

# AN APPLICATION OF RESIDUE NUMBER SYSTEM (RNS) TO MOLECULAR BIOLOGY

Joshua Apigagua Akanbasiam<sup>1,2</sup>, Kwame Osei Boateng<sup>2</sup> and Matthew Glover<sup>2</sup>  
Addo

<sup>1</sup>Department of Electrical/Electronics Engineering, DHLTU, Wa - Ghana

<sup>2</sup>Department of Computer Engineering, KNUST, Kumasi – Ghana

## ABSTRACT

*George Gamow introduced the mathematical angle that led to a better understanding of the number of possible codons of the genetic code. This builds a harmonious relationship between mathematics and molecular biology. Although very unsuccessful in his quest to provide the representation of the genetic code, with the diamond code, but his mathematical postulation of the genetic code being a 4 by 4 by 4 codon representation of 64 codons was correct. Bioinformatics ascribe digits to nitrogenous bases, binary digits, but there has been proponents for a quaternary number system representation of the genetic code since the code is made of the four (4) nitrogenous bases. The unique representation of Gamow offers the opportunity to represent the genetic code as a 3-moduli set RNS. This sets up further considerations of the genetic code (code of life) in the field of Residue Number System (RNS). AnRNS-Genetic code has the potential of opening new extraordinary vistas of potential progress in contemporary molecular biology. This work explores the feasibility of RNS as a number system of choice for generating the genetic code and further respond to the quest for a quaternary number system for the genetic code.*

## KEYWORDS

RNS, Genetic Code, RNS-Genetic Code, Big-O, DNA, RNA

## 1. INTRODUCTION

“In their recent publication, “The Death of Death,” Jose Luis Cordeiro and David Wood claim that by 2045 or earlier humans will only die in accidents and not due to natural causes or illness[1]”. They are pushing for old age to be classified as an “illness”. Their claim can partly be attributed to how far the field of molecular biology has come and how quick it is evolving, with the help of mathematics and computers [2]. A better understanding of the code of life and its manipulations (genetic engineering)[3] would help stir the world towards realising this thinking. An interdisciplinary coordinated effort is needed and the success of this quest could be traced back to the very foundation of the human life, thus the genetic code – code of life[4].The study of genes, genetic variations, and heredity in living things is known as genetics[5].Although traditionally considered a branch of biology, it also has strong connections to other life sciences and the study of information systems. Genetic information exists in a sequence of nucleotides on each strand of DNA. DNA is composed of a chain of nucleotides, of which there are four types, Adenine (A), Guanine (G), Thymine (T) and Cytosine (C)[6]. Deoxyribonucleic Acid (DNA) are used to produce Ribonucleic Acids (RNA) in a process called transcription and in RNA the nucleotides are, Adenine (A), Guanine (G), Cytosine (C), and Uracil (U). The Thymine (T) in the case of DNA is replaced with Uracil (U) for RNA. RNAs are used to produce proteins in a process called translations and this is generally tagged the central dogma of molecular biology[7], seen in Figure 1.



Figure 1: The Central Dogma of Molecular Biology

The genetic code is the way by which DNA stores the genetic information and it consists of “codons” of three nucleotides[4]. The 20 distinct amino acids found in all living things are specified by the three nucleotides' in a  $4^3 = 64$  possible combinations. These amino acids form proteins, usually the proteins are made of 100's or 1000's of amino acids. The order of the amino acids determines the protein's structure and function. Each proteins has a unique shape which is determined by the order in which the amino acid are arranged. The genetic code is nearly universal [8]thus almost all organisms use the same code – this is good for both genetic engineering and biotechnology which forms the direction of contemporary science. Additionally, it is purposefully redundant, eliminating the majority of coding errors, mutations, involving single bases. A code requires the code to have rules and meaning of its symbols to accomplish some task. Thus, a code must have a syntax and semantics. The standard genetic code represents an extension of the of the four-letter bases of DNA; Adenine (A), Guanine (G), Cytosine (C), and Thymine (T). It could also be written in RNA where Thymine is replaced with Uracil (U). The information in DNA is in sequence of bases and a DNA word are three (3) letters long (triplets). This forms the bases for a three moduli set consideration (representation) of the genetic code. Number representation is arguably the most important topic in computer arithmetic and digital applications[9]. The choice of number representation affects the implementation cost and delay of any digital applications [10].

