Gibón vol. IX (2011) pp. 51–69
Regular Research Article

ISSN 1655-7247

On Generalized Choral Sequences

Joel R. Noche
Department of Mathematics
Ateneo de Naga University

Abstract

Generalized choral sequences are infinite binary words $(c_n)_{n\geq 0}$ defined by $c_{3i+r_0} = \texttt{0}$, $c_{3i+r_1} = \texttt{1}$, and $c_{3i+r_c} = c_i$ (where the r's are distinct fixed elements of $\{0, 1, 2\}$) for all non-negative integers i. We present some of their properties. In particular, we show how each is generated by a deterministic finite automaton with output and how each is a fixed point of a uniform morphism. We also look at their subword complexity and Lyndon factorizations.

Keywords: combinatorics on words

We use definitions and notation from Allouche and Shallit (2003). Denote by $\mathbb{N}$ the set of non-negative integers and by $\mathbb{Z}^+$ the set of positive integers. A *word* is a concatenation (a sequence) of *letters* (symbols) chosen from an *alphabet* (a non-empty set of letters). A *finite word* has a finite *length* (the number of letters it contains) and is denoted by a lowercase italic letter. The length of a finite word w is denoted by $|w|$. The *empty word*, denoted by ϵ, has a length of zero. A one-sided right-infinite word (which we will simply call an *infinite word*) is a map from $\mathbb{N}$ to an alphabet and is denoted by a lowercase boldface letter. We call a word over the alphabet $\{\texttt{0},\texttt{1}\}$ a *binary word*.

Concatenation of words is denoted by the juxtaposition of their symbols. For example, if $w = \texttt{0}$ and $x = \texttt{01}$, then $wx = \texttt{001}$, $xw = \texttt{010}$, $w\texttt{1} = x$, and $wwxwx = w(wx)^2 = \texttt{0}^3\texttt{10}^2\texttt{1}$. We define $w^1 = w$ and $w^0 = \epsilon$ for any finite word w. If x is a finite non-empty word, then x^ω (the superscript is omega) is the infinite word $xxx\cdots$.

A word y is a *subword* of w if there exist x and z such that

$w = xyz$. If $x = \epsilon$, then y is a *prefix* of w. If $z = \epsilon$, then y is a *suffix* of w. If $x = \epsilon$ and $z \neq \epsilon$, then y is a *proper prefix* of w. If $x \neq \epsilon$ and $z = \epsilon$, then y is a *proper suffix* of w.

We extend these definitions to infinite words. An infinite word $\mathbf{w}$ can be written as an infinite sequence of finite subwords $(w_n)_{n\geq 0} = w_0 w_1 w_2 \cdots$. A word y is a subword of $\mathbf{w}$ if there exist x and $\mathbf{z}$ such that $\mathbf{w} = xy\mathbf{z}$. If $x = \epsilon$, then y is a prefix (and a proper prefix) of $\mathbf{w}$. A word $\mathbf{z}$ is a subword (and a suffix) of $\mathbf{w}$ if there exists a y such that $\mathbf{w} = y\mathbf{z}$. If $y \neq \epsilon$, then $\mathbf{z}$ is a proper suffix of $\mathbf{w}$.

The set of all finite words made up of letters chosen from an alphabet Σ is denoted by Σ^*. Note that $\epsilon \in \Sigma^*$. If $a \in \Sigma$ and $w \in \Sigma^*$, then $|w|_a$ denotes the number of occurrences of the letter a in the word w. The frequency of a letter a in an infinite word $\mathbf{w} = (w_n)_{n\geq 0}$ (where the w's are letters), denoted by $\mathrm{Freq}_{\mathbf{w}}(a)$, is $\lim_{n\to\infty} \frac{1}{n}|w_0 w_1 \cdots w_{n-1}|_a$, if this limit exists.

Let Σ and Δ be alphabets. A *morphism* is a map μ from Σ^* to Δ^* that obeys the identity $\mu(wx) = \mu(w)\mu(x)$ for all words $w, x \in \Sigma^*$. If $\Sigma = \Delta$, then the application of a morphism can be iterated. For example, if μ is the morphism mapping $\mathtt{0}$ to $\mathtt{1}$ and $\mathtt{1}$ to $\mathtt{10}$, then $\mu(\mathtt{0}) = \mathtt{1}$, $\mu^2(\mathtt{0}) = \mu(\mu(\mathtt{0})) = \mu(\mathtt{1}) = \mathtt{10}$, and so on. We define $\mu^1(w) = \mu(w)$ and $\mu^0(w) = w$ for any word w.

A morphism $\mu : \Sigma^* \to \Delta^*$ is *k-uniform* if there is a constant k such that $|\mu(a)| = k$ for all $a \in \Sigma$. A *coding* is a 1-uniform morphism.

A *fixed point* of a morphism $\mu : \Sigma^* \to \Sigma^*$ is a finite word w (or infinite word $\mathbf{w}$) such that $\mu(w) = w$ (or $\mu(\mathbf{w}) = \mathbf{w}$). If there exists a letter $a \in \Sigma$ such that $\mu(a) = ax$ and x is a word composed of letters $x_i \in \Sigma$ such that $\mu^m(x_i) \neq \epsilon$ for any $m \in \mathbb{Z}^+$, then the morphism μ is *prolongable* on the letter a. If so, then $\lim_{m\to\infty} \mu^m(a)$ (denoted by $\mu^\omega(a)$) is the fixed point of μ iterated on a, where the length of the iterates from the letter a tends to infinity.

Let $\overline{w}$ denote the *complement* of the finite binary word w (and $\overline{\mathbf{w}}$ denote the complement of the infinite binary word $\mathbf{w}$) where the overbar represents the morphism mapping $\mathtt{0} \to \mathtt{1}$ and $\mathtt{1} \to \mathtt{0}$. For example, if $w = \mathtt{001}$, then $\overline{w} = \mathtt{110}$.

Given a finite word $w = a_0 a_1 \cdots a_n$, where the a's are letters, its *reversal*, denoted by w^R, is $a_n \cdots a_1 a_0$.

Definition 1. *A* generalized choral sequence *is an infinite binary word* $\mathbf{c}(r_0, r_1, r_c, z) = (c_n)_{n\geq 0}$ *defined by* $c_{3i+r_0} = \mathtt{0}$, $c_{3i+r_1} = \mathtt{1}$, $c_{3i+r_c} = c_i$, *and* $c_0 = z$ *(where the* r*'s are distinct fixed elements of*

$\{0,1,2\}$ *and* $z = \mathtt{0}$ *if* $r_0 = 0$, $z = \mathtt{1}$ *if* $r_1 = 0$, *and* z *could either be* $\mathtt{0}$ *or* $\mathtt{1}$ *if* $r_c = 0$*) for all* $i \in \mathbb{N}$.

There are eight distinct choral sequences: (Spaces have been inserted to improve readability.)

$\mathbf{c}(0,2,1,\mathtt{0}) = \mathtt{001\,001\,011\ \ 001\,001\,011\ \ 001\,011\,011} \cdots$
$\mathbf{c}(1,2,0,\mathtt{0}) = \mathtt{001\,001\,101\ \ 001\,001\,101\ \ 101\,001\,101} \cdots$
$\mathbf{c}(0,1,2,\mathtt{0}) = \mathtt{010\,011\,010\ \ 010\,011\,011\ \ 010\,011\,010} \cdots$
$\mathbf{c}(2,1,0,\mathtt{0}) = \mathtt{010\,110\,010\ \ 110\,110\,010\ \ 010\,110\,010} \cdots$
$\mathbf{c}(1,2,0,\mathtt{1}) = \mathtt{101\,001\,101\ \ 001\,001\,101\ \ 101\,001\,101} \cdots$
$\mathbf{c}(1,0,2,\mathtt{1}) = \mathtt{101\,100\,101\ \ 101\,100\,100\ \ 101\,100\,101} \cdots$
$\mathbf{c}(2,1,0,\mathtt{1}) = \mathtt{110\,110\,010\ \ 110\,110\,010\ \ 010\,110\,010} \cdots$
$\mathbf{c}(2,0,1,\mathtt{1}) = \mathtt{110\,110\,100\ \ 110\,110\,100\ \ 110\,100\,100} \cdots$

Sequence $\mathbf{c}(0,2,1,\mathtt{0})$ is Stewart's choral sequence. (Stewart (1995) presented the sequence $(c_n)_{n\geq 1}$ and not $(c_n)_{n\geq 0}$.) Sequence $\mathbf{c}(0,1,2,\mathtt{0})$ is from Berstel and Karhumäki (2003).

Definition 2. *A generalized choral sequence* $\mathbf{c}(r_0, r_1, r_c, z)$ *is called a* type-012 sequence *if* (r_0, r_1, r_c) *is a circular permutation of* $(0,1,2)$. *Otherwise (if* (r_0, r_1, r_c) *is a circular permutation of* $(2,1,0)$*), it is called a* type-210 sequence.

Some Properties

A previous work (Noche, 2008a) presented a characteristic function for generalized choral sequences as well as proofs of the following two theorems.

Theorem 1. *A generalized choral sequence is* cube-free*, that is, it does not contain any subword of the form* xxx*, where* x *is a non-empty finite subword.*

Theorem 2. *Given two generalized choral sequences, if they are both type-012 (or if they are both type-210) then any finite subword of one is a subword of the other. Otherwise, any finite subword of one is the complement of a subword of the other.*

The following theorem is related to the previous theorem and their proofs are similar.

