



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Identification of characteristics frequency and hot-spots in protein sequence of COVID-19 disease

Vikas Pathak^{a,b,*}, Satyasai Jagannath Nanda^a, Amit Mahesh Joshi^a, Sitanshu Sekhar Sahu^c

^a Malaviya National Institute of Technology, Jaipur 302017, Rajasthan, India

^b Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur 302017, Rajasthan, India

^c Birla Institute of Technology, Mesra, Ranchi 835215, Jharkhand, India

ARTICLE INFO

Keywords:

Hot-spots identification
SARS CoV2
IIR digital filter
COVID-19

ABSTRACT

COVID-19 has threatened the whole world since December 2019 and has also infected millions of people around the globe. It has been transmitted through the SARS CoV-2 virus. Various proteins of the SARS CoV-2 virus have an important role in its interaction with human cells. Specifically, the interaction of S-protein with human ACE-2 protein helps in entering of SARS CoV-2 virus into a human cell. This interaction take-place at some specific amino-acid locations called as hot-spots. Understanding of this interaction is helpful for drug designing and vaccine development for new variants of COVID-19 disease. An attempt has been made in this paper for understanding this interaction by finding the characteristics frequency of SARS-related protein families using the resonance recognition model (RRM). Hardware implementation of Bandpass notch (BPN) lattice IIR filter system architecture is also carried out, which is used for hot-spots identification in SARS CoV-2 proteins. Various signal processing techniques like retiming, pipelining, etc. are explored for performance improvement. Synthesis of proposed BPN filter system has been done using Xilinx ISE EDA tool on Zynq-series (Zybo-board) FPGA family. It is found that retimed and pipelined architecture of hardware-implemented BPN lattice IIR filter-based hot-spots detection system improves the speed (computational time) by 14 to 31 times for different SARS CoV2 related proteins as compared to its MATLAB simulation with similar functionality.

1. Introduction

COVID-19 (Corona Virus Disease) is a viral respiratory disease, which came in 2019 and is caused due to a SARS (Severe Acute Respiratory Syndrome) type novel Coronavirus [1]. The virus has been spread among the people through the droplets from the infected person as well as with close contact with the corona positive case. It can also be transmitted by the spitting, coughing, breathing, and sneezing of the COVID-19 infected patient [2]. The novel coronavirus has a severe impact on persons with underneath health diseases such as diabetes, cardiovascular disease, hypertension, chronic respiratory disease, etc. [3–5]. Now more than 26.67 crore cases and 52.78 lakhs deaths [6] are reported by this virus throughout the world.

The Coronaviruses are envelop viruses with single-positive stranded RNA molecules of a large genome size of 30K longer, which belong to the family Coronaviridae and subfamily Coronavirinae [7]. SARS CoV belongs to β -coronavirus. Coronaviruses have mainly two types of proteins: structural and non-structural proteins, which are encoded from 6 open reading frames (ORFs) codes for viral replication [8–10]. Structural proteins are mainly four types (shown by SARS CoV-2

structure in Fig. 1): (i) Spike (S) glycoprotein (ii) Membrane (M) protein (iii) Nucleocapsid (N) protein (iv) Envelope (E) protein. All these proteins play their own role in entering the virus into the human cell and further spreading it from one person to other.

Structures of the SARS CoV-2 virus signify that the E, S and M proteins collectively construct the envelope of the SARS virus [10]. The shape of the envelope mostly depends on M protein. The E protein is the smallest structural protein, which can combine several molecules to form an oligomer and generate an ion channel. The E protein has multiple roles in the viral replication cycle like (1) virion release, (2) viral assembly and (3) viral pathogenesis. The M and S proteins belong to the trans-membrane proteins, which are involved in virus assembly during replication [12]. M protein interacts with the Nucleocapsid, Envelope, Spike and Membrane glycoprotein itself throughout the virus particle assembly process. It has been demonstrated that M protein is more common within the virus membrane, and it is significant for the promising process of coronaviruses [12]. Polymers of S proteins implanted in the envelope giving it a crown-like appearance, thus this virus is named as coronavirus. S-protein mediates the interaction of

* Corresponding author.

E-mail address: 2013rec9567@mnit.ac.in (V. Pathak).