## 2. GENETIC ENGINEERING AND BIOTECHNOLOGY

The term "biotechnology" refers to the utilization of biological processes, organisms, cells, and cellular components in the creation of new technology. The main aim is to develop new tools and products that are useful in research, agriculture, industry and medicine. Some of the earlier application of biotechnology is the process of fermenting bear or preparing bread using yeast. The field of biotechnology has expanded over the years to include sub-fields like genomics and genetic engineering bringing into bear the relevance of the genetic code in biotechnology. A biotechnological application known as genetic engineering uses methods that enable the exchange of genes across different animals. This is possible because the genetic code is universal. Discovering the genetic code made it possible to think of organisms as information systems. Humans are consistently trying to produce better crops, livestock and medicine. This is achieved through deliberate modification of an organisms characteristics by manipulating its genetic information (DNA), thus by taking a specific gene from one organism and placing them in another organism genome. Genetic engineering has led to the production of organisms that have acquired one or more genes by artificial means. Another way to think of genetic engineering is as the process of changing an organism's genetic makeup using recombinant DNA technologies. The manipulation of genomes has led to the control of breeding and production of offspring with desired traits. There are several genetic disorders that can be passed on from one generation to another and genetic engineering (gene editing) offers the potential to disable target genes, or correct harmful mutations in plants and animals. With the help of genetic engineering, it is possible to create plants that are more nutritious, can withstand extreme environmental conditions

(herbicides), or have increased insect resistance without the need of pesticides. Bacteria that produce human insulin; chicken more resistant to infection; cows that produce leaner meat; rice enhanced with vitamin A (Golden Rice); trees that grow faster than normal,; bananas containing the vaccine for disease such as cholera and hepatitis; are achievements of genetic engineering and consequently biotechnology and this is made possible due to a better understanding and digitization of the genetic code.

### 3. FEATURES OF THE GENETIC CODE

The canonical genetic code basically has sixty-four (64) codons of which sixty-one (61) code for amino acids[11]. Three (3) codons serve as termination signals or are generally referred to as stop codons for protein synthesis. One (1) codon, methionine, serves as initiation signal or generally referred to as start codon thus it signal for the beginning of protein synthesis in the mRNA chain. The promoter or start codon can be counted as part of the protein sequence or polypeptide in which case it is not serving as a promoter codon. The canonical genetic code is degenerate, thus different codons can code for the same amino acids;Leucine, Arginine and Serine have 6 codons; Valine, Proline, Threonine, Alanine and Glycine have 4 codons; Isoleucine, and the terminator (stop) have 3 codons; Phenylalanine, Tyrosine, Histidine, Glutamine, Asparagine, Lysine, Aspartic acid, Glutamic acid and Cysteine each have 2 codons; Tryptophan and Methionine are the codons with only one codons each[12]. The genetic code has other features as being polar or nonpolar, aliphatic or aromatic, uncharged or charged (positive or negative). Figure 2 illustrates the canonical genetic code table and the some of its features.

		1st base					
		U	C	A	G		
2nd base	U	UUU Phenylalanine UUC Phenylalanine UUA Leucine UUG Leucine	UCU Serine UCC Serine UCA Serine UCG Serine	UAU Tyrosine UAC Tyrosine UAA Stop UAG Stop	UGU Cysteine UGC Cysteine UGA Stop UGG Tryptophan	U C A G	
	C	CUU Leucine CUC Leucine CUA Leucine CUG Leucine	CCU Proline CCC Proline CCA Proline CCG Proline	CAU Histidine CAC Histidine CAA Glutamine CAG Glutamine	CGU Arginine CGC Arginine CGA Arginine CGG Arginine	U C A G	
	A	AUU Isoleucine AUC Isoleucine AUA Isoleucine AUG Methionine (Start)	ACU Threonine ACC Threonine ACA Threonine ACG Threonine	AAU Asparagine AAC Asparagine AAA Lysine AAG Lysine	AGU Serine AGC Serine AGA Arginine AGG Arginine	U C A G	
	G	GUU Valine GUC Valine GUA Valine GUG Valine	GCU Alanine GCC Alanine GCA Alanine GCG Alanine	GAU Aspartic Acid GAC Aspartic Acid GAA Glutamic Acid GAG Glutamic Acid	GGU Glycine GGC Glycine GGA Glycine GGG Glycine	U C A G	

Nonpolar, aliphatic  
 Polar, uncharged  
 Aromatic  
 Positively charged  
 Negatively charged

Figure 2: The canonical genetic code and some relevant features[13]

### 4. NUMBER SYSTEM AND BINARY NUMBER SYSTEM

Leibniz published his paper “Explanation of Binary Arithmetic” he used the I ching which dates from the 9th century BC in china to support his work. He interpreted the hexagons of the I ching as evidence of binary calculus[14]. The primary number system humans used has been the decimal number system but computers and digital technological advancement fuelled the adoption of a more sophisticated number system, binary number system. The discovery of binary number system and its digital application marked an exciting development in digital electronics, computing and telecommunications[10]. This number system is used in all digital computing applications. Number system is the form and basics of all digital applications, it ranks in the echelon of all digital consideration thus it determines efficiency, speed, space, error control

