Enhancement of Exon Regions Recognition in Gene Sequences Using a Radix -4 Multi-valued Logic with DSP Approach

D. Venkat Reddy, D. V. Paradesi Rao And E.G.Rajan

Abstract-**Numerous levels of concepts perform logical designand logical representations in an efficient manner. In typical and quantum theories of computation, Binary logic and Boolean algebra occupies an imperative place. But they havethe limitation of representing signals or sequences by using either binary '1' or '0'. This has major drawbacks that the neutralities or any intermediate values are ignored which are essential in most of the applications. Because of the occurrence of such situations it is the need of the hour to look into other alternative logics in order to fulfill the necessities of the user in their respective applications. The binary logic can be replaced by Multi-Valued Logic (MVL), which grabs the positions of the major applications because of the ability to provide representation by using more than two values. As most of the significant applications are based on the logical sequences, the multi-valued logic shines because of its thriving feature. Genomic signal processing, a novel research area in bioin-formatics, is one of the foremost applications which involve the operations of logical sequences. It is concerned with the digital signal representations and analysis of genomic data.Determination of the coding region in DNA sequence is one of the genomic op-erations. This leads to the identification of the characteristics of the gene which in turn finds out an individual's behavior. In order to extract the coding regions on the basis of logical sequences a number of techniques have been proposed by researchers. But most of the works utilized binary logic, which lead to the problem of losing some of the coding regions and incorrectly recognizing non-coding regions as the coding regions. Hereby, we are proposing an approach for recognizing the exon regions from a gene sequence based on the multi-valued logic. In this approach, we have utilized fourlevel logical system, termed as quaternary logic for the representation of gene sequences and so that we recognize theexon regions from the DNA sequence.**

Keywords-Multi-valued logic (MVL), Quaternary logic, spectral component, gene sequence, exon regions, exon regions recognition

I. INTRODUCTION

Sundry number systems have been developed rightfrom the evolution of computers. These systems concentrate on simplifying the basic mathematical operations so as to assist in making the computer more powerful [6]. Complex mathematical operations in computing systems are carried out in terms of logical operations that make the computation easier. For any signal xin the circuit, we indicate its logic values before and after a clock transition as $x(t)$ and $x(t0)$

About-Associate   Professor   Department   of   E.C.E
venkatreddyphd@gmail.com

respectively [3]. Logic design is carried out in an effective way at a variety of levels of abstraction. In general it starts with a "word" description of the design problem, which is later represented as a truth table from which a logic function is formulated. The majority of the logic design has been on the basis of the binary logic because of its intuitive relationship to the binary states of electronic switches, ON and OFF or abstractly 0 and 1. The similar mind set has been carried over to logic design by means of bio-molecules in spite of them being free from this restriction [2]. On the basis of the number of states of the technology (components) used [2] the type of logic used will depend [2]. Digital circuit design has traditionally been linked with binary logic where the two logic levels are represented by two discrete values of current, voltage or charge [5]. Due to the intuitive relationship to the binary states of electronic switches, ON and OFF or abstractly 0 and 1[2] a large amount of the logic design are based on the binary logic.Binary logic and Boolean algebra plays a vital role in typical and quantum theories of computation. But, in Binary logic there is a possibility for only two outputs which denotes either a true condition (1) or a false condition (0). When a neutral or an intermediate value has to be represented as an output in the application then it is referred to a partially correct value either 1 or 0. It is that binary will submit a value or in any way fade and die i.e., the neutralities are ignored in this logic. These types of situations have raised the alarm to search for other alternative logics in order to satisfy the user requirement in their respective applications. The MVL is considered to be a potential substitute to Binary logic [9]. For the last couple of decades, multiple-valued logic (MVL) has attracted significant attention, particularly among circuit and system designers. MVL circuits allow more than two levels of logic and depending on the number of allowed levels, we may have ternary (base = 3) or quaternary (base =4) logic styles. MVL seek out to enhance the information processing efficiency of a circuit by transmitting more information on each signal line than simple binary logic and by implementing complex functions of the inputs in a single gate [4]. MVL circuits can minimize the number of operations necessary to implement a specific mathematical function and further, have an advantage in terms of reduced area [5], which in turn reduces parasitics linked with routing

and offers a higher speed of operation [1]. Compared with the conventional binary logic, the information density in MVL is much higher [1]. Some of the recent research works related to the MVL is given as follows. Yi Jin et al. [11] have proposed a theory known to be the Decrease-Radix Design. And on the basis of their theory, the regulations of building multi-valued logic operation units were offered. The theory laid down a solid foundation forthe design of re-constructible logic units in ternary optical computers in addition to any other multi-valued computers. The key of the theory was that if the physical states to represent information incorporated a special state "D", then any of the $n(n \times n)$ n-valued LUs could be realized by means of the combination of the $n \times n \times (n-1)$ operationbasic-units (OBU) according to the DRD theory, where "D" is a special physical state named by the authors. It would result in A while the state D operated with any state A. M.H. Nodine et al. [12] have provided an overview of methods proposed in order to implement multi-valued logic in CMOS and then portrayed Intrinsity's patented Fast reg Technology as a mature methodology for silicon implementation of multi-valued logic. Fast Technology was on the basis of three fundamental characteristics comprising the utilization of (1) footed NMOS transistor domino logic, (2) multi-phased overlapping clocks, and (3) 1-of-N encoding of MVL signals. In order to offer additional opportunities for power optimization, the concepts of null value as well as mutex properties were introduced, presenting additional challenges for MVL representation as well as synthesis. A high-level design method of multiple-valued arithmetic circuits is proposed by Y. Watanabe et al. [13]. Their method utilized a cell-based approach with a dedicated hardware description language called ARITH. By means of ARITH, they could portray and verify any binary/multiplevalued arithmetic circuits in a formal manner. The ARITH description may perhaps be transformed into a technologydependent netlist in binary/multiple-valued fused logic. The process of transforming the netlist into a physical layout pattern was automatically carried out by an off-theshelfplace-and-route tool. They presented a particular cell library comprising a multiple-valued signed-digit adder and its related circuits with a 0.35mum CMOS technology, and illustrated that their method possibly will synthesize a 32times 32-bit parallel multiplier in multiple-valued currentmode logic from an ARITH description. M.H.A. Khan et al. [14] have offered a heuristic algorithm for concurrent variable ordering as well as quaternary Galois field expansion selection in order to build optimal quaternary Galois field decision diagram (QGFDD). They also give you an idea about way of flattening the QGFDD in order to generate QGFSOP expression. They have written Java program for the purpose of building QGFDD for multioutput quaternary functions and offered experimental results. Navi et al. [9] have offered two versions of current mode 1-bit adder. The related propagation delays of them are 70ps, 150ps respectively. Those adders made use of only 6 transistors. The chip density advantage of the multi-valued approach