<https://doi.org/10.1016/j.bspc.2022.103909>

Received 8 December 2021; Received in revised form 28 April 2022; Accepted 12 June 2022

Available online 16 June 2022

1746-8094/© 2022 Elsevier Ltd. All rights reserved.

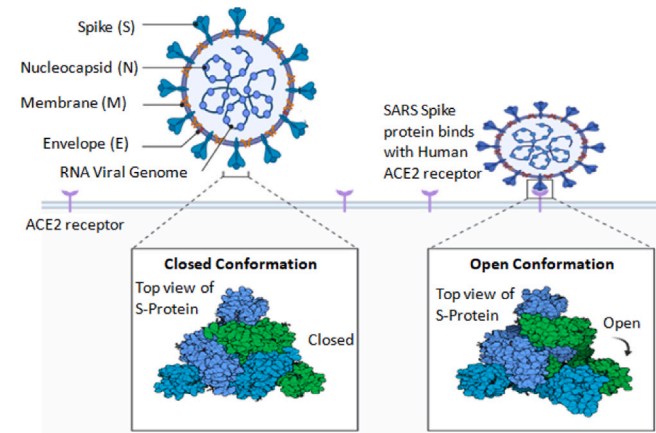


Fig. 1. Structure of SARS CoV-2 virus and its interaction/binding with Human ACE2 target protein [11].

the virus to the host cell [13]. Spike glycoprotein consists of S1 and S2 subunits [14]. The S2 subunit of the SARS CoV-2 virus is highly preserved and shares 99% similarity with those of the two bats human SARS-CoV and SARS-like CoVs. The S1 subunit has 70% similarity with these CoVs, but the core receptor binding domain is very much preserved. These amino-acid variations are accountable for the direct interaction of SARS CoV-2 spike protein with the host receptor of humans.

Structural proteins are responsible for binding the virus with host cells by interacting with receptor proteins. Specifically SARS CoV-2 virus binds with human (according to Fig. 1) by interaction of Spike protein with the Angiotensin-Converting Enzyme 2 (ACE2) receptor proteins [15]. This protein has a dense edge due to which the virus is attached more strongly as compared to other viruses of the equivalent origin. S protein encourages the entry of the SARS CoV-2 virus into human cells and is the main target of antibodies.

Besides these proteins, other proteins named as RNA dependent RNA polymerase (RdRp, also named nsp12) protein is the essential component of coronaviral replication/transcription machinery and seems to be a primary target for the antiviral drug, remdesivir [1,16–18]. RdRp proteins of SARS-CoV-2 showcase the structural resemblance with different amino acid residues preserved in the active site. This resemblance makes drug designing an efficient strategy that can reduce the drug development time as compared to that of De Novo drug discovery.

These proteins (of a similar functional group) interact with each other at a particular amino acid location called as hot-spot region. So identification of these hot spot locations is very important for their actual interaction and finding their functionality.

From our previous work [19,20], it is observed that proteins of the same functional group share a common characteristics frequency. Thus, the characteristics frequencies of these important protein families are identified. Then hot-spots of SARS CoV-2 virus proteins are found out by designing and tuning the bandpass notch (BPN) filter according to the characteristics frequency of that particular protein family. In work [19], the direct form-II structure of the BPN (ANF) IIR filter was used for the identification of hot-spots in various proteins, which has a higher area due to the requirement of two ANF filters. This area is further reduced in work [20] by using only a single ANF filter for the same application with more number of protein data-sets. Different signal processing techniques including retiming, pipelining and unfolding are also explored for a speed improvement in lattice hardware structures of ANF IIR filter [21,22], which was used for exon region identification in eukaryotic genes. In this manuscript, an area-optimized VLSI hardware architecture of the BPN lattice IIR filter-based hot-spots detection system is proposed, which is further optimized

Table 1

EIIP values of amino acids.