coding among others. This research, therefore presents RNS as an ideal notation (representation) of the code of life (genetic code). Modern computer or digital applications operate on the principles of two states “ON” and “OFF”. This corresponds to electrical current being present or not, or being low or high and have the collective adoption and prominence of the binary number system. RNS was introduced as a result of the difficulties the binary number system presented to digital applications. Due to the rapid design speed, reduced space requirements, and energy savings, residue number system (RNS) has recently been proposed as a substitute method. In addition to lowering network overhead, RNS has been utilized in wireless sensor networks to boost dependability and use less transmission energy. In a software-defined network, it has also been demonstrated to improve switching operations and replace lookup tables with straightforward modular processes. Additionally, it has received some interest from researchers working on digital image processing and promises an image coding scheme that allows safe image processing on a high-speed, low-power VLSI platform. RNS is one of the core tools of contemporary high-performance computing systems because it integrates well with contemporary parallel processing environments and platforms. The beneficial properties of RNS allows for the reduction in computational complexity of cryptographic primitives, digital signal processing algorithms, and artificial neural networks. In addition, due to its unique combination of properties, RNS can be used to construct error correction codes and distributed storage schemes.

## 5. INTRODUCTION TO RNS AND RRNS

RNS representation solves the carry propagation problem of weighted number systems (binary)[15]. The residues are quite smaller and makes it easier to implement arithmetic (Adder, subtractor, or multipliers) by direct look-up tables. Since each digit is a relatively smaller number, the structure of RNS arithmetic makes operations much fast and simple. This speed and simplicity are the primary advantages of RNS. RNS is a system for encoding integers by their residues, or remainders, after division by a predetermined range of substantially prime integers[16]. These residues can be subjected to modular arithmetic operations like addition, subtraction, and multiplication in separate channels, making the system carry-free. The fundamental drawback of RNS is the extensive set of processes required for scaling and conversion back to RNS. A large integer is represented by a residue number system as a collection of smaller integers.

Let  $Y = \{m_1, m_2, \dots, m_n\}$  be a set of pairwise relatively prime integers.  
The dynamic range

$$M = (m_1 \times m_2 \times \dots \times m_n) = \prod_{i=1}^n m_i$$

Then for any integer in the residue class  $R_m$  has a unique n-tuple RNS representation given as;  
 $X \rightarrow (x_1, x_2, \dots, x_n)$  Where  $x_i = |X|_{m_i}$  or  $X \bmod m_i$  and is called the  $i_{th}$  residue

### 5.1. Conversion [Forward and Reverse]

The two bidirectional consideration of conversion is from standard notation to RNS or RNS to standard representations of numbers[17]. Two classes of conversion methods can be found in literature: one that is based directly on the Chinese remainder theorem and the other that uses an intermediary Mixed Radix representation. RNS has demonstrated some difficulties in magnitude comparison or sign detection, division, base extension, scaling etc[10]. The research community is constantly working and in no time will have widespread implementation of RNS in many fields.

## 6. THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

The entirety of an organism's genetic information, including its structure and function, is encoded in its DNA. [18]. Proteins are formed using the genetic code of the DNA. The central dogma is among one of two hypothesis of Francis Crick, aside the sequence hypothesis, which better explained protein synthesis. The central dogma is a unidirectional flow equation that represents the fundamental law of molecular biology[7]. It is the mechanism whereby inherited information (traits) are used to create actual objects, namely enzymes and proteins. Central to the formation of proteins is a four (4) letter nucleic acid, a three (3) letter concatenation of these letters constituting a codon or amino acid and 20 different codons or amino acids combine to produce protein. The hypothesized directions for genetic information transfer are indicated by the arrows. The one surrounding DNA denotes that DNA serves as the blueprint for its own self-replication. A DNA template controls RNA production through a procedure known as transcription, as seen by the arrow between DNA and RNA. In line with this, an RNA template controls the process of protein synthesis known as translation. It is noteworthy that the latter two arrows are portrayed as unidirectional, meaning that neither RNA sequences nor DNA are ever produced on protein templates.

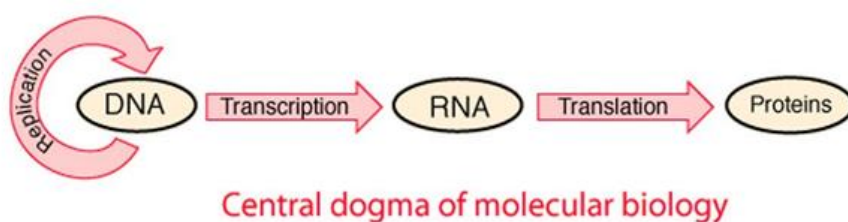


Figure 3. The central dogma according to Watson[19]

### 6.1. DNA

Friedrich Miescher, a Swiss researcher, discovered a cellular substance in 1869 which he termed nuclein – this was latter to be known as DNA[20]. Also the discovery of the structure of DNA by Crick and Watson was a momentous event in life science[21], an event that would shape the discourse of the study of nature. It is a repository of genetic information thus it carries the genetic blueprint used in the growth, development, functioning and reproduction of all known living organisms. It is generally considered to be inherently stable. DNA is a double helix and is made of a Sugar, phosphate, and nitrogenous base (purines and Pyrimidines). DNA regulates every chemical reaction that occurs within cells, determining the kind of cells that are produced—muscle, blood, nerve, etc.—as well as which organisms are created—humans, bees, monkeys, etc. DNA replicate to produce two identical DNA and also serve as a template for synthesizing RNA.