was significant. They carried out simulations through the HSPICE in a 0.18μm technology at 27 centigrade; with 1.8 volt supply voltage. The simulation results demonstrated that they attained a remarkable improvement in terms of transistor count, chip area as well as propagation delay. The minimum number of transistors reported in current mode was 11 and it is 10 in voltage mode. That is they accomplished 40% performance in terms of transistor count and the improvement in speed is about 2.5%. B.J. Falkowski et al. [10] have offered classification of novel fastest quaternary linearly independent transforms. They were defined recursively and also had consistent formulas linking their forward and inverse transform matrices. Their transform matrices' properties and calculation example were revealed. The computational costs of the calculation for offered transforms as well were discussed. The experimental results were made known and compared with the well known quaternary arithmetic transform.The architecture of Chinese abacus adder is offered by Shu-Chung Yi et al. [15]. As high radix of adder might minimize the number of carry propagation, their Chinese abacus adder might achieve high-speed operation. The simulation results of their works were compared with CLA (Carry Lookahead) adder. The delay of the 8-bit abacus adders are 22%, 17%, and 14% less than those of CLA adders for 0.35μm, 0.25μm, and 0.18μm technologies, respectively. The power consumption of the abacus adders were 30%,34%, and 60% less than those of CLA adders for 0.35μm, 0.25μm, and 0.18μm technologies, respectively. The utilization of Chinese abacus approach resulted in a competitive technique with regard to conventional fast adder. Luis E. Cordova et al. [16] have presented an approach in order to generate molecular electronics systems by means of introducing a multi-valued programmable logic (MVPL) block for the purpose of modeling the characteristics lately observed experimentally in few molecular structures so as to achieve circuit robustness on a molecular substrate. They demonstrated that given the experimentally observed characteristics of few types of molecular electronics substrates, they possibly will adopt a multi-valued logic system for the logic blocks, by this means partially avoiding the necessity to incorporate specialized fault detection or fault tolerant circuitry at the molecular level that would be essential in binary logic circuits. They evaluated their MVPL model against the other work presented in the literature with respect to the efficiency of the formulation in order to understand computing architectures in the face of high defects in the self-assembled substrate. MVL has a number of advantages when compared to the existing binary systems. Improving binary logic levels to ternary, penta and quad levels, more advanced processing values are obtained in a variety of computing applications. Genomic signal processing is one of the most noteworthy application areas where MVL can be implemented to overcome drawbacks faced while computed with binary logic. A major challenge for genomic research is to establish a relationship among sequences structures and functions of genes. DNA sequence

is an emblematic string of letters 'A', 'C', 'G' and 'T'. The segments of DNA molecule, called genes are recognized to carry valuable information in their protein coding regions (exons) and are responsible for protein synthesis. In eukaryotes, exon regions are segmented by non-coding regions (introns), while in prokaryotes these regions are continuous [7]. The pivotal problem of gene identification in eukaryotes is distinguishing exons, from introns and intergenic regions [8]. Enormous selections of techniques are employed in separating the sequence from the DNA, until now binary logic is being utilized more normally to extract thissequence. In the mean time, while computing with binary logic the sequence obtained may contain noises impregnated with them which make it difficult to distinguish between the exons and introns. Even after the noise being removed, the coding regions are only recognized with a stress. Also when the magnitude of exon is improved in order to enhance the strength of the signal it will alternately increase the magnitude of the intron which also causes a drawback in the identification of the exons from the introns. Hereby, we are proposing an approach which overcomes the mentioned drawbacks in identifying the protein coding regions based on binary logic. The proposed approach utilizes quaternary logic, radix-4 multi-valued logic which replaces the binary logic. Utilizing the four-level logic in the conventional technique of coding region identification, we obtain the exons more clearly. In order to replace the binary by quaternary values, we are proposing a simple, but more effectual logical conversion technique which converts the binary represented gene sequence into quaternary indicators. Hence, the intention of extracting the information region from any of the gene sequence can be achieved by our proposed quaternary level based coding region identification. The rest of the paper is organized as follows: Section 2 gives a brief introduction about the multi-valued logic and the quaternary logic which is utilized in our approach and Section 3 briefs some fundamental ideas about the genomic signal processing. Section 4 succinct the conventional binary logic based DSPmethod for exon regions identification and Section 5 details the proposed quaternary logic based DSP method with sufficient formulations and illustrations. Section 6 discusses about the implementation results and section 7 concludes the paper.

