamming metric. We then propose two more tractable alternatives to the new MLD, based on list decoding of classical codes. Section IV demonstrates the empirical performance of the new decoders, and develops a method for analyzing the performance using finite block-length random linear codes.

II. Channel Capacity of $W(p,q)$

We start the treatment of the dual-parameter barrier channel by deriving its channel capacity. The derivation is given for the special case $Q = 3$, but can be extended to the general case. This generalizes the capacity of the single-parameter special case $W(q/2, q)$, derived in [3].

To simplify the following expressions, we define the functions $\beta_{p,q}(\varphi) \triangleq q + (1-p-q)\varphi$ and $\gamma(a) \triangleq h_2(a) + a$, where $h_2(\cdot)$ is the binary entropy function.

Theorem 1: The capacity of $W(p,q)$, for $p+q < 1$, is given by

$$\gamma\left(\beta_{p,q}(\varphi^*)\right) - \varphi^* h_2(p) - (1-\varphi^*)\gamma\left(q\right), \tag{2}$$

where $\varphi^* \triangleq \min\left[\frac{1-q-\left(1+2^{-\frac{\gamma(p)-1-h_2(q)}{1-p-q}}\right)^{-1}}{1-p-q}, 1\right]$.

Before we prove the theorem, we show that the capacity-achieving input distribution (CAID) of the channel has the symmetry $\Pr\{X = 1\} = \Pr\{X = 2\}$.

Lemma 1: Let Φ^*_X be the CAID of $W(p,q)$ with arbitrary p, q. Then there exists $0 \leq \varphi \leq 1$ such that

$$\Phi^*_X(x) = \begin{cases} \varphi/2, & x \neq 0 \\ 1-\varphi, & x = 0 \end{cases} \tag{3}$$

Proof: Let Φ_X be an input distribution, and define the compact notations $\varphi_x \triangleq \Phi_X(x)$. Denote $\varphi \triangleq \varphi_1 + \varphi_2$. Since $H(Y|X) = (1-\varphi)(h_2(q) + q) + \varphi h_2(p)$ is a function of φ alone, the claim follows by proving that $H(Y)$ is maximized with $\varphi_1 = \varphi_2$. Define Z as an indicator of the event $Y \neq 0$. Since $H(Y) = H(Z) + H(Y|Z)$, while $H(Z)$ is a function of φ alone, it remains to maximize $H(Y|Z)$. Now,

$$H(Y|Z) = \Pr\{Z = 1\} H(Y|Z = 1), \tag{4}$$

which is maximized when $\Pr\{Y = 1\} = \Pr\{Y = 2\}$. Since $X = 1$ and $X = 2$ have the same transition probability p, setting $\varphi_1 = \varphi_2$ results in $\Pr\{Y = 1\} = \Pr\{Y = 2\}$. ■

We now prove Theorem 1.

Proof: Let Φ^*_X be the CAID as defined in Lemma 1. The capacity is given by $H(Y) - H(Y|X)$. Calculating $\Pr\{Y = y\}$ for every $y \in \{0,1,2\}$ leads directly to $H(Y) = h_2(1 - \beta_{p,q}(\varphi^*)) + \beta_{p,q}(\varphi^*)$. Due to the symmetry of $h_2(\cdot)$ around $1/2$, we can substitute $1 - \beta_{p,q}(\varphi^*)$ with $\beta_{p,q}(\varphi^*)$ in its argument, and get $H(Y) = \gamma\left(\beta_{p,q}(\varphi^*)\right)$. It is also straightforward to see that the conditional entropy is $H(Y|X) = \varphi^* h_2(p) + (1-\varphi^*)\gamma(q)$.

