Complexity of error-correcting codes derived from combinatorial games*

Aviezri S. Fraenkel and Ofer Rahat
Department of Computer Science and Applied Mathematics
The Weizmann Institute of Science
Rehovot 76100, Israel
fraenkel, ofer@wisdom.weizmann.ac.il
http://www.wisdom.weizmann.ac.il/~fraenkel, ~ofer

Abstract

The main result of the present paper is to establish an algorithm for computing a linear error-correcting code called *lexicode* in $O(n^{d-1})$ steps, where n and d are the size and distance of the code, respectively. This is done by using the theory of combinatorial games, specifically, two-player cellular automata games. Previous algorithms were exponential in n.

1 Introduction

For a linear error-correcting code, denote by n the common size of the codewords, by k its dimension, and by d the minimum distance of any two codewords. The main purpose of this paper is to present an $O(n^{d-1})$ algorithm for computing a basis for such a code, called *lexicode*, using the theory of combinatorial games. Previous methods [5], [4] required $O(2^{2n})$ steps. In [19], [20], an iterative algorithm for constructing the lexicode with time complexity $O(2^{n-k+1})$ is given. In [3] Theorem 3.5, the asymptotic lower bound $k \geq (2n)/d$ is given. In [19] the slightly tighter asymptotic lower bound $k \geq (3n)/d$ is given.

When comparing the dimension of the binary lexicode calculated in [5] Table VII with those of the best linear codes listed in [2] for $n \leq 44$ and $k \leq 10$, it turns out that the lexicode has always either the parameters (n, k, d) of the best code known, or its dimension k is within one or two of the best ones. In particular, the lexicode contains both Hamming, Golay and other codes. For details see e.g., [5], [18], §2. The Golay code was called "... probably the most important of all codes, for both practical and theoretical reasons" in [17], Ch. 2.

*Part of the material in this paper was presented at "Computer Games 2002", Edmonton, Alberta, Canada, July 2002.

The present results stem from a program to compute codes efficiently, using the theory of combinatorial games. See [8], end of Introduction, and Remark 4.2 in the penultimate paragraph of that paper. It partly motivated the invention of the *Cellata* games [12], [10], [11], a special case of which will feature below, where they are called *Celata* games.

In §2 some necessary background concepts are collected, §3 contains a definition of a special case of a Celata game and of a *Celopt* game, where some of their properties are explored. It also contains a definition of the lexicode. An important connection between a Celopt game and a lexicode is stated in Theorem 1. The key results of the present paper are presented in §4, based on further properties of Celopt games, derived from properties of Celata games. The latter results stem from previous results on *annihilation* games, which are a special case of Celata games [7], [14], [15], [16], [21]. For annihilation games with the last player losing, see [6]. Another key result is Theorem 6 in §5, which leads to the algorithm for computing the basis, each of whose members has weight $\geq d$. The second part of Theorem 6 points out that it can be arranged that all basis elements have weight precisely d. A brief wrap-up is presented in the final §6.

2 Preliminaries

We are concerned with games played on a digraph $G = (V, E)$. Tokens are initially distributed on a subset of V. A move consists of selecting a token and moving it to a neighboring vertex, unoccupied or occupied. The player making the last move wins, the opponent loses. If there is no last move, the outcome is a draw.

The corresponding *game-graph* is a digraph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ where $\boldsymbol{V}$ is the set of all collections of token distributions on G, and $(\boldsymbol{u}, \boldsymbol{v}) \in \boldsymbol{E}$ if there is a move from $\boldsymbol{u}$ to $\boldsymbol{v}$. Note that for the case of a single token on V we have $\boldsymbol{G} = G$.

For this and the following notations and basic facts, see e.g., [9].

A vertex $\boldsymbol{u} \in \boldsymbol{V}$ (game-position) is labeled N, if the *N*ext player (the player moving from $\boldsymbol{u}$) has a winning move; by P if the *P*revious player (who landed on $\boldsymbol{u}$) can win.

Denote by $\mathcal{P}$ and by $\mathcal{N}$ the set of all P-positions and all N-positions respectively.

For any digraph $G = (V, E)$, the set $F(u)$ of *followers* of $u \in V$ is defined by $F(u) = \{v \in V : (u, v) \in E\}$.

It can then be shown that

$$\boldsymbol{u} \in \mathcal{P} \text{ if and only if } F(\boldsymbol{u}) \subseteq \mathcal{N}, \tag{1}$$

$$\boldsymbol{u} \in \mathcal{N} \text{ if and only if } F(\boldsymbol{u}) \cap \mathcal{P} \neq \emptyset. \tag{2}$$

Further, if G is acyclic, then the sets $\mathcal{P}$, $\mathcal{N}$ partition V: $\mathcal{P} \cup \mathcal{N} = V$, $\mathcal{P} \cap \mathcal{N} = \emptyset$.

Let $S \subset \mathbb{Z}_{\geq 0}$, $S \neq \mathbb{Z}_{\geq 0}$, and $\overline{S} = \mathbb{Z}_{\geq 0} \setminus S$. The *minimum excluded value* of S is

$$\operatorname{mex} S = \min \overline{S} = \text{least nonnegative integer not in } S.$$

A (classical) *Sprague-Grundy function*, or simply *g-function* $g\colon V \to \mathbb{Z}_{\geq 0}$ on a digraph $G = (V, E)$ is defined recursively by

$$g(u) = \operatorname{mex}\, g\bigl(F(u)\bigr),$$

where for any set T and any function h on T,

$$h(T) = \{h(t)\colon t \in T\}.$$

We then have

$$\mathcal{P} = \{\boldsymbol{u} : g(\boldsymbol{u}) = 0\}, \quad \mathcal{N} = \{\boldsymbol{u} : g(\boldsymbol{u}) \neq 0\}.$$

We denote the sum of the integers $a, b \in \mathbb{Z}_{\geq 0}$ over GF(2) by $a \oplus b$, also known as XOR, or *Nim-sum*. The XOR of $a_1, \ldots, a_m$ may also be denoted by $a_1 \oplus \ldots \oplus a_m = \sum'^{m}_{i=1} a_i$, where the apostrophe denotes XOR.