### 6.2. RNA

An important component of protein synthesis is ribonucleic acid (RNA), a single-stranded polymer. Adenine (A), Guanine (G), Cytosine (C), and Uracil (U) are nitrogenous bases that make up RNA [22]. RNA at a glance looks like DNA but has some variation with DNA. It has ribose sugar, accounting for its name, "ribonucleic". It is also single stranded thus it is formed by splitting the double helix of DNA and each one forming the template for forming RNA, mRNA.

This feature makes RNA a good intermediary for the formation of protein but less stable compared with DNA as a store of genetic information. There are a number of RNA's and they all have different functions in aiding protein synthesis. Notably among them are; messenger RNA, mRNA, the ribosomal RNA, rRNA and the transfer RNA, tRNA. Messenger RNA (mRNA) functions as messenger carrying information in a gene to protein synthesizing machinery. The mRNA is coded such that every three (3) nucleotides, a codon of the genetic code, correspond to one amino acid. The smallest of the three RNA molecules, transfer RNA (tRNA), transports amino acids from the cytoplasm to the machinery responsible for synthesising proteins. Because tRNA serves as an adapter for the translation of a given amino acid's nucleotide sequence from mRNA, they are also known as adapter molecules. Thus the genetic code generally can be represented in either DNA or RNA where in the case of RNA the letter T is replaced with U[23]. The progress made in biotechnology, genetic engineering and bioinformatics is digitizing these fundamental units by ascribing digits to the nitrogenous bases. This allows for sophisticated algorithms and lately artificial intelligence to help advance the course of biotechnology, genetic engineering and personalized medicine. Binary notation has been the most propagated, not without challenges, but the ideal notation for the genetic code is a four (4) digit system (quaternary) since there are four nitrogenous bases.

### 6.3. Protein

The most significant biological component, proteins, were first discovered by Gerardus Johannes Mulder in 1930. Protein is derived from the Greek word proteios, which means "of the first importance". Proteins are polymers of monomer units called amino acids joined together by peptide bonds[24]. They play major role as essential parts of the structure and function of the human body. They could act as catalyst to certain reactions, form the most part of hormones and enzymes or provide energy in the absence of carbohydrates. Proteins are made from amino acids which are made from nitrogenous bases in triplets called codons – constituents of the genetic code. Proteins are constructed from a set of twenty (20) amino acids linked by peptide bonds. They are one of the essential building blocks of the human body. Proteins are constructed from a set of twenty (20) amino acids. They are used to make antibodies, enzymes, hormones, and also form part of contractile muscles, structural, storage and transport.

## 7. METHODOLOGY

The genetic code is composed of three-letter combinations drawn from a base alphabet with four letters[25]. For each of the 20 amino acids, these base triplets serve as a codon. George Gamow postulated that for a three letter codons to be generated out of four possible alphabets of DNA, there will be  $(4^3)$ , thus there are 64 possibilities to achieve the whole genetic code[26]. Since the canonical genetic code itself is derived from 20 amino acids it implies the genetic code is redundant, most amino acids correspond to more than one base sequence, but in a purposeful way. The redundancy takes care of the majority of single-base errors in coding. With this understanding of the genetic code this work looks at where RNS perfectly fits in the generation of the canonical genetic code. RNS is basically a concatenation of the various digits of the moduli sets under consideration, thus a two moduli set gives a two digit concatenation, three moduli sets is made of a three digit concatenation and so on. The canonical genetic code is the set of rules by which genetic information is translated into proteins from codons i.e. triplets of nucleotides, to amino acids. These codon emanate from four (4) letters as mentioned and are combined in triplets know as codons or more generally amino acids. Given that  $m_1m_2m_3$  are the triplet of bases making up an amino-acid, in the concept of RNS each of  $m_1m_2m_3$  represent the first, second, and third bases respectively in an amino acid representation. The combination of the four (4) bases gives rise to a total of sixty-four (64) triplet codons in a  $4 \times 4 \times 4$  matrix or  $4^3$  index fashion. The

model proposed, hence algorithm allows for at least a three (3) moduli sets where  $m_1$ ,  $m_2$  and  $m_3$  must be greater than or equal to four (4) in which case a product of the three moduli sets representing the dynamic range is greater than or equal 64. This will suffice in representing the genetic code thus allows for a perfect coding of the genetic code in RNS.