It remains to find φ^*. Calculating the derivative of the mutual information $\mathcal{I}(X;Y)$ with respect to φ, we get

$$\frac{d}{d\varphi}\mathcal{I}\left(X;Y\right) = (1-q-p)\log\left(\frac{1-\beta_{p,q}(\varphi)}{\beta_{p,q}(\varphi)}\right) + 1 + h_2(q) - \gamma(p). \tag{5}$$

It can be observed that $\mathcal{I}(X;Y)$ is concave in φ, because the derivative of the log in (5) is negative from the fact that $\beta_{p,q}(\varphi)$ is monotone increasing. Since $\mathcal{I}\left(X;Y\right) = 0$ for $\varphi = 0$, φ^* equals the maximum point φ^{MAX} of $\mathcal{I}\left(X;Y\right)$ if $\varphi^{\text{MAX}} \leq 1$, and $\varphi^* = 1$ otherwise. By equating (5) to 0, we get

$$\varphi^{\text{MAX}} = \frac{1-q-\left(1+2^{-\frac{\gamma(p)-1-h_2(q)}{1-p-q}}\right)^{-1}}{1-p-q}, \tag{6}$$

and the theorem statement follows. ■

Remark 1: The capacity can also be derived with similar arguments for the (less interesting) case of $p+q > 1$, but we omit this derivation. In the special case of $p+q = 1$, $\beta_{p,q}(\varphi)$ does not depend on φ, which leads to $\varphi^* = 1$ for every p, and capacity of $1-p$.

Fig. 3 depicts the channel capacity as a function of $p+q$, for several relations between p and q. Note that the channel obtained by not using the barrier level is simply BEC(p). It can be seen in Fig. 3 that its capacity is significantly lower than the capacity of ternary barrier channels with various values of p and q. The plot also motivates specifically the study of the *dual*-parameter version of the channel, by showing that the known special case $p = q/2$ has large gaps (both upward and downward) to other potentially interesting cases.

III. Decoding Algorithms

A. Preliminaries and known results

We first define the type of errors occurring in barrier channels, assuming state 0 is the barrier state.

Definition 2: Let $\mathbf{c} \in \mathbb{Z}_Q^n$ be a codeword sent over the barrier channel, and $\mathbf{r} \in \mathbb{Z}_Q^n$ be the word received at the channel output. A barrier error occurred in index i if one of the following holds:

1) $c_i = 0$ and $r_i \neq 0$.
2) $c_i \neq 0$ and $r_i = 0$.

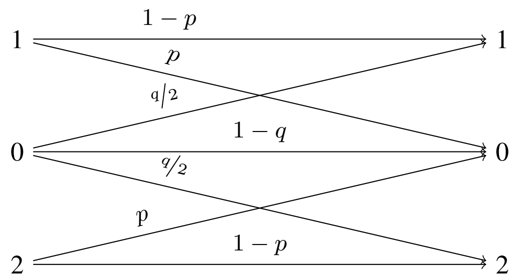

Fig. 2: The dual-parameter barrier channel for $Q = 3$.

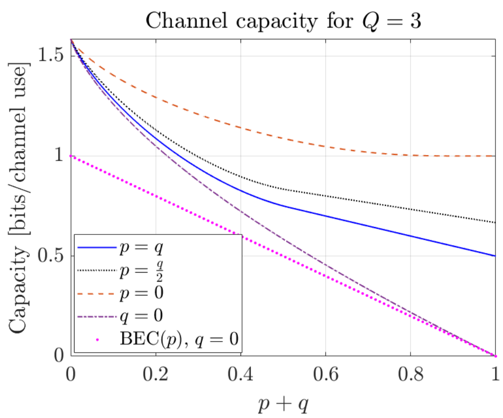

Fig. 3: Capacity of the ternary dual-parameter barrier channel $W(p,q)$, for several relations between p and q.

In this subsection we review the important results previously derived on correction of barrier errors. [2] derived the size of codes that guarantee correction of t barrier errors, and proposed a construction for $t = 1$. The main contribution of [3] is a construction method for ternary codes correcting t barrier errors, using a pair of binary Hamming-metric constituent codes. We review this construction method for completeness of the presentation. Let us construct each Q-ary codeword $\mathbf{c} \in \mathcal{C}$ based on codewords from two codes: (1) $\boldsymbol{\theta} \in \Theta$ where Θ is a binary code with length n and minimum *Hamming* distance $2t+1$; and (2) $\boldsymbol{\lambda} \in \Lambda_{w_H(\boldsymbol{\theta})}$ where Λ_w is a $(Q-1)$-ary code with length w and minimum *Hamming* distance $t+1$.