Theorem 3. *Given a type-012 sequence and a type-210 sequence, any finite subword of one is the reversal of a subword of the other.*

Proof. Let $\mathbf{v} = (v_n)_{n\geq 0}$ be the given type-012 sequence and $\mathbf{w} = (w_n)_{n\geq 0}$ be the given type-210 sequence. We will show that any finite subword of $\mathbf{v}$ is the reversal of a subword of $\mathbf{w}$. (The proof that any finite subword of $\mathbf{w}$ is the reversal of a subword of $\mathbf{v}$ is similar.)

Because any generalized choral sequence has all the subwords `001`, `010`, `011`, `100`, `101`, and `110` (Noche, 2008a), any length-3 subword of $\mathbf{v}$ is the reversal of a subword of $\mathbf{w}$. That is, for a given $j \in \mathbb{N}$, there exists a $k \in \mathbb{N}$ such that $v_j v_{j+1} v_{j+2} = w_{k+2} w_{k+1} w_k$.

By Definition 2, there exists a subword $v_{3j+r_c} v_{3j+r_c+1} v_{3j+r_c+2} v_{3j+r_c+3} v_{3j+r_c+4} v_{3j+r_c+5} v_{3j+r_c+6} = v_j \mathtt{01} v_{j+1} \mathtt{01} v_{j+2}$ and a subword $w_{3k+r_c} w_{3k+r_c+1} w_{3k+r_c+2} w_{3k+r_c+3} w_{3k+r_c+4} w_{3k+r_c+5} w_{3k+r_c+6} = w_k \mathtt{10} w_{k+1} \mathtt{10} w_{k+2}$ such that $v_j \mathtt{01} v_{j+1} \mathtt{01} v_{j+2} = w_{k+2} \mathtt{01} w_{k+1} \mathtt{01} w_k$. Any length-5 subword of $\mathbf{v}$ is a subword of a length-7 subword of the form $v_{3j+r_c} v_{3j+r_c+1}\ v_{3j+r_c+2} v_{3j+r_c+3} v_{3j+r_c+4} v_{3j+r_c+5} v_{3j+r_c+6}$. Thus, any length-5 subword of $\mathbf{v}$ is the reversal of a subword of $\mathbf{w}$. (This is also true for subwords of length less than 5.)

Extending this reasoning to arbitrarily long finite subwords of similar form yields the result that any finite subword of $\mathbf{v}$ is the reversal of a subword of $\mathbf{w}$. □

Remark 1. *For any generalized choral sequence* $\mathbf{c}$*,* $\mathrm{Freq}_{\mathbf{c}}(\mathtt{0}) = \frac{1}{2}$ *and* $\mathrm{Freq}_{\mathbf{c}}(\mathtt{1}) = \frac{1}{2}$.

We first show that $\mathrm{Freq}_{\mathbf{c}}(\mathtt{0}) = \frac{1}{2}$. Let $\mathbf{c} = (c_n)_{n\geq 0}$. Consider the subword $c_0 c_1 \cdots c_{n-1}$ with length n. As $n \to \infty$, around $\frac{1}{3}$ of the letters of $c_0 c_1 \cdots c_{n-1}$ will be due to the subwords c_{3i+r_0}, around $\frac{1}{3}$ will be due to c_{3i+r_1}, and around $\frac{1}{3}$ will be due to c_{3i+r_c}, for $i \in \mathbb{N}$ such that $3i + r < n$.

Since $c_{3i+r_0} = \mathtt{0}$, the number of occurrences of `0` due to the subwords c_{3i+r_0} approaches $\frac{n}{3}$ as $n \to \infty$. Since $c_{3i+r_1} = \mathtt{1}$, the number of occurrences of `0` due to the subwords c_{3i+r_0} and c_{3i+r_1} still approaches $\frac{n}{3}$ as $n \to \infty$.

Now consider the subwords c_{3i+r_c}. Since $c_{3i+r_c} = c_i$, one third of the subwords c_{3i+r_c} will be `0`, and one third will be `1`. Of the remaining subwords, one third will be `0`, one third will be `1` and so on.

Thus, the number of occurrences of `0` in the subword $c_0 c_1 \cdots c_{n-1}$ as $n \to \infty$ approaches $\frac{n}{3}\left(1 + \frac{1}{3}\left(1 + \frac{1}{3}\left(1 + \cdots\right)\right)\right) =$

$\frac{n}{3} + \frac{n}{3^2} + \frac{n}{3^3} + \cdots$ and $\mathrm{Freq}_{\mathbf{c}}(\mathtt{0}) = \lim_{n\to\infty} \frac{1}{n}\left(\frac{n}{3} + \frac{n}{3^2} + \frac{n}{3^3} + \cdots\right) = \sum_{i=1}^{\infty} \frac{1}{3^i} = -1 + \sum_{i=1}^{\infty} \left(\frac{1}{3}\right)^{i-1} = -1 + \frac{1}{1-\frac{1}{3}} = \frac{1}{2}$.

Using similar reasoning, we can show that $\mathrm{Freq}_{\mathbf{c}}(\mathtt{1}) = \frac{1}{2}$.

Proposition 1. $\mathbf{c}(1,2,0,z) = z\mathbf{c}(0,1,2,\mathtt{0})$ *and* $\mathbf{c}(2,1,0,z) = z\mathbf{c}(1,0,2,\mathtt{1})$

Proof. We first show that $\mathbf{c}(1,2,0,z) = z\mathbf{c}(0,1,2,\mathtt{0})$. Let $(a_n)_{n\geq 0} = \mathbf{c}(1,2,0,z)$, $(b_n)_{n\geq 0} = \mathbf{c}(0,1,2,\mathtt{0})$, and $(c_n)_{n\geq 0} = z\mathbf{c}(0,1,2,\mathtt{0})$. Thus, $a_{3i+1} = \mathtt{0}$, $a_{3i+2} = \mathtt{1}$, $a_{3i} = a_i$, $b_{3i} = \mathtt{0}$, $b_{3i+1} = \mathtt{1}$, and $b_{3i+2} = b_i$ for all $i \in \mathbb{N}$. Also, $c_{i+1} = b_i$ for all $i \in \mathbb{N}$ so $c_{3i+1} = b_{3i} = \mathtt{0}$, $c_{3i+2} = b_{3i+1} = \mathtt{1}$, and $c_{3i+3} = b_{3i+2} = b_i = c_{i+1}$ for all $i \in \mathbb{N}$. Since $c_{3(i+1)} = c_{i+1}$ for $i \in \mathbb{N}$ and $c_{3\cdot 0} = c_0$, $c_{3i} = c_i$ for all $i \in \mathbb{N}$. Also, $a_{3i+1} = \mathtt{0} = c_{3i+1}$ and $a_{3i+2} = \mathtt{1} = c_{3i+2}$ for all $i \in \mathbb{N}$.

We will prove that $(a_n)_{n\geq 0} = (c_n)_{n\geq 0}$ using induction. Assume $a_n = c_n$ for all $n \leq k$, where $n, k \in \mathbb{N}$ and k is fixed. There are three possibilities: $k \equiv 0 \pmod 3$, $k \equiv 1 \pmod 3$, and $k \equiv 2 \pmod 3$. If $k \equiv 0 \pmod 3$, then $k+1 = 3i+1$ for some $i \in \mathbb{N}$ and $a_{k+1} = a_{3i+1} = c_{3i+1} = c_{k+1}$. If $k \equiv 1 \pmod 3$, then $k+1 = 3i+2$ for some $i \in \mathbb{N}$ and $a_{k+1} = a_{3i+2} = c_{3i+2} = c_{k+1}$. If $k \equiv 2 \pmod 3$, then $k+1 = 3i$ for some $i \in \mathbb{Z}^+$ and $a_{k+1} = a_{3i} = a_i = c_i = c_{3i} = c_{k+1}$ (where $a_i = c_i$ because $i \geq 1$ and $i \leq 3i-1 = k$). Thus, if $a_n = c_n$ for all $n \leq k$, then $a_n = c_n$ for all $n \leq k+1$. Since $a_0 = c_0 = z$, it follows that $a_n = c_n$ for all $n \in \mathbb{N}$.

The proof that $\mathbf{c}(2,1,0,z) = z\mathbf{c}(1,0,2,\mathtt{1})$ is similar. □

Proposition 2. *The complement of a generalized choral sequence is also a generalized choral sequence.*

Proof. Let $\mathbf{c}(r_0,r_1,r_c,z) = (a_n)_{n\geq 0}$. Thus, $a_{3i+r_0} = \mathtt{0}$, $a_{3i+r_1} = \mathtt{1}$, and $a_{3i+r_c} = a_i$ for all $i \in \mathbb{N}$. Also, $a_0 = z$ where $z = \mathtt{0}$ if $r_0 = 0$, $z = \mathtt{1}$ if $r_1 = 0$, and $z = \mathtt{0}$ or $z = \mathtt{1}$ if $r_c = 0$. Let $(b_n)_{n\geq 0} = \overline{(a_n)_{n\geq 0}}$ so that $b_i = \overline{a_i}$ for all $i \in \mathbb{N}$. Thus, $b_{3i+r_0} = \overline{a_{3i+r_0}} = \mathtt{1}$, $b_{3i+r_1} = \overline{a_{3i+r_1}} = \mathtt{0}$, and $b_{3i+r_c} = \overline{a_{3i+r_c}} = \overline{a_i} = b_i$ for all $i \in \mathbb{N}$. Also, $b_0 = \overline{a_0} = \overline{z}$ where $\overline{z} = \mathtt{1}$ if $r_0 = 0$, $\overline{z} = \mathtt{0}$ if $r_1 = 0$, and $\overline{z} = \mathtt{1}$ or $\overline{z} = \mathtt{0}$ if $r_c = 0$. Thus, $\overline{\mathbf{c}(r_0,r_1,r_c,z)} = \mathbf{c}(r_0',r_1',r_c',z')$ where $r_0' = r_1$, $r_1' = r_0$, $r_c' = r_c$, and $z' = \mathtt{0}$ if $r_0' = 0$, $z' = \mathtt{1}$ if $r_1' = 0$ and $z' = \mathtt{0}$ or $z' = \mathtt{1}$ if $r_c' = 0$. □

We follow Cassaigne and Karhumäki (1997). Let Σ be an alphabet and ? be a letter not in Σ. For a word $w = xy$ with $x \in \Sigma$

and $y \in (\Sigma \cup \{?\})^*$, let $T_0(w) = w^\omega$ and let $T_i(w)$ for $i \in \mathbb{Z}^+$ be the word obtained from $T_{i-1}(w)$ by replacing the first occurrence of ? in $T_{i-1}(w)$ by the i-th letter of $T_{i-1}(w)$. The *Toeplitz word determined by the pattern* w is $T(w) = \lim_{i\to\infty} T_i(w)$, an infinite word over Σ.