II.   MULTI-VALUED LOGIC (MVL)

Traditional calculi are merely two valued for any proposition. MVL are logical calculi wherein there are above two truth values [22]. Numerous forward-thinking efforts dedicated to the MVL synthesis have been made in recent years, in specific, however an effective methodology for MVL design is until now an open challenge [19]. Being able to minimize the number of interconnection lines or nets and enhance their information content, MVL turn out to be quite attractive [17]. In multi-valued logic, the connectives as well as the rules for building formula are those utilized in classical logic, and the disjunction, conjunction and negation of formulas are defined by max, min operations in addition to the complementation to 1, respectively [18]. There are numerous advantages of such logic systems as well as circuits when compared with the binary ones. The major advantages of MV digital and computer systems are: enhanced speed of arithmetic operations realization, superior density of memorized information, improved usage of transmission paths, minimizing of interconnections complexity and interconnections area, lessening of pin number of integrated circuits and printed boards, possibilities for easier testing [20]. One can attain a more cost-effective way of exploiting interconnections by means of a larger set of signals over the same area in MVL devices, permitting easy implementation of circuits. In MVL devices, the noise advantage of binary logic is preserved. The higher radix in use is the ternary (radix-3) as well as the quaternary (radix-4) [21].

1) Quaternary logic

Quaternary logic is very appropriate for encoded realization of binary logic functions by grouping 2-bits together into quaternary digits. This sort of quaternary encoded reversible realization of binary logic function necessitates half times input/output lines than the original binary reversible realization. As the number of input/output lines is minimized, this quaternary encoded realization of binary logic function causes the circuit more compact and manageable, in particular for the quantum technology, where the cost of qudit (quantum digit) realization and qubit (quantum bit) realization are almost similar [23]. Quaternary identity logic circuit works as a buffer and is utilized in designing quaternary D flip-flop. Quaternary identity logic circuit is composed of thermometer code circuit, EXOR gate, 2 bias inverters, Ntype transmission gate, and P-type transmission gate [24]. Quaternary logic is of more interest when compared to other types of multi-valued logic because of the simplicity of signal grouping by two bits [19].Quaternary Signed Digit numbers are represented by means of 3-bit 2's complement notation. Every number can be represented by

$$X = \sum_{i}^{n} x_i 4^i$$

where xi can be any value from the set {3, 2,1,0,1,2,3} in order to produce an apt decimal representation [25]. The quaternary logic makes use of 0, 1, 2 and 3 logic levels. Figure 1 illustrates quaternary logic levels
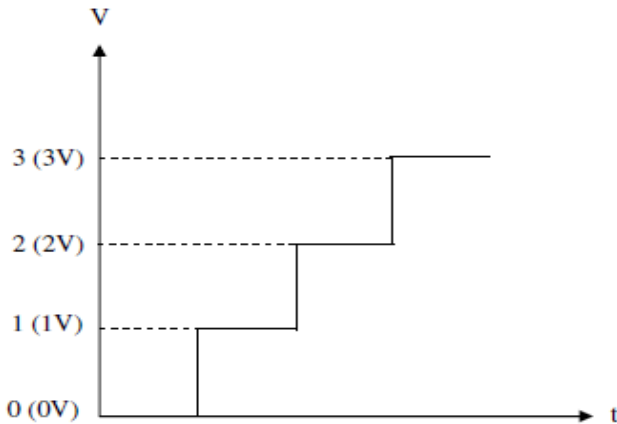
Fig. 1: 3V supply level for quaternary logic representationWhen the binary logic is used to represent any number, only

$2i$ combinations are possible (when i number of bits is used for representation). But in Quaternary logic, the bit level representations are up to i 4 combinations. While applying the value of $i = 3$, the total no of combinations that can be represented is 64. This makes the quaternary logic more strong and advantageous when applied for signal or sequence representations. Taking all the features into concern, we have employed the logic for gene sequence representation in the application of identification of protein coding regions.

III. GENOMIC SIGNAL PROCESSING

With the vast amount of genomic and proteomic data that is available in the public domain, it is becoming more and more significant to be able to process this information in ways that are helpful to humankind. In this context, raditional as well as modern signal processing methods have played a significant role in these fields [26]. Gene identification is one of the most vital tasks in the study of genomes [31]. The engineering discipline that studies the processing of genomic signals is broadly classified as Genomic signal processing (GSP) [30]. The definition of the Genomic Signal Processing can be given as the analysis, processing and the utilization of genomic signals in order to obtain biological knowledge as well as with that knowledge system-based application are devised. Due to the most important role played in genomics by transcriptional signaling and the related pathway modeling, it is only natural that the theory of signal processing be supposed to be utilized in both structural and functional understanding.

The aim of GSP is to combine the theory and methods of signal processing with the global understanding of functional genomics, with special importance on genomic regulation [27]. DNA is the building block of all life on this planet, from single cell microscopic bacteria to more advanced creatures like humans [29]. DNA topology is of primary importance for an extensive range of biological processes. Because of the topological state of genomic DNA is of importance for its replication, recombination and transcription, there is an immediate interest to acquire

information regarding the supercoiled state from sequence periodicities. Identification of dominant periodicities in DNA sequence will help understand the significant role of coherent structures in genome sequence organization [30]. Genome sequencing is figuring out the order of DNA nucleotides, or bases, in a genome that frame an organism's DNA. Genome sequences are big in size and are capable of ranging from several million base pairs in prokaryotes to billions of base pairs in eukaryotes [29]. The DNA is made up of long sequences of four kinds of nitrogen containing bases {A, C, G, T}. These sequences are grouped in coding regions – exons (of the eukaryotic genes) and non-coding regions – introns (a variety of regulatory regions such as promoters, enhancers, silencers, long repeats with apparently no function, etc.). The coding regions are translated into proteins, whereas the immense majority of the non-coding regions appear to have no biological function whatsoever [28]. Accurate prediction of exon regions is a research problem at present being addressed [7]. So as to be successful, a gene finding algorithm has to incorporate good indices for the protein coding regions [31]. Additionally processing and analyzing this information are of major importance. The volume of genomic data is expanding at an enormous as well as still growing rate, while its basic properties and relationships are not so far fully understood and are subject to continuous revision. This data is stored, managed, and analyzed on a huge diversity of computing systems, from small personal computers which makes uses many disk files to supercomputers operating on large commercial databases [32].