The locations of zeros in the Q-ary codeword $\mathbf{c}$ are determined by the locations of zeros in $\boldsymbol{\theta}$. That is,

$$c_i = 0, \forall i : \theta_i = 0. \tag{7}$$

For the remaining positions, symbols are set according to $\boldsymbol{\lambda}$:

$$c_i = \psi\left(\lambda_{\sigma_i(\boldsymbol{\theta})}\right), \forall i : \theta_i = 1, \tag{8}$$

where $\sigma_i(\mathbf{x})$ is the number of indices $i' \leq i$ that have non-zeros in $\mathbf{x}$, and $\psi(x_i) = x_i + 1$.

A code constructed as above is denoted $\mathcal{C} = \Theta \otimes \{\Lambda_w\}_w$, and it is proved [3] to guarantee correction of up to t barrier errors.

B. Reduction of MLD to MLD of the simpler constituent codes

In [3], an ML decoder (for the special case $p = q/2$) is defined through a distance metric on the ternary alphabet $\mathbb{Z}_3$. For more efficient decoding, we show in this subsection a reduction of ML decoding to simpler ML decoders of the constituent codes Θ and Λ_w. The advantage of this reduction is that the constituent codes are traditional Hamming-metric codes over lower-order alphabets (Θ is a binary code, and when $Q = 3$ so are Λ_w). Let $\mathbf{r} = (r_1, \dots, r_n) \in \mathbb{Z}_Q^n$ be the word output from the channel $W(p,q)$. An MLD for a code $\mathcal{C} = \Theta \otimes \{\Lambda_w\}_w$ needs to find a pair of codewords $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{\lambda} \in \Lambda_{w_H(\boldsymbol{\theta})}$ that jointly maximize the likelihood of observing $\mathbf{r}$. We now decompose this task using individual decoders for Θ, Λ_w, and a rule for combining the individual decoder outputs.

1) MLD for the code Θ: We first define a mapping from the channel alphabet to the binary alphabet of Θ.

Definition 3: Let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{Z}_Q^n$. The **indicator mapping** $\imath(\mathbf{x}) = (\imath(x_1), \dots, \imath(x_n))$ is defined as

$$\imath(x_j) = \begin{cases} 1, & x_j \in \mathbb{Z}_Q \setminus \{0\} \\ 0, & x_j = 0 \end{cases}. \tag{9}$$

Now we define the following decoder for Θ, invoked on the input $\imath(\mathbf{r})$.

Definition 4: The **ML indicator decoder** for Θ outputs the codeword

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta}\{\mu_1 \imath(\mathbf{r})\, \boldsymbol{\theta}^T - \mu_2 w_H(\boldsymbol{\theta})\}, \tag{10}$$

where $\mu_1 \triangleq \log\left(\frac{(Q-1)(1-p)(1-q)}{pq}\right)$ and $\mu_2 \triangleq \log\left(\frac{1-q}{p}\right)$. We will later show that $\hat{\boldsymbol{\theta}}$ in (10) maximizes the indicator vector's likelihood function $\Pr\{\imath(\mathbf{r})\,|\boldsymbol{\theta}\}$, hence the ML qualification in Definition 4.

2) MLD for the codes Λ_w: We first define a mapping from the length-n channel output to the decoder input.