Using this definition of Toeplitz words and Definition 1, we see that some generalized choral sequences are Toeplitz words over $\{\mathtt{0},\mathtt{1}\}$ (Cassaigne & Karhumäki, 1997, Example 4).

Remark 2. *The sequences* $\mathbf{c}(0, 2, 1, \mathtt{0})$*,* $\mathbf{c}(0, 1, 2, \mathtt{0})$*,* $\mathbf{c}(1, 0, 2, \mathtt{1})$*, and* $\mathbf{c}(2, 0, 1, \mathtt{1})$ *are the Toeplitz words* $T(\mathtt{0?1})$*,* $T(\mathtt{01?})$*,* $T(\mathtt{10?})$*, and* $T(\mathtt{1?0})$*, respectively.*

The other generalized choral sequences are not Toeplitz words (unless we relax the condition that w start with a letter from Σ (Cassaigne & Karhumäki, 1997, Example 2)).

Automatic Sequence

We follow Allouche and Shallit (2003) again. A *deterministic finite automaton with output* (DFAO) is a model of computation defined by a finite set of states Q, a finite input alphabet Σ, a transition function $\delta : Q \times \Sigma \to Q$ (which we extend to $\delta : Q \times \Sigma^* \to Q$ so that $\delta(q, \epsilon) = q$ and $\delta(q, xa) = \delta(\delta(q, x), a)$ for all $q \in Q$, $x \in \Sigma^*$, and $a \in \Sigma$ (Allouche & Shallit, 2003, p. 129)), an initial state $q_0 \in Q$, a finite output alphabet Δ, and an output function $\tau : Q \to \Delta$. A DFAO has a word $w \in \Sigma^*$ as input and it moves from state to state according to δ while reading the letters of w (in order from left to right). When the end of w is reached, the automaton halts in a state q and outputs the letter $\tau(q)$.

A DFAO with representations of base-k numbers as input is called a k-DFAO. A *finite-state function* $f : \Sigma^* \to \Delta$ is one that can be computed by a DFAO such that $f(w) = \tau(\delta(q_0, w))$. Informally, a word $\mathbf{w} = (w_n)_{n\geq 0}$ is k*-automatic* if w_n is a finite-state function of the base-k digits of n (starting with the most significant digit). (Note that the input can have an arbitrary finite number of leading zeros.) A k-automatic infinite word has an associated k-DFAO.

Proposition 3. *A generalized choral sequence is 3-automatic.*

Proof. The 3-DFAO of a generalized choral sequence $\mathbf{c}(r_0, r_1, r_c, z)$ is shown in Table 1 and in Figure 1, where $Q = \{q_{r_0,\mathtt{0}}, q_{r_1,\mathtt{1}}, q_{r_c,\mathtt{0}}, q_{r_c,\mathtt{1}}\}$, $\Sigma = \{0, 1, 2\}$, $\Delta = \{\mathtt{0}, \mathtt{1}\}$, and the initial state is $q_{0,z}$.

Table 1
Deterministic finite automaton with output of a generalized choral sequence

q	$\delta(q, r_0)$	$\delta(q, r_1)$	$\delta(q, r_c)$	$\tau(q)$
$q_{r_0,\mathsf{0}}$	$q_{r_0,\mathsf{0}}$	$q_{r_1,\mathsf{1}}$	$q_{r_c,\mathsf{0}}$	$\mathsf{0}$
$q_{r_1,\mathsf{1}}$	$q_{r_0,\mathsf{0}}$	$q_{r_1,\mathsf{1}}$	$q_{r_c,\mathsf{1}}$	$\mathsf{1}$
$q_{r_c,\mathsf{0}}$	$q_{r_0,\mathsf{0}}$	$q_{r_1,\mathsf{1}}$	$q_{r_c,\mathsf{0}}$	$\mathsf{0}$
$q_{r_c,\mathsf{1}}$	$q_{r_0,\mathsf{0}}$	$q_{r_1,\mathsf{1}}$	$q_{r_c,\mathsf{1}}$	$\mathsf{1}$

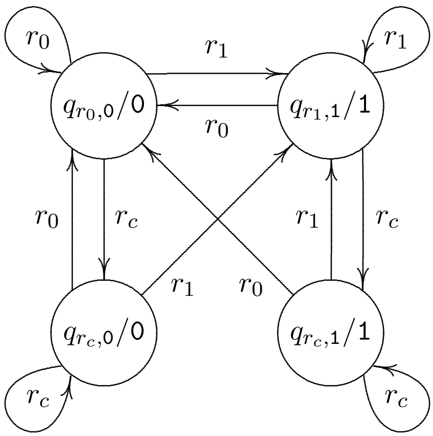

Figure 1. Transition diagram of Table 1

Let $\mathbf{w} = (w_n)_{n\geq 0}$ be the word generated by this automaton. Each w_n is a finite-state function of $s = s_0 s_1 \cdots s_{k-1}$, a base-3 representation of n. (Because leading zeros are allowed, this representation is not unique.) We will show that $\mathbf{w} = \mathbf{c}(r_0, r_1, r_c, z)$.

Let i be the number of leading zeros of s. Thus, $0 \leq i \leq k$ and $s = 0^i s_i \cdots s_{k-1}$. The automaton starts at state $q_{0,z}$. Note that $\delta(q_{0,z}, 0^i) = q_{0,z}$, that is, leading zeros cause the automaton to remain in the initial state. If $i = k$, then $n = 0$ and $w_n = w_0 = \tau(\delta(q_{0,z}, s)) = \tau(\delta(q_{0,z}, 0^k)) = \tau(q_{0,z}) = z$.

Note that $\delta(q, r_0) = q_{r_0,\mathsf{0}}$ for any $q \in Q$. Thus, if $s_{k-1} = r_0$, then $w_n = \tau(\delta(q_{0,z}, s)) = \tau(\delta(\delta(q_{0,z}, s_0 s_1 \cdots s_{k-2}), s_{k-1})) = \tau(\delta(\delta(q_{0,z}, s_0 s_1 \cdots s_{k-2}), r_0)) = \tau(q_{r_0,\mathsf{0}}) = \mathsf{0}$. If the last ternary digit of n is r_0, then $n \equiv r_0 \pmod 3$, that is, $n = 3i + r_0$ for a unique $i \in \mathbb{N}$ (since $n \geq 0$ and $r_0 \geq 0$). Thus, $w_{3i+r_0} = \mathtt{0}$ for all $i \in \mathbb{N}$.

Also, since $\delta(q, r_1) = q_{r_1,1}$ for any $q \in Q$, we see using the same reasoning that $w_{3i+r_1} = \mathtt{1}$ for all $i \in \mathbb{N}$.

If the last ternary digit of n is r_c, that is, if $s_{k-1} = r_c$, then $n = 3i + r_c$ for some $i \in \mathbb{N}$. Since $s_0 s_1 \cdots s_{k-2} s_{k-1} = s_0 s_1 \cdots s_{k-2} r_c$ is the base-3 representation of $3i + r_c$, then $s_0 s_1 \cdots s_{k-2}$ is the base-3 representation of i. Let $q' = \delta(q_{0,z}, s_0 s_1 \cdots s_{k-2})$ so that $w_i = \tau(\delta(q_{0,z}, s_0 s_1 \cdots s_{k-2})) = \tau(q')$ and $w_{3i+r_c} = \tau(\delta(q_{0,z}, s_0 s_1 \cdots s_{k-2} r_c)) = \tau(\delta(\delta(q_{0,z}, s_0 s_1 \cdots s_{k-2}), r_c)) = \tau(\delta(q', r_c))$. Note that $\tau(q') = \tau(\delta(q', r_c))$. Thus, $w_{3i+r_c} = w_i$ for all $i \in \mathbb{N}$.

By Definition 1, $\mathbf{w} = \mathbf{c}(r_0, r_1, r_c, z)$. □

Fixed Point of a Uniform Morphism

Since a generalized choral sequence is 3-automatic, then (by Cobham's theorem (Allouche & Shallit, 2003, p. 175)) it is the image, under a coding τ, of a fixed point of a 3-uniform morphism μ. That is, $\mathbf{c} = \tau(\mathbf{w})$ where $\mathbf{w} = \mu(\mathbf{w})$. The following theorem is a generalization of one by Noche (2008b) stating the well-known fact that $\mathbf{c}(0, 2, 1, \mathtt{0})$ is the fixed point of the morphism $\mathtt{0} \to \mathtt{001}$ and $\mathtt{1} \to \mathtt{011}$ iterated on $\mathtt{0}$.