### 7.1. Design Algorithms

<b>RNS (<math>m_1, m_2, m_3</math>)</b>			
Input: Three relatively prime numbers $m_1, m_2, m_3$			
Output: RNS-Genetic code Reference Table			
1.	Array RNSCode[ $m_1, m_2, m_3$ ]		
2.	for $i \leftarrow 0$ to $m_1 - 1$		
3.		for $j \leftarrow 0$ to $m_2 - 1$	
4.			for $k \leftarrow 0$ to $m_3 - 1$
5.			RNSCode[ $i, j, k$ ] $\leftarrow$ $ijk$
6.	return RNSCode		

Figure 4: Algorithm for RNS representation

<b>RNSGeneticCode(<math>m_1, m_2, m_3</math>)</b>			
Input: Three relatively prime numbers $m_1, m_2, m_3$			
Output: RNS-Genetic code Reference Table			
1.	Array GeneticCode[4, 4, 4]		
2.	for $i \leftarrow 0$ to $m_1 - 1$		
3.		for $j \leftarrow 0$ to $m_2 - 1$	
4.			for $k \leftarrow 0$ to $m_3 - 1$
5.			if ( $(i < 4) \ \&\& \ (j < 4) \ \&\& \ (k < 4)$ )
6.			GeneticCode[ $i, j, k$ ] $\leftarrow$ $ijk$
7.	return GeneticCode		

Figure 5: Algorithm for the RNS-Genetic Coding

The algorithm in figure 4 allows for the publishing of any three moduli set RNS. It allows the user to input three relatively prime moduli which prints out all the possible RNS representation for the set of moduli entered. This is the first approach to generating all possible digits of any three moduli set RNS digits and concatenating these digits. The algorithm in figure 5 further helps truncate the RNS digits generated to four (4) which suites the definition of the four (4) nitrogenous bases. Thus this generates a reference table that help sorts out all legitimates codons of the RNS digits generated. Those digits that do meet the demand of the checks in line 5 of figure 6 are truncated and do not form part of the consideration of the RNS-Genetic code.

## 8. DISCUSSION OF RESULTS

Any three (3) moduli sets RNS will adequately generate the genetic code and would have the same digit for the 64 codons. Figure xx shows the look of the RNS capable of generating the Genetic Code. A two moduli set system will be inadequate in generating the genetic code since each codon is made of triplet of basis an in this case digits. The table below highlights the legitimate RNS codons in green and the illegitimate codons in plane cells. Since there are four

nitrogenous bases the digit representation are; 0, 1, 2 and 3 and higher digits thus greater than 3 are truncated. Also it is worthy to note that each moduli set must be greater than or equal to four (4) to adequately represent each digits of the codon bases, the first, second and third base positions. When the moduli sets are greater than three, then a declaration of some to the moduli as redundant will suffice in generating the genetic code, and this fits in the frame of redundant residue number systems, RRNS.