Definition 5: Given $\boldsymbol{\theta} \in \{0,1\}^n$, the **residual mapping** maps a vector $\mathbf{r} \in \mathbb{Z}_Q^n$ to a vector $\rho^{(\boldsymbol{\theta})}(\mathbf{r}) = \left(\rho_1^{(\boldsymbol{\theta})}(\mathbf{r}), \dots, \rho_{w_H(\boldsymbol{\theta})}^{(\boldsymbol{\theta})}(\mathbf{r})\right)$ such that for every $1 \leq j \leq n$ with $\theta_j = 1$,

$$\rho_{\sigma_j(\boldsymbol{\theta})}^{(\boldsymbol{\theta})}(\mathbf{r}) = \begin{cases} ?, & r_j = 0 \\ r_j - 1, & \textit{otherwise} \end{cases} \tag{11}$$

Note that the alphabet of the elements of $\rho^{(\boldsymbol{\theta})}(\mathbf{r})$ is $\{\mathbb{Z}_{Q-1} \cup \{?\}\}$. Now we define the following decoder for $\Lambda_{w_H(\boldsymbol{\theta})}$, invoked on the input $\rho^{(\boldsymbol{\theta})}(\mathbf{r})$.

Definition 6: Given $\boldsymbol{\theta} \in \{0,1\}^n$, the **ML residual decoder** for $\Lambda_{w_H(\boldsymbol{\theta})}$ first finds all the codewords $\boldsymbol{\lambda}$ such that $\lambda_i = \rho_i^{(\boldsymbol{\theta})}(\mathbf{r})$ for every i with $\rho_i^{(\boldsymbol{\theta})}(\mathbf{r}) \neq ?$. Then the decoder outputs $\hat{\boldsymbol{\lambda}}$ if unique, "fail" if multiple codewords were found, and "reject" if none were found.

3) Combining the individual MLDs: Let $\mathbf{r}$ be the channel output. We are now ready to define the combined decoder. Note that $\Psi(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})$ in the last line marks the mapping performed by the construction in (7),(8).

Algorithm 1 MLD for $\mathcal{C} = \Theta \otimes \{\Lambda_w\}_w$:

<u>Input</u> : $\mathbf{r} \in \mathbb{Z}_Q^n$ - channel output
<u>Output</u> : $\hat{\mathbf{c}} \in \mathbb{Z}_Q^n$ - decoded codeword
Initialize: $\Theta' \leftarrow \Theta$
while not returned **do**
  set $\hat{\boldsymbol{\theta}}$ to the output of indicator MLD with input $\imath(\mathbf{r})$ and code Θ'
  invoke residual MLD with input $\rho^{(\hat{\boldsymbol{\theta}})}(\mathbf{r})$ and code $\Lambda_{w_H(\hat{\boldsymbol{\theta}})}$
  if "reject" **then**
    $\Theta' \leftarrow \Theta' \setminus \hat{\boldsymbol{\theta}}$
  else if "fail" **then**
    return *decoding failure*
  else
    return $\hat{\mathbf{c}} = \Psi(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})$
  end if
end while

---

4) Proving the ML property of Algorithm 1: Given a channel output $\mathbf{r}$ and a candidate codeword $\boldsymbol{c}$, we can partition the n coordinates to 5 disjoint sets: $\mathcal{S}_0$ where $r_i = c_i = 0$, $\mathcal{S}_1$ where $r_i = c_i \neq 0$, $\mathcal{S}_2$ where $r_i = 0, c_i \neq 0$, $\mathcal{S}_3$ where $r_i \neq 0, c_i = 0$, $\mathcal{S}_4$ where $r_i \neq 0, c_i \neq 0, r_i \neq c_i$. Thus the ML codeword is the one that maximizes $\sum_{s=0}^{3} \ell_s |\mathcal{S}_s|$, subject to $|\mathcal{S}_4| = 0$, where $\ell_0 = \log(1-q)$, $\ell_1 = \log(1-p)$, $\ell_2 = \log(p)$, and $\ell_3 = \log(q/(Q-1))$. This follows from the channel definition and taking the log of the likelihood function $\Pr\{\mathbf{r}|\mathbf{c}\}$.

Proposition 1: For any $\mathbf{r}$, let $\hat{\mathbf{c}}^{(ML)}(\mathbf{r})$ be a unique ML codeword from $\mathcal{C} = \Theta \otimes \{\Lambda_w\}_w$. Then the output of Algorithm 1 equals $\hat{\mathbf{c}}^{(ML)}(\mathbf{r})$.