Proposition 4. *A generalized choral sequence* $\mathbf{c}(r_0, r_1, r_c, z) = (c_n)_{n \geq 0}$ *is the fixed point of the morphism* μ *iterated on* z *mapping* $0 \to a_0 a_1 a_2$ *and* $1 \to b_0 b_1 b_2$, *where* $a_{r_0} = b_{r_0} = \mathtt{0}$, $a_{r_1} = b_{r_1} = \mathtt{1}$, $a_{r_c} = \mathtt{0}$, *and* $b_{r_c} = \mathtt{1}$.

Proof. Recall that $z = \mathtt{0}$ if $r_0 = 0$, $z = \mathtt{1}$ if $r_1 = 0$, and z could either be $\mathtt{0}$ or $\mathtt{1}$ if $r_c = 0$. If $z = \mathtt{0}$, then either $r_0 = 0$ or $r_c = 0$. Either way, $a_0 = \mathtt{0}$ and $\mathtt{0}$ maps to $\mathtt{0} a_1 a_2$. If $z = \mathtt{1}$, then either $r_1 = 0$ or $r_c = 0$. Either way, $b_0 = \mathtt{1}$ and $\mathtt{1}$ maps to $\mathtt{1} b_1 b_2$. Since $a_1, a_2, b_1, b_2 \in \{\mathtt{0}, \mathtt{1}\}$, the morphism μ is prolongable on the letter z. Thus, μ iterated on z has a unique fixed point $\mu^\omega(z)$ which we will call $\mathbf{w} = (w_n)_{n \geq 0}$, where the w's are letters. Note that $w_0 = z$.

The morphism μ is 3-uniform (because $\mu(\mathtt{0}) = a_0 a_1 a_2$ and $\mu(\mathtt{1}) = b_0 b_1 b_2$). Thus (by a lemma in (Allouche & Shallit, 2003, p. 174)), $\mu(w_i) = w_{3i+0} w_{3i+1} w_{3i+2}$ for all $i \in \mathbb{N}$. If $r_0 = 0$, then $\mu(\mathtt{0}) = \mathtt{0} a_1 a_2$ and $\mu(\mathtt{1}) = \mathtt{0} b_1 b_2$. If $r_0 = 1$, then $\mu(\mathtt{0}) = a_0 \mathtt{0} a_2$ and $\mu(\mathtt{1}) = b_0 \mathtt{0} b_2$. If $r_0 = 2$, then $\mu(\mathtt{0}) = a_0 a_1 \mathtt{0}$ and $\mu(\mathtt{1}) = b_0 b_1 \mathtt{0}$. In any case, $w_{3i+r_0} = \mathtt{0}$ for all $i \in \mathbb{N}$. Similarly, it can be seen that $w_{3i+r_1} = \mathtt{1}$ for all $i \in \mathbb{N}$. If $r_c = 0$, then $\mu(\mathtt{0}) = \mathtt{0}a_1a_2$ and $\mu(\mathtt{1}) = \mathtt{1}b_1b_2$. If $r_c = 1$, then $\mu(\mathtt{0}) = a_0\mathtt{0}a_2$ and $\mu(\mathtt{1}) = b_0\mathtt{1}b_2$. If $r_c = 2$, then $\mu(\mathtt{0}) = a_0a_1\mathtt{0}$ and $\mu(\mathtt{1}) = b_0b_1\mathtt{1}$. In any case, $w_{3i+r_c} = w_i$ for all $i \in \mathbb{N}$. By Definition 1, $\mathbf{c} = \mathbf{w} = \mu^\omega(z)$. □

A morphism $\mu : \Sigma^* \to \Sigma^*$ is said to be *primitive* if there exists an integer $n \geq 1$ such that for all $a, b \in \Sigma$, a occurs in $\mu^n(b)$. The morphism μ in Proposition 4 is primitive because $\mathtt{0}$ occurs in $\mu^n(\mathtt{1})$ and $\mathtt{1}$ occurs in $\mu^n(\mathtt{0})$ for any $n \in \mathbb{Z}^+$.

An infinite word $\mathbf{w}$ is *uniformly recurrent* if, for every finite subword y of $\mathbf{w}$, there exists an integer k such that every subword of length k of $\mathbf{w}$ contains y.

If μ is a primitive morphism prolongable on z, then $\mu^\omega(z)$ is uniformly recurrent (Allouche & Shallit, 2003, Theorem 10.9.5).

Remark 3. *A generalized choral sequence is uniformly recurrent.*

Subword Complexity

Let $\mathrm{Sub}_{\mathbf{w}}(n)$ denote the set of all subwords of length n of an infinite word $\mathbf{w}$ and $p_{\mathbf{w}}(n)$ denote the *subword complexity* function of $\mathbf{w}$, the function counting the number of distinct length-n subwords of $\mathbf{w}$.

For a generalized choral sequence $\mathbf{c}$, $\mathrm{Sub}_{\mathbf{c}}(1) = \{\mathtt{0}, \mathtt{1}\}$, $\mathrm{Sub}_{\mathbf{c}}(2) = \{\mathtt{00}, \mathtt{01}, \mathtt{10}, \mathtt{11}\}$, and $\mathrm{Sub}_{\mathbf{c}}(3) = \{\mathtt{001}, \mathtt{010}, \mathtt{011}, \mathtt{100}, \mathtt{101}, \mathtt{110}\}$. Thus, $p_{\mathbf{c}}(1) = 2$, $p_{\mathbf{c}}(2) = 4$, and $p_{\mathbf{c}}(3) = 6$. If $\mathbf{c}$ is a type-012 sequence, then $\mathrm{Sub}_{\mathbf{c}}(4) = \{\mathtt{0010}, \mathtt{0011}, \mathtt{0100}, \mathtt{0110}, \mathtt{1001}, \mathtt{1010}, \mathtt{1011}, \mathtt{1101}\}$; if $\mathbf{c}$ is a type-210 sequence, then $\mathrm{Sub}_{\mathbf{c}}(4) = \{\mathtt{0010}, \mathtt{0100}, \mathtt{0101}, \mathtt{0110}, \mathtt{1001}, \mathtt{1011}, \mathtt{1100}, \mathtt{1101}\}$. In either case, $p_{\mathbf{c}}(4) = 8$.

Theorem 4. *For a generalized choral sequence* $\mathbf{c}$, $p_{\mathbf{c}}(n) = 2n$ *for all* $n \in \mathbb{Z}^+$.

Proof. Let $S_n = \{c_jc_{j+1}\cdots c_{j+n-1} : j \in \mathbb{N}\}$ be the set of all length-n subwords of a generalized choral sequence $\mathbf{c}(r_0, r_1, r_c, z) = (c_n)_{n\geq 0}$. The set S_n can be partitioned into three subsets: $S_{n,0}$ contains only all the subwords with initial index $j \equiv r_0 \pmod 3$, $S_{n,1}$ contains those with $j \equiv r_1 \pmod 3$, and $S_{n,c}$ contains those with $j \equiv r_c \pmod 3$. Let $|S|$ denote the number of elements in a set S. Thus, $p_{\mathbf{c}}(n) = |S_n| = |S_{n,0}| + |S_{n,1}| + |S_{n,c}|$.

The proof is by induction: given $p_{\mathbf{c}}(i) = 2i$ for all $i \leq k$ where $i, k \in \mathbb{Z}^+$ and k is fixed, we will show that $p_{\mathbf{c}}(i) = 2i$ for all $i \leq k+1$.

It is easy to see that $p_{\mathbf{c}}(i) = 2i$ for all $i \leq 3$, so we only need to look at $k \geq 3$.

If $\mathbf{c}$ is, say, a type-012 sequence then length-k subwords of $\mathbf{c}$ can be visualized as shown below, where the bottom overline shows a subword from $S_{k,0}$, the middle overline shows a subword from $S_{k,1}$, and the top overline shows a subword from $S_{k,c}$. (If $\mathbf{c}$ is a type-210 sequence, then the subwords would be of the form $\cdots 10c_i10 \cdots$; the end result is the same.)

If $k \equiv 0 \pmod 3$:

$$\cdots\quad 0\quad 1\quad c_i\quad 0\quad 1\quad \cdots\quad c_{i+m-1}\quad 0\quad 1\quad c_{i+m}\quad 0\quad 1\quad \cdots$$

If $k \equiv 1 \pmod 3$:

$$\cdots\quad 0\quad 1\quad c_i\quad 0\quad 1\quad \cdots\quad c_{i+m-1}\quad 0\quad 1\quad c_{i+m}\quad 0\quad 1\quad \cdots$$

If $k \equiv 2 \pmod 3$:

$$\cdots\quad 0\quad 1\quad c_i\quad 0\quad 1\quad \cdots\quad c_{i+m-1}\quad 0\quad 1\quad c_{i+m}\quad 0\quad 1\quad \cdots$$

Note that $0 < 3m \leq k$ since $k \geq 3$. Thus, $m < k$ and $m+1 < k$.

We are given $|S_{k,0}|+|S_{k,1}|+|S_{k,c}| = 2k$. Consider what happens when $k \equiv 0 \pmod 3$ and we now look at subwords of length $k+1$:

$$\cdots\quad 0\quad 1\quad c_i\quad 0\quad 1\quad \cdots\quad c_{i+m-1}\quad 0\quad 1\quad c_{i+m}\quad 0\quad 1\quad \cdots$$

All the subwords in $S_{k+1,0}$ are just the subwords in $S_{k,0}$ with the letter 0 concatenated at the end; the number of subwords remains the same. That is, $|S_{k+1,0}| = |S_{k,0}|$. All the subwords in $S_{k+1,1}$ are just those in $S_{k,1}$ with a 1 concatenated at the end. Thus, $|S_{k+1,1}| = |S_{k,1}|$.

Note that $|S_{k,c}|$ is equal to the number of distinct subwords $c_ic_{i+1}\cdots c_{i+m-1}$, that is, $|S_{k,c}| = |S_{m,c}|$. Since $m < k$, $|S_{m,c}| = 2m$. Also, $|S_{k+1,c}|$ is equal to the number of distinct subwords $c_ic_{i+1}\cdots c_{i+m}$, that is, $|S_{k+1,c}| = |S_{m+1,c}| = 2(m+1) = 2m+2 = |S_{m,c}|+2 = |S_{k,c}|+2$. (Since $m+1 < k$, $|S_{m+1,c}| = 2(m+1)$.) Thus, $|S_{k+1}| = |S_{k+1,0}| + |S_{k+1,1}| + |S_{k+1,c}| = |S_{k,0}| + |S_{k,1}| + (|S_{k,c}| + 2) = |S_k| + 2 = 2k+2 = 2(k+1)$, that is, $p_{\mathbf{c}}(k+1) = 2(k+1)$. Thus,

$p_{\mathbf{c}}(i) = 2i$ for all $i \leq k+1$.

Using similar reasoning, we find that when $k \equiv 1 \pmod 3$, $|S_{k+1,0}| + |S_{k+1,1}| + |S_{k+1,c}| = |S_{k,0}| + (|S_{k,1}| + 2) + |S_{k,c}|$; when $k \equiv 2 \pmod 3$, $|S_{k+1,0}| + |S_{k+1,1}| + |S_{k+1,c}| = (|S_{k,0}| + 2) + |S_{k,1}| + |S_{k,c}|$. In any case, the end result is the same.

It is easy to see that $p_{\mathbf{c}}(i) = 2i$ for all $i \leq 3$, $i \in \mathbb{Z}^+$. Thus, $p_{\mathbf{c}}(n) = 2n$ for all $n \in \mathbb{Z}^+$. □

Lyndon Factorization

We now follow Richomme (2003). Words may be ordered lexicographically. Let the alphabet $\{\mathtt{0},\mathtt{1}\}$ be ordered such that $\mathtt{0} < \mathtt{1}$. We say that $v \leq w$ (or $w \geq v$) if and only if either v is a prefix of w or there exist words x, y, z and letters a, b such that $v = xay$, $w = xbz$, and $a < b$. We say that $v < w$ (or $w > v$) if $v \leq w$ and $v \neq w$.

Remark 4. (Chen, Fox, & Lyndon, 1958, p. 82) *Let a, b, c, and d be finite words over an ordered alphabet. If $a < b$ and $|a| \geq |b|$, then $ac < bd$.*

We extend lexicographic order to the set of finite or infinite words (Melançon, 1996). We say that $v < \mathbf{w}$ if and only if either v is a prefix of $\mathbf{w}$ or there exist words $x, y, \mathbf{z}$ and letters a, b such that $v = xay$, $\mathbf{w} = xb\mathbf{z}$, and $a < b$. We say that $\mathbf{v} < w$ if and only if there exist words $x, \mathbf{y}, z$ and letters a, b such that $\mathbf{v} = xa\mathbf{y}$, $w = xbz$, and $a < b$. We say that $\mathbf{v} < \mathbf{w}$ if and only if there exist words $x, \mathbf{y}, \mathbf{z}$ and letters a, b such that $\mathbf{v} = xa\mathbf{y}$, $\mathbf{w} = xb\mathbf{z}$, and $a < b$.

Chen et al. (1958) introduced what they called *standard sequences* but which are now called *Lyndon words*. A Lyndon word is a word that is less than any of its non-empty proper suffixes. Lyndon words were originally defined as finite words (Chen et al., 1958) but the definition was eventually extended to include infinite words (Siromoney, Mathew, Dare, & Subramanian, 1994, Proposition 2.2). For example, letters are Lyndon words; $\mathtt{01011}$ is a finite Lyndon word while $\mathtt{01101}$ is not; and $\mathtt{01}^{\omega} = \mathtt{0111}\cdots$ is an infinite Lyndon word while $(\mathtt{01})^{\omega} = \mathtt{010101}\cdots$ is not.

Lemma 1. (Siromoney et al., 1994) *An infinite word is an infinite Lyndon word if and only if it has an infinite number of prefixes which are Lyndon words.*

A morphism μ over an ordered alphabet Σ is *order-preserving* if for all $u, v \in \Sigma^*$, $u \le v$ implies $\mu(u) \le \mu(v)$ (Richomme, 2003). For an order-preserving morphism μ, if $u < v$ then $\mu(u) < \mu(v)$ (Richomme, 2003, Lemma 3.2).

Remark 5. *For an order-preserving morphism μ, if $u < v$ then $\mu^n(u) < \mu^n(v)$ for $n \in \mathbb{Z}^+$.*

A morphism is a *Lyndon morphism* (Richomme, 2003) if it preserves (finite) Lyndon words. A morphism μ on $\{\mathtt{0},\mathtt{1}\}$ such that $\mathtt{0} < \mathtt{1}$ is a Lyndon morphism if and only if $\mu(\mathtt{0})$ and $\mu(\mathtt{1})$ are Lyndon words and $\mu(\mathtt{0}) < \mu(\mathtt{1})$ (Richomme, 2003, Proposition 4.7).

Remark 6. *The morphism μ mapping $\mathtt{0} \to \mathtt{001}$ and $\mathtt{1} \to \mathtt{011}$ is a Lyndon morphism.*

Lemma 2. *For the morphism μ mapping $\mathtt{0} \to \mathtt{001}$ and $\mathtt{1} \to \mathtt{101}$, $\mu^n(\mathtt{0}^i\mathtt{1}^j)$ for $n, j \in \mathbb{N}$, $i \in \mathbb{Z}^+$ is a Lyndon word.*

Proof. We first show that $\mu^n(\mathtt{0})$ for $n \in \mathbb{N}$ is a Lyndon word (the case where $i = 1$ and $j = 0$).

The words $\mu(\mathtt{0}) = \mathtt{001}$ and $\mu(\mathtt{1}) = \mathtt{101}$ differ only in the first letter. The words $\mu^2(\mathtt{0}) = \mu(\mathtt{0})\mu(\mathtt{0})\mu(\mathtt{1})$ and $\mu^2(\mathtt{1}) = \mu(\mathtt{1})\mu(\mathtt{0})\mu(\mathtt{1})$ also differ only in the first letter because $\mu(\mathtt{0})$ and $\mu(\mathtt{1})$ differ only in the first letter. Continuing this reasoning, it can be seen that $\mu^n(\mathtt{0})$ and $\mu^n(\mathtt{1})$ for $n \in \mathbb{Z}^+$ differ only in the first letter and that if $\mu^n(\mathtt{0}) = \mathtt{0}x$ then $\mu^n(\mathtt{1}) = \mathtt{1}x$.

Assume that $\mu^k(\mathtt{0})$ is a Lyndon word for some $k \in \mathbb{Z}^+$. Now, $\mu^{k+1}(\mathtt{0}) = \mu^k(\mathtt{0})\mu^k(\mathtt{0})\mu^k(\mathtt{1}) = \mathtt{0}x\mathtt{0}x\mathtt{1}x$ for some $x = a_1a_2\cdots a_n$ where the a's are letters. Since $\mu^k(\mathtt{0})$ is a Lyndon word, it is less than any of its non-empty proper suffixes. Thus, $\mathtt{0}x < a_1a_2\cdots a_n$, $\mathtt{0}x < a_2\cdots a_n$, and so on up to $\mathtt{0}x < a_n$.

Using Remark 4 and starting from $\mathtt{0}x < x$, we get $\mathtt{0}x\mathtt{0}x\mathtt{1}x < x\mathtt{0}x\mathtt{1}x$. From $\mathtt{0}x < a_2\cdots a_n$, we get $\mathtt{0}x\mathtt{0}x\mathtt{1}x < a_2\cdots a_n\mathtt{0}x\mathtt{1}x$. Similarly, we go on up to $\mathtt{0}x\mathtt{0}x\mathtt{1}x < a_n\mathtt{0}x\mathtt{1}x$.

Clearly, $\mathtt{0}x\mathtt{0}x\mathtt{1}x < \mathtt{0}x\mathtt{1}x$. Using Remark 4 and starting from $\mathtt{0}x < x$, we get $\mathtt{0}x\mathtt{0}x\mathtt{1}x < x\mathtt{1}x$. Similarly, $\mathtt{0}x\mathtt{0}x\mathtt{1}x < a_2\cdots a_n\mathtt{1}x$ and so on up to $\mathtt{0}x\mathtt{0}x\mathtt{1}x < a_n\mathtt{1}x$.

Clearly, $\mathtt{0}x\mathtt{0}x\mathtt{1}x < \mathtt{1}x$. Using Remark 4 and starting from $\mathtt{0}x < x$, we get $\mathtt{0}x\mathtt{0}x\mathtt{1}x < x$. Similarly, $\mathtt{0}x\mathtt{0}x\mathtt{1}x < a_2\cdots a_n$ and so on up to $\mathtt{0}x\mathtt{0}x\mathtt{1}x < a_n$.

Because $0x0x1x$ is less than any of its non-empty proper suffixes, $\mu^{k+1}(0)$ is a Lyndon word. Now, $\mu^0(0) = 0$ is a Lyndon word. Also, $\mu^k(0)$ is a Lyndon word for $k = 1$. Thus, $\mu^n(0)$ is a Lyndon word for any $n \in \mathbb{N}$.

The proof that $\mu^n(0^i1^j)$ for $n, i, j \in \mathbb{Z}^+$ is a Lyndon word is similar, but now we use $\mu^k(0^i1^j) = (\mu^k(0))^i(\mu^k(1))^j = (0x)^i(1x)^j$. □

Lemma 3. *For the morphism μ mapping $0 \to 001$ and $1 \to 011$, $\mu^n(1) > \mu^{n+1}(1)$ for $n \in \mathbb{N}$.*

Proof. By Remark 6, μ is a Lyndon morphism. Thus, it is order-preserving (Richomme, 2003, Proposition 4.2). From Remark 5, since $011 < 1$, then $\mu^n(011) = \mu^{n+1}(1) < \mu^n(1)$ for $n \in \mathbb{N}$. □

Lemma 4. *For the morphism μ mapping $0 \to 001$ and $1 \to 101$, $\mu^n(01^m) > \mu^{n+1}(01^m)$ for $n, m \in \mathbb{Z}^+$.*

Proof. The morphism μ is order-preserving because $\mu(01) < \mu(1)$ (Richomme, 2003, Lemma 3.13). Fix $m \in \mathbb{Z}^+$. From Remark 5, since $001(101)^m < 01^m$, then $\mu^n(001(101)^m) = \mu^{n+1}(01^m) < \mu^n(01^m)$ for $n \in \mathbb{Z}^+$. □

To *factorize* a word is to express it as a sequence of subwords. Theorems 5 and 6 describe *Lyndon factorization.*

Theorem 5. (Chen et al., 1958; Lothaire, 1997, Theorem 5.1.5) *Any non-empty finite word w may be uniquely factorized as a non-increasing finite sequence of finite Lyndon words $(\ell_k)_{0 \le k \le n}$. That is, $w = \ell_0\ell_1 \cdots \ell_n$ where $\ell_0 \ge \ell_1 \ge \cdots \ge \ell_n$.*

For example, the Lyndon factorization of 01011 is (01011) while that of 01101 is $(011)(01)$.

Theorem 6. (Siromoney et al., 1994, Theorem 2.3) *Any infinite word may $\mathbf{w}$ be uniquely factorized as either a non-increasing infinite sequence of finite Lyndon words $(\ell_k)_{k \ge 0}$ or a non-increasing finite sequence of finite Lyndon words $(\ell_k)_{0 \le k \le n}$ followed by an infinite Lyndon word $\mathbf{x} \le \ell_n$. That is, either $\mathbf{w} = \ell_0\ell_1\ell_2 \cdots$ where $\ell_0 \ge \ell_1 \ge \ell_2 \cdots$ or $\mathbf{w} = \ell_0\ell_1 \cdots \ell_n\mathbf{x}$ where $\ell_0 \ge \ell_1 \ge \cdots \ge \ell_n \ge \mathbf{x}$.*

For example, the Lyndon factorization of $\mathtt{01}^\omega = \mathtt{0111}\cdots$ is $(\mathtt{0111}\cdots)$ while that of $(\mathtt{01})^\omega = \mathtt{010101}\cdots$ is $(\mathtt{01})(\mathtt{01})(\mathtt{01})\cdots$.

We will show that the Lyndon factorizations of the generalized choral sequences are as follows:

$$\begin{aligned}
\mathbf{c}(0,2,1,\mathtt{0}) &= (\mathtt{0010010110010010110010110 11}\cdots)\\
\mathbf{c}(1,2,0,\mathtt{0}) &= (\mathtt{001001101001001101101001101}\cdots)\\
\mathbf{c}(0,1,2,\mathtt{0}) &= (\mathtt{01})(\mathtt{001101})(\mathtt{001001101101001101})\cdots\\
\mathbf{c}(2,1,0,\mathtt{0}) &= (\mathtt{01011})(\mathtt{001011011})\\
&\quad\ (\mathtt{001001011001011011001011011})\cdots\\
\mathbf{c}(1,2,0,\mathtt{1}) &= (\mathtt{1})(\mathtt{01})(\mathtt{001101})(\mathtt{001001101101001101})\cdots\\
\mathbf{c}(1,0,2,\mathtt{1}) &= (\mathtt{1})(\mathtt{011})(\mathtt{001011011})\\
&\quad\ (\mathtt{001001011001011011001011011})\cdots\\
\mathbf{c}(2,1,0,\mathtt{1}) &= (\mathtt{1})(\mathtt{1})(\mathtt{011})(\mathtt{001011011})\\
&\quad\ (\mathtt{001001011001011011001011011})\cdots\\
\mathbf{c}(2,0,1,\mathtt{1}) &= (\mathtt{1})(\mathtt{1})(\mathtt{011})(\mathtt{01})(\mathtt{001101101})(\mathtt{001101})\cdots
\end{aligned}$$

Proposition 5. *The sequence* $\mathbf{c}(0,2,1,\mathtt{0})$ *is an infinite Lyndon word.*

Proof. By Proposition 4, $\mathbf{c}(0,2,1,\mathtt{0}) = \mu^\omega(\mathtt{0})$ where μ maps $\mathtt{0} \to \mathtt{001}$ and $\mathtt{1} \to \mathtt{011}$. Note that $\mu^k(\mathtt{0}) = c_0c_1\cdots c_{3^k-1}$ for $k \in \mathbb{N}$. From Remark 6, μ is a Lyndon morphism. Thus, since $\mathtt{0}$ is a Lyndon word, then $\mu^k(\mathtt{0})$ is a Lyndon word for $k \in \mathbb{N}$. The proper prefixes $c_0c_1\cdots c_{3^k-1}$ for $k \in \mathbb{N}$ are all Lyndon words and by Lemma 1, $\mathbf{c}(0,2,1,\mathtt{0})$ is an infinite Lyndon word. □

Proposition 6. *The sequence* $\mathbf{c}(1,2,0,\mathtt{0})$ *is an infinite Lyndon word.*

Proof. By Proposition 4, $\mathbf{c}(1,2,0,\mathtt{0}) = \mu^\omega(\mathtt{0})$ where μ maps $\mathtt{0} \to \mathtt{001}$ and $\mathtt{1} \to \mathtt{101}$. Note that $\mu^k(\mathtt{0}) = c_0c_1\cdots c_{3^k-1}$ for $k \in \mathbb{N}$. From Lemma 2, $\mu^k(\mathtt{0})$ is a Lyndon word for $k \in \mathbb{N}$. The proper prefixes $c_0c_1\cdots c_{3^k-1}$ for $k \in \mathbb{N}$ are all Lyndon words and by Lemma 1, $\mathbf{c}(1,2,0,\mathtt{0})$ is an infinite Lyndon word. □

We define $\sum_{n=a}^{b} f(n)$ to be 0 if $b < a$.

Proposition 7. *The sequence* $\mathbf{c}(0,1,2,\mathtt{0})$ *is an infinite non-increasing sequence of finite Lyndon words* $(\ell_k)_{k\geq 0}$ *with* $\ell_0 = \mathtt{01}$ *and* $\ell_k = \mu^k(\ell_0)$ *for* $k \in \mathbb{Z}^+$, *where* μ *is the morphism mapping* $\mathtt{0} \to \mathtt{001}$ *and* $\mathtt{1} \to \mathtt{101}$.

Proof. Let $\mathbf{c}(0,1,2,\mathtt{0}) = (c_n)_{n\geq 0}$ and $(\ell_k)_{k\geq 0} = (w_n)_{n\geq 0}$ where $c_n, w_n \in \{\mathtt{0},\mathtt{1}\}$.

From Definition 1, $c_{3i} = \texttt{0}$, $c_{3i+1} = \texttt{1}$, and $c_{3i+2} = c_i$ for $i \in \mathbb{N}$.

From $\ell_k = \mu^k(\texttt{01})$ for $k \in \mathbb{Z}^+$ and given that μ is 3-uniform, we get $|\ell_k| = 2 \cdot 3^k$. Thus, $\ell_k = w_m w_{m+1} \cdots w_{m+2\cdot 3^k - 1}$ where $m = \sum_{j=1}^{k} 2 \cdot 3^{j-1}$. From Definition 1 and Proposition 4, $w_{m+3p} = w_{m'+p}$, $w_{m+3p+1} = \texttt{0}$, and $w_{m+3p+2} = \texttt{1}$ where $m' = \sum_{j=1}^{k-1} 2 \cdot 3^{j-1}$ and p goes from 0 to $2 \cdot 3^{k-1} - 1$. Note that as k goes from 1 onwards, the indices $m+3p$, $m+3p+1$, and $m+3p+2$ cover all the integers from 2 onwards.

Thus, the w_n's are characterized by $w_0 w_1 w_2 = \texttt{010}$, $w_{m+3p} = w_{m'+p}$, $w_{m+3p+1} = \texttt{0}$, and $w_{m+3p+2} = \texttt{1}$ where $m = \sum_{j=1}^{k} 2 \cdot 3^{j-1}$ and $m' = \sum_{j=1}^{k-1} 2 \cdot 3^{j-1}$, for all $p \in \mathbb{N}$ such that $p \le 2 \cdot 3^{k-1} - 1$, and for all $k \in \mathbb{Z}^+$.

We show that $(c_n)_{n\ge 0} = (w_n)_{n \ge 0}$. Clearly, $c_0 c_1 = w_0 w_1$, so we now consider $n \ge 2$. Since $m = 2 + \sum_{j=2}^{k} 2 \cdot 3^{j-1}$, if follows that $m \equiv 2 \pmod 3$ for $k \ge 1$ (that is, for $n \ge 2$). If $n \equiv 0 \pmod 3$, then $c_n = c_{3i}$ for some $i \ge 0$ and $w_n = w_{m+3p+1}$ for some m and p for $k \ge 1$. If $n \equiv 1 \pmod 3$, then $c_n = c_{3i+1}$ and $w_n = w_{m+3p+2}$. If $n \equiv 2 \pmod 3$, then $c_n = c_{3i+2}$ and $w_n = w_{m+3p}$.

Now, $c_{3i} = \texttt{0} = w_{m+3p+1}$ and $c_{3i+1} = \texttt{1} = w_{m+3p+2}$, so all we need to show is that $c_{3i+2} = w_{m+3p}$. Consider the letter $c_n = c_{3i+2}$ where $i \ge 0$ and $n \ge 2$. It is part of a subword ℓ_k for some $k \ge 1$, that is, since $\ell_k = w_m w_{m+1} \cdots w_{m+2\cdot 3^k - 1}$, then $m \le 3i+2 \le m+2\cdot 3^k - 1$. But $3i+2 \equiv 2 \pmod 3$, so $m \le 3i+2 \le m + 2\cdot 3^k$. Thus, $(m-2)/3 \le i \le (m + 2 \cdot 3^k - 2)/3$. But $(m-2)/3 = (\sum_{j=2}^{k} 2 \cdot 3^{j-1})/3 = \sum_{j=2}^{k} 2 \cdot 3^{j-2} = \sum_{j=1}^{k-1} 2 \cdot 3^{j-1} = m'$ and $(m + 2 \cdot 3^k - 2)/3 = m' + 2 \cdot 3^{k-1}$. Thus, $m' \le i \le m' + 2 \cdot 3^{k-1}$, and c_i is part of the subword ℓ_{k-1}, that is, $c_i = w_{m'+p}$. Finally, $c_{3i+2} = c_i = w_{m'+p} = w_{m+3p}$.

From Lemma 2, $\ell_k = \mu^k(\texttt{01})$ is a Lyndon word for $k \in \mathbb{N}$. Clearly, $\ell_0 = \texttt{01} > \ell_1 = \texttt{001101}$. From Lemma 4, $\ell_1 > \ell_2 > \ell_3 > \cdots$. □

Proposition 8. *The sequence* $\mathbf{c}(2,1,0,\texttt{0})$ *is an infinite non-increasing sequence of finite Lyndon words* $(\ell_k)_{k\ge 0}$ *with* $\ell_0 = \texttt{01011}$ *and* $\ell_k = \mu^k(\texttt{011})$ *for* $k \in \mathbb{Z}^+$*, where* μ *is the morphism mapping* $\texttt{0} \to \texttt{001}$ *and* $\texttt{1} \to \texttt{011}$*.*

Proof. Let $\mathbf{c}(2,1,0,\texttt{0}) = (c_n)_{n\ge 0}$ and $(\ell_k)_{k \ge 0} = (w_n)_{n\ge 0}$ where $c_n, w_n \in \{\texttt{0}, \texttt{1}\}$. Now, $c_{3i} = c_i$, $c_{3i+1} = \texttt{1}$, $c_{3i+2} = \texttt{0}$, and $c_0 = \texttt{0}$ for $i \in \mathbb{N}$. The w_n's are characterized by $w_0 w_1 w_2 w_3 w_4 = \texttt{01011}$,

$w_{m+3p} = \mathtt{0}$, $w_{m+3p+1} = w_{m'+p}$, and $w_{m+3p+2} = \mathtt{1}$ where $m = 2+\sum_{j=1}^{k} 3^j$ and $m' = 2+\sum_{j=1}^{k-1} 3^j$, for all $p \in \mathbb{N}$ such that $p \leq 3^k - 1$, and for all $k \in \mathbb{Z}^+$. Clearly, $c_0c_1c_2c_3c_4 = w_0w_1w_2w_3w_4$. Because $m \equiv 2 \pmod 3$, it follows that $c_{3i} = w_{m+3p+1}$, $c_{3i+1} = w_{m+3p+2}$, and $c_{3i+2} = w_{m+3p}$.

Subword ℓ_0 is a Lyndon word and so is each ℓ_k for $k \in \mathbb{Z}^+$ because μ is a Lyndon morphism (Remark 6) and $\mathtt{011}$ is a Lyndon word. Clearly, $\ell_0 = \mathtt{01011} > \ell_1 = \mathtt{001011011}$. Note that $\ell_k = \mu^{k+1}(\mathtt{1})$. From Lemma 3, $\ell_1 > \ell_2 > \ell_3 > \cdots$. □

Proposition 9. *The sequence* $\mathbf{c}(1, 2, 0, \mathtt{1})$ *is an infinite non-increasing sequence of finite Lyndon words* $(\ell_k)_{k\geq 0}$ *with* $\ell_0 = \mathtt{1}$*,* $\ell_1 = \mathtt{01}$*, and* $\ell_k = \mu^{k-1}(\ell_1)$ *for* $k \geq 2$*, where* μ *is the morphism mapping* $\mathtt{0} \to \mathtt{001}$ *and* $\mathtt{1} \to \mathtt{101}$*.*

Proof. Let $\mathbf{c}(1, 2, 0, \mathtt{1}) = (c_n)_{n\geq 0}$ and $(\ell_k)_{k\geq 0} = (w_n)_{n\geq 0}$ where $c_n, w_n \in \{\mathtt{0}, \mathtt{1}\}$. Now, $c_{3i} = c_i$, $c_{3i+1} = \mathtt{0}$, $c_{3i+2} = \mathtt{1}$, and $c_0 = \mathtt{1}$ for $i \in \mathbb{N}$. The w_n's are characterized by $w_0w_1w_2w_3 = \mathtt{1010}$, $w_{m+3p} = w_{m'+p}$, $w_{m+3p+1} = \mathtt{0}$, and $w_{m+3p+2} = \mathtt{1}$ where $m = 1 + \sum_{j=2}^{k} 2 \cdot 3^{j-2}$ and $m' = 1 + \sum_{j=2}^{k-1} 2 \cdot 3^{j-2}$, for all $p \in \mathbb{N}$ such that $p \leq 2 \cdot 3^{k-2} - 1$, and for all $k \geq 2$. Clearly, $c_0c_1c_2 = w_0w_1w_2$. Because $m = 3 + \sum_{j=3}^{k} 2 \cdot 3^{j-2}$ and $m \equiv 0 \pmod 3$, it follows that $c_{3i} = w_{m+3p}$, $c_{3i+1} = w_{m+3p+1}$, and $c_{3i+2} = w_{m+3p+2}$.

Subword ℓ_0 is a Lyndon word and, by Lemma 2, so is each $\ell_{k+1} = \mu^k(\mathtt{01})$ for $k \in \mathbb{N}$. Clearly, $\ell_0 = \mathtt{1} > \ell_1 = \mathtt{01} > \ell_2 = \mathtt{001101}$. From Lemma 4, $\ell_2 > \ell_3 > \ell_4 > \cdots$. □

Proposition 10. *The sequence* $\mathbf{c}(1, 0, 2, \mathtt{1})$ *is an infinite non-increasing sequence of finite Lyndon words* $(\ell_k)_{k\geq 0}$ *with* $\ell_0 = \mathtt{1}$ *and* $\ell_k = \mu^k(\ell_0)$ *for* $k \in \mathbb{Z}^+$*, where* μ *is the morphism mapping* $\mathtt{0} \to \mathtt{001}$ *and* $\mathtt{1} \to \mathtt{011}$*.*

Proof. Let $\mathbf{c}(1, 0, 2, \mathtt{1}) = (c_n)_{n\geq 0}$ and $(\ell_k)_{k\geq 0} = (w_n)_{n\geq 0}$ where $c_n, w_n \in \{\mathtt{0}, \mathtt{1}\}$. Now, $c_{3i} = \mathtt{1}$, $c_{3i+1} = \mathtt{0}$, and $c_{3i+2} = c_i$ for $i \in \mathbb{N}$. The w_n's are characterized by $w_0 = \mathtt{1}$, $w_{m+3p} = \mathtt{0}$, $w_{m+3p+1} = w_{m'+p}$, and $w_{m+3p+2} = \mathtt{1}$ where $m = \sum_{j=1}^{k} 3^{j-1}$ and $m' = \sum_{j=1}^{k-1} 3^{j-1}$, for all $p \in \mathbb{N}$ such that $p \leq 3^{k-1} - 1$, and for all $k \in \mathbb{Z}^+$. Clearly, $c_0 = w_0$. Because $m = 1 + \sum_{j=2}^{k} 3^{j-1}$ and $m \equiv 1 \pmod 3$, it follows that $c_{3i} = w_{m+3p+2}$, $c_{3i+1} = w_{m+3p}$, and $c_{3i+2} = w_{m+3p+1}$.

