Codon Characterization Based on Electrical Response

Monalisa Dutta
Institute of Radio Physics and Electronics
University of Calcutta
92 A.P.C. Road, Kolkata-700009, India

Soma Barman
Institute of Radio Physics and Electronics
University of Calcutta
92 A.P.C. Road, Kolkata-700009, India

ABSTRACT
Electrical circuit modeling of biological system is a significant research topic in recent years. Codon is an important element in DNA sequence which is responsible for formation of protein in genes. 64 codons of genes are modeled using MOSFET from elementary level and characterized them based on purine and pyrimidine property. The model is realized in Spice domain and a new genetic code classification table is described based on simulated transient voltage and current responses of purine and pyrimidine bases. The effect of base position is investigated using frequency domain analysis of voltage and phase characteristics of codon electrical circuit.

Keywords
Codon, DNA, nucleotides electrical model, MOSFET.

1. INTRODUCTION
One of the most important scientific discoveries of the 20th century by Crick and Watson is the double-helix structure of DNA (Deoxyribonucleic acid) and the mechanism of base pairing by which genetic information is stored in living organisms [1]. DNA is blueprint of every living organism [2]. DNA is string of genetic code (or codon). George Gamow, 1953, was first proposed that the sets of three bases must be employed to encode the 20 standard amino acids used by living cells to build proteins [3]. From several decades, the researchers from different fields show their interest on study of electrical response of DNA, theoretically as well as experimentally [4]. Many researchers are studied charge migration through DNA molecules [5, 6] and the electrical conductivity of DNA molecules for its possible application in molecular electronics in different ways [7-11] and created a revolution in traditional pharmaceutical industries [4]. In order to study the electrical characteristic theoretically, different type of electrical equivalent circuit of DNA molecules were modeled by many researchers from different fields. DNA consists of a long chain of codon and each codon of DNA was modeled as a R-C ladder circuit to study its nature [11]. DNA also was modeled as L-C oscillator based on chemical structure of DNA molecules [12]. The PSpice model of electrical behavior of DNA used in nanoelectronics was described by Vedrana Hodzic and also studied the nonlinear I-V characteristic [13]. Researchers also proposed diode like behavior of codons [14]. H-Bond has capacitive property proposed by many researchers [15].

The equivalent electrical circuits using passive circuit components become very bulky and consume high power compared to active circuit components. The equivalent electrical circuit design using MOSFET is a challenging research for molecular device designers [17]. In present paper the authors have described equivalent electrical circuit realization of each codon using MOSFET and study their transient and frequency response characteristic. Based on transient responses, the codons are grouped and a new genetic code table is designed. The frequency response of the equivalent electrical circuit of codon is analyzed to study the effect of position of nucleotide base in codon.

The rest of the paper is organized as background of DNA, methods of equivalent electrical circuit modeling of codon, time and frequency domain response of codon equivalent electrical circuit and conclusion.

2. BACKGROUND
Deoxyribonucleic acid, commonly known as DNA, has two strands, running in opposite directions. The backbone of strands made up by sugar phosphate. Since each strand of a DNA is represented by an alphabet of four elements, named as adenine (A), guanine (G), cytosine (C) and thymine (T) (Fig. 1a) [18]. A small portion of single strands given below.

ATCCAGCTAGGGCAAGTAGGCAAATATCATGATAGG

The nucleotide bases C and T are known as pyrimidine (aromatic heterocyclic organic compound) and A and G are known as purine (heterocyclic aromatic organic compound and It consists of two pyrimidine ring). The nucleotide A, G, C and T of a single strand connected with T, C, G and A respectively by hydrogen bonds. There are two and three hydrogen bonds between the base pair A, T and pair C, G shown in Figure 1b. Three adjacent nucleotides form a codon (Figure 2). This provides genetic code information for a particular amino acid. Different combination of four bases form 64 codons. The 64 codons and their corresponding amino acids shown in table I [19].

Fig.1: (a) Double helix structure of DNA. A small portion of sugar phosphate bond and A-T and G-C base pair are highlighted in box
Fig1: (b) The detailed structure of base pair of A, T, C, G.

Fig2: A model of mapping of three adjacent nucleotides into an amino acid.

At first, the authors are realized a model of each nucleotide A, G, C, T of DNA using MOSFET and cascading them to form an equivalent electrical model of codons. Designing the electrical model of codon, consideration is given on the number of benzene rings present in pyrimidine and purine structure, physiochemical property of hydrogen bond and sugar phosphate bond of each nucleotide base. The main objectives of the paper are:

1. Realized equivalent electrical model of 64 codons.
2. Study of the transient and frequency characteristic of 64 codon electrical circuit in Spice domain.
3. A new genetic code table is designed based on transient behavior of codon electrical model.
4. Analyze the effect of base position in codon based on frequency characteristic.