Proof: We first prove that the indicator decoder maximizes $\sum_{s=0}^{3} \ell_s |\mathcal{S}_s| + \ell_1 |\mathcal{S}_4|$. This can be seen by substituting in this sum the identities $|\mathcal{S}_1| + |\mathcal{S}_4| = \imath(\mathbf{r})\boldsymbol{\theta}^T$, $|\mathcal{S}_3| = w_H(\imath(\mathbf{r})) - \imath(\mathbf{r})\boldsymbol{\theta}^T$, $|\mathcal{S}_2| = w_H(\boldsymbol{\theta}) - \imath(\mathbf{r})\boldsymbol{\theta}^T$, and $|\mathcal{S}_0| = n - w_H(\imath(\mathbf{r})) - w_H(\boldsymbol{\theta}) + \imath(\mathbf{r})\boldsymbol{\theta}^T$. By expressing all set sizes as functions of $w_H(\boldsymbol{\theta})$, $\imath(\mathbf{r})\boldsymbol{\theta}^T$, and ignoring all terms that do not depend on $\boldsymbol{\theta}$, we get (10).

Next observe that for any $\mathbf{c} = \Psi(\boldsymbol{\theta}, \boldsymbol{\lambda})$, $|\mathcal{S}_4| = 0$ if and only if $\boldsymbol{\lambda}$ satisfies the condition in Definition 6 that $\lambda_i = \rho_i^{(\boldsymbol{\theta})}(\mathbf{r})$ for every i with $\rho_i^{(\boldsymbol{\theta})}(\mathbf{r}) \neq ?$. To complete the proof, assume that the pair $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}$ output by Algorithm 1 is different from the unique ML pair $\hat{\boldsymbol{\theta}}^{(ML)}, \hat{\boldsymbol{\lambda}}^{(ML)}$. Case 1 $\hat{\boldsymbol{\theta}} \neq \hat{\boldsymbol{\theta}}^{(ML)}$: since both Q-ary codewords have $|\mathcal{S}_4| = 0$ (the former by not being rejected and the latter by the Q-ary ML condition), a higher Q-ary likelihood implies a higher value in (10); a contradiction. Case 2 $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(ML)}$: also cannot happen because it implies that $\Psi(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}^{(ML)})$, $\Psi(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})$ are two distinct ML codewords, in contradiction to uniqueness. Similarly, the uniqueness also excludes the case that Algorithm 1 returns decoding failure. ■

The joint ML decoder in Algorithm 1 improves over the separate bounded-distance decoders suggested in [3] for $\Theta \otimes \{\Lambda_w\}_w$ codes. Its main advantage is that the two decoders cooperate, allowing the residual decoder to reject wrong Θ codewords with high indicator likelihoods. This feature is important when the errors exceed the unique-decoding capabilities of Θ.

C. Cooperative List Decoding (CLD)

Algorithm 1 simplifies ML decoding compared to the Q-ary barrier ML decoder, by reduction to decoding of binary codes in the Hamming metric and erasure decoding. However, ML decoding of binary codes is still in general a computationally hard problem. To mitigate this hardness, we propose a simplification of Algorithm 1 using *list decoding*, which is a more tractable computational task than ML decoding. We define two variants of this proposition.

1) Cooperative list decoder (CLD): A list decoder for the code Θ outputs a list of likely codewords $\{\hat{\boldsymbol{\theta}}_l\}_{l=1}^{L}$, where L is the list size. Then the CLD for the code $\mathcal{C} = \Theta \otimes \{\Lambda_w\}_w$ is obtained by running Algorithm 1 with initializing $\Theta' \leftarrow \{\hat{\boldsymbol{\theta}}_l\}_{l=1}^{L}$ instead of $\Theta' \leftarrow \Theta$. This way, the indicator MLD only needs to search over the L list codewords, with the benefit that $L \ll |\Theta|$.