S. no.	Amino acid	Character code		EIIP value
		3 letter	1 letter	
1.	Leucine	Leu	L	0.0000
2.	Isoleucine	Ile	I	0.0000
3.	Asparagine	Asn	N	0.0036
4.	Glycine	Gly	G	0.0050
5.	Valine	Val	V	0.0057
6.	Glutamic acid	Glu	E	0.0058
7.	Proline	Pro	P	0.0198
8.	Histidine	His	H	0.0242
9.	Lysine	Lys	K	0.0371
10.	Alanine	Ala	A	0.0373
11.	Tyrosine	Tyr	Y	0.0516
12.	Tryptophan	Trp	W	0.0548
13.	Glutamine	Gln	Q	0.0761
14.	Methionine	Met	M	0.0823
15.	Serine	Ser	S	0.0829
16.	Cysteine	Cys	C	0.0829
17.	Threonine	Thr	T	0.0941
18.	Phenylalanine	Phe	F	0.0946
19.	Arginine	Arg	R	0.0959
20.	Aspartic acid	Asp	D	0.1263

for performance improvement using retiming and pipelining (unlike unfolding in paper [21] due to higher area requirement). Results are also validated through MATLAB simulation. It has been observed that hardware simulation of proposed hardware architecture is 14 to 31 times faster for hot-spots identification in various SARS CoV2 proteins compared to MATLAB simulation with its identical behavior. Newly identified hot-spots are responsible for interactions of various SARS CoV2-related proteins, which can be helpful for drug designing of new variants of SARS CoV2 virus-related COVID-19 disease.

The remaining part of the paper is summarized as: Section 2 provides the basics of characteristics frequency and hot-spots identification system in SARS CoV2 proteins. The proposed optimized hardware architecture of the BPN lattice IIR filter system for hot-spots detection in SARS CoV2 proteins is covered in Section 3. Various simulation and synthesis results with their detailed discussions are detailed in Section 4. Finally, Section 5 concludes the paper.

2. RRM model and designing of BPN filter for characteristics frequency and hot-spots identification system in SARS CoV-2 proteins

From literature [19,20,23–25], it has been inferred that proteins of the same functional group (or same protein family), interact with each other and other target molecules at particular active sites. Identification of these active sites is very essential for these types of interactions. Thus those locations of amino-acids need to be determined, which are responsible for this interaction. These amino-acid locations (which define the protein functionality) are called as hot-spots, due to which stability of active sites is ensured.

RRM model is very helpful for applying any digital signal processing (DSP) technique for finding the hot-spots for protein interaction. In RRM model, character sequence of amino acids is converted into numerical sequence according to the electron–ion interaction potential (EIIP) values [19] as per the Table 1. EIIP values represent the average energy of valence electrons in amino acids and are relevant to protein's biological properties. Then consensus spectrum (DFT multiplication of protein sequences) is calculated as per the following equation:

$$H(\omega) = |I_1(\omega) \parallel I_2(\omega) \parallel I_3(\omega) \parallel I_4(\omega) \parallel \dots \parallel I_N(\omega)| \quad (1)$$

where $I_1(\omega), I_2(\omega), \dots, I_N(\omega)$ denotes the DFTs of N-proteins. According to the RRM model, proteins (of the same family) interact with their target molecules at a particular frequency known as characteristics frequency, which is determined by a distinct peak in the plot ($H(\omega)$) of the consensus spectrum.

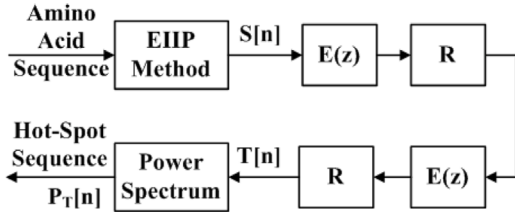


Fig. 2. The complete model of IIR digital filter based system for hot spot detection in SARS proteins.

After determining the characteristics frequency (ω_0) using the RRM model, hot-spots are determined using a narrow bandpass notch (BPN) digital IIR filter. The BPN filter model (including zero-phase filtering) (which is shown in Fig. 2) is used for finding the hot-spots of proteins of the SARS CoV-2 virus.

The power spectrum (represented by $P_T[n]$) of the final filter output is calculated as follows:

$$P_T[n] = |T[n]|^2 \quad (2)$$

The hot-spot locations of proteins of a particular protein family is determined from the peaks in the plot of power spectrum ($P_T[n]$).

Initially, inverse-Chebyshev filters were used by [23,24] for designing of BPN IIR filter, which has higher filter order (of 8). Here IIR filter is preferred as compared to the FIR filter due to its necessity of lower filter order to obtain the same selectivity for effective identification of hot-spots in SARS CoV-2 related proteins.