Table 1 – RNS with moduli set 5, 6 and 7 - Coded and Un-coded moduli digits

	4	5	7	Amino Acid	4	5	7	Amino Acid	4	5	7	Amino Acid		
0	0	0	0	TTT	47	3	2	5	94	2	4	3		
1	1	1	1	CCC	48	0	3	6	95	3	0	4		
2	2	2	2	AAA	49	1	4	0	96	0	1	5		
3	3	3	3	GGG	50	2	0	1	ATC	97	1	2	6	
4	0	4	4		51	3	1	2	GCA	98	2	3	0	AGT
5	1	0	5		52	0	2	3	TAG	99	3	4	1	
6	2	1	6		53	1	3	4		100	0	0	2	TTA
7	3	2	0	GAT	54	2	4	5		101	1	1	3	CCG
8	0	3	1	TGC	55	3	0	6		102	2	2	4	
9	1	4	2		56	0	1	0	TCT	103	3	3	5	
10	2	0	3	ATG	57	1	2	1	CAC	104	0	4	6	
11	3	1	4		58	2	3	2	AGA	105	1	0	0	CIT
12	0	2	5		59	3	4	3		106	2	1	1	ACC
13	1	3	6		60	0	0	4		107	3	2	2	GCC
14	2	4	0		61	1	1	5		108	0	3	3	TGG
15	3	0	1	GTC	62	2	2	6		109	1	4	4	
16	0	1	2	TCA	63	3	3	0	GGT	110	2	0	5	
17	1	2	3	CAG	64	0	4	1		111	3	1	6	
18	2	3	4		65	1	0	2	CTA	112	0	2	0	TAT
19	3	4	5		66	2	1	3	ACG	113	1	3	1	CGC
20	0	0	6		67	3	2	4		114	2	4	2	
21	1	1	0	CCT	68	0	3	5		115	3	0	3	GTG
22	2	2	1	AAC	69	1	4	6		116	0	1	4	
23	3	3	2	GGA	70	2	0	0	ATT	117	1	2	5	
24	0	4	3		71	3	1	1	GCC	118	2	3	6	
25	1	0	4		72	0	2	2	TAA	119	3	4	0	
26	2	1	5		73	1	3	3	CGG	120	0	0	1	TTC
27	3	2	6		74	2	4	4		121	1	1	2	CCA
28	0	3	0	TGT	75	3	0	5		122	2	2	3	AAG
29	1	4	1		76	0	1	6		123	3	3	4	
30	2	0	2	ATA	77	1	2	0	CAT	124	0	4	5	
31	3	1	3	GCG	78	2	3	1	AGC	125	1	0	6	
32	0	2	4		79	3	4	2		126	2	1	0	ACT
33	1	3	5		80	0	0	3	TTG	127	3	2	1	GAC
34	2	4	6		81	1	1	4		128	0	3	2	TGA
35	3	0	0	GTT	82	2	2	5		129	1	4	3	
36	0	1	1	TCC	83	3	3	6		130	2	0	4	
37	1	2	2	CAA	84	0	4	0		131	3	1	5	
38	2	3	3	AGG	85	1	0	1	CTC	132	0	2	6	
39	3	4	4		86	2	1	2	ACA	133	1	3	0	CGT
40	0	0	5		87	3	2	3	GAG	134	2	4	1	
41	1	1	6		88	0	3	4		135	3	0	2	GTA
42	2	2	0	AAT	89	1	4	5		136	0	1	3	TCG
43	3	3	1	GGC	90	2	0	6		137	1	2	4	
44	0	4	2		91	3	1	0	GCT	138	2	3	5	
45	1	0	3	CTG	92	0	2	1	TAC	139	3	4	6	
46	2	1	4		93	1	3	2	CGA	140				

Thus methionine would always be coded as [203] irrespective of the moduli sets chosen this is seen in figure 6. The code AGU represent Serine would always have the RNS digits [230]. This offers some generality for RNS being an ideal number of choice for generating the genetic code (code of life). In addition to this generality all representation depending on the moduli set chosen has a decimal representation for each codon, since each concatenated three digit moduli set represent a particular decimal value. These values could be exploited for other research considerations relating to the genetic code. All residues digits greater than or equal to three (3) are legitimate digits for generating the RNS-Genetic code, thus all other residue digits other than these are illegitimate and hence are truncated in any consideration in generating the RNS-Genetic code.

000	010	020	030	UUU	UCU	UAU	UGU
001	011	021	031	UUC	UCC	UAC	UGC
002	012	022	032	UUA	UCA	UAA	UGA
003	013	023	033	UUG	UCG	UAG	UGG
100	110	120	130	CUU	CCU	CAU	CGU
101	111	121	131	CUC	CCC	CAC	CGC
102	112	122	132	CUA	CCA	CAA	CGA
103	113	123	133	CUG	CCG	CAG	CGG
200	210	220	230	AUU	ACU	AAU	AGU
201	211	221	231	AUC	ACC	AAC	AGC
202	212	222	232	AUA	ACA	AAA	AGA
203	213	223	233	AUG	ACG	AAG	AGG
300	310	320	330	GUU	GCU	GAU	GGU
301	311	321	331	GUC	GCC	GAC	GGC
302	312	322	332	GUA	GCA	GAA	GGA
303	313	323	333	GUG	GCG	GAG	GGG



Figure 6: The RNS-Genetic Code

### 8.1. Big-O Analysis

The first algorithm publishes any three (3) moduli set residue numbers with moduli sets  $m_1$ ,  $m_2$  and  $m_3$ . It is deemed to have an exponential big-O algorithm and more specially as can be seen in the analysis above Big-O of  $O(n^3)$ . Although a cubic big-O algorithms is not entirely efficient it is worthy to know that the big-O focuses on the worst case scenario. The choice of moduli sets will play a major role in the algorithms sorting out 64 codons out a large set of numbers. That is to say that a careful choice of smaller moduli set like 4, 5 and 7 sets the dynamic range at 140 and hence will execute quiet quickly since the 64 codons will be sorted out of 140 set of numbers.

ANALYSIS	
1	1
2	$1 + m_1$
3	$1 + m_2$
4	$1 + m_3$
5	$5 (m_1 * m_2 * m_3)$
6	$2 (m_1 * m_2 * m_3)$
7	1
$t(n) = 1 + 1 + m_1 + 1 + m_2 + 1 + m_3 + 5(m_1 * m_2 * m_3) + 2(m_1 * m_2 * m_3)$	
$t(n) = 2 + 3n + 7n^3$	
As $n \rightarrow \infty$	
$t(n) = O(2 + 3n + 7n^3)$	
$t(n) = O(n^3)$	

### 9. CONCLUSION

In this work, we have successfully presented a three (3) moduli RNS-Genetic code. This creates a very strong relationship between the mathematics of RNS and the foundation of molecular biology- genetic code. This generation of the genetic code using RNS is universal thus to say that any moduli sets chosen can be used to generate the RNS-Genetic code. Hence, the choice of RNS moduli set, which has a direct correlation of efficient implementation of RNS, and is a strong research consideration would only go a long way to aid in an efficient generation of the RNS-genetic code proposed. This research work also buttress the quest for quaternary number system for molecular biological analysis. It will also serve other future considerations like intron excision and exon splicing and the codon anticodon table for protein synthesis.