2) Persistent cooperative list decoder (PCLD): Identical to CLD, but instead of returning "decoding failure" when residual MLD fails, it continues to the next codeword in the list (this amounts to merging the 'else if' of Algorithm 1 into the 'if' statement, which will now be: if "reject" or "fail").

It is clear that with CLD and/or PCLD, similarly to the MLD of Algorithm 1, a decoding success can occur even if the correct $\boldsymbol{\theta}$ is not the ML codeword given the channel output's indicator word. Thanks to its persistence, PCLD may succeed in finding the correct codeword further down the list, while CLD stops the search upon residual-decoding failure. Note that neither CLD nor PCLD are equivalent to MLD, because the list decoder of Θ in general does not guarantee finding the ML $\boldsymbol{\theta}$ codeword. The special case of CLD with $L = 1$ is the previously suggested decoder based on unique decoding of Θ. Our generalization to $L > 1$ is more powerful not only thanks to more opportunities to find the most likely indicator codeword, but also in its ability to reject wrong indicator codewords using information from the residual decoder.

3) Success conditions for PCLD: The PCLD succeeds in all instances where the following conditions are met: (1) The correct codeword $\boldsymbol{\theta} \in \Theta$ is in the list $\{\hat{\boldsymbol{\theta}}_l\}_{l=1}^{L}$, (2) all other codewords in the list that have higher or equal likelihoods than

$\boldsymbol{\theta}$ given $\imath(\mathbf{r})$ reject or fail by the residual decoder, and (3) residual decoding of $\rho^{(\boldsymbol{\theta})}(\mathbf{r})$ does not return decoding failure.

IV. Performance Evaluation

A. Reed-Muller indicator codes and BCH residual codes

We evaluate the PCLD performance for ternary ($Q = 3$) barrier errors using widely adopted binary codes for Θ and Λ_w: Reed-Muller ($\mathcal{RM}$) codes for the former and modified (shortened/lengthened) BCH codes for the latter. To facilitate comparison with unique decoding, we design both codes such that $\mathcal{C} = \Theta \otimes \{\Lambda_w\}_w$ guarantees correction of t barrier errors (as in Definition 2). Toward this end, we take Θ to be a $\mathcal{RM}(r, m)$ code such that $d_{\mathcal{RM}} = 2^{m-r} \geq 2t + 1$, and construct the family of codes $\{\Lambda_w\}_{w=0}^{n}$ by shortening or lengthening primitive binary BCH codes with $d_{\text{BCH}} \geq t + 1$ to fit lengths $w \in \{0, \ldots, n\}$.

When the number of barrier errors is $t+\tau$, for several values of $0 \leq \tau \leq n-t$, we evaluate the decoding-success probability of random codewords from $\mathcal{C}$, averaged over all patterns of barrier errors. We compare the results of three decoders: (1) **unique** $\mathcal{RM}$ decoder (based on the recursive decoder in [4]) followed by BCH erasure decoding, (2) **list** $\mathcal{RM}$ decoder (based on the recursive algorithm in [5] and its implementation in [6]) followed by BCH erasure decoding of the list's closest codeword to $\imath(\mathbf{r})$ (in the Hamming metric), and (3) the **PCLD** decoder that uses the same $\mathcal{RM}$ list decoder, but iterates on the list codewords (as described in Section III-C2) in increasing Hamming distances to $\imath(\mathbf{r})$. Figure 4 compares the decoding success of the three decoders, for $\Theta = \mathcal{RM}(3, 7)$ (block length $n = 128$). It can be observed that PCLD has substantial success probability even for $\tau = 3, 4$ errors beyond the guaranteed $t = 7$, while the two other decoders degrade much faster with τ. The performance of PCLD improves with increasing L, while the non-cooperative list alternative is significantly inferior even with the maximal list size $L = 32$. This demonstrates that the power of PCLD comes from the cooperation with the residual code, and not merely from the larger list sizes.