In [25] second-order BPN IIR filter is implemented due to its lower filter order to obtain the same selectivity for effective identification of hot-spots in SARS CoV-2 related proteins. This filter design technique needs more iterations (of 32-order), which results in larger computational time.

This computational time is further reduced by decreasing the filter design time using another method [26]. This filter design process also results in more computational time because of the N-times designing of BPN filters. Alternatively, the BPN filter can also be designed using all pass-filter (APF) based anti-notch filter (ANF) technique [27].

For drawing the lattice structures, consider the general form of second-order APF [28] as follows:

$$A(z) = \frac{k_1 + k_2(1 + k_1)z^{-1} + z^{-2}}{1 + k_2(1 + k_1)z^{-1} + k_1 z^{-2}} \quad (3)$$

Here, $k_1 = R^2$ and the value of k_2 is given by:

$$k_2 = \frac{-2R\cos\phi}{1 + k_1} = \frac{-2R\cos\phi}{1 + R^2} \quad (4)$$

After simplification the value of $k_2 = -\cos\omega$.

These lattice filter coefficients k_1 and k_2 are responsible for finding the self tuning of parameters Ω (3-dB attenuation bandwidth) and ω (notch frequency) respectively. Here Ω and ω determine the quality and anti-notch frequency of the ANF filter respectively. It is found that quality (Ω) and anti-notch frequency (ω) can be easily changed by adjusting the ANF lattice filter coefficients k_1 and k_2 , without altering each other. Hence, multiplier coefficients k_1 and k_2 can be independently adjusted for controlling the quality and anti-notch frequency of the ANF filter [21].

Hence for better quality (Ω) control assuming pole radius $R = 0.99$, then $k_1 = R^2 = (0.99)^2 = 0.9840$ and $k_2 = -\cos\omega_0$. It is clear from these values that multiplier coefficient k_1 is constant and k_2 can be easily determined just by calculating the value of $-\cos\omega_0$, where ω_0 denotes the anti-notch frequency (i.e. normalized characteristics frequency (in radian) of particular protein family, whose hot-spots are going to be detected). Therefore, ANF (BPN) filter can be simply designed and tuned according to normalized characteristics frequency (of SARS CoV-2 protein family) just by varying the anti-notch frequency (ω_0) for identification of hot-spots in SARS CoV-2 proteins.

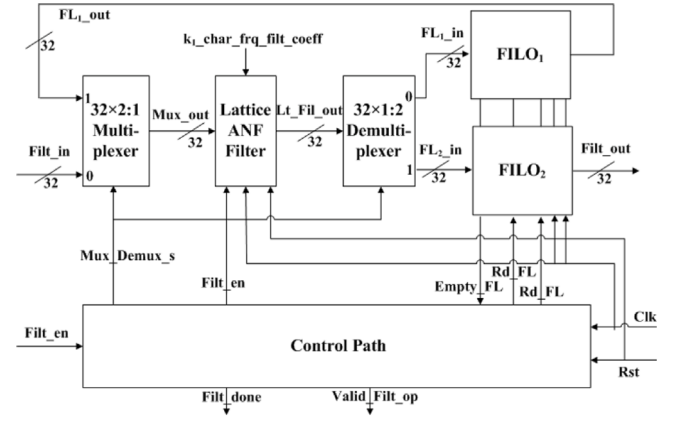


Fig. 3. Data path of optimized BPN lattice IIR digital filter system for hot-spots detection in SARS CoV2.

3. Hardware architecture of BPN digital lattice IIR filter system for hot-spots detections in SARS CoV-2 proteins

BPN filter based hot-spots detection system has two parts: Data Path and Control path.

3.1. Data path of proposed BPN filter based hot-spots detection system

The block diagram of the complete BPN filter system (indicated in Fig. 2) is shown in Fig. 3, which denotes the data-path of the 32-bit data flow from various blocks. This VLSI hardware architecture clearly shows the following blocks:

1. Lattice ANF Filter
2. Multiplexer (MUX)
3. Demultiplexer (DEMUX)
4. First-in Last-out (FILO)
5. Control Path

Block diagram of this data path is the area optimized version of the complete IