

## Minimisation of Switching Functions for Effective Biological Sequence Analysis

Zubairu Abdul, Ibrahim AA, Ayuba Muh'd Hussaini And Sabo Ayuba Khayi

<sup>1,3</sup>Shehu Shagari College of Education Sokoto, Nigeria

<sup>2</sup>State University, Sokoto State, Nigeria

<sup>4</sup>Nuhu Bamallima Polytechnic Zaria, Kaduna State, Nigeria

---

**Abstract:** Application of algebraic theoretic methods in study of DNA have been proposed recently. For instance the Boolean algebra of the four DNA bases and the Boolean algebra of the genetic code (Sánchez et al., 2004a, 2004b and 2005a). The Boolean structures were defined using the most elementary physiochemical properties of bases, the hydrogen bond numberings and base chemical types (both purines or pyrimidines). Also from the Boolean lattice of four DNA bases, a vector space and a Lie algebra was established using a new algebraic framework of the genetic code. Moreso a novel lie algebra of the genetic code over the Galois field of four DNA base was established (Robersy Sanchez, Ricardo Garau 1985) there connecting the algebraic relationships among codons to physiochemical properties of amino acids, codon assignment and single point mutation in genes. The proposed work aim at minimization of switching functions in the computations and analysis of biological sequence, thereby reducing unwanted literals.

---

### I. Introduction

Switching expressions and their manipulations provides effective tool for sequence analysis, coding as well as DNA activation (Zhiming Fang 2010). According to him many real-world processes such as chemical processes, air traffic management, manufacturing systems, automotive systems, biological systems etc can be modelled as switched systems. He further maintained that switched systems consists of sub-systems and a switching law orchestrating the active subsystem at each time instant. By a switching expression is meant the combination of a finite number of switching variables (x, y, zetc) and constants (0, 1) by means of switching operations (+, ., and ^). More precisely, any switching constant or variable is a switching expression, and if  $S_1$  and  $S_2$  are switching expressions then so are  $S_1 \wedge S_2$ ,  $S_1 + S_2$  and  $S_1 S_2$ .

Mathematical modelling of biological process provides deep insight into complex cellular systems (Rui-Sheng Wang et al 2002). While quantitative and continuous models such as differential equation have been widely used, their use is obstructed in system where the knowledge of mechanistic detail and biological parameters is scarce. On the other hand a wealth of molecular level qualitative data on individual components and interaction can be obtained from the experimental literature and high-throughput technologists, making qualitative approaches such as switching function extremely desirable.

### Objective

The objective is to minimize variables and constants during the study and computation of biological sequence using switching function technique with the sole aim of eradicating or reducing unwanted literals.

### Definition of some related term

#### 1. Switching algebra

The basic postulate of switching algebra is the existence of two-valued switching variable that can take either of the two distinct values, 0 and 1. Thus, if x is a switching variable, then

$x \neq 0$  if and only if  $x = 1$

$x \neq 1$  if and only if  $x = 0$

A switching algebra is an algebra system consisting of the set  $\{0,1\}$ , two binary operations called OR and AND, denoted by the symbol "+" and "." respectively, and one unary operation called NOT, denoted by a prime.

The definition of the OR and AND operations are as follows:

OR operation

$$0 + 0 = 0$$

$$0 + 1 = 1$$

$$1 + 0 = 1$$

$$1 + 1 = 1$$

**AND operation**

- 0 . 0 = 0
- 0 . 1 = 0
- 1 . 0 = 0
- 1 . 1 = 1

Thus the OR combination of two switching variables  $x + y$  is equal to 1 if the value of either  $x$  or  $y$  is 1 or if the values of both  $x$  and  $y$  are 1. The AND combination of these variables  $x . y$  is equal to 1 if and only if the values of  $x$  and  $y$  are both equal to 1. The NOT operation, which is known as complementation, is define as follows:

- 0' = 1
- 1' = 0

**2. Biological sequence**

In analysis when talking about sequences one will generally consider sequence of the form  $(x_1, x_2, \dots \dots \dots)$  or  $(x_0, x_1, \dots \dots \dots)$  which is to say infinite sequence of elements indexed by natural numbers.

A biological sequence is a sequence of characters from an alphabet. For DNA sequence, character alphabet is {A, C, G, T}, for RNA sequence, alphabet is {A, C, G, U}, and for protein sequence, character set is {A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V}.

3. Set - A set is an unordered collection of distinct objects. We used the notation  $x \in S$  to mean  $x$  is an element of  $S$  and  $x \notin S$  to mean  $x$  is not an element of  $S$ . Given two subsets of the universal set  $U$ ,  $X$  and  $Y$ , we say  $X$  is a subset of  $Y$ , written  $X \subseteq Y$ , if  $x \in X$  impie that  $x \in y$ . Alternatively we may say that  $Y$  is a superset of  $X$ ,  $X \subseteq Y$ . Also two subset  $X$  and  $Y$  of  $U$  are equal if  $X \subseteq Y$  and  $Y \subseteq X$ . A set with  $K$  elements is called a  $K$  set or the set with cardinality  $K$  denoted by  $/K/$
- 4.. Function – If  $A$  and  $B$  are sets, a function from  $A$  to  $B$  is rule that assign for each  $a \in A$  a unique  $b \in B$ . We write  $f(a) = b$  and say that  $f$  maps  $a$  to  $b$ . We also say the value of  $f$  at  $a$  is  $b$ . We write  $f: A \rightarrow B$  to indicate that  $f$  is a function from  $A$  to  $B$ . We call the set  $A$  the domain of  $f$  and the set  $B$  the range or equivalently codomain of  $f$
5. Boolean function – A Boolean function is a function  $f$ , from the Cartesian product  $x^n\{0, 1\}$ to  $\{0, 1\}$  . Alternatively we write  $f: x^n\{0, 1\} \rightarrow \{0, 1\}$ . The set  $x^n\{0, 1\}$ , by definition, the set of all  $n$ -tuples  $(x_1, x_2, \dots \dots \dots x_n)$  where each  $x_i$  is either 0 or 1 is called the domain of  $f$ . The set  $\{0, 1\}$  is called the codomain (range) of  $f$ .
6. Boolean algebra - a set of elements  $a, b, c \dots \dots \dots$  and binary operations “+” and “.”, that satisfies the idempotent, commutative, absorption and associative laws, and are mutually distributive.

**7. Huntinton’s postulate**

(i) Closure: There exist a domain  $B$  having at least two distinct elements and two binary operators  $(+)$  and  $(.)$  Such that

- a) a) If  $x$  and  $y$  are elements, then  $x + y$  is an element.

The operation performed by  $(+)$  is called logical addition.

- b) If  $x$  and  $y$  are elements then  $x.y$  is an element.

The operation performed by  $(.)$  is called multiplication.

(ii) Identity element: Let  $x$  be an element in the domain  $B$

- a) There exist an element  $0 \in B$  called the identity element with respect to  $(+)$  having the property  $x + 0 = x$ .
- b) There exist an element  $1 \in B$  called the identity element with respect to  $(.)$  having the property that  $x . 1 = x$

**(iii) Commutative law**

- a) Commutative law with respect to addition:  $x + y = y + x$
- b) Commutative law with respect to multiplication:  $x . y = y . x$

**iv) Distributive law:**

- a) Multiplication is distributive over addition  $x . (y + z = (x . y) + (x . z))$
- b) Addition is distributive over multiplication:  $x + (y . z) = (x + y) . (x + z)$
- (i) Complementation: If  $x$  is an element in the domain  $B$ , then there exist another element  $x'$  the complement of  $x$ , satisfying the properties :

$$a) x + x' = 1 \quad (b) x . x' = 0$$

**Notations**

Henceforth the following notation are adopted:

BCDE	Biological concept diagram editor
OR	Denoted by “+”
AND	Denoted by “.”
NOT	Denoted by “-“
$x, y$	Denote a switching variable
$x + y$	The OR combination of two switching variable
$x . y$	The AND combination of two switching variable
DNA	DeoxyriboNucleic Acid
{A, C, G, T}	DNA sequence
{A, C, G, U}	RNA sequence
{A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V}	Protein sequence
$a \in A$	Element a belongs to A
$X \subseteq Y$	X is a subset of Y. Or Y is a superset of X
/K/	Cardinality of K
$f(a) = b$	f maps a to b
$f: A \rightarrow B$	Indicates f is a function from A to B
$f: x^n\{0, 1\} \rightarrow \{0, 1\}$ .	A Boolean function from the Cartesian product $x^n\{0, 1\}$ to $\{0, 1\}$

## II. Methodology

Any biological sequence is a sequence of characters drawn from an alphabet. For DNA sequence, character set is {A, C, G, T} for RNA sequence, character set is {A, C, G, U}, for protein sequence, character set is: {A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V} (Nivit Gill 2011)  
 In the case of the DNA sequence, convert the given biological sequence {A, C, G, T} into binary form, transform them to switching function and computational process can then apply.

### Assumptions

**Assumption 1**                      Let's the four nucleotides A, C, G, T be represented by  
 $\{0, 0, 0\}, \{0, 0, 1\}, \{0, 1, 0\}, \{0, 1, 1\}$ , So that  
 $T\{A, C, G, T\} = A'C'GT' + A'CGT + A'CG'T$   
 $T\{A, C, G, T\} = A'[C'GT' + C'GT + CG'T]$   
 $= A'[C'(GT' + GT) + CG'T]$   
 $= A'[C'(0 + 1) + CG'T]$   
 $= A'[C'] = 0$   
 So T is 0 independent of the value of the variables of A, C, G, T  
 The above simplification illustrate minimization of switching function.

**Assumption 2**                      Let  $T(x, y, z) = x'y'z + yz + xz$  be a switching expression, by eliminating redundant literals  
 $x'y'z + yz + xz = z(x'y + y + x)$   
 $= z(x' + y + x)$   
 $= z(y + 1)$   
 $= z(1)$

Hence  $T(x, y, z)$  is actually independent of the values of x and y and depends only on z.

## III. Conclusions

The above minimization technique illustrates how DNA sequence can be coded using 3 bits codes and using switching functions computation could then follow in order to minimized the function so that we could obtain equivalent form of the function for effective machine processing as it is well known that redundancy needs not to be processed in the hardware.

## References

- [1]. JD Bashford and PD Jarvis, Systematics of the genetic code and anticodone International press 1998
- [2]. Robersey Sanchez, Ricardo Grau, A Novel lie algebra over the galois field of our DNA bases, publication, ActaBoitheoretica, march 2006, volume 54 issue
- [3]. J.D Bashford, I. Tsohamtjis, P.D Jarvis, Asuper symmetric model of the evolution of the genetic code, publication, National Academy of science USA nuvember 13 issue
- [4]. ZviKohaviNirajKjha, Switching and finite automata theory, 2010 Cambridge University press. USA.