### REFERENCES

[1] “2045: An immortality Odyssey?– The New Indian Express.” <https://www.newindianexpress.com/opinions/2018/jun/16/2045-an-immortality-odyssey-1828876.html> (accessed Apr. 05, 2022).

[2] I. Trenchev, M. Traykov, M. Trencheva, M. Stoyanova, and I. Trenchev, “Mathematical Models for Studying the Properties of the Genetic Code Mathematical Modeling of Optimal TGCs,” pp. 40–54, 2017.

[3] FAO, “Biosafety Resource Book,” no. January 2011, p. vi + 80 pp., 2011, [Online]. Available: <http://www.fao.org/docrep/014/i1905e/i1905e.pdf>.

[4] B. Hayes, “The Invention of the Genetic Code,” *Am. Sci.*, vol. 86, no. 1, pp. 8–14, 1998, doi: 10.1511/1998.17.3338.

[5] D. Counsell, “Bioinformatics and Molecular Evolution,” *Comp. Funct. Genomics*, vol. 6, no. 5–6,

- pp. 317–319, 2005, doi: 10.1002/cfg.486.
- [6] J. D. W. with A. Berry, *DNA - The secret life*. 2003.
- [7] M. Morange, “The Central Dogma of molecular biology,” *Resonance*, vol. 14, no. 3, pp. 236–247, 2009, doi: 10.1007/s12045-009-0024-6.
- [8] E. V. Koonin and A. S. Novozhilov, “Origin and evolution of the genetic code: The universal enigma,” *IUBMB Life*, vol. 61, no. 2, pp. 99–111, 2009, doi: 10.1002/iub.146.
- [9] R. Lal, “Number system,” *Infosys Sci. Found. Ser. Math. Sci.*, no. October, pp. 55–91, 2017, doi: 10.1007/978-981-10-4253-9\_3.
- [10] O. Amos and B. Premkumar, *Residue Number System Theory and Implementation*. 2007.
- [11] K. Watanabe and T. Suzuki, “Universal Genetic Code and its Natural Variations,” *Encycl. Life Sci.*, pp. 1–8, 2008, doi: 10.1002/9780470015902.a0000810.pub2.
- [12] J. D. Watson and F. H. C. Crick, “‘On Protein Synthesis’ Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid,” *Nature*, vol. 171, no. 4356, pp. 737–738, 1953, doi: 10.1038/171737a0.
- [13] B. Ostash and M. Anisimova, *Visualizing Codon Usage Within and Across Genomes: Concepts and Tools*, no. May. 2020.
- [14] S. P. Fedotov, “Simple association of the genetic code with hexagrams of the Book of Changes (I Ching),” *Cardiometry*, no. 9, pp. 32–43, 2016, doi: 10.12710/cardiometry.2016.9.3243.
- [15] K. A. Gbolagade, *Effective Reverse Conversion in Residue Number System Processors*. 2010.
- [16] E. K. Bankas, K. A. Gbolagade, and S. Dan Cotofana, “An effective New CRT based reverse converter for a novel moduli set  $\{2^{2n+1} - 1, 2^{2n+1}, 2^{2n} - 1\}$ ,” *Proc. Int. Conf. Appl. Syst. Archit. Process.*, pp. 142–146, 2013, doi: 10.1109/ASAP.2013.6567567.
- [17] A. Persson and L. Bengtsson, “Forward and reverse converters and moduli set selection in signed-digit residue number systems,” *J. Signal Process. Syst.*, vol. 56, no. 1, pp. 1–15, 2009, doi: 10.1007/s11265-008-0249-8.
- [18] S. Damodaran, “Amino acids, peptides, and proteins,” *Fennema’s Food Chem.*, pp. 235–356, 2017, doi: 10.1201/9781315372914.
- [19] “DNA.” <http://hyperphysics.phy-astr.gsu.edu/hbase/Organic/dogma.html> (accessed Apr. 06, 2022).
- [20] L. A. Pray and A. A. Aa, “Discovery of DNA Structure and Function: Watson and Crick The First Piece of the Puzzle: Miescher Discovers DNA Citation: Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. Nature Education 1(1):100,” 2008, Accessed: Oct. 12, 2021. [Online]. Available: <https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397>.
- [21] C. De Duve, “The second genetic code,” *Nature*, vol. 333, no. 6169, pp. 117–118, 1988, doi: 10.1038/333117a0.
- [22] A. V. Lobanov, A. A. Turanov, D. L. Hatfield, and V. N. Gladyshev, “Dual functions of codons in the genetic code,” *Critical Reviews in Biochemistry and Molecular Biology*, vol. 45, no. 4, Taylor & Francis, pp. 257–265, Aug. 2010, doi: 10.3109/10409231003786094.
- [23] “Darwin1.” <http://web.furman.edu/~wworthen/bio111/code.htm> (accessed Apr. 06, 2022).
- [24] “Protein structure: Primary, secondary, tertiary & quaternary (article) | Khan Academy.” <https://www.khanacademy.org/science/biology/macromolecules/proteins-and-amino-acids/a/orders-of-protein-structure> (accessed Dec. 31, 2020).
- [25] N. N. Kozlov, “The Study of the Secrets of the Genetic Code,” *J. Comput. Commun.*, vol. 06, no. 07, pp. 64–83, 2018, doi: 10.4236/jcc.2018.67007.
- [26] V. Nanjundiah, “George Gamow and the genetic code,” *Resonance*, vol. 9, no. 7, pp. 44–49, 2004, doi: 10.1007/bf02903575.