IV. CONVENTIONAL DSP APPROACH FOR EXO REGIONS RECOGNITION

As we have talked about previously, there are four nucleotides (or bases) comprised in the strands of DNA. They are designated by the characters A, T, C,and G. A haracter string composes of these four bases. And such a character string can be mapped to four signals. The conventional DSP methods for coding region identification utilize the binary signals to represent the signals. To be brief, the string which is composed of four bases is mapped into four binary signals. The value of „1' is taken by the signal bA(n) in the case if A is present in the DNA sequence at index n . But if it is not the case that is if A is absent at index n the value of „0' is taken. For instance, bA(n) for the DNA segment „CGTCGTGGAA' is given as 0000000011. In the same manner the signals bT (n) ,bG (n) and bC (n) can be acquired. After that the DFT of bA(n) , BA(f) over W samples is found. . In the same manner it is possible to obtain the DFT of bT (n) ,bG (n) and bC (n) , termed as BT (f) , BG (f) and BC (f) respectively. Period-three behavior is noticed in several genes and it is also found that is very much helpful in recognizing the coding regions [33]. In addition, several researchers have observed that the period-3 property to be a good (preliminary) indicator of gene location [26]. For this reason, the (f = N / 3) –DFT coefficient magnitude is

frequently considerably larger when compared to the surrounding DFT coefficient magnitudes. And this corresponds to a coding region inside the gene. Based upon the gene [33] this effect differs and be

able to be fairly pronounced or fairly weak. A figure that can be utilized in order to measure the total spectral content S(f) of a DNA character string at frequency f is defined as the sum of the magnitude of the DFT values of the four binary nucleotide sequences. Observe that a calculation of the DT at the single point f = N / 3 is adequate. The window can after that be slid by one or more bases and S(N / 3) recalculated. Therefore, we obtain a picture of how S(N / 3) evolves along the length of the DNA sequence. It is essential that the window length W be adequately large (typical window sizes are a few hundreds, e.g., 351, to a few thousands). On the other hand a long window implies longer computation time, and in addition compromises the base-domain resolution in predicting the exon location [26]. On the other hand the non-coding regions in the DNA spectrum at3 2p are not wholly suppressed by the conventional DSP.

Therefore, a non-coding region may be mistakenly recognized as a coding region [34]. To overcome these shortcomings, we have replaced the binary logic by MVL for mapping the sequences in the proposed approach and so that exact coding regions are identified effectively.

V.   PROPOSED QUATERNARY LOGIC BASED EFFECTIVE DSP APPROACH

In this paper, MVL based processing of genomic signals in order to recognize the exon regions are proposed. Here, MVL used is four-level logic i.e. quaternary logic for identifying the coding regions in a gene sequence. The steps involved in the quaternary logic based coding region identification are depicted in the figure 2. The approach replaces the binary logic by quaternary logic by means of a conversion technique. Thus the obtained quaternary indicators are subjected for the further process of exon identification in the gene sequence.
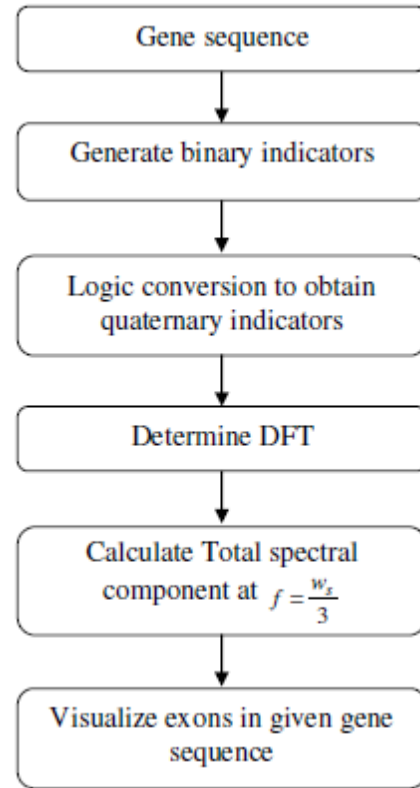

Fig 2: Proposed Quaternary logic based DSP approach for exon regions recognition

As an initial process which is depicted in the figure 2, we are utilizing the binary logic representation of the gene sequences. Let the sequence be gs , which is the combination of nucleotide bases A (Adenine), G (Guanine), C (Cytosine) and T (Thiamine) having the length of l . The binary indication of the any such sequences is given as

$$b_T(n) = \begin{cases} 1; & if \ g_s(n) = 'T' \\ 0; & else \end{cases} \qquad (1.b)$$

$$b_A(n) = \begin{cases} 1; & if \ g_s(n) = 'A' \\ 0; & else \end{cases} \qquad (1.a)$$

$$b_G(n) = \begin{cases} 1; & if \ g_s(n) = 'G' \\ 0; & else \end{cases} \qquad (1.c)$$

$$b_C(n) = \begin{cases} 1; & if \ g_s(n) = 'C' \\ 0; & else \end{cases} \qquad (1.d)$$

In equation (1) ,b_A (n), b_T (n), b_G (n) and b_C (n) are the binary indications of the DNA sequence g_s representing the nucleotide bases 'A', 'T', 'G' respectively where j= 1,2,3…l Then the binary indicators are applied to a logical conversion which converts

the binary indicators into quaternary indicators. Logic conversion: The conversion technique, heart of the proposed approach presented here converts the binary indicators into quaternary indicators. The conversion procedure takes a few steps which are given as follows. Initially the indices of the binary indicators are obtained as

$[A_P] \ll \Delta\, b_A\,(n)\,;\quad$ if $\quad b_A\,(n) = 1 \qquad\qquad$ (2.a)

$[A'_P] \ll \Delta\, b_A\,(n)\,;\quad$ if $\quad b_A\,(n) = 0 \qquad\qquad$ (2.b)

In equation (2), A_P and A'_P are the index vectors of $b_A\,(n)$ which have the index values of logical '1's and '0's respectively. Then the four different logical levels are assigned to each of the binary values constituted by the indicators as

$$A_V = \begin{cases} V_1\,; & if\ \ A_P(n)\,\%\,2 = 0 \\ V_3\,; & else \end{cases} \qquad (3.a)$$

$$A'_V = \begin{cases} V_0\,; & if\ \ A'_P(n)\,\%\,2 = 0 \\ V_2\,; & else \end{cases} \qquad (3.b)$$

Equation (3) represents the conversion of the binary indicator values to the quaternary logical values. V_0, V_1, V_2 And V_3 are the four logical levels utilized in our approach which directly represents the values 0,1,2 and 3 respectively. $A_V\,(x)$ and $A'_V\,(x)$ are the vectors which have he quaternary values of logical '1' and logical '0' of the binary indicator respectively. As per the index taken from the binary indicator, the values should be concatenated in order to get the final quaternary indicators of the DNA sequences. This can be performed as

$$A_1(n) = \begin{cases} A_V(x)\,; & if\ \ A_P(x) = n \\ 0 & ;\ \ else \end{cases} \qquad (4.a)$$

$$A_2(n) = \begin{cases} A'_V(x)\,; & if\ \ A'_P(x) = n \\ 0 & ;\ \ else \end{cases} \qquad (4.b)$$

In equation (4), if $A_P\,(x) = n$ and $A'_P\,(x)$ n are not satisfied, then the value of x is not incremented. It remains until the condition gets satisfied. Then by performing the addition operation between the vectors A1 and A2 , the final quaternary indicator for the nucleotide base 'A' is obtained. This can be given as

qA (n) = A1 + A2 $\qquad\qquad\qquad$ (5)

where, qA(n) is the quaternary indicator for the nucleotide base which is converted from the corresponding binary indicator . In similar fashion, the quaternary indicators for other nucleotide bases qT (n) , qG (n) and qC (n) are obtained. Henceforth, we use the quaternary indicators for the purpose of the identifying the protein coding region. After obtaining the four quaternary indicators, the spectral content is calculated. The spectral content calculation and the further process of coding region identification are performed for a window of sequences, which is sliding in nature. Let, the window performs a single sliding movement

having the size of ws . Then the total number of windows used in our approach is given by

$n_{W =}$ $\,l\,$ --- $Ws + 1$ $\qquad\qquad\qquad$ (6)

where, nw is the total number of windows as we have chosen single sliding movement and ws as window size. For each further process, the sequence covering by the window is applied. For spectral component calculation, we have to determine the DFT for $q_A^{(z)}(m)$ the quaternary indicator sequence at z^{th} window which can be given as

$$Q_A^{(z)}(f) = \sum_{m=0}^{w_s-1} q_A^{(z)}(m)\exp\left(-j\frac{2\pi f}{w_s}m\right);$$
$$0 \le f \le w_s - 1 \qquad (7)$$

Here, $q_A^{(z)}(m)$ represents the quaternary indicator sequence covered by the window z of size W_s, $1 \le z \le n_w$. Hence the DFT of the quaternary indicator, $Q_A^{(z)}(f)$ representing the base 'A' is determined for all the windowed sequences which are sliding. Likely, the DFT will be determined for all the other quaternary indicators, $q_T^{(z)}(n)$ AND $q_G^{(z)}(n)$ so that we obtain $q_C^{(z)}(n)$

$Q_G^{(z)}(f)$ And $Q_C^{(z)}(f)$ Then the total spectral content of the sequence at certain frequency f is given as

$$S_z(f) = |Q_A^{(z)}(f)|^2 + |Q_T^{(z)}(f)|^2 +$$
$$|Q_G^{(z)}(f)|^2 + |Q_C^{(z)}(f)|^2 \qquad (8)$$

The spectral content, hereby, calculated using the equation (8) considers all the nucleotide bases of DNA sequences. Then, with the aid of the period-3 behavior, the exons are identified from the spectral content. This is due to the fact that the distribution of bases in the exons, which are the integer multiple of 3 exhibit the period-3 behavior. The reason of the period-3 behavior is because of the presence of 3-nucleotide code structure in protein coding region. The period-3 behavior of coding protein region refers to the maximum of Fourier power spectrum (FPS) at the position Of 1/3 frequency [35]. Therefore, it can be decided that the Value of $S_Z(f)$ is maximum at $f = w_s / 3$ while a coding region is there. As a result, after obtaining the total spectral content for all the gene sequences based on the window size, the period-3 behavior is taken. The peaking magnitudes appearing in the total spectral components $S_Z(f)$ give the exon regions. Hence , by the approach we can easily identify most of the exon regions occupying in the gene sequences mainly because of the thriving contribution of the quaternary logic. As the usage of quaternary logic makes the exon regions more dominating in the spectrum rather than the usage of binary logic, the exon

regions are not affected by any of the disturbances. However, the proposed approach struggles in identifying the exon which does not exhibits period-3 behavior. Though, the approach has this drawback, it is very strong in identifying the exact exon regions instead of mistakenly taking the introns because of the effective work of quaternary logic.

VI.   RESULTS AND DISCUSSION

We have implemented the proposed quaternary logic based coding region recognition in the working platform of MATLAB (version 7.8) from which we have visualized the performance of the approach. For performance visualization, we have utilized DNA sequences of two different organisms, namely, Brucella Suis and Caenorhabditis elegans (C. elegans). As discussed earlier, we have replaced
the binary logic by quaternary logic in the conventional DSP
method of coding region recognition from the mentioned gene sequences. The size of the sliding window, we have chosen for taking the sample sequence is 351 (i.e.ws =351). Then, as per the approach procedure, the spectral components are determined for each window of sequences. As the gene sequences are very huge, it takes too much time to recognize all the coding regions throughout the sequences. Hence we have taken two different samples of sequences and so we have recognized the exon regions within the given sequences. The spectral component for the two different samples of sequences, 5000 and 20,000 for the gene Brucella suis is shown in the figure 3(a) and figure 3(b)
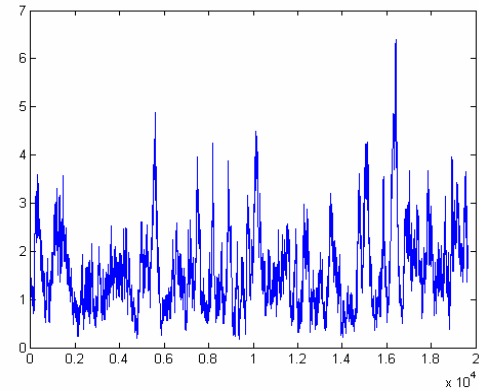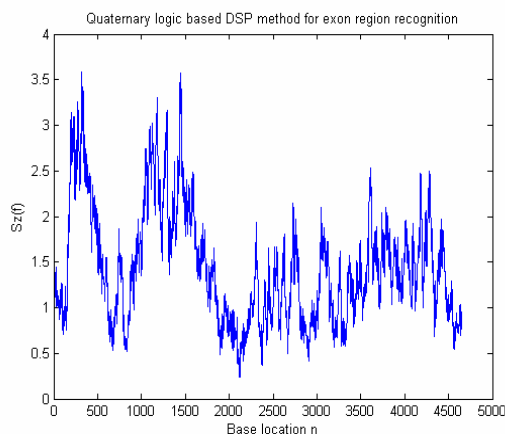respectively.

Fig 3: Normalized spectral content of the gene sequence Brucella suis for samples (a) 5000 and (b) 20,000

From the figure 3, the exon regions of the gene Brucella suis
has been visualized clearly. As discussed earlier, the exon regions are nothing but the peaking magnitudes of the spectral content of the gene sequences. It is clear that the proposed approach increases the magnitude of the exon regions rather than the intron regions and so we have obtained all the exon regions which exhibits period-3 behavior. Thus the proposed approach overcomes the major drawback of incapability of recognizing the exons in binary logic based coding region recognition. We have tested the approach using the chromosome III of C. elegans also for seeking different results. The spectral results obtained for C. elegans chromosome III is illustrated in the figure 4.

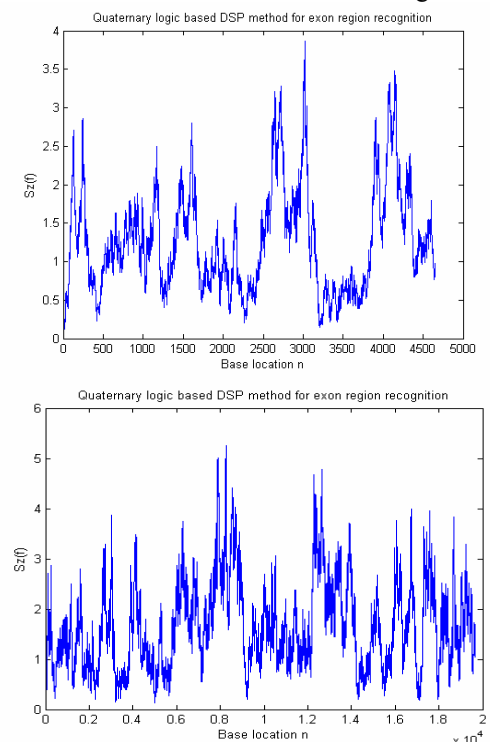Fig 4: Normalized spectral content of the gene sequence C. Elegans chromo some III for samples (a) 5000 and (b)20,000

In similar fashion, we have obtained the spectral component regions of the C. elegans gene sequence which is constituted by the exon as well as the intron regions. The regions at which the magnitude of the sequence peaks have been recognized as the exon regions exhibiting period-3 behavior.

Figure 4 (a) gives the exon regions of the sequence for a sample of 5000 and figure 4 (b) gives the exon regions for a sample of 20,000. A comparison is made between both the binary and the proposed quaternary based DSP method for exon regions recognition for the two gene sequences, Brucella suis and C. elegans. The comparison plot of spectral component distribution between the conventional and the proposed approach has been given in the figure 5.
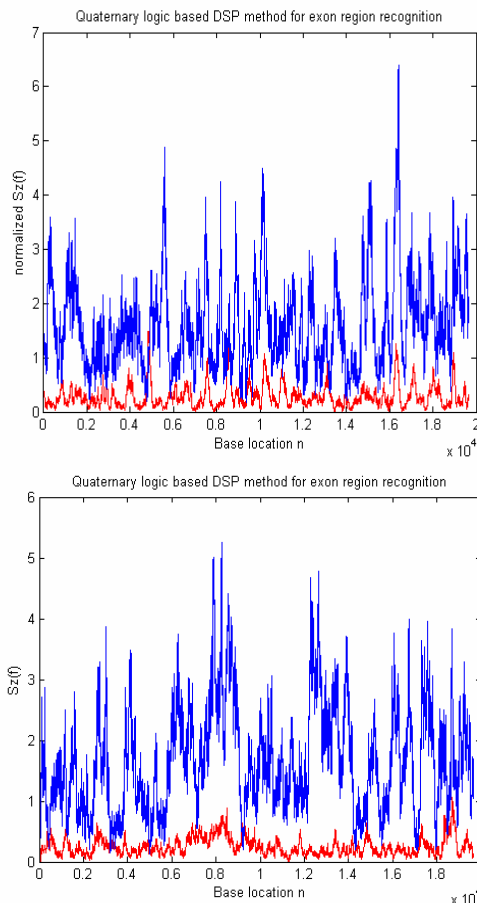

Fig 5: A comparison between the spectral components obtained from the conventional binary logic based DSP method and quaternary logic based DSP method for coding region recognition in the gene (i) Brucella suis and (ii) C. Elegans chromosome III .

The comparison provided between the conventional binary based DSP method and the proposed quaternary based DSP method clearly illustrates the performance difference in recognizing the exon regions. Conventional method shows only a very small spike in the exon regions, but the proposed approach makes a huge peak in the same. This makes clear that the proposed approach is more effective in recognizing the exon regions mainly because of the performance of the quaternary logic.

VII.   CONCLUSION

Being an alternative to binary logic, in this paper, we have proposed a radix-4 MVL based approach for exact identification of exons from the gene sequence. With the aid of the fundamental DSP technique, we have developed the quaternary logic based DSP approach for exon regions recognition in gene sequences. The proposed approach just replaces the binary logic by quaternary logic in the conventional DSP method to identify the coding region in DNA sequence. Because of the utilization of quaternary logic in DSP method of coding region recognition, the magnitude of the coding regions has been increased heavily. This makes the identification of coding region from the gene sequences more comfortable. It has been well known that the coding region exhibits period-3 behavior and so it peaks when the gene sequences are applied for spectral content calculation. The approach has performed more effectively that it identified the exact exon regions and restricted the introns from domination. This makes the incorrect decision of taking the introns as exons have been mitigated. Hence, the proposed quaternary logic based DSP approach for recognizing the coding region in DNA sequence is more effective rather than binary logic based approach and so we can identify the exact coding regions and not no-coding regions.

VIII.   REFERENCES

1) Wu Gang, Cai Li, and Li Qin, "Ternary logic circuitdesign based on single electron transistors", Journal of Semiconductors, Vol 30, Issue 2, February 2009.

2) H.A. Aleem and F. Mavituna, "Multiple Valued Logic as a Design Tool for Synthetic Biology",University of Manchester, UK.

3) Xunweiwu, Bangyuan Chen and Massoud Pedram, "Power Estimation in Binary CMOS Circuits Based on Multiple-valued Logic", Multi. Val. Logic., Vol 00, pp. 1-17, 2000.

4) H. L. Chan, S. Mohan, Pinaki Mazumder, and George I. Haddad, "Compact Multiple-valued Multiplexers Using Negative Differential Resistance Devices",IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 31, NO. 8, AUGUST 1996.

5) Arijit Raychowdhury and Kaushik Roy, "A NovelMultiple-Valued Logic Design Using Ballistic Carbon Nanotube FETs", Proceedings of 34th International Symposium on Multiple-Valued Logic, pp: 14-19, 2004.

6) Mehdi Hosseinzadeh, Somayyeh Jafarali Jassbi, and Keivan Navi, "A Novel Multiple Valued Logic OHRNS Modulo rn Adder Circuit", International Journal of Electronics, Circuits and Systems, 2007.

7) Mahmood Akhtar, "Comparison of Gene and Exon

Prediction Techniques for Detection of Short CodingRegions", International Journal of Information Technology, Vol. 11, Issue 8, pp: 26-35, 2005.

8) Achuthsankar S. Nair and Sivarama PillaiSreenadhan, "A coding measure scheme employingelectron-ion interaction pseudopotential (EIIP)",Bioinformation, Vol 1 Issue 6, pp: 197–202, 2006.

9) Navi, Foroutan, Mazloomnejad, Bahrololoumi,Hashemipour and Haghparast, "A Six Transistors Full Adder", World Applied Sciences Journal, Vol.4,No.1, pp.142-149, 2008.

10) B.J. Falkowski, Cheng Fu, "Classification of Fastest Quaternary Linearly Independent Arithmetic Transforms," Proceedings of 38th International Symposium on Multiple Valued Logic, pp. 169-173,22-24 May, Dallas, TX, 2008.

11) Yi Jin, Jun-Yong Yan and Kai-zhong Zuo, "Decrease-radix design principle for carrying/borrowing free multi-valued and application in ternary optical computer," Science in China SeriesF: Information Sciences, Vol.51, No.10, pp. 1415-1426, October 2008.

12) M.H. Nodine and C.M. Files, "A MatureMethodology for Implementing Multi-Valued Logicin Silicon," Proceedings of 38th InternationalSymposium on Valued Logic, pp. 2 - 7, 22-24 May,2008.

13) Y. Watanabe, N. Homma, K. Degawa, T. Aoki and T.Higuchi, "High-Level Design of Multiple-ValuedArithmetic Circuits Based on Arithmetic DescriptionLanguage," Proceedings of 38th InternationalSymposium on Multiple Valued Logic, pp.112-117,22-24 May, Dallas, TX, 2008.

14) M.H.A. Khan, N.K. Siddika and M.A. Perkowski,"Minimization of Quaternary Galois Field Sum ofProducts Expression for Multi-Output QuaternaryLogic Function Using Quaternary Galois FieldDecision Diagram," Proceedings of 38th InternationalSymposium on Multiple Valued Logic, pp. 125-130,22-24 May, Dallas, TX, 2008.

15) Shu-Chung Yi, Kun-Tse Lee, Jin-Jia Chen, Chien-Hung Lin and Chuen-Ching Wang," The ne architecture of radix-4 Chinese abacus adder", I proceedings of IEEE International symposium onMultivalue logic, pp. 12, 17-20 May, 2006.

16) Luis E. Cordova, and James P. Davis, "TowardsAchieving Molecular Circuit Robustness Using a
Multi-Valued Programmable Logic Model,"Proceedings of IEEE International Workshop on
Design and Test of Defect-Tolerant NanoscaleArchitectures, May 1-2, Palm Springs, CA, USA,2005.

17) Chung-Yu WU and Hong-Yi Huang, "Design andApplication of Pipelined Dynamic CMOS TernaryLogic and Simple Ternary Differential Logic", IEEEJournal of Solid-State Circuits, Vol. 28, Issue. 8, pp:895-906, 1993.

18) Xiang Li and Baoding Liu, "Hybrid Logic andUncertain Logic", Journal of Uncertain Systems Vol.3, Issue. 2, pp: 83-94, 2009.

19) Ashur Rafiev, Julian P. Murphy, Danil Sokolov andAlex Yakovlev, "Conversion Driven Design ofBinary to Mixed Radix Circuits", IEEE InternationalConference on Computer Design, pp: 410-415, 2008.

20) Dusanka Bundalo, Zlatko Bundalo, Aleksandar Iliskovic and Branimir Djordjevic, "Architecture andDesign of Multiple Valued Digital and Computer Systems", 1'st Balkan Conference in Informatics,2003.

21) S.L. Hurst, "Multiple-valued logic – its status and itsfuture," IEEE Transactions on Computers, vol. C-33,December 1984, pp. 1160-1179.

22) D. Venkat Reddy, Ch.D.V. Paradesi Rao, E. G.Rajan, "Sequential Circuits in the Framework O (2n+1)-ary Discrete Logic", IJCSNS International Journal of Computer Science and Network Security,Vol. 8 Issue.7, July 2008.

23) Mozammel H. A. Khan, "Reversible Realization of
Quaternary Decoder, Multiplexer, and DemultiplexerCircuits", in proc. of 38th IEEE Intl. Symposium on Multiple Valued Logic, pp: 208-213, 22- 24 May 2008, Doi: 10.1109/ISMVL.2008.33.

24) Sangmi Shim, Gisoo Na, Seungwoo Park andSeunghong Hong, "Design of Q-IDEN D Flip-FlopUsing RS-latch", in proc. of International Journal onComputer Science and Network Security, Vol. 6,Issue. 9A, September 2006.

25) Abdul Ahad S.Awwal, Syed M. Munir, A.T.MShafiqul Khalid, Howard E. Michel and Oscar N.
Garcia, "Multivalued Optic Parallel Computation Using an Optical Programmable Logic Array",
Journal of Informatica, Vol 24, pp: 467-473, 2000.

26) P. P. Vaidyanathan and Byung-Jun Yoon, "The roleof signal-processing concepts in genomics andproteomics", in proc. of Journal on Franklin Institute,vol. 341, no. 1-2, pp: 111- 135, 2004.

27) Edward R. Dougherty, Ilya Shmulevich and MichaelL. Bittner, "Genomic Signal Processing: The SalientIssues", EURASIP Journal on Applied SignalProcessing, vol. 2004, pp: 146- 153, Jan. 2004.

28) Ioan Oprea, Sergiu Pasca and Vlad Gavrila, "Metho of DNA Analysis Using the Estimation of theAlgorithmic Complexity", in proc. of LeonardoElectronic Journal of Practices and

Technologies,no.5, pp: 53-66, July-December 2004.

29) Sara Nasser, Adrienne Breland, Frederick C. HarrisJr., Monica Nicolescu, and Gregory L. Vert, "FuzzyGenome Sequence Assembly for Single and Environmental Genomes", in proceeding of Fuzzy Systems in Bioinformatics and Computational Biology , vol. 242, Springer, pp: 19-44, 2009.

30) A.M. Selvam, "Quantumlike Chaos in the FrequencyDistributions of Bases A, C, G, T in Human Chromosome1 DNA", Apeiron, vol. 11, no. 3, July2004.

31) Jianbo Gao, Yan Qi, Yinhe Cao, and Wen-wen Tung,"Protein Coding Sequence Identification by Simultaneously Characterizing the Periodic and Random Features of DNA Sequences", Journal of Biomedicine and Biotechnology, Vol 2, pp: 139–146,2005.

32) Edward R. Dougherty, Ilya Shmulevich, Jie Chen, and Z. Jane Wang, "Genomic Signal Processing andStatistics", EURASIP Book Series on Signal Processing and Communications, Vol 2, 2005.

33) S. Tiwari, S. Ramachandran, A. Bhattacharya, S. Bhattacharya, and R. Ramaswamy, "Prediction of probable genes by Fourier analysis of genomic sequences," Comput. Appl. Biosci., vol. 13, no. 3, pp.263–270, 1997. http://bioinformatics.oxfordjournals.org/cgi/reprint/13/3/263

34) Trevor W. Fox, Alex Carreira, "A digital signal processing method for gene prediction with improvednoise suppression", EURASIP Journal on Applied Signal Processing, Volume 2004, Pages: 108 – 114, 2004, ISSN:1110-8657, DOI 10.1155/S1110865704309285.

35) Yuan Xin TIAN, Chao CHEN, Xiao Yong ZOU, JianDing QIU, Pei Xiang CAI1, Jin Yuan MO,"FourierPower Spectrum Analysis of Exons for the Period-3Behavior", Chinese Chemical Letters,Vol.16,No.7,pp939942,2005.http://www.imm.ac.cn/journal/ccl.ht