We shall also be concerned with vectors over $\mathrm{GF}(2)^n$ (n-dimensional binary vectors). The *numerical value* of a vector $\boldsymbol{u} = (u^0, \ldots, u^{n-1}) \in \mathrm{GF}(2)^n$ is $|\boldsymbol{u}| := \sum_{i=0}^{n-1} u^i 2^i$. The *weight* of $\boldsymbol{u}$ is $w(\boldsymbol{u}) = \sum_{i=0}^{n-1} u^i =$ number of 1-bits of $\boldsymbol{u}$. The *parity weight* of $\boldsymbol{u}$ is $w'(\boldsymbol{u}) = \sum'^{n-1}_{i=0} u^i$.

The Hamming distance between two vectors $\boldsymbol{u}$, $\boldsymbol{v}$ over GF(2) is defined by $H(\boldsymbol{u}, \boldsymbol{v}) = w(\boldsymbol{u} \oplus \boldsymbol{v})$. It is the number of positions in which $\boldsymbol{u}$ and $\boldsymbol{v}$ differ.

Following [1], a vector $\boldsymbol{u}$ with $w(\boldsymbol{u})$ even is said to be an *evil* number, and one with $w(\boldsymbol{u})$ odd is an *odious* number.

The *lexicode* on binary vectors (words) with n bits and distance $d \geq 2$, is defined by applying the greedy algorithm to the sequence of lexicographically ordered vectors over GF(2): adjoin $\Phi = (0, \ldots, 0)$, and at each step adjoin the vector $\boldsymbol{u}$ with smallest numerical value which satisfies $H(\boldsymbol{u}, \boldsymbol{v}) \geq d$ for every previously adjoined $\boldsymbol{v}$. The result is a linear code. See [5], [4], [18], [8].

The basic fact used in this paper is that the set $\mathcal{P}$ of any game Γ forms a linear code: The P-positions constitute a (graph-theoretic) independent set of the game-graph of Γ. If its P-positions are represented as vectors, there is a certain minimum Hamming distance between them.

3 Celata and Celopt Games

Two-player cellular automata games, or *Celata* games for short, were defined in [12]. See also [10], [11]. Below is a brief description of a class of Celata games.

Given a finite digraph $G = (V, E)$, also called *groundgraph*, order V ($|V| = n$) in some way, say

$$V = \{z_0, \ldots, z_{n-1}\}.$$

This ordering is assumed throughout.

The special case of Celata games on G of interest here, depends on a parameter $s \in \mathbb{Z}_{>0}$. A move from an occupied vertex z_k consists of *firing* (see below) z_k and some neighborhood of z_k of size q_k, where

$$q = q_k = \min(s, |F(z_k)|). \tag{3}$$

It will be called *Celata game with parameter s*, or simply *Celata* game.

Firing z_k means to *complement* z_k and a q_k-neighborhood of z_k: the token on z_k is removed, and the token-occupancy of q of its neighbors is complemented.

Let $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ denote the following game-graph of the Celata game played on $G = (V, E)$. The digraph $\boldsymbol{G}$ is also called the *Celata graph* of G. Any vertex in $\boldsymbol{G}$ can be described in the form $\boldsymbol{u} = (u^0, \ldots, u^{n-1})$ over the field GF(2), where $u^k = 1$ if z_k is occupied by a token, $u^k = 0$ if z_k is unoccupied. In particular, $\Phi = (0, \ldots, 0)$ is a leaf of $\boldsymbol{V}$, and $|\boldsymbol{V}| = 2^n$.

Note that $\boldsymbol{V}$ is an abelian group under the addition $\oplus$ of GF(2), also called Nim-addition, with identity Φ. Every nonzero element has order 2. Moreover, $\boldsymbol{V}$ is a vector space over GF(2) satisfying $1\boldsymbol{u} = \boldsymbol{u}$ for all $\boldsymbol{u} \in \boldsymbol{V}$. For $i \in \{0, \ldots, n-1\}$, define unit vectors $\boldsymbol{z}_i = (z_i^0, \ldots, z_i^{n-1})$ with $z_i^j = 1$ if $i = j$; $z_i^j = 0$ otherwise. They span the vector space. In particular, for any $\boldsymbol{u} = (u^0, \ldots, u^{n-1}) \in \boldsymbol{V}$, we can write $\boldsymbol{u} = \sum_{i=0}^{n-1} u^i \boldsymbol{z}_i = \sum_{i=0}^{\prime\, n-1} u^i \boldsymbol{z}_i$.

For defining $\boldsymbol{E}$, let $\boldsymbol{u} \in \boldsymbol{V}$ and let $0 \le k \le n-1$. Let $F^q(z_k) \subseteq F(z_k)$ be any subset of $F(z_k)$ satisfying

$$|F^q(z_k)| = q. \tag{4}$$

Define,

$$(\boldsymbol{u}, \boldsymbol{v}) \in \boldsymbol{E} \ \text{ if } u^k = 1, \ q > 0, \ \text{and } \boldsymbol{v} = \boldsymbol{u} \oplus \boldsymbol{z}_k \oplus \sum\nolimits'_{z_\ell \in F^q(z_k)} \boldsymbol{z}_\ell, \tag{5}$$

for every $F^q(z_k)$ satisfying (4).

Informally, an edge $(\boldsymbol{u}, \boldsymbol{v})$ reflects the firing of u^k in $\boldsymbol{u}$ (with $u^k = 1$), i.e., the complementing of the tokens on z_k and $F^q(z_k)$. Such an edge exists for every $F^q(z_k)$ satisfying (4). Note that if $z_k \in G$ is a leaf, then there is no move from $\boldsymbol{z}_k$, since then $q = 0$.

If (5) holds, we also write $\boldsymbol{v} = F_k^q(\boldsymbol{u})$. The set of all followers of $\boldsymbol{u}$ is

$$F(\boldsymbol{u}) = \bigcup_{u^k=1} \ \bigcup_{F^q(z_k) \subseteq F(z_k)} F_k^q(\boldsymbol{u}).$$

A *Celopt*$_s$ = *Celopt* game with parameter $s \in \mathbb{Z}_{>0}$ and game-graph called *Celopt-graph*, is defined the same way as a Celata game, except that we *opt* for $p \in \{0, \ldots, s\}$ followers, i.e., (5) is replaced by

$$(\boldsymbol{u}, \boldsymbol{v}) \in \boldsymbol{E} \ \text{ if } u^k = 1, \ q > 0, \ \text{and } \boldsymbol{v} = \boldsymbol{u} \oplus \boldsymbol{z}_k \oplus \sum\nolimits'_{z_\ell \in F^p(z_k)} \boldsymbol{z}_\ell, \tag{6}$$

for all $p \in \{0, \ldots, s\}$, where $F^p(z_k)$ — the only difference between (5) and (6) — is any subset of p followers of z_k.