B. Analysis with residual random linear codes (RLC)

Toward a systematic design of the residual codes $\{\Lambda_w\}_{w=0}^{n}$, we analyze the outcomes of the residual decoder assuming *random linear codes* (RLC) with a prescribed redundancy. Throughout this sub-section we focus on binary codes Λ_w (corresponding to $Q = 3$), but extension to non-binary codes is possible. Recall from Section III-C3 that two events related to the residual codes cause failure of the PCLD: 1) not rejecting or failing on a wrong codeword $\hat{\boldsymbol{\theta}}$, and 2) decoding failure of the correct codeword $\boldsymbol{\theta}$. We want to analyze the probabilities of these events when the residual codes are RLC, defined next.

Definition 7: Define $\Lambda[w, w - r]$ as a linear random code defined by a parity-check matrix H with dimensions $r \times n$, where each entry of H is an independent and identically distributed (i.i.d.) Bernoulli random variable with parameter $1/2$.

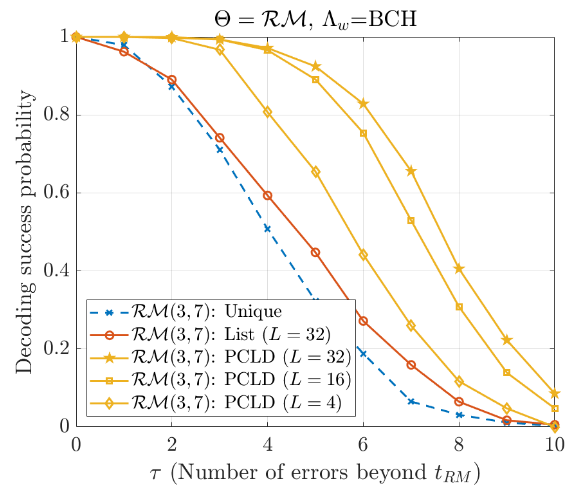

Fig. 4: $\mathcal{RM}$-BCH scheme: Probability of successful decoding

Suppose the residual decoder is invoked on $\rho^{(\boldsymbol{\theta})}(\mathbf{r})$ of the correct codeword $\boldsymbol{\theta}$. Then the probability of it returning decoding failure is found by the following well known result.

Proposition 2: Let $\Lambda[w, w - r]$ be a random linear code, and $\boldsymbol{\lambda} \in \Lambda[w, w - r]$ be any codeword. The failure probability of ML-decoding $\boldsymbol{\lambda}$ with $1 \leq e \leq r$ erasures, denoted $f_1(r, e)$, is

$$f_1(r, e) = 1 - \prod_{i=0}^{e-1} \left(1 - 2^{i-r}\right). \tag{12}$$

Proof: Immediate from evaluating the probability that the erasure-positions sub-matrix of H is full rank, and taking the complement. See e.g. Eq. (3.2) in [7]. ■

For PCLD failures due to a wrong codeword $\hat{\boldsymbol{\theta}}$ not rejected or failed by the residual decoder, we have the following.

Proposition 3: Let $\Lambda[w, w - r]$ be a random linear code, and $\boldsymbol{\rho} \in \{0, 1, ?\}^w$ be a length-w word drawn uniformly from $\{0, 1\}^w$ and erased in $1 \leq e \leq r$ positions. The probability that ML-decoding $\boldsymbol{\rho}$ neither rejects nor fails, denoted $f_2(r, e)$, equals to

$$f_2(r, e) = [1 - f_1(r, e)] \cdot 2^{-(r-e)}. \tag{13}$$

Proof: Not failing and not rejecting means that there is a unique way to complete the erasures of $\boldsymbol{\rho}$ into a codeword. This requires that H_{e}, the erasure-positions sub-matrix of H, is full rank, whose probability is the first term in (13). Given a full-rank H_{e}, the probability of not rejecting $\boldsymbol{\rho}$ is the probability that the vector $\mathbf{b}$ obtained by multiplying the non-erasure-position sub-matrix $H_{\bar{\mathsf{e}}}$ by the non-erased symbols of $\boldsymbol{\rho}$ is in the column span of H_{e}. Since $\boldsymbol{\rho}$ is drawn uniformly independent of $H_{\bar{\mathsf{e}}}$, $\mathbf{b}$ is uniformly distributed in $\{0, 1\}^r$. The second term in (13) follows as the probability that a uniform vector in $\{0, 1\}^r$ is in the span of e linearly-independent random size-r columns.

■

Propositions 2 and 3 can be used to analyze the success probability of PCLD once we know the distributions of lengths

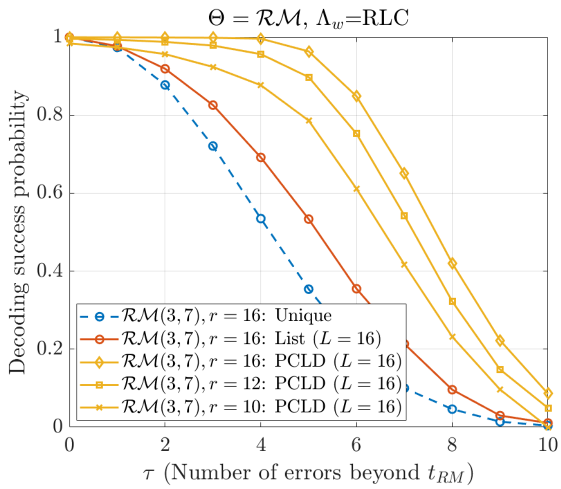

Fig. 5: Decoding-success probability using RLC analysis in conjunction with $\mathcal{RM}$ indicator codes.

(w) and erasure counts (e) induced by the output list of the indicator decoder. We show the outcome of this analysis in Fig. 5, comparing three different values of the residual-code redundancy (r).

V. Conclusion

The work presented in this paper commenced with a generalization of a non-binary non-symmetric channel, in which all errors are either to or from a specific alphabet symbol. We characterized the channel theoretically by deriving its capacity and maximum-likelihood decision rule. On the practical side, we devised a general decoding algorithm that enhances decoding performance compared to previously suggested decoders. We exemplified the performance improvement in block-error probability using exhaustive simulations for a code constructed from Reed-Muller and BCH constituent codes. Finally, we developed a framework for analyzing our new decoder using linear random codes.

A straightforward direction for future research is the analysis and evaluation of the newly suggested decoding algorithm using additional codes for Θ. Specifically, designing the constituent codes according to the channel parameters may further optimize the scheme. In addition, decoding soft-inputs from the channel can strengthen the decoder, and allow information from both codes to be utilized jointly (instead of sequentially as in CLD/PCLD).

VI. Acknowledgment

This work was supported in part by the US-Israel Binational Science Foundation and in part by the Israel Science Foundation.

References

[1] Y. Telepinsky, V. Mor, M. Schultz, Y. M. Hung, A. D. Kent and L. Klein, "Toward a six-state magnetic memory element," Applied Physics Letters, vol. 108 no. 18, 2016.

[2] R. Ahlswede and H. Aydinian, "Error control codes for parallel asymmetric channels," IEEE Trans. Inf. Theory, vol. 54, no. 2, pp. 831–836, 2008.

[3] N. Bitouze, A. Graell i Amat and E. Rosnes, "Error correcting coding for a nonsymmetric ternary channel," IEEE Trans. Comm., vol. 56, no. 11, pp. 5715–5729, 2010.

[4] I. Dumer, "Soft-decision decoding of Reed-Muller codes: a simplified algorithm," IEEE Trans. Inf. Theory, vol. 52, no. 3, pp. 954–963, 2006.

[5] I. Dumer and K. Shabunov, "Soft decision decoding of Reed-Muller codes: recursive lists," IEEE Trans. Inf. Theory, vol. 52, no. 3, pp. 1260–1266, 2006.

[6] K. Shabunov, "ECCLab," (2020), Available: https://github.com/kshabunov/ecclab.

[7] C. Di, D. Proietti, E. Telatar, T.s J. Richardson and R. L. Urbanke, "Finite-length analysis of low-density parity-check Codes on the binary erasure channel," IEEE Trans. Inf. Theory, vol. 48, no. 6, pp. 1570–1579, 2002.