## AUTHORS

**Ing. Joshua Apigagua Akanbasiam** is currently a PhD. Candidate in the Department of Computer Engineering, KNUST, and he holds a Masters in Telecommunications Engineering and BSc. in Computer Engineering, KNUST. He is a Senior Lecturer and currently the head of the department of Electrical/Electronics Engineering, Dr. Hilla Limann Technical University formerly Wa Polytechnic - Ghana. Prior to Lecturing he served as the Maintenance Manager of Mass Telecom Innovation (MTI) a company that maintained Telecom Power for MTN and Airtel cell sites in the three Northern Regions. He is a member of GhIE and IEEE. His research interest is in Broadband services, Computer Arithmetic and Applications, RNS and its application, Digital Signal Processing, and Bioinformatics.

**Prof. Ing. Kwame Osei Boateng** holds a Doctorate in Systems Engineering from the Graduate School of Engineering and Science of Ehime University, Japan, Masters in Computer Science from the same university and BSc. (Hons.) degree in Electrical/Electronic Engineering from the University of Science and Technology (now Kwame Nkrumah University of Science and Technology), Kumasi Ghana. Since March 2003, Prof. Ing. Boateng has been working for Kwame Nkrumah University of Science and Technology and has risen through the ranks as a Senior Lecturer, an Associate Professor and currently a Professor. He has held several positions as head of department of Computer Engineering, the dean of Faculty of Electrical and Computer Engineering of the College of Engineering and director of the Institute of Distance Learning and ICT consultant. Prof. Ing. Boateng is a member of the GhIE, IEEE and IEEE Computer Society. He has served on technical programme committee for the 11th IEEE (ATS'02) and IEEE VLSI Test Symposium (VTS), ICAST 2012, WTS 2013, and ESTE 2015. His present research interests are in the areas of design, test and diagnosis of logic and VLSI circuits, test of mixed-signal circuits, network security protocols, applications of residual number systems (RNS), image processing and smart metering.

**Prof. Matthew Glover Addo;** Mathew Glover Addo is a Professor of Microbiology/Molecular Biology at the Department of Theoretical and Applied Biology, Faculty of Biosciences at the Kwame Nkrumah University of Science and Technology (KNUST). Currently, he is the Director of the Institute of Distance Learning, KNUST. Prof Addo holds a BSc in Biological Sciences from KNUST-Ghana, an MSc in Biotechnology from the University of Bergen, Norway and Doctor of Science/Doctor of Philosophy degrees at the Université Paris Sud IX, Paris, France and KNUST, Kumasi. He completed the doctorate programme in May, 2011. During the doctorate research programme, Prof. Addo designed an efficient screening method for the identification of genes involved in the mitochondrial genome stability using *Caenorhabditis elegans* as a model organism. He worked with the Functions and Dysfunction of Mitochondrial group of the Institute of Genetics and Microbiology where for the first time, he identified four (4) new nuclear genes (Y105E8A.23, dnj-10, atad-3, and phi-37) involved in mitochondrial stability. Prof. Addo has has 20 years of teaching, collaborative research and consultancy experience. His specialisation is in Gene Expression, Clinical Infectious Microorganisms and Food and Water Microbiology. He has served the international and local communities in a number of capacities and has several peer reviewed publications in reputable international journals. He has also assessed a number of MSc/MPhil and PhD theses. Prof. Addo is a visiting Lecturer/Assessor at the ISA-Lille University, Lille, France. Before his appointment as Director of IDL, Prof Addo was the Dean of Faculty of Biosciences.