Two Batch Search with Lie Cost

Rudolf Ahlswede*, Ferdinando Cicalese¶, Christian Deppe*, and Ugo Vaccaro‡
* Department of Mathematics, Bielefeld University, Germany. E-mail: {ahlswede,cdeppe}@math.uni-bielefeld.de
¶Faculty of Technology, Bielefeld University, Germany. E-mail: nando@cebitec.uni-bielefeld.de.
‡Department of Computer Science and Applications, University of Salerno, Italy. E-mail: uv@dia.unisa.it.

***Abstract*—We consider the problem of searching for an unknown number in the search space $\mathcal{U} = \{0, \ldots, M-1\}$. q-ary questions can be asked and some of the answers may be wrong. An arbitrary integer weighted bipartite graph Γ is given, stipulating the cost $\Gamma(i,j)$ of each answer $j \neq i$ when the correct answer is i, i.e., the cost of a wrong answer. Correct answers are supposed to be cost-less. It is assumed that a maximum cost e for the sum of the cost of all wrong answers can be afforded by the responder during the whole search. We provide tight upper and lower bounds for the largest size $M = M(q, e, \Gamma, n)$ for which it is possible to find an unknown number $x^* \in \mathcal{U}$ with n q-ary questions and maximum lie cost e. Our results improve the bounds in [9] and [3]. The questions in our strategies can be asked in two batches of non-adaptive questions. Finally, we remark that our results can be further generalized to a wider class of error models including also unidirectional errors.**

***Index Terms*—Coding with feedback, Searching with lie cost $\Gamma(i,j)$ for answer j, if i is correct, Compound unidirectional errors**

I. Introduction, Basic Definitions, and Results

The problem described in the abstract can be considered equivalently as a coding problem for a channel with noiseless feedback, an adaptive search problem, or a game with two players (see [13]). Already Rényi in [18] (Bar Kochba Game), Berlekamp in [6] (Quiet-Question-Noisy-Answer Game), and Ulam in [20] used the game notation, which we also did in [9], [3], and here again. Readers who are more familiar to coding theory should read the final section "Comments and Discussions", in which we formulate the problem and result in the language of coding theory.

The game is played by two persons, called Paul and Carole. Carole chooses a number $x^* \in \mathcal{U} = \{0, \ldots, M-1\}$. Paul has to find x^* asking n q-ary questions. By a q-ary question we mean a question of the type *"Which set among $T_0, T_1, \ldots, T_{q-1}$ does the secret number x^* belong to?"*, where $\mathbf{T} = (T_0, \ldots, T_{q-1})$ is a partition of the set $\mathcal{U}$. Carole's answer is just an index $k \in \mathcal{Q} = \{0, 1, \ldots, q-1\}$, saying that x^* belongs to T_k. Carole may decide to answer incorrectly.

Now a cost matrix $\Gamma : \mathcal{Q} \times \mathcal{Q} \to \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ is given such that $\Gamma(i,i) = 0$ for each i and $\Gamma(i,j) > 0$ for each $i \neq j$. We shall also refer to Γ as "channel", a name suggested by the information theoretical model of the problem. Γ is used to assign cost to Carole's answers. If Carole answers j to a question whose correct answer is i, then this answer has *individual cost* $\Gamma(i,j)$. Every correct answer has cost 0. The total cost of Carole's n answers is not allowed to be larger than the given total cost e known to both players.

Let $w = \min\{\Gamma(i,j) : i \neq j\}$. In words, w is the cost of the *cheapest* possible lie for Carole.

We define the set consisting of all possible sequences of $\lfloor e/w \rfloor$ lies, with total cost not larger than e as

$$F = \left\{ \big((a_1, b_1), \cdots, (a_{\lfloor e/w \rfloor}, b_{\lfloor e/w \rfloor})\big) \right.$$
$$\left. \text{such that } \forall j \ a_j \neq b_j \text{ and } \sum_{j=1}^{\lfloor e/w \rfloor} \Gamma(a_j, b_j) \leq e \right\}.$$

Note that these are the longest allowed sequences of lies. Intuitively, these are the sequences of lies on which Carole has the largest number of possible alternatives. It turns out that, asymptotically, only the number of such sequences counts, regardless of the actual structure of the channel.

For any q, e, Γ, and n we want to estimate the largest $M = M(q, e, \Gamma, n)$ for which it is possible to find an unknown number $x^* \in \mathcal{U}$ with n q-ary questions and maximum lie cost not larger than e. We study the asymptotics of $M(q, e, \Gamma, n)$ for $n \to \infty$.

We refer to the survey papers [17] and [13] for the most significant results in the huge literature on the Rényi-Berlekamp-Ulam game. Here, we limit ourselves to cite the main results that are related to our paper, using our formulation of the problem as a game with a cost matrix.

The problem was studied in [3], where upper and lower bounds were given for all e and w, which coincide when w is a divisor of e.

Special cases of our model have been considered before.

In [19] the binary game with asymmetric error was introduced, that is, the channel $\Gamma_0 = \begin{pmatrix} 0 & e+1 \\ 1 & 0 \end{pmatrix}$. For $e = 1$, $M(2, 1, \Gamma_0, n)$ was determined for n in [10] up to a factor 2. More generally, games with $q \geq 2$, $e \geq 0$, and $\Gamma(i,j) \in \{1, e+1\}$ for all $i \neq j$ were considered in [14], where asymptotically the exponential growth in n was determined. This is a special case of the work [3].

Other studies concern strategies with little adaptiveness. It was shown in [11], [12] that two-batch strategies can be as powerful as the fully adaptive ones for $\Gamma = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and

¶Supported by the Sofja Kovalevskaja Award of the Alexander von Humboldt Foundation and The Bundesministerium für Bildung and Forschung.
*Supported by the DFG-project "Informationstheorie und Kombinatorik" of the German Research Foundation.

for its q-ary generalization. Subsequently in [15] the results of [14] were also obtained with two-batch strategies.

In the present paper we close the gap between the bounds of [3] and provide matching upper and lower bounds also if w is not a divisor of e. Our main result states that, for any fixed q, e, and Γ and for $n \to \infty$ it holds that

$$M(q,e,\Gamma,n) = \left(q^{n+\lfloor e/w \rfloor}/|F|\binom{n}{\lfloor e/w \rfloor}\right)(1+o(1)).$$

Moreover, our search strategies consist of two batches of non-adaptive questions.

Finally, we give some further generalizations for a multi-channel version of the game that can be used to model the case of unidirectional errors (for more information see [1], [7], [8]).

II. Searching with lies with a general channel

We remind the reader of the definitions of q, Γ ,w ,e ,n , M, and F. They enter statements, lemmas, and theorems without further explanations. We study $M(q,\Gamma,e,n)$ for fixed q, Γ , e as function of n.

At the stage of the game, when questions $\mathbf{T}_1,\ldots,\mathbf{T}_t$ have been asked and t answers have been received, Paul's *state* of knowledge is represented by an $(e+1)$-tuple $(A_0,A_1,A_2,\ldots,A_e)$ of pairwise disjoint subsets of $\mathcal{U}$, where for each $i=0,1,\ldots,e$ the subset A_i consists of all the elements of $\mathcal{U}$ which could possibly coincide with the unknown number x^*, supposing that the sum of the individual cost of Carole's answers equals i. In particular, the *initial* state is given by $(\mathcal{U},\emptyset,\emptyset,\ldots,\emptyset)$. A state is final if $|\cup_i A_i| \le 1$, i.e., if there is only one element that is candidate to be the unknown number x^* or if no number exists that complies with Carole's answers taking into account the maximum amount of lying she can do.

We describe a *strategy $\mathcal{S}$ with n questions* by a q-ary tree of depth n, where each node ν is mapped into a question $\mathbf{T}_\nu$, and the q edges stemming from ν are labelled with 0, 1, $\ldots, q-1$ and represent Carole's possible answers to $\mathbf{T}_\nu$. Any possible *play* of the game coincides with a root-to-leaf path in the tree. The traversed nodes are the questions asked by Paul and the traversed edges are the corresponding answers of Carole. We say that strategy $\mathcal{S}$ is *winning* iff for every path in the tree the state associated with the leaf reached by the path is final. A strategy is said to be *non-adaptive* if all nodes at the same depth of the tree are mapped into the same question. This captures the fact that in a non-adaptive strategy, Paul cannot use the information gained from the previous answers of Carole before formulating the next question. In fact, he has to formulate all the questions at once before asking Carole to answer them.

III. Some Useful Lemmas

In this section we shall recall some known facts that will be used in this paper.

The following is a well-known application of Chernov's bound.

Lemma 1: Let $X_1,\ldots,X_n$ be independent 0-1 random variables. Let $X=\sum_{i=1}^n X_i$ and $\mu=\mathbf{E}[X]$. For $0<\delta<1$,

$$Pr(|X-\mu|>\delta\mu) \le 2e^{-\mu\delta^2/3}.$$

We shall now introduce two types of regularities for sequences in $\mathcal{Q}^n$. They will be used repeatedly in the following sections.

Definition 1: [Regular and P-wise regular sequences] Let n be a positive integer and $\vec{s}=s_1,\ldots,s_n$ be a sequence in $\mathcal{Q}^n$. We say that $\vec{s}$ is *regular* if for each $a\in\mathcal{Q}$ the number of occurrences of a in $\vec{s}$ differs from n/q by at most $n^{5/7}$.

Given a positive integer P, we say that $\vec{s}$ is *P-wise regular* if, for each $j=0,\ldots,P-1$ and for each $c\in\mathcal{Q}$, at least $\frac{n}{qP}(1-\frac{1}{P})$ occurrences of c appear in the sequence $s^{(j)}$ $(j=0,\ldots,P-1)$, defined by

$$s^{(j)} = s_{j\lfloor n/P\rfloor+1}\ldots s_{(j+1)\lfloor n/P\rfloor} \text{ for } j \le P-1-(n \bmod P)$$

$$s^{(j)} = s_{n-(P-j)\lceil n/P\rceil+1}\ldots s_{n-(P-j-1)\lceil n/P\rceil} \text{ otherwise.}$$

Intuitively, if we split a P-wise regular sequence into P consecutive subsequences as evenly as possible, each subsequence contains almost the same number of occurrences of each character.

Lemma 2: For any real number $\delta\in(0,1)$ there exists n_0 such that for all integers $n\ge n_0$ the number of sequences in $\mathcal{Q}^n$ that are not regular is bounded from above by δq^n.

Proof: Let $r=r(n)=n^{5/7}$ and fix an integer $c\in\mathcal{Q}$. The number of sequences that are not regular because the number of occurrences of c does not respect the required bound can be computed as $q^n\times Pr(|\sum_{i=1}^n X_i - n/q|>r)$, where X_i is the random variable which takes value 1 if the ith element in a randomly chosen sequence—with each element chosen independently and uniformly in $\mathcal{Q}$—is c, i.e., $Pr(X_i=1)=1/q$.

Thus, the desired result follows directly from Lemma 1. ∎

We shall use the following result from [15]. It allows us to restrict our analysis to instances where the search space cardinality has a special form. Intuitively, these are the worst case search space cardinalities for a given number of questions, where M is expressed almost as a power of q. The case when the search space cardinality is a power of q has a special symmetrical structure and via the following lemma it will be possible to exploit such symmetry also in the general case.

Lemma 3: [15] Fix three real numbers δ,α,α' such that $\delta\in(0,1)$ and $0<\alpha<\alpha'<\frac{q^{\lfloor e/w\rfloor}}{|F|}$. Then, there exists n_0 such that for any $n\ge n_0$, for any $M\le\alpha\frac{q^n}{\binom{n}{\lfloor e/w\rfloor}}$, there exist $a\in(q^T,q^{T+1}]\cap\mathbb{N}$, and a non-negative integer m such that

$$M \le (1-\delta)aq^m < \alpha'\frac{q^n}{\binom{n}{\lfloor e/w\rfloor}},$$

where T is an integer that only depends on δ and q.

IV. The Winning Strategy

In this section we shall prove that, asymptotically in n, for any $M\le\frac{q^n}{\binom{n}{\lfloor e/w\rfloor}}(1+o(1))$, Paul has a winning strategy in a game with search space of cardinality M.

Let $M \leq \alpha \frac{q^n}{\binom{n}{\lfloor e/w \rfloor}}$ for some $\alpha < \frac{q^{\lfloor e/w \rfloor}}{|F|}$. In particular, since we are interested in the worst case bound, we can assume $M = \lfloor \alpha \frac{q^n}{\binom{n}{\lfloor e/w \rfloor}} \rfloor$.

Then there exist α' and δ such that $\frac{q^{\lfloor e/w \rfloor}}{|F|} > \alpha' > \alpha$ and also $\frac{\alpha'}{1-\delta} < \frac{q^{\lfloor e/w \rfloor}}{|F|}$. Furthermore, by Lemma 3, we can find m and a such that $M \leq aq^m < \frac{\alpha'}{1-\delta} \frac{q^n}{\binom{n}{\lfloor e/w \rfloor}}$. This, together with $M = \lfloor \alpha \frac{q^n}{\binom{n}{\lfloor e/w \rfloor}} \rfloor$, gives

$$m \geq \log \frac{M}{a} = \log \frac{\lfloor \alpha \frac{q^n}{\binom{n}{\lfloor e/w \rfloor}} \rfloor}{a},$$

and also

$$n - m \geq \log \frac{a(1-\delta)}{\alpha'} + \log \binom{n}{\lfloor e/w \rfloor}.$$

Therefore, $n \to \infty$ implies $m \to \infty$ and also $n - m \to \infty$.

By Lemma 2 there are enough regular sequences in the space $\mathcal{Q}^m$ so that Paul can define an injective function f that maps elements of the search space $\mathcal{U}$ to pairs $(i, \vec{w})$, such that $i \in \{1, 2, \ldots, a\}$ and $\vec{w} \in \mathcal{Q}^m$ is regular.

The first batch of questions. Let x^* be the searched element and $f(x^*) = (i^*, \vec{w}^*)$. Paul asks a first batch of m questions, where the ith question is "What is the ith component of $\vec{w}^*$?"

Let $\vec{w}'$ be the sequence of Carole's answers. Let $r = r(n) = n^{5/7}$. We can immediately observe that for each $i \in \mathcal{Q}$ the number of occurrences of i in $\vec{w}'$ is at most $m/q + r + \lfloor e/w \rfloor$.

For each $j = 0, 1, \ldots, \lfloor e/w \rfloor$ and $k = 0, 1, \ldots, j$, let $F_k^{(j)}$ denote the set of possible sequences of k lies $(a_1, b_1), \ldots, (a_k, b_k)$ such that the total cost $\left(\sum_{\ell=1}^k \Gamma(a_\ell, b_\ell)\right)$ is in the set $\{jw, jw+1, \ldots, j(w+1)-1\} \cap \{0, 1, \ldots, e\}$.

We have the following.

Lemma 4: Let $(A_0, \ldots, A_e)$ be the state resulting from Carole's answers to Paul's first batch of questions. For $j = 0, 1, \ldots, \lfloor e/w \rfloor - 1$ we have

$$\frac{1}{a} \sum_{i=0}^{w-1} |A_{jw+i}| \leq \sum_{k=0}^{j} \frac{|F_k^{(j)}|}{k!} \left(\frac{m}{q} + r + \left\lfloor \frac{e}{w} \right\rfloor \right)^k.$$

Moreover, we have

$$\frac{1}{a} \sum_{i=\lfloor e/w \rfloor w}^{e} |A_i| \leq \sum_{k=0}^{\lfloor e/w \rfloor} \frac{|F_k^{(\lfloor e/w \rfloor)}|}{k!} (\frac{m}{q} + r + \left\lfloor \frac{e}{w} \right\rfloor)^k.$$

Proof: We shall limit ourselves to explain the first statement. The second statement follows by repeating the same argument with $j = \lfloor e/w \rfloor$ and observing that in this case we are only interested in lie-patterns and sequences of answers with total cost not exceeding e.

We have to count the number of possible different sequences of correct answers that can have lead Carole to answer the way she did.

First we notice that in a sequence of Carole's answers with total cost between jw and $j(w+1)-1$ there can be at most j lies.

Given $0 \leq k \leq j$, there are $|F_k^{(j)}|$ possible sequences of k lies of total cost between jw and $j(w+1)-1$. Let $(a_1, b_1), \ldots, (a_k, b_k)$ represent one such sequence. Each position a_i can be chosen in at most $(\frac{m}{q} + r + \left\lfloor \frac{e}{w} \right\rfloor)$ ways, due to the regularity property and the maximum possible number of deviations from regularity given by the allowed maximum number of wrong answers.

This gives us at most $|F_k^{(j)}|(\frac{m}{q} + r + \left\lfloor \frac{e}{w} \right\rfloor)^k$ mendacious sequences. Finally, we notice that each distinct sequence appears in this counting once for each possible permutation of the positions for the a_i's. Hence, we divide by the factor $k!$.

Summing up over all possible k's gives us the desired result for each $j = 0, 1, \ldots, \lfloor e/w \rfloor - 1$. ∎

The next two results prove that, starting from the position reached after the first m questions, Paul can encode the remaining candidates in order to successfully finish the game with a second batch of non-adaptive questions. The key point is the following lemma, which is a generalization of Lemma 2 in [3]. The difference is that the conditions (8) and (9) in [3] are replaced by Lemma 4 and then in the estimation of equations (13)-(17) in Lemma 2 in [3] we use the cardinality of F instead of the cardinality of $E = \{(j,k) : \Gamma(j,k) = w\}$.

Lemma 5: There exists an n_0 such that for all $n \geq n_0$ we have the inequality

$$q^{n-m} \geq \left(\sum_{j=0}^{w\lfloor e/w \rfloor - 1} |A_j| \right) \left(\sum_{i=0}^{2\lfloor e/w \rfloor} \binom{n-m}{i} q^i \right) + \sum_{j=w\lfloor e/w \rfloor}^{e} |A_j|,$$

where $(A_0, \ldots, A_e)$ is the state resulting from Carole's answers to Paul's first batch of questions.

Proof: We use the fact that $m = O(q^{(n-m)/\lfloor e/w \rfloor})$, which follows directly from $\frac{M}{1-\delta} \leq aq^m \leq \frac{\alpha'}{1-\delta} \frac{q^n}{\binom{n}{\lfloor e/w \rfloor}}$ as given by Lemma 3.

By the previous lemma and asymptotically with n (hence also with m) we obtain that the sum of the cardinalities of the sets $A_1, \ldots, A_{w\lfloor e/w \rfloor - 1}$ is bounded by

$$O\left(m^{\lfloor e/w \rfloor - 1}\right) = O\left(q^{(n-m)\frac{\lfloor e/w \rfloor - 1}{\lfloor e/w \rfloor}}\right) = o(q^{n-m}).$$

Moreover, the factor $\sum_{i=0}^{2\lfloor e/w \rfloor} \binom{n-m}{i} q^i$ accompanying $|A_j|$ is polynomial in $n - m$. Therefore, it can be absorbed in the above calculation.

Regarding the sets $A_{w\lfloor e/w \rfloor}, \ldots, A_e$ we need to be just a little bit more careful. By the previous lemma, the sum of the cardinalities of these sets is bounded by $O(\frac{|F|}{q^{\lfloor e/w \rfloor}} \frac{n^{\lfloor e/w \rfloor}}{\lfloor e/w \rfloor!})$.

By using Lemma 3 this can be bounded by $O(q^{n-m-\lfloor e/w \rfloor} |F| \alpha') = q^{n-m}(1 - \Omega(1))$. This completes the proof. ∎

Theorem 1: Starting from the state $(A_0, \ldots, A_e)$ resulting from Carole's answers to Paul's first batch of questions, there exists a non-adaptive winning strategy of size $n - m$ with the channel Γ with total cost e.

Proof: As a consequence of the previous lemma, there exists a mapping θ sending elements of $\bigcup_{i=0}^{w\lfloor e/w \rfloor - 1} A_i$ one-to-one onto a set $\mathcal{C}_1 \subseteq \mathcal{Q}^{n-m}$ and elements of $\bigcup_{j=w\lfloor e/w \rfloor}^{e} A_j$ one-to-one onto a set $\mathcal{C}_2 \subseteq \mathcal{Q}^{n-m}$ in such a way that

(i) for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}_1$, $d_H(\mathbf{x}_1, \mathbf{x}_2) \geq 2\lfloor e/w \rfloor + 1$,
(ii) for all $\mathbf{x}_1 \in \mathcal{C}_1, \mathbf{x}_2 \in \mathcal{C}_2$, $d_H(\mathbf{x}_1, \mathbf{x}_2) \geq \lfloor e/w \rfloor + 1$,

where $d_H(\cdot,\cdot)$ is the Hamming distance between q-ary sequences.

The following simple algorithm accomplishes the above task. Start with $\mathcal{C}_1 = \mathcal{C}_2 = \mathcal{A} = \emptyset$. 1. Pick up an arbitrary element $\mathbf{x} \in \mathcal{Q}^{n-m} \setminus \mathcal{A}$ and add it to $\mathcal{C}_1$. 2. Add to $\mathcal{A}$ the set $\{\mathbf{y} \in \mathcal{Q}^{n-m} \setminus \mathcal{A} \ : \ d_H(\mathbf{x},\mathbf{y}) \leq 2\lfloor e/w \rfloor\}$. Repeat 1. and 2. until $\mathcal{C}_1$ reaches the desired cardinality. Finally, pick up $|\bigcup_{j=w\lfloor e/w \rfloor}^{e} A_j|$ elements from $\mathcal{Q}^{n-m} \setminus \mathcal{A}$ and put them in $\mathcal{C}_2$.

Lemma 5 guarantees that one can extend $\mathcal{C}_1$ up to the desired cardinality and be left with enough elements in $\mathcal{Q}^{n-m}\setminus\mathcal{A}$ to accommodate the desired set $\mathcal{C}_2$.

By construction, it is also clear that $\mathcal{C}_1$ and $\mathcal{C}_2$ satisfy the desired distance constraints.

The second batch of questions. Paul arbitrarily fixes the map θ. Then, he asks Carole "What is the jth component of the element of $\mathcal{Q}^{n-m}$ onto which the secret number x^* is mapped by θ?", for $j = 1, 2, \ldots, n-m$.

The constraints on the Hamming distance between any two sequences of length $(n-m)$ in $\mathcal{C}_1$ and $\mathcal{C}_2$ ensure that Paul, independently of Carole's lies, will be able to find x^* as the number $a \in \bigcup_{i=1}^{e} A_j$ of minimum Hamming distance from the sequence of length $(n-m)$ defined by the answers of Carole.

Notice that, due to the constraints on the lie-cost, in this second batch Carole cannot lie at all if the secret number is one of the elements in $\bigcup_{j=w\lfloor e/w \rfloor}^{e} A_j$. Alternatively, she cannot lie more than $\lfloor e/w \rfloor$ times if the secret number is one of the elements in $\bigcup_{i=0}^{w\lfloor e/w \rfloor - 1} A_i$.

Therefore, if $x^* \in \bigcup_{j=w\lfloor e/w \rfloor}^{e} A_j$, the sequence of her answers will be exactly one of the sequences in $\mathcal{C}_2$. Hence, by inverting θ, Paul will correctly find the secret number.

Conversely, if $x^* \in \bigcup_{i=0}^{w\lfloor e/w \rfloor - 1} A_i$, the sequence of Carole's answers will not differ from $\theta(x^*)$ in more than $\lfloor e/w \rfloor$ places. For any $y \in \bigcup_{i=0}^{e} A_i$, $y \neq x^*$ we have $d_H(\theta(y), \theta(x^*)) \geq 2e+1$. Thus, $\theta(y)$ differs from the sequence of Carole's answers in more than e places. Whence, by choosing a such that $\theta(a)$ has minimum distance from the sequence of length $(n-m)$ of Carole's answers, Paul correctly find x^*. ∎

V. The upper bound

We shall now give an upper bound on the largest integer M such that Paul has a strategy of size n to determine Carole's secret number in the Rényi-Berlekamp-Ulam game with the channel Γ with total cost e, over a search space of cardinality M.

Definition 2: [Paths and Bushes] Given a strategy $\mathcal{S}$, and integers $x \in \mathcal{U}$ and $0 \leq i \leq e$, an (i,x)-path in the strategy $\mathcal{S}$ is a root to leaf path $\vec{\pi}$ such that the final state $(A_0, \ldots, A_e)$ associated with the leaf reached by $\vec{\pi}$ satisfies $|A_i| = \{x\}$ and $A_j = \emptyset$ for all $j \neq i$. A $(0,x)$-path is also called the *sincere* path for x, since it describes a play in which Carole's secret number is x and she always answers sincerely. For each $x \in \mathcal{U}$, the x-bush is the union over all $i = 0, 1, \ldots, e$ of all (i,x)-paths.

Let $\mathcal{S}$ be a winning strategy for Paul with n questions. Fix $x \in \mathcal{U}$ and let $\vec{\eta} = \eta_1, \ldots, \eta_n$ be the $(0,x)$-path in $\mathcal{S}$. For $i = 1, \ldots, n$, let b_i be the label of the edge η_i, i.e., the answer to the ith question, when the play is the one given by $\vec{\eta}$. For each $i = 1, 2, \ldots, n$ and for each $j \neq b_i$ such that $\Gamma(b_i, j) \leq e$, there must exist a $(\Gamma(b_i,j), x)$-path in $\mathcal{S}$ which coincides with $\vec{\eta}$ in the first $i-1$ components and whose ith component is j. In fact, this is the path that coincides with a play of the game in which Carole chooses the number x and decides to lie (only) at the ith question by answering j instead of b_i.

Let $\vec{\pi} = \pi_1, \ldots, \pi_n$ be such a path for a fixed choice of $i \in \{1, \ldots, n\}$ and $j \in \mathcal{Q} \setminus \{b_i\}$. Let $c_1, \ldots, c_n$ be the label of the edges of $\vec{\pi}$. By definition, $\vec{\pi}$ coincides with $\vec{\eta}$ in the first $i-1$ edges. In analogy with what we observed for $\vec{\eta}$, we have that for each $k = i+1, i+2, \ldots, n$, and each $\ell \neq c_k$ such that $\Gamma(b_i, j) + \Gamma(c_k, \ell) \leq e$, the strategy $\mathcal{S}$ must include a $(\Gamma(b_i,j) + \Gamma(c_k,\ell), x)$-path that coincides with $\vec{\pi}$ in its first k components and whose kth edge is labeled by ℓ. These are the paths describing plays of the game in which Carole chooses the number x and lies exactly twice: once at the ith question, by answering j instead of b_i and once at the kth question, by answering ℓ instead of c_k. Of course, the argument can be iterated until there is still room left for Carole to lie.

We shall now turn the above observation into a practical way to count the paths in the x-bush for some number x. For this purpose we need to consider the structure of the paths involved. In fact, by knowing the labels on the edges of the sincere path one can count the number of additional paths needed to accommodate Carole's possible strategies that are based on exactly one lie. Then, once these paths have been given, based on their structure, one can count the number of paths necessary to accommodate Carole's possible strategies that are based on exactly two lies; and so on.

For any n, such a counting would immediately give us an upper bound on the size of the search space where Paul can successfully search with a strategy with n questions. In fact, Paul has to accommodate M bushes (one for each number in the search space) into a tree with q^n paths. In order to achieve this, he must first accurately choose the sincere paths so that there is space for the paths taking care of Carole's strategies that are based on exactly one lie. Then he must fix these paths so that—according to the above counting—there remains space for the paths necessary to accommodate the resulting possible strategies of Carole using exactly two lies. Continuing this way, it must be possible to accommodate with the q^n paths available all strategies of Carole that include lies of total cost $\leq e$. Conversely, Paul has no winning strategy of size n for a search space of cardinality M, if it turns out that there is no way for him to choose the M sincere paths and the 1-lie paths and so on, in such a way that the *necessarily* resulting paths are in total at most q^n.

We shall show that, for all sufficiently large n, in a strategy of size n, almost all bushes include $\binom{n}{\lfloor e/w \rfloor} \frac{|F|}{q^{\lfloor e/w \rfloor}} (1 + o(1))$ paths. In fact, the number of bushes that might violate this bound is negligible.

In the following we shall identify a path by its associated sequence of labels. Let P be an integer whose value will be determined in the course of the following analysis. We shall need one more definition.

Definition 3: A path in a strategy of Paul of size n is called

P-wise regular if it coincides with a P-wise regular sequence in $\mathcal{Q}^n$. A bush is called *regular* if each path it consists of is P-wise regular.

We are now ready to prove the following upper bound.

Theorem 2: For all $\epsilon > 0$ there exists an integer n_0, such that for all integers $n > n_0$, if Paul has a strategy of size n to determine a number $x \in \mathcal{U}$ in the q-ary Rényi-Berlekamp-Ulam game with lies with total cost e with the channel Γ, then

$$M \leq \left(\frac{q^{\lfloor \frac{e}{w} \rfloor}}{|F|} + \epsilon \right) \frac{q^n}{\binom{n}{\lfloor \frac{e}{w} \rfloor}}.$$

We shall bound the number of bushes that Paul can pack in a tree/strategy of size n. We shall observe that for all sufficiently large n, in a strategy of size n almost all bushes are regular. Intuitively, this means that most paths in a winning strategy of Paul represent plays in which Carole has a lot of freedom in choosing the position for her lies. The regularity of almost all bushes will also give us the possibility to bound the size of a bush from below and, hence, the number of such bushes that can be contained in a tree with q^n leaves. We shall use the following generalization of [14, Lemma 4.1].

Lemma 6: For each $\epsilon > 0$ there exists n_0, such that for all $n \geq n_0$ the size of a regular bush in a winning strategy for Paul with n questions is at least $\binom{n}{\lfloor \frac{e}{w} \rfloor} \left(\frac{q^{\lfloor \frac{e}{w} \rfloor}}{|F|} + \frac{\epsilon}{2} \right)^{-1}$.

Proof: Let $x \in \mathcal{U}$ be such that the x-bush in Paul's strategy is regular. Let us think of each path in this regular bush as divided into P parts, each one of size $\lceil n/P \rceil$ or $\lfloor n/P \rfloor$ (as explained in Definition 1). By definition, in each one of such parts, each label occurs at least $\frac{n}{qP}(1 - \frac{1}{P})$ times. Assume that Carole's secret number is x and she has accepted to stick to the following rule: for each $j = 1, 2, \ldots, P$, the sequence of answers to the questions from the (jn/P)th one to the $((j+1)n/P - 1)$th one will contain at most one wrong answer.

We shall now count the total number of (i, x)-paths, with $i \in \{w \lfloor e/w \rfloor, \ldots, e\}$. These are all and only the paths followed by the game when Carole uses exactly $\lfloor e/w \rfloor$ lies. There are $|F|$ possible sequences of $\lfloor e/w \rfloor$ lies, such that the total lie cost is not larger than e. According to the deal above, Carole can choose the parts in which she will answer incorrectly in $\binom{P}{\lfloor e/w \rfloor}$ ways. Finally, in each of these parts she can place the lie in at least $\frac{n}{qP}(1 - \frac{1}{P})$ ways, due to the number of occurrences of each label in each Pth fraction of a P-wise regular path. Summarizing, there are at least

$$\binom{P}{\lfloor e/w \rfloor} |F| \left(\frac{n}{qP}(1 - \frac{1}{P}) \right)^{\lfloor e/w \rfloor}$$

paths in the x-bush. For each ϵ there is a (large) P such that

$$\binom{P}{\lfloor e/w \rfloor} |F| (\frac{n}{qP}(1 - \frac{1}{P}))^{\lfloor e/w \rfloor} > \binom{n}{\lfloor \frac{e}{w} \rfloor} (\frac{q^{\lfloor \frac{e}{w} \rfloor}}{|F|} + \frac{\epsilon}{2})^{-1},$$

which completes the proof. ∎

We are now ready to prove the theorem.

Proof of Theorem 2: By the previous lemma, it follows that Paul can have at most

$$\frac{q^n}{\binom{n}{\lfloor e/w \rfloor}} \left(\frac{q^{\lfloor e/w \rfloor}}{|F|} + \frac{\epsilon}{2} \right)$$

regular bushes in his strategy.

On the other hand, he cannot have room for more than $\frac{q^n}{\binom{n}{\lfloor e/w \rfloor}} \frac{\epsilon}{2}$ non-regular bushes, since there are at most so many paths that are not P-wise regular. In fact, by using for P-wise regular sequences Lemma 1 in a way analogous to its use in Lemma 2, one can easily bound from above the number of sequences that are not P-wise regular by $q^{n(1-c)}$ for some constant c only depending on P and q. Moreover, for any constant c and for any $\epsilon > 0$ we have $q^{n(1-c)} \leq \frac{q^n}{\binom{n}{\lfloor e/w \rfloor}} \frac{\epsilon}{2}$ for all sufficiently large n.

Summing up we have that a strategy of Paul of size n cannot contain M bushes if $M > \frac{q^n}{\binom{n}{\lfloor e/w \rfloor}} \left(\frac{q^{\lfloor e/w \rfloor}}{|F|} + \epsilon \right)$. The proof of Theorem 2 is complete. ∎

Now we can formulate our main result, which follows directly from Lemma 4, Theorem 1, and Theorem 2.

Theorem 3: Asymptotically in n we have

$$M(q, e, \Gamma, n) = \frac{q^{n + \lfloor e/w \rfloor}}{|F| \binom{n}{\lfloor e/w \rfloor}} (1 + o(1)).$$

Moreover, optimal search strategies can be implemented in two batches of non-adaptive questions.

VI. Unidirectional Errors and More

We discuss in this section another well-studied model of errors: unidirectional errors. In order to simplify the discussion, let us assume that the channel is completely symmetric and all lies have cost 1. We say that the game is on unidirectional errors if at the beginning of the game Carole has also to decide the direction of all her lies. More precisely, if she chooses *increasing* (*decreasing*) lies, she agrees on the following: she decides that when the correct answer to a question is i, she only chooses the lies among the j's such that $j > i$ ($j < i$.) This information is kept secret to Paul. In other words we can think of Carole and Paul stipulating that the game can be played with two different channels, one that only includes errors of the type $i \to j$ ($j > i$) and another that only includes errors of the type $i \to j$ ($j < i$). At the beginning, Carole can choose the channel she prefers to use, but then she will only use that one. Her choice is secret to Paul. We expect this assumption to restrict the number of possible lie patterns that Carole can use. Having tighter bounds for such situations is important in those applications where engineering constraints may allow to assume that errors in the same transmission block only occur in one direction.

By using an argument analogous to the one described above, we can prove the following general result where we assume the parameter e to be given.

Theorem 4: Let $\Gamma_1, \Gamma_2, \ldots, \Gamma_t$ be channels. For $j = 1, \ldots, t$, let $w_j = w^{\Gamma_j}$. Let $\tilde{w} = \min_{j=1}^{t} w_j$ and $\tilde{e} = \lfloor e/\tilde{w} \rfloor$. Let $G_i = G^{\Gamma_i}$ be defined as

$$G_i = \{((a_1,b_1),\cdots,(a_{\tilde{e}},b_{\tilde{e}})) : a_j \neq b_j \ \forall j, \sum_{j=1}^{\tilde{e}} \Gamma_i(a_j,b_j) \leq e\}$$

Let $M(n)$ be the largest size of a search space where Paul can find a secret number in the variant of the game where Carole is allowed to choose the Γ_i she wants to play with, and keep it secret to Paul.

Then, for all sufficiently large n, we have

$$M(n) = \left(\frac{q^{n+\tilde{e}}}{\binom{n}{\tilde{e}}} \left| \bigcup_{i:w_i=\tilde{w}} G_i \right|^{-1} \right) (1+o(1)).$$

Moreover, optimal strategies can be implemented in two batches of non-adaptive questions.

VII. Comments and Discussions

We finish our paper with an alternative formulation of the problem and the main theorem which is closer to the classical model used in coding theory. We shall then reformulate our problem as a channel model with feedback. Already Berlekamp in 1964 described the problem in both ways in his PhD-thesis.

A sender wants to transmit a message $x \in \mathcal{U}$ over a noisy q-ary channel with feedback. $\mathcal{U} = \{0,\ldots,M-1\}$ denotes the set of possible messages and $\mathcal{Q} = \{0,\ldots,q-1\}$ the coding alphabet. We have a passive feedback, that means that the sender always knows what has been received. The codewords (blockcodes) are elements of $\mathcal{Q}^n$ and the encoding function is defined by : $c(x) = \left(c_1(x),\ldots,c_n\left(x,y^{n-1}\right)\right)$, where $c_i : \mathcal{U} \times \mathcal{Q}^{i-1} \to \mathcal{Q}$ is a function for the i-th letter which depends on the message we want to transmit and the $(i-1)$ letters which have been received before. In the basic model one assumes that the noise changes less than e symbols. We generalized this model in the following way. We assume that for each pair of code symbols, $c_1, c_2 \in \mathcal{Q}$, there is a fixed amount of noise $\Gamma(c_1,c_2)$ that has to be produced by the channel in order to change c_1 into c_2. Moreover, we assume that the channel can produce a maximum amount of noise per codeword sent. This amount is defined by the parameter e.

The noise function Γ characterizing the channel can be represented as a $q \times q$ matrix, where the entry in the i-th column and j-th row denotes the value of $\Gamma(i,j)$. For example the channel with binary input and output alphabet and bounded number of errors corresponds to the matrix $\Gamma = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

The goal of the paper is to find upper and lower bounds for $M(q,e,\Gamma,n)$, which denotes the maximal number of messages, which can be transmitted over a q-ary channel with feedback described by the function Γ and the total cost bound e with a blockcode of length n.

We show in this paper that given q, e, and Γ asymptotically in n we have

$$M(q,e,\Gamma,n) = \frac{q^{n+\lfloor e/w \rfloor}}{|F|\binom{n}{\lfloor e/w \rfloor}}(1+o(1)).$$

We also give a strategy, which achieves this bound. Furthermore, this strategy does not use the full feedback. Let m be as defined in Lemma 3. The sender transmits a first batch of m symbols without intermediate feedback, then he receives in one shot the symbols actually delivered to the receiver. Depending on them he sends the remaining $n-m$ symbols. Therefore we show that for big n the two-batch strategy is as good as a full adaptive strategy. An open question is how to minimize the number of symbols in the first batch, since this means to reduce the use of the feedback channel.

Our Theorem 2 is a generalization of Berlekamp's Volume Bound, which is a generalization of the Sphere Packing Bound to feedback. It also holds without feedback and therefore gives a Sphere Packing Bound for a q-ary input, q-ary output channel with error function Γ.

In the second batch of our strategy we use a Gilbert-Varshamov construction, which in our case leads to an optimal strategy. If one uses such construction without feedback it does not give an optimal solution, because the packing is not optimal and one gets a lot of not used sequences. In our case the situation is different. We have a lot of messages for which we know that no error can occur any more and we can use the sequences, which in the Gilbert-Varshamov construction without feedback are not used. A related result for the case of the classical q-ary channel without costs can be found in [5].

We remark that for the equivalent channel without feedback there does not exists such a general combinatorial model. For simple cases we can provide some comparisons between the models with and without feedback. For q-ary codes with errors the Hamming Bound holds with and without feedback. A code is called perfect, if it achieves the Hamming Bound and it is called quasi-perfect if the difference between its number of messages and the Hamming Bound is 1. For every given q and e it is shown that there exists an n_0 such that for all $n \geq n_0$ all code strategies with feedback are perfect or quasi-perfect. And the code strategies are known. Without feedback for all $n > 1$ perfect codes are only known for $q \leq 3$. These are the following cases: Hamming Codes for $e = 1$, Repetition Codes for $e = \frac{n-1}{2}$ (n odd), the binary Golay-Code $(23,12,7)$, and the two ternary Golay-Codes $(11,6,5)$ and $(23,11,5)$. Another good comparison of error-correcting codes with and without feedback is the capacity error function C, the supremum of the rates achievable for all $\tau = \frac{e}{n}$ and all large n. This problem without feedback is still open for the binary case. It is completely known for the binary case with feedback ([6], [21]) and almost known for the q-ary case ([4]). In all these cases the capacity error function with feedback C_f is bigger than the capacity error function without feedback C for all $\tau > 0$, if $C_f > 0$.

The new idea of a game with a channel was developed during the project "General Theory of Information Transfer and Combinatorics" at the ZIF (Center for Interdisciplinary Research) in Bielefeld from 2001-2004 (see [2]). During this project the papers [9] and [3] were written. In [9] the authors analyzed the special case where the channel satisfies for every $k \in \mathcal{Q}$ $|\{(j,k) : \Gamma(j,k) = w, j \in \mathcal{Q}\}| = d$ and the cardinality of the search space is a power of q, i.e. $M = q^m$. Upper and lower bounds were given. However, there is a problem with the lower bound of [9]. In fact, it only holds if w is a divisor of e (in this case the bound is tight) and unfortunately this is

not stated in the formulation of the main theorem in [9].

ACKNOWLEDGEMENT

We thank Joel Spencer and Robert Ellis for fruitful discussions in Victoria in 2006. This discussion helped us to remove an earlier divisibility assumption $d|n$ in Theorem 3 and on the other hand influenced Robert Ellis to get a theorem for bounded error strings in [16].

His result can now also be used to prove Theorem 3 by applying it to our set F.

REFERENCES

[1] R. Ahlswede, H. Aydinian, L.H. Khachatrian, and L.M. Tolhuizen, On q-ary codes correcting all unidirectional errors of a limited magnitude, Special issue dedicated to the memory of Varshamov, Abstract included in Proceedings of the International workshop on Algebraic and Combinatorial Coding Theory (ACCT), Kranevo, Bulgaria, June 19 - 25, 2004, Preprint in Arxiv, CS.IT-0607132, 2006.

[2] R. Ahlswede, L. Bäumer, N. Cai, H. Aydinian, V. Blinovsky, C. Deppe, and H. Mashurian (Eds.), General Theory of Information Transfer and Combinatorics, LNCS, Vol. 4123, Springer Verlag, 2006.

[3] R. Ahlswede, F. Cicalese, and C. Deppe, Searching with lies under error cost constraints, General Theory of Information Transfer and Combinatorics, Special Volume of Discrete Applied Mathematics, Vol. 156, No. 9, 1444-1460, 2008.

[4] R. Ahlswede, C. Deppe, and V. Lebedev, Nonbinary error correcting codes with noiseless feedback, localized errors or both, Annals of European Academy of Sciences, No. 1, 285 - 309, 2005.

[5] L.A. Bassalygo, Nonbinary error-correcting codes with one-time error-free feedback, Problems of Information Transmission, 41, 2, 125–129, 2005. Translated from Problemy Peredachi Informatsii, 2, 63-67, 2005.

[6] E. R. Berlekamp. Block coding for the binary symmetric channel with noiseless, delayless feedback. In: Error-correcting Codes, H.B. Mann (Editor), Wiley, New York, 61-88, 1968.

[7] M. Blaum, Codes for Detecting and Correcting Unidirectional Errors, IEEE Computer Society Press, Los Alamitos, CA, 1993.

[8] B. Bose and T.R.N. Rao, Theory of unidirectional error correcting/detecting codes, IEEE Trans. Comput. 31, No. 6, 521–530, 1982.

[9] F. Cicalese, C. Deppe, and D. Mundici, *Q*-ary Ulam-Rényi game with weighted constrained lies, LNCS, Vol. 3106, 82-91, 2004.

[10] F. Cicalese and D. Mundici, Optimal coding with one asymmetric error: below the sphere packing bound, In Proc. of COCOON 2000, LNCS, Vol. 1858, 159-169, 2000.

[11] F. Cicalese and D. Mundici, Optimal binary search with two unreliable tests and minimum adaptiveness, Proc. of ESA '99, LNCS, Vol. 1643, 257-266, 1999.

[12] F. Cicalese, D. Mundici, and U. Vaccaro, Least adaptive optimal search with unreliable tests, Theor. Computer Science, 270, 877-893, 2002.

[13] C. Deppe, Coding with feedback and searching with lies, Entropy, Search, Complexity, Series Bolyai Society Mathematical Studies , Vol. 16, 27-70, 2007.

[14] I. Dumitriu and J. Spencer, The liar game over an arbitrary channel, Combinatorica, 25, 537-559, 2005.

[15] I. Dumitriu and J. Spencer, The two-batch liar game over an arbitrary channel, SIAM J. of Discrete Mathematics, 19, 1056-1064, 2006.

[16] R. Ellis and K.L. Nyman, Two batch liar games on a general bounded channel, manuscript submitted to Combinatorica.

[17] A. Pelc. Searching games with errors – fifty years of coping with liars, Theoret. Comput. Sci., 270, 71–109, 2002.

[18] A. Rényi, On a problem of information theory, MTA Mat. Kut. Int. Kozl., 6B, 505–516, 1961.

[19] R. L. Rivest, A. R. Meyer, D. J. Kleitman, K. Winklmann, and J. Spencer, Coping with errors in binary search procedures, Journal of Computer and System Sciences, 20, 396-404, 1980.

[20] S.M. Ulam, Adventures of a Mathematician, Scribner's, New York, 1976.

[21] K.Sh. Zigangirov, Number of correctable errors for transmission over a binary symmetrical channel with feedback, *Problems Inform. Transmission*, vol. 12, 85-97, 1976.