3. METHODS
For designing 64 codons electrical equivalent circuit using MOSFET, pyrimidine and purine are considered as a resistor because of their constant ring characteristic[11]. The net atomic charge in H bond produces a screening effect or atomic shielding which is viewed as a capacitor[15]. Intermolecular distance between two consecutive base pair is much shorter, so hopping mechanism is considered for charge transfer from one molecule to another. This hopping phenomenon is equivalently represented by diode [14]. The schematic is realized using Tanner S-Edit tool (version 16.0).

3.1 Design specification for equivalent resistors, capacitor and diode
For resistor modeling, values are taken 100 ohms and 200 ohms for pyrimidine (one ring) and purine (two ring) base respectively. In triode region the drain current (I_D) is linear function of drain to source voltage (V_DS), and in this region MOSFET behaves as a resistor [20,21], the equation for drain current is shown below:

$$I_D = \frac{W}{L} \frac{V_{DS}}{2} \left( V_{GS} - V_{TN} \right) \left( V_{DS} - \frac{V_{DS}^2}{2} \right) (1)$$

For triode region V_DS < (V_GS - V_TN) and V_GS ≥ V_TN.

The threshold voltage V_TN is calculated as

$$V_{TN} = V_{TH} + \gamma \left( \sqrt{2 \Phi_f} \right) + \sqrt{2 \Phi_e}$$

Where K_P, V_TO, γ, Φ, λ, Φ_f, W_eff, L_eff, and V_GS-V_TN are represent transconductance parameter, zero bias threshold voltage, body effect parameter, surface inversion potential (2θ_f), channel length modulation, oxide thickness, effective channel width, effective channel length, threshold voltage, and overdrive voltage respectively.

- W/L=15μ/0.60μ, V_GS=2.67volt for R_{OX} = 100 ohms
- W/L=15μ/0.60μ, V_GS=1.68volt for R_{OX} = 200 ohms.

Adenine (A) and Thymine (T) have two hydrogen bonds and Guanine (G) and Cytosine (C) have three hydrogen bonds, so 20fF and 30fF capacitance is considered respectively. To model bias-independent overlap capacitor we consider equations (3). The equation for drain current is shown below:

$$C_{MOS} = W \cdot L_{eff} \cdot C_{OX} + 2W \cdot L_{OV} \cdot C_{OX}$$

Where, V_{GO}> V_{TP} and L_{ov}, L_{D}=L_{ov} represent overlapping length, effective channel length respectively.

- W/L=5.1364μ/0.60μ for C_{MOS} = 20fF.
- W/L=15μ/0.60μ for C_{MOS} = 30fF.

In order to represent the charge hopping phenomena between two adjacent nucleotides, the authors short the gate and drain terminal which behaves like a diode. Based on the specifications, the equivalent electrical circuit of 64 codons are designed in the paper. Out of the 64 electrical equivalent circuits of codons only GTA is shown in figure 3.

4. RESULT AND DISCUSSION
Here a new genetic table is designed based on electrical responses of codon circuit. The electrical codon circuits are excited by an input voltage 900 mV and frequency 100 kHz. In Table II, the authors grouped 64 codons based on transient voltage and current response which correlates the ring structure of nucleotide. The effect of base position in codon is studied by observing frequency response of all the codon electrical circuit.

The authors observed all the codons belong to the group ‘2**’, show a stable nature in voltage as well as stable nature in current. But the codons belong to one group differ in voltage and current values from other groups. For example the triplets which are from the group ‘211’ (such as GCC, GCT, ACC, ACT, GTC, GTT, ATC, ATT) has uniformly same current and voltage values, but they are different from other groups i.e. ‘212’ (such as GCA, GGC etc), ‘221’ (such as GGC, GGT etc), ‘222’ (such as GG, GGA etc) (Figure 3a, b).
Fig 3: The electrical equivalent circuit of GTA.

Same phenomena is observed for the codon under the group ‘1**’ they exhibit same nature in electrical behavior. For example all the codons from the group ‘111’ (such as CCC, CCT, TCC, TTT, CTG, TTT, TTT, TTT) has uniformly same current or voltage values with respect to time, but they are different in value from other groups such as ‘112’ (such as CCG, CCA etc.), ‘121’ (such as CGG, CGT etc.), ‘122’ (such as CGG, CGA etc). All ‘1**’ groups show stable nature in voltage but oscillatory nature in current. (Figure 4c, 4d).

The transient voltage response show exponentially increasing in nature for codon whose first base is purine. Whereas the codon start with pyrimidine base exhibit exponentially decreasing voltage response.

Transient current response oscillatory in nature for codon begins with Pyrimidine base, whereas the codon start with purine base is stable. Therefore, transient characteristics of codon equivalent circuit can easily classify the total 64 codons into two groups based on transient voltage and current response. For voltage response either exponentially increasing or decreasing in nature and for current response either stable or unstable in nature.

Based on transient response of codon equivalent circuit, a new genetic table is proposed by authors which is depicted in Table II and III. The authors broadly divided genetic table into two group, one ring and two ringings, each group divided into four subgroups i.e. 11, 12, 21, 22 shown in Table II. The dimension of original genetic code table is 4×4 and based on electrical response which is transformed into 2×4 genetic code table.

The frequency domain response is analyzed to examine the effect of base position in codon. The codon which starts with pyrimidine are very weak in nature but purines are very strong base, it is also observed that the effect of third base position is negligible in codon. Figure 5 represents a comparisons between the group 111 and 112 (change in third base position).
Fig 4: (c), (d) Shows transient characteristic of 111 groups

TABLE I. The standard genetic code and corresponding amino acids

<table>
<thead>
<tr>
<th>Base</th>
<th>TTT</th>
<th>TTC</th>
<th>TTA</th>
<th>TTG</th>
<th>CTT</th>
<th>CTC</th>
<th>CTA</th>
<th>CTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>Phe</td>
<td>Val</td>
<td>Leu</td>
<td>Cys</td>
<td>Leu</td>
<td>Pro</td>
<td>Leu</td>
<td>Arg</td>
</tr>
<tr>
<td>C</td>
<td>TAT</td>
<td>TAC</td>
<td>TAA</td>
<td>TAG</td>
<td>CAT</td>
<td>CAC</td>
<td>CCA</td>
<td>CCG</td>
</tr>
<tr>
<td>A</td>
<td>Tyr</td>
<td>Thr</td>
<td>Stop</td>
<td>Stop</td>
<td>His</td>
<td>Gln</td>
<td>Gln</td>
<td>GGG</td>
</tr>
<tr>
<td>G</td>
<td>TGT</td>
<td>TGC</td>
<td>TGA</td>
<td>TGG</td>
<td>GTG</td>
<td>GTC</td>
<td>GTG</td>
<td>GTG</td>
</tr>
</tbody>
</table>

Table 2. New genetic code table based on transient response of electrical equivalent circuits of codons

<table>
<thead>
<tr>
<th>The number of ring present in 2nd and 3rd base position</th>
<th>11</th>
<th>12</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Number of ring present in 1st base position</td>
<td>111</td>
<td>112</td>
<td>121</td>
<td>122</td>
</tr>
<tr>
<td>1</td>
<td>211</td>
<td>212</td>
<td>221</td>
<td>222</td>
</tr>
</tbody>
</table>
Table 3. Illustrated view of proposed table

<table>
<thead>
<tr>
<th>Group</th>
<th>CC(G/A)</th>
<th>TC(G/A)</th>
<th>CT(G/A)</th>
<th>TT(G/A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>122</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>211</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>212</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>222</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig 5: The average amplitude and phase response for the codons of the group 111 and 112.

Fig 6: The average amplitude and phase response for the codons of the group 121 and 121.

Fig 7: The average amplitude and phase response for the codons of the group 222 and 221.
where the change in amplitude and phase is very negligible. But the change is very significant for the groups 111 and 121, shown in figure 6, because the second base position is varying.

But in Figure 7 and 8, no significant change occurs because the codon start with strong base purine, therefore change in 2nd and 3rd base have a negligible effect on codon characteristic.

5. CONCLUSION
A novel approach is considered in the present paper to design a new genetic code table based on the electrical response of codon. The effect of base position in codon is investigated and analyzed using frequency response of codon electrical circuit. It is a new concept which may further be extended in amino acids classification and protein characterization. The work is significant for biologists who are struggling to produce proteins for drugs, biofuel and more. The work may help the biotechnologist to design devices that could be implanted in human body and programmed with genetic code. The results of this study are important in forensics, disease diagnosis, protein identification, etc.

6. ACKNOWLEDGMENTS
The authors wish to thank DST, Science and Engineering Research Board (SERB/F/4504/2013-2014), Govt. of India for funding support of research work.

7. REFERENCE