Subword ℓ_0 is a Lyndon word and so is each ℓ_k for $k \in \mathbb{Z}^+$ because μ is a Lyndon morphism (Remark 6). From Lemma 3, $\ell_0 > \ell_1 > \ell_2 > \cdots$. □

Proposition 11. *The sequence* $\mathbf{c}(2,1,0,\mathtt{1})$ *is an infinite non-increasing sequence of finite Lyndon words* $(\ell_k)_{k\geq 0}$ *with* $\ell_0 = \ell_1 = \mathtt{1}$ *and* $\ell_k = \mu^{k-1}(\ell_1)$ *for* $k \geq 2$, *where* μ *is the morphism mapping* $\mathtt{0} \to \mathtt{001}$ *and* $\mathtt{1} \to \mathtt{011}$.

Proof. Let $\mathbf{c}(2,1,0,\mathtt{1}) = (c_n)_{n\geq 0}$ and $(\ell_k)_{k\geq 0} = (w_n)_{n\geq 0}$ where $c_n, w_n \in \{\mathtt{0},\mathtt{1}\}$. Now, $c_{3i} = c_i$, $c_{3i+1} = \mathtt{1}$, $c_{3i+2} = \mathtt{0}$, and $c_0 = \mathtt{1}$ for $i \in \mathbb{N}$. The w_n's are characterized by $w_0 w_1 = \mathtt{11}$, $w_{m+3p} = \mathtt{0}$, $w_{m+3p+1} = w_{m'+p}$, and $w_{m+3p+2} = \mathtt{1}$ where $m = 1 + \sum_{j=2}^{k} 3^{j-2}$ and $m' = 1 + \sum_{j=2}^{k-1} 3^{j-2}$, for all $p \in \mathbb{N}$ such that $p \leq 3^{k-2} - 1$, and for all $k \geq 2$. Clearly, $c_0 c_1 = w_0 w_1$. Because $m = 2 + \sum_{j=3}^{k} 3^{j-2}$ and $m \equiv 2 \pmod 3$, it follows that $c_{3i} = w_{m+3p+1}$, $c_{3i+1} = w_{m+3p+2}$, and $c_{3i+2} = w_{m+3p}$.

Subwords ℓ_0 and ℓ_1 are Lyndon words and so is each ℓ_k for $k \geq 2$ because μ is a Lyndon morphism (Remark 6). Note that $\ell_0 = \ell_1 = \mathtt{1}$. From Lemma 3, $\ell_1 > \ell_2 > \ell_3 > \cdots$. □

Proposition 12. *The sequence* $\mathbf{c}(2,0,1,\mathtt{1})$ *is an infinite non-increasing sequence of finite Lyndon words* $(\ell_k)_{k\geq 0}$ *with* $\ell_0 = \ell_1 = \mathtt{1}$, $\ell_2 = \mathtt{011}$, $\ell_3 = \mathtt{01}$, *and* $\ell_k = \mu(\ell_{k-2})$ *for* $k \geq 4$, *where* μ *is the morphism mapping* $\mathtt{0} \to \mathtt{001}$ *and* $\mathtt{1} \to \mathtt{101}$.

Proof. Let $\mathbf{c}(2,0,1,\mathtt{1}) = (c_n)_{n\geq 0}$ and $(\ell_k)_{k\geq 0} = (w_n)_{n\geq 0}$ where $c_n, w_n \in \{\mathtt{0},\mathtt{1}\}$. Now, $c_{3i} = \mathtt{1}$, $c_{3i+1} = c_i$, and $c_{3i+2} = \mathtt{0}$ for $i \in \mathbb{N}$.

Note that for $k = 2j+2$ and $j \in \mathbb{Z}^+$, k is an even number greater than or equal to 4. For $k = 2j+3$ and $j \in \mathbb{Z}^+$, k is an odd number greater than or equal to 5.

From $\ell_{2j+2} = \mu^j(\mathtt{011})$ for $j \in \mathbb{Z}^+$, we get $|\ell_{2j+2}| = 3^{j+1}$. Thus, $\ell_{2j+2} = w_q w_{q+1} \cdots w_{q+3^{j+1}-1}$ where $q = 2 + \sum_{i=1}^{j}(3^i + 2\cdot 3^{i-1})$. Also, $w_{q+3p} = w_{q'+p}$, $w_{q+3p+1} = \mathtt{0}$, and $w_{q+3p+2} = \mathtt{1}$ where $q' = 2 + \sum_{i=1}^{j-1}(3^i + 2\cdot 3^{i-1})$ and p goes from 0 to $3^j - 1$.

From $\ell_{2j+3} = \mu^j(\mathtt{01})$ for $j \in \mathbb{Z}^+$, we get $|\ell_{2j+3}| = 2\cdot 3^j$. Thus, $\ell_{2j+3} = w_r w_{r+1} \cdots w_{r+2\cdot 3^j-1}$ where $r = 5 + \sum_{i=1}^{j}(3^{i+1} + 2\cdot 3^{i-1})$. Also, $w_{r+3p} = w_{r'+p}$, $w_{r+3p+1} = \mathtt{0}$, and $w_{r+3p+2} = \mathtt{1}$ where $r' = 5 + \sum_{i=1}^{j-1}(3^{i+1} + 2\cdot 3^{i-1})$ and p goes from 0 to $2\cdot 3^{j-1} - 1$.

Now, $c_0c_1c_2c_3c_4c_5c_6 = w_0w_1w_2w_3w_4w_5w_6$. Note that $q = 7 + \sum_{i=2}^{j}(3^i + 2 \cdot 3^{i-1})$ and $r = 16 + \sum_{i=2}^{j}(3^{i+1} + 2 \cdot 3^{i-1})$ so that $q \equiv r \equiv 1 \pmod 3$. Thus, it follows that $c_{3i} = w_{q+3p+2} = w_{r+3p+2}$, $c_{3i+1} = w_{q+3p} = w_{r+3p}$, and $c_{3i+2} = w_{q+3p+1} = w_{r+3p+1}$.

Subwords ℓ_0 and ℓ_1 are Lyndon words and, by Lemma 2, so is each $\ell_{2j+2} = \mu^j(\mathtt{011})$ and each $\ell_{2j+3} = \mu^j(\mathtt{01})$ for $j \in \mathbb{N}$. Clearly, $\ell_0 = \ell_1 = \mathtt{1} > \ell_2 = \mathtt{011}$. Note that $\mu^j(\mathtt{01})\mu^j(\mathtt{1}) = \mu^j(\mathtt{011}) > \mu^j(\mathtt{01})$ for $j \in \mathbb{N}$. Thus, $\ell_{2j+2} > \ell_{2j+3}$ for $j \in \mathbb{N}$, that is, $\ell_2 > \ell_3$, $\ell_4 > \ell_5$, and so on. The morphism μ is order-preserving because $\mu(\mathtt{01}) < \mu(\mathtt{1})$ (Richomme, 2003, Lemma 3.13). From Remark 5, since $\mathtt{01} > \mathtt{001101101}$, then $\mu^j(\mathtt{01}) > \mu^j(\mathtt{001101101}) = \mu^{j+1}(\mathtt{011})$ for $j \in \mathbb{N}$. Thus, $\ell_{2j+3} > \ell_{2j+4}$ for $j \in \mathbb{N}$, that is, $\ell_3 > \ell_4$, $\ell_5 > \ell_6$, and so on. We then get $\ell_2 > \ell_3 > \ell_4 > \ell_5 > \ell_6 > \cdots$. □

Acknowledgment

I thank an anonymous colleague for pointing out the connection to Toeplitz words. I thank an anonymous referee for his or her comments.

References

Allouche, J.-P., & Shallit, J. (2003). *Automatic sequences: Theory, applications, generalizations.* Cambridge University Press.

Berstel, J., & Karhumäki, J. (2003). Combinatorics on words—A tutorial. *Bulletin of the European Association of Theoretical Computer Science*(79), 178–228.

Cassaigne, J., & Karhumäki, J. (1997). Toeplitz words, generalized periodicity and periodically iterated morphisms. *European Journal of Combinatorics*, *18*, 497–510.

Chen, K. T., Fox, R. H., & Lyndon, R. C. (1958). Free differential calculus, IV. The quotient groups of the lower central series. *Annals of Mathematics*, *68*(1), 81–95.

Lothaire, M. (1997). *Combinatorics on words* (2nd ed.). Cambridge University Press.

Melançon, G. (1996). Lyndon factorization of infinite words. In *Stacs 96* (Vol. 1046/1996, pp. 147–154). Springer Berlin/Heidelberg.

Noche, J. R. (2008a). Generalized choral sequences. *Matimyás Matematika*, *31*(1–3), 25–28.

Noche, J. R. (2008b). On Stewart's choral sequence. *Gibón*, *8*(1), 1–5.

Richomme, G. (2003). Lyndon morphisms. *Bulletin of the Belgian Mathematical Society*, *10*, 761–785.

Siromoney, R., Mathew, L., Dare, V. R., & Subramanian, K. G. (1994). Infinite Lyndon words. *Information Processing Letters*, *50*(2), 101–104.

Stewart, I. (1995, October). Mathematical recreations: The never-ending chess game. *Scientific American*, *273*(4), 158, 160.

Joel R. Noche (jrnoche@mbox.adnu.edu.ph) is currently the managing editor of *Gibón* and was previously the assistant editor of *Kamawotan*, the journal of the College of Arts and Sciences, Ateneo de Naga University. He obtained his BS in Electrical Engineering (*cum laude*) and MS in Electrical Engineering (microelectronics option) degrees from the University of the Philippines at Diliman. He is currently finishing his dissertation for a PhD in Education (mathematics education) from the University of the Philippines Open University.