In other words, the edges (moves) describe the firing of z_k and p followers in $F^q(z_k)$ for all $p \in \{0, \ldots, s\}$. A single move consists of choosing one of the values $p \in \{0, \ldots, s\}$ and firing z_k and p followers. Note that $p = 0$ means to complement z_k only. This is a legal move unless z_k is a leaf, in which case $q = 0$, and then there is no move from z_k, by convention.

Lemma 1. *Every Celopt$_s$ game on a digraph $G = (V, E)$ reduces to a Celata game on $G' = (V', E')$, where $|V'| = |V| + s$, $|E| \le |E'| \le |E| + s|V|$, and the complexity of the strategy for both games is the same.*

Proof. Put $V' = V \cup \{\ell_1 \ldots, \ell_s\}$, where the ℓ_i are leaves. For $k \in \{0, \ldots, n-1\}$, denote by L_k the number of edges leading from z_k to leaves in G. If $L_k < q_k$, adjoin an edge (z_k, ℓ_i) for every $i \in \{1 \ldots, q_k - L_k\}$. Thus every non-leaf vertex z_k of G' satisfying $L_k < q_k$, has q_k edges to leaves. It is now clear that Celopt on G is the same as Celata on G': firing a neighborhood of size p of z_k on G corresponds to firing the same neighborhood on G' as well as $q - p$ leaves. ■

Definition 1. The *Nim-graph* of size n is a digraph $G = (V, E)$ with $V = \{z_1, \ldots, z_n\}$, $E = \{(z_j, z_i) : 1 \le i < j\}$ for $j = 1, \ldots, n$. A Celopt$_s$ game (of size n) on a Nim-graph (of size n) will be called a *Nim-Celopt$_s$* game.

Remark 1. (i) Note that the indices are in the integer interval $[1, n]$, rather than $[0, n-1]$ as before. This happens to be more convenient for the present case. Therefore we also use here the notation $\boldsymbol{u} = (u^1, \ldots, u^n)$. When $n = 0$, the Nim-graph is empty.
(ii) The vertex z_1 is not a leaf. When $p = 0$, a token on z_i, for any $i \in \{1, \ldots, n\}$ can be removed; in particular from z_1.
(iii) From any vertex z_i, every follower is accessible. It follows that $g(\boldsymbol{z}_i)$ is monotonically increasing: $1 = g(\boldsymbol{z}_1) < 2 = g(\boldsymbol{z}_2) < g(\boldsymbol{z}_3) < \ldots$.

Theorem 1. *Let $\mathcal{P}$ be the set of P-positions of Nim-Celopt$_s$. Let $\mathcal{L}$ be the set of codewords of the lexicode with Hamming distance $d \ge s + 2$. Then $\mathcal{P} = \mathcal{L}$.*

Proof. Induction on the numerical value $|\boldsymbol{u}|$ of the vector $\boldsymbol{u}$. Clearly, $\Phi \in \mathcal{P}$ and $\Phi \in \mathcal{L}$. Let $|\boldsymbol{u}| \ge 1$. The induction hypothesis implies that for every $\boldsymbol{v}$ with $|\boldsymbol{v}| < |\boldsymbol{u}|$ we have $\boldsymbol{u} \in \mathcal{P}$ if and only if $\boldsymbol{u} \in \mathcal{L}$. We show $\boldsymbol{u} \in \mathcal{P}$ if and only if $u \in \mathcal{L}$.

Let $\boldsymbol{u} \in \mathcal{P}$. We have to show that for every $|\boldsymbol{v}| < |\boldsymbol{u}|$, if $\boldsymbol{v} \in \mathcal{L}$, then $H(\boldsymbol{u}, \boldsymbol{v}) \ge d$. Equivalently, for every $|\boldsymbol{v}| < |\boldsymbol{u}|$, if $H(\boldsymbol{u}, \boldsymbol{v}) \le d - 1 = s + 1$, then $\boldsymbol{v} \notin \mathcal{L}$. Indeed, the last two conditions imply that $\boldsymbol{v} \in F(\boldsymbol{u})$ in Celopt. Moreover, $\boldsymbol{v} \in \mathcal{N}$ since $\boldsymbol{u} \in \mathcal{P}$, so $\boldsymbol{v} \notin \mathcal{L}$ by the induction hypothesis.

Conversely, suppose that $\boldsymbol{u} \notin \mathcal{P}$. Then $\boldsymbol{u} \in \mathcal{N}$. Thus there exists $\boldsymbol{v} \in F(\boldsymbol{u}) \cap \mathcal{P}$. The rules of Celopt imply $|\boldsymbol{v}| < |\boldsymbol{u}|$, and $H(\boldsymbol{u}, \boldsymbol{v}) \le s + 1 < d$. By the induction hypothesis, $\boldsymbol{v} \in \mathcal{L}$. Hence $\boldsymbol{u} \notin \mathcal{L}$. ■

In the next section we shall deal with Nim-Celopt games, for which, by Lemma 1, the results of [12] apply. There, the generalized Sprague-Grundy function γ had to be used because the groundgraph contained cycles. Here we deal only with acyclic digraphs, so the classical Sprague-Grundy function g suffices.

4 The Sprague-Grundy Function of Nim-Celopt games

Calculation of the γ values of a digraph similar to Nim-Celopt appears in Example 1, §3 of [12]. Further examples appear in [10], [11].

We shall compute the g-values of the game-graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ of Nim-Celopt in order to get its P-positions. We make use of the following result.

Theorem 2. *Let $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ be the game-graph of any Celopt game. For $\boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{G}$ we have, $g(\boldsymbol{u} \oplus \boldsymbol{v}) = g(\boldsymbol{u}) \oplus g(\boldsymbol{v})$.*

Proof. Theorem 1 of [12] implies that $g(\boldsymbol{u} \oplus \boldsymbol{v}) = g(\boldsymbol{u}) \oplus g(\boldsymbol{v})$ for any Celata game. The same holds for any Celopt game by Lemma 1. ■

Lemma 2. *For the game-graph of Nim-Celopt$_s$, $g(\boldsymbol{z}_k) = g_s(\boldsymbol{z}_k) = \mathrm{mex}\{g(\boldsymbol{z}_{i_1}) \oplus \ldots \oplus g(\boldsymbol{z}_{i_j}) : 1 \le i_1 < \ldots < i_j < k, \quad j \le s\}$.*

Proof. By definition, $g(\boldsymbol{z}_k) = \mathrm{mex}\{g(\boldsymbol{z}_{i_1} \oplus \ldots \oplus \boldsymbol{z}_{i_j}) : 0 \le i_1 < \ldots < i_j < k, \quad j \le s\}$. The result now follows from Theorem 2. ■

Example 1. We have,

$g_1(\boldsymbol{z}_i)_{i \ge 1} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \ldots\}$,
$g_2(\boldsymbol{z}_i)_{i \ge 1} = \{1, 2, 4, 7, 8, 11, 13, 14, 16, \ldots\}$,
$g_3(\boldsymbol{z}_i)_{i \ge 1} = \{1, 2, 4, 8, 15, 16, 32, 51, 64, \ldots\}$.

Definition 2. The *left-shift* of $\boldsymbol{u} = (u^1, \ldots, u^n)$ is $\mathrm{L}(\boldsymbol{u}) = (u^2, \ldots, u^n)$. The *right-shift* with *parity bit* $w'(\boldsymbol{u})$ is $\mathrm{R}(\boldsymbol{u}) = (w'(\boldsymbol{u}), u^1 \ldots, u^n)$.

Remark 2. Note that $\mathrm{R}(\boldsymbol{u})$ is always evil. In terms of numerical values, $|\mathrm{L}(\boldsymbol{u})| = \lfloor |\boldsymbol{u}|/2 \rfloor$, $|\boldsymbol{u}| = \lfloor |\mathrm{R}(\boldsymbol{u})|/2 \rfloor$, $\mathrm{L}(\mathrm{R}(\boldsymbol{u})) = \boldsymbol{u}$; and $\mathrm{R}(\mathrm{L}(\boldsymbol{u})) = \boldsymbol{u}$ if and only if $\boldsymbol{u}$ is evil.

Denote by $\mathcal{P}_s$, $\mathcal{N}_s$, $F_s(\boldsymbol{u})$ the set of P-positions, N-positions, followers of $\boldsymbol{u}$ in Nim-Celopt$_s$ respectively.

Theorem 3. *Let $s \in \mathbb{Z}_{>0}$ be odd, $s + 1 \le n$. Then $\mathcal{P}_{s+1} = \{R(\boldsymbol{u}) : \boldsymbol{u} \in \mathcal{P}_s\}$.*

Notation 1. It is sometimes convenient to denote vectors $\boldsymbol{u} = (u^1, \ldots, u^n)$ by the superscripts i for which $u^i = 1$. Thus $\boldsymbol{u} = (3, 7)$ denotes a vector for which $u^3 = u^7 = 1$, and $u^i = 0$ for all $i \notin \{3, 7\}$. This notation is used, inter alia, in Example 2 and in the proof of Theorem 6.

Example 2.

n	1	2	3	4	5	6	7	8	9	10	11
$g_1(\boldsymbol{z}_n)$	1	2	3	4	5	6	7	8	9	10	11
$g_2(\boldsymbol{z}_n)$	1	2	4	7	8	11	13	14	16	19	21

We have

$$\mathcal{P}_1 = \{(2,4,6),(2,3,4,5),\ldots\}, \quad \mathcal{P}_2 = \{(1,3,5,7),(3,4,5,6),\ldots\},$$

where $\mathrm{R}(2,4,6) = (1,3,5,7)$, with $g_2(1)\oplus g_2(3)\oplus g_2(5)\oplus g_2(7) = 1\oplus 4\oplus 8\oplus 13 = 0$; and $\mathrm{R}(2,3,4,5) = (3,4,5,6)$, with $g_2(3)\oplus g_2(4)\oplus g_2(5)\oplus g_2(6) = 4\oplus 7\oplus 8\oplus 11 = 0$.

Proof of Theorem 3. Let $\mathcal{P}'_{s+1} = \{\mathrm{R}(\boldsymbol{u}) : \boldsymbol{u} \in \mathcal{P}_s\}$, $\mathcal{N}'_{s+1} = V \setminus \mathcal{P}'_{s+1}$. Since the groundgraph of Nim-Celopt is acyclic, its P- and N-positions are unique. By (1), (2) it therefore suffices to show the following: if $\boldsymbol{u} \in \mathcal{P}'_{s+1}$, then $F_{s+1}(\boldsymbol{u}) \subseteq \mathcal{N}'_{s+1}$, and if $\boldsymbol{u} \in \mathcal{N}'_{s+1}$, then $F_{s+1}(\boldsymbol{u}) \cap \mathcal{P}'_{s+1} \neq \emptyset$.

Let $\boldsymbol{u} \in \mathcal{P}'_{s+1}$, $\boldsymbol{v} \in F_{s+1}(\boldsymbol{u})$. The move from $\boldsymbol{u}$ to $\boldsymbol{v}$ is done by selecting an occupied z_k and firing it together with $p \leq s+1$ followers of z_k. Remark 2 implies that $\mathcal{P}'_{s+1}$ has only evil vectors. Assume $\boldsymbol{v} \in \mathcal{P}'_{s+1}$. Since s is odd, we have $p \leq s$. Thus there exists of a move from $\mathrm{L}(\boldsymbol{u}) \in \mathcal{P}_s$ to $\mathrm{L}(\boldsymbol{v}) \in \mathcal{P}_s$, a contradiction.

Now let $\boldsymbol{u} \in \mathcal{N}'_{s+1}$. Put $\boldsymbol{w} = \mathrm{L}(\boldsymbol{u})$. We consider two cases.

(i) $\boldsymbol{w} \in \mathcal{N}_s$. Then there exists $\boldsymbol{z} \in F_s(\boldsymbol{w}) \cap \mathcal{P}_s$. Let $\boldsymbol{v} = \mathrm{R}(\boldsymbol{z})$. Then $\boldsymbol{v} \in \mathcal{P}'_{s+1}$. The move from $\boldsymbol{w}$ to $\boldsymbol{z}$ involves complementing some occupied vertex z_k and a p-neighborhood of z_k, for some $p \leq s$. Note that the occupied vertex z_k in $\boldsymbol{w}$, is the occupied vertex z_{k+1} in $\boldsymbol{u}$. Hence it is possible to move from $\boldsymbol{u}$ to $\boldsymbol{v}$ directly, by complementing z_{k+1} and a suitable p-neighborhood of z_{k+1} for some $p \leq s$, and by controlling the occupancy of z_0. See Fig. 1. Therefore $\boldsymbol{v} \in F_{s+1}(\boldsymbol{u}) \cap \mathcal{P}'_{s+1}$.

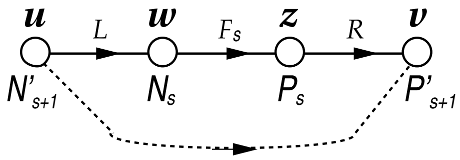

Figure 1: Case (i) of the proof.

(ii) $\boldsymbol{w} \in \mathcal{P}_s$. If $\boldsymbol{w} = \Phi$ we are done, since then $\boldsymbol{w} \in F_{s+1}(\boldsymbol{u}) \cap \mathcal{P}'_{s+1}$. Hence we may assume that there is a token on $\boldsymbol{w}$. Without loss of generality, assume that $w^0 = 1$, where $\boldsymbol{w} = (w^0,\ldots,w^{n-1})$. Put $\boldsymbol{w}' = (1-w^0,\ldots,w^{n-1})$. Then $\boldsymbol{w}' \in F_s(\boldsymbol{w}) \cap \mathcal{N}_s$. Thus there exists $\boldsymbol{z} \in F_s(\boldsymbol{w}') \cap \mathcal{P}_s$. Put $\boldsymbol{v} = \mathrm{R}(\boldsymbol{z})$. Then $\boldsymbol{v} \in \mathcal{P}'_{s+1}$. Note that we can move from $\boldsymbol{u}$ to $\boldsymbol{v}$ directly (see Fig. 2): it involves complementing at most $s+3$ vertices z_i. Now if $\boldsymbol{u}$ were evil, then Remark 2 would imply that $\boldsymbol{u} = \mathrm{R}(\boldsymbol{w}) \in \mathcal{P}'_{s+1}$. Hence $\boldsymbol{u}$ is odious. Also $\boldsymbol{v}$ is evil. Therefore at most $s+2$ vertices are complemented in moving from $\boldsymbol{u}$ to $\boldsymbol{v}$. Thus $\boldsymbol{v} \in F_{s+1}(\boldsymbol{u}) \cap \mathcal{P}'_{s+1}$. ∎

Corollary 1. *If $\boldsymbol{u} \in \mathcal{P}$ in a Nim-Celopt$_s$ game with s even, then $\boldsymbol{u}$ is evil.*

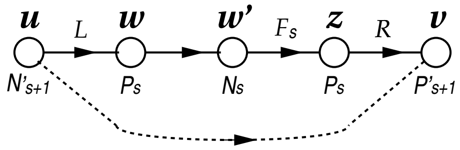

Figure 2: Case (ii) of the proof.

Proof. Follows directly from the statement of Theorem 3 and from Remark 2. ■

Definition 3. (i) For any Nim-Celopt game, if $g(\boldsymbol{z}_k)$ ($\in \mathbb{Z}_{\geq 0}$) is not a power of 2, then $g(\boldsymbol{z}_k)$ is called a *seed.*
(ii) Denote by $t = t(G) \in \mathbb{Z}_{\geq 0}$ the smallest number satisfying $g(\boldsymbol{u}) \leq 2^t - 1$ for every $\boldsymbol{u}$ in a Celopt game.
(iii) In a Celopt graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, put $\boldsymbol{V}_i = \{\boldsymbol{u} \in \boldsymbol{V} : g(\boldsymbol{u}) = i\}$, for $0 \leq i < 2^t$.

Note that, for any $k \in \mathbb{Z}_{\geq 0}$, $g(\boldsymbol{z}_k)$ is a Nim-sum of lower powers of two if and only if $g(\boldsymbol{z}_k)$ is a seed.

Theorem 4. *Let $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ be the Celopt graph of the finite acyclic digraph $G = (V, E)$. Then g is a homomorphism from $\boldsymbol{V}$ onto* $\mathrm{GF}(2)^t$ for t given in Definition 3, with kernel $\boldsymbol{V}_0$ and quotient space $\boldsymbol{V}/\boldsymbol{V}_0 = \{\boldsymbol{V}_i : 0 \leq i < 2^t\}$, $\dim(\boldsymbol{V}) = m + t$, where $m = \dim(\boldsymbol{V}_0)$.

Proof. This is a special case of Theorem 2 of [12] for the case of Celata played on an acyclic digraph. The same holds for any Celopt game by Lemma 1. ■

Remark 3. (i) Theorem 4 implies directly that every g-value in $\{0, \ldots, 2^t - 1\}$ is assumed in a Celopt game, in particular in a Nim-Celopt game. Further, $t = \lfloor \log_2 g(\boldsymbol{u}) \rfloor + 1$ for every $\boldsymbol{u}$ for which $g(\boldsymbol{u}) > 2^{t-1}$.
(ii) Note that $\mathcal{P} = \boldsymbol{V}_0$, so $\mathcal{P}$ is a linear subspace of $\boldsymbol{V}$.
(iii) Every seed $g(\boldsymbol{u})$ produces a new P-position in a Nim-Celopt game: it is the (Nim)-sum σ of smaller powers of 2, so $g(\boldsymbol{u}) \oplus \sigma = 0$.
(iv) Theorem 4 and Lemma 2 imply that for every $k \in \{0, \ldots, t-1\}$, 2^k is assumed by some $g(\boldsymbol{z}_i)$.
(v) In stating in Theorem 4 that g is a homomorphism, we also identified the numerical value of g with the binary vector representation of that value. We shall do so also below.

Theorem 5. *For all $n \in \mathbb{Z}_{>1}$, if s is odd, then $g_s(\boldsymbol{z}_{n-1}) = \lfloor g_{s+1}(\boldsymbol{z}_n)/2 \rfloor$, and $g_{s+1}(\boldsymbol{z}_n)$ is an odious number.*

Proof. The first part is proved by induction on n. For $n = 2$, $g_s(\boldsymbol{z}_1) = 1$ for all $s \in \mathbb{Z}_{>0}$, and $\lfloor g_{s+1}(\boldsymbol{z}_2)/2 \rfloor = \lfloor 2/2 \rfloor = 1$. So we may assume $n \geq 3$. We consider two cases.

(i) $g_{s+1}(\boldsymbol{z}_n)$ is a seed. Then Remark 3(iv) implies that $g_{s+1}(\boldsymbol{z}_n)$ can be written in the form $(u^1, \ldots, u^n)$ such that

$$g_{s+1}(\boldsymbol{z}_n) = \sum_{i=1}^{n} u^i g_{s+1}(\boldsymbol{z}_i), \tag{7}$$

where $u^n = 0$, and $u^i \neq 0 \implies g_{s+1}(\boldsymbol{z}_i)$ is a nonnegative power of 2 for all $i \in \{1, \ldots, n\}$. Since $g_{s+1}(\boldsymbol{z}_1) = 1$ we thus have,

$$\left\lfloor \frac{g_{s+1}(\boldsymbol{z}_n)}{2} \right\rfloor = \left\lfloor \frac{\sum_{i=1}^{n} u^i g_{s+1}(\boldsymbol{z}_i)}{2} \right\rfloor = \sum_{i=2}^{n} u^i \ \frac{g_{s+1}(\boldsymbol{z}_i)}{2} \ = \sum\nolimits'_{i \in \{2,\ldots,n-1\}} u^i \ \frac{g_{s+1}(\boldsymbol{z}_i)}{2}.$$

Note that (7) implies $(u^1, \ldots, u^{n-1}, 1) \in \mathcal{P}_{s+1}$. By Corollary 1, $(u^1, \ldots, u^{n-1}, 1)$ is evil, hence $g_{s+1}(\boldsymbol{z}_n) = (u^1, \ldots, u^{n-1}, u^n)$ is odious $(u^n = 0)$. Further, Theorem 3 implies $\mathrm{L}(u^1, \ldots, u^{n-1}, 1) = (u^2, \ldots, u^{n-1}, 1) \in \mathcal{P}_s$, where the rightmost 1-bit is at position $n-1$ of the vector. This and the induction hypothesis then imply,

$$g_s(\boldsymbol{z}_{n-1}) = \sum\nolimits'_{i \in \{2,\ldots,n-1\}} u^i g_s(\boldsymbol{z}_{i-1}) = \sum\nolimits'_{i \in \{2,\ldots,n-1\}} u^i \ \frac{g_{s+1}(\boldsymbol{z}_i)}{2}.$$

Thus $g_s(\boldsymbol{z}_{n-1}) = \lfloor g_{s+1}(\boldsymbol{z}_n)/2 \rfloor$.

(ii) $g_{s+1}(\boldsymbol{z}_n)$ is a power of 2, i.e., $g_{s+1}(\boldsymbol{z}_n) = 2^t$, where $t \in \mathbb{Z}_{\geq 0}$ is the smallest number satisfying $g_{s+1}(\boldsymbol{z}_{n-1}) \leq 2^t - 1$. In particular, $g_{s+1}(\boldsymbol{z}_n)$ is odious. By monotonicity and Remark 3(iv), $g_{s+1}(\boldsymbol{z}_{n-1}) \geq 2^{t-1}$. Since a power of 2 is not the Nim-sum of lower powers of 2, there is no binary vector $(u^1, \ldots, u^{n-1}, 1) \in \mathcal{P}_{s+1}$. Hence by Theorem 3, there is no binary vector $(u^2, \ldots, u^{n-1}, 1) \in \mathcal{P}_s$, so also $g_s(\boldsymbol{z}_{n-1}) = 2^r$, where r is the smallest nonnegative integer satisfying $g_s(\boldsymbol{z}_{n-2}) \leq 2^r - 1$. By the induction hypothesis, $2^{t-2} \leq \lfloor g_{s+1}(z_{n-1})/2 \rfloor = g_s(z_{n-2}) < 2^{t-1}$. It follows that $r = t-1$, so $g_s(\boldsymbol{z}_{n-1}) = 2^{t-1} = g_{s+1}(\boldsymbol{z}_n)/2$. ∎

5 An Efficient Computation of the Lexicode

As was remarked at the end of §2, the P-positions of a game constitute a code. In this section we present an algorithm for computing a basis for the P-positions of Nim-Celopt$_s$, played on the Nim-graph $G = (V, E)$, where $V = (z_0, \ldots, z_n)$. By Theorem 1, it is the lexicode with distance $d = s + 2$. We let $t \in \mathbb{Z}_{\geq 0}$ be the smallest number satisfying $g_s(\boldsymbol{z}_n) \leq 2^t - 1$.

Theorem 6. *For any Nim-Celopt$_s$ game,*

(a) $\dim \mathcal{P}_s$ *is the number of seeds* $\leq 2^t - 1$.

(b) *If* $g_s(z_m)$ *is a seed, then there is* $\boldsymbol{u} \in \mathcal{P}_s$ *with* $u^m = 1$; *and* $w(\boldsymbol{u}) = s + 2$.

Example 3. Illustration of Theorems 5 and 6.

n	1	2	3	4	5	6	7	8	9	10	11
$g_1(\boldsymbol{z}_n)$	1	2	3	4	5	6	7	8	9	10	11
$g_2(\boldsymbol{z}_n)$	1	2	4	7	8	11	13	14	16	19	21
$t = \lfloor log_2\ g_2(\boldsymbol{z}_n) \rfloor + 1$	1	2	3	3	4	4	4	4	5	5	5
$\dim \mathcal{P}_2 = n - t$	0	0	0	1	1	2	3	4	4	5	6

Proof of Theorem 6. (a) Any P-position can be written in the form

$$\boldsymbol{u} = ((i_1, \ldots, i_l, m) : i_1 < \ldots < i_l < m,\ g(\boldsymbol{z}_{i_1}) \oplus \ldots \oplus g(\boldsymbol{z}_{i_l}) \oplus g(\boldsymbol{z}_m) = 0)$$

for suitable $m \le n$ (we used Notation 1 for $\boldsymbol{u} = (u^0, \ldots, u^n)$). By Remark 1(ii), $g(\boldsymbol{z}_{i_1}) < \ldots < g(\boldsymbol{z}_{i_l}) < g(\boldsymbol{z}_m)$. If $g(\boldsymbol{z}_m) = 2^k$ is a power of 2, then $\boldsymbol{u}$ cannot be a P-position, since 2^k is not the sum of distinct nonnegative integers $< 2^k$. Let $T := \{g(\boldsymbol{z}_j) : 0 < j < m,\ g(\boldsymbol{z}_j) \text{ is a power of } 2\}$. Whenever $g(\boldsymbol{z}_m)$ is a seed, there clearly are powers of 2 such that $g(\boldsymbol{z}_m) \oplus \sum'_{g(\boldsymbol{z}_j) \in R} g(\boldsymbol{z}_j) = 0$, where R is a suitable subset of T. Thus, $\boldsymbol{v} := g(\boldsymbol{z}_m) \oplus \sum'_{g(\boldsymbol{z}_j) \in R} \in \mathcal{P}$. Denote by S the set of all such $\boldsymbol{v}$. All $\boldsymbol{v} \in S$ are linearly independent, since every $\boldsymbol{z}_m$ such that $g(\boldsymbol{z}_m)$ is a seed appears in S only once, and then $u^k = 0$ for all $k > m$. Thus S constitutes a basis for $\mathcal{P}$.

(b) The first part was already proved in (a). For $\boldsymbol{u} \in S$, Lemma 2 implies that $g(\boldsymbol{z}_m)$ is the *smallest* nonnegative integer which is not the Nim-sum of p of its followers, for all $p \in \{0, \ldots, q\}$. Further, $g(\boldsymbol{z}_m) = g(\boldsymbol{z}_{i_l}) \oplus x$, where $x := \sum'^{l-1}_{j=1} g(\boldsymbol{z}_{i_j})$. Clearly, $x < g(\boldsymbol{z}_{i_l}) < g(\boldsymbol{z}_m)$. Hence there are $0 < k_1 < \ldots < k_p$ with $p \le s$, such that $x = \sum'^{p}_{j=1} g(\boldsymbol{z}_{k_j})$. Therefore $g(\boldsymbol{z}_m) = g(\boldsymbol{z}_{i_l}) \oplus \sum'^{p}_{j=1} g(\boldsymbol{z}_{k_j})$ is the Nim-sum of at most $s+1$ summands, so for $\boldsymbol{u} = \boldsymbol{z}_m \oplus \boldsymbol{z}_{i_l} \oplus \sum'^{p}_{j=1} g(\boldsymbol{z}_{k_j})$, we have $\boldsymbol{u} \in \mathcal{P}$ and $w(\boldsymbol{u}) \le s+2$. Since Theorem 1 implies that $\mathcal{P}$ is a set of codewords of Hamming distance $d \ge s+2$, so in particular $w(\boldsymbol{u}) \ge s+2$, we have $w(\boldsymbol{u}) = s+2$. ■

The following is an algorithm for computing $\mathcal{P}_s \cap \{0,1\}^n$, i.e., the basis for a linear code (LinCod).

Algorithm LinCod

1. For $m \in \{1, \ldots, n\}$, compute $g_s(\boldsymbol{z}_m) = \text{mex}\{g(\boldsymbol{z}_{i_1}) \oplus \ldots \oplus g(\boldsymbol{z}_{i_j}) : 1 \le i_1 < \ldots < i_j < m, \quad j \le s\}$.

2. Compute a basis member of $\mathcal{P}$ for each seed: it is the seed and all the powers of two appearing in it. (Thus, if the seed is 13, then the basis member induced by 13 is $\{13, 8, 4, 1\}$.) End.

Verification and Complexity of Algorithm LinCod

Step 1 takes time $n \sum_{i=1}^{s} \binom{n}{i} = O\left(n(n + n^2 + \ldots + n^s)\right) = O(n^{s+1})$. This dominates the complexity of step 2. Then Theorem 6(b) and its proof show how to compute the basis members. By Theorem 1, the time complexity is

$O(n^{d-1})$. The space complexity is $O(n^{d-2})$. It is possible to formulate a different algorithm, which takes time $O(n^{2d})$ and space $O(n)$, by doing an exhaustive search for a vector u with the properties of Theorem 6(b).

Remark 4. (i) If we are only interested in computing the dimension of $\mathcal{P}$, it suffices, by Theorem 6(a), to count the number of seeds, deleting step 2.
(ii) For s even, Theorem 5 implies that the complexity is $O(n^s) = O(n^{d-2})$. For the special case $s = 2$, the algorithm is linear in n, since for $s = 1$ (ordinary Nim — see also Example 3), we have $g(n) = n$.

Example 4. Let $n = 8$, $s = 3$. Then $d = 5$ by Theorem 1. We get $\left(g(\boldsymbol{z}_i)_{i=1}^8\right) = (1, 2, 4, 8, 15, 16, 32, 51)$. The seeds are $15, 51$, hence a basis is given by

$$\{11111000, 11000111\}.$$

The set of all P-positions is the linear span of the basis vectors, so $\mathcal{P} = \{\Phi, 11111000, 11000111, 00111111\}$.

6 Epilogue

We note that the *Turning Turtles* and *Mock Turtles* games [1] Ch. 14, have been mentioned in connection with lexicodes in [5], [18]. It turns out that the P-positions of these games also constitute the lexicode. This can be proved the same way as Theorem 1 above. Initial segments of the g-values for various values of s (in a different notation) of Turning Turtles are given in Table 3 in [1], Ch. 14. It was observed there that for $s = 2$ the g-values are odious; but it was not mentioned there that the g-values are in fact odious for all even s (Theorem 5 above).

In [13], two classes of games were defined:

(i) Games People Play (*PlayGames*): games that are challenging to the point that people will purchase them and play them.

(ii) Games Mathematicians Play (*MathGames*): games that are challenging to a mathematician or other scientist to play with and ponder about, but not necessarily to "the man in the street".

Examples of PlayGames are chess, go, hex, reversi; of MathGames: Nim-type games, Wythoff games, annihilation games, octal games.

Some "rule of thumb" properties, which seem to hold for the majority of PlayGames and MathGames were listed there. One of the referees of this paper asked us to state that Celata and Celopt games belong to the class of MathGames.

Theorem 3 shows how to get a code with parameters $(n + 1, k, d + 1)$ from a code with parameters (n, k, d), for d odd, by adding the "parity check bit" $w'(\boldsymbol{u})$. This phenomenon is well-known in coding theory. It is equivalent to the Mock Turtle Theorem in [1], Ch. 14. See also [18].

To summarize, the central result of this paper is an $O(n^{d-1})$ algorithm for computing the basis elements of a lexicode. The previously known algorithms have complexity exponential in n: $O(2^{2n})$ ([5]), $O(2^{n(1-3/d)+1})$ ([19]). All the basis elements which the algorithm produces have weight $\geq d$. The results were obtained from the theory of combinatorial games; specifically, from two-player games on cellular automata. There aren't many meaningful applications of the theory of impartial combinatorial games to areas outside of game theory. This is one of them.

Acknowledgment. We thank Michael Langberg for inspiring conversations, which led to an improvement in the complexity of the algorithm.

References

[1] E. R. Berlekamp, J. H. Conway and R. K. Guy [1982], *Winning Ways for your Mathematical Plays* (two volumes), Academic Press, London.

[2] Bounds on the minimum distance of linear codes http://www.win.tue.nl/~aeb/voorlincod.html

[3] R. A. Brualdi and V. S. Pless [1993], Greedy codes, *J. Combin. Theory*, (Ser. A) **64**, 10–30.

[4] J. H. Conway [1990], Integral lexicographic codes, *Discrete Math.* **83**, 219–235.

[5] J. H. Conway and N. J. A. Sloane [1986], Lexicographic codes: error-correcting codes from game theory, *IEEE Trans. Inform. Theory* **IT-32**, 337–348.

[6] T. S. Ferguson [1984], Misère annihilation games, *J. Combin. Theory* (Ser. A) **37**, 205–230.

[7] A. S. Fraenkel [1974], Combinatorial games with an annihilation rule, in: *The Influence of Computing on Mathematical Research and Education*, Missoula MT, August 1973, Proc. Symp. Appl. Math., (J. P. LaSalle, ed.), Vol. 20, Amer. Math. Soc., Providence, RI, pp. 87–91.

[8] A. S. Fraenkel [1996], Error-correcting codes derived from combinatorial games, in: *Games of No Chance*, Proc. MSRI Workshop on Combinatorial Games, July, 1994, Berkeley, CA (R. J. Nowakowski, ed.), MSRI Publ. Vol. 29, Cambridge University Press, Cambridge, pp. 417–431.

[9] A. S. Fraenkel [1996], Scenic trails ascending from sea-level Nim to alpine chess, in: *Games of No Chance*, Proc. MSRI Workshop on

Combinatorial Games, July, 1994, Berkeley, CA (R. J. Nowakowski, ed.), MSRI Publ. Vol. 29, Cambridge University Press, Cambridge, pp. 13–42.

[10] A. S. Fraenkel [2001], Virus versus mankind, *Proc. 2nd Intern. Conference on Computers and Games CG'2000* (T. Marsland and I. Frank, eds.), Vol. 2063, Hamamatsu, Japan, Oct. 2000, Lecture Notes in Computer Science, Springer, pp. 204–213.

[11] A. S. Fraenkel [2002], Mathematical chats between two physicists, in: *Puzzler's Tribute: a Feast for the Mind*, honoring Martin Gardner (D. Wolfe and T. Rodgers, eds.), A K Peters, Natick, MA, pp. 383–386.

[12] A. S. Fraenkel [2002], Two-player games on cellular automata, in: *More Games of No Chance*, Proc. MSRI Workshop on Combinatorial Games, July, 2000, Berkeley, CA, MSRI Publ. (R. J. Nowakowski, ed.), Vol. 42, Cambridge University Press, Cambridge, pp. 279–306.

[13] A. S. Fraenkel [2002], Complexity, appeal and challenges of combinatorial games, expanded version of a keynote address at Dagstuhl Seminar, Feb. 17–22, 2002. To appear in *Theoret. Comput. Sci.*, special issue on Algorithmic Combinatorial Game Theory. Preprint at http://www.wisdom.weizmann.ac.il/~fraenkel

[14] A. S. Fraenkel, U. Tassa and Y. Yesha [1978], Three annihilation games, *Math. Mag.* **51**, 13–17, special issue on Recreational Math.

[15] A. S. Fraenkel and Y. Yesha [1976], Theory of annihilation games, *Bull. Amer. Math. Soc.* **82**, 775–777.

[16] A. S. Fraenkel and Y. Yesha [1982], Theory of annihilation games — I, *J. Combin. Theory* (Ser. B) **33**, 60–86.

[17] F. J. MacWilliams and N. J. A. Sloane [1978], *The Theory of Error-Correcting Codes*, North-Holland, Elsevier, Amsterdam.

[18] V. Pless [1991], Games and codes, in: *Combinatorial Games*, Proc. Symp. Appl. Math. (R.K. Guy, ed.), Vol. 43, Amer. Math. Soc., Providence, RI, pp. 101–110.

[19] A. Trachtenberg [2002], Designing lexicographic codes with a given Trellis complexity, *IEEE Trans. Inform. Theory* **48**, 89–100.

[20] A. Trachtenberg and A. Vardy [1997], *Lexicographic Codes:Constructions, Bounds, and Trellis Complexity*, Proc. 31st Annual Conference on Information Sciences and Systems. (http://citeseer.nj.nec.com/448945.html)

[21] Y. Yesha, Theory of annihilation games [1978], Ph.D. Thesis, Weizmann Institute of Science, Rehovot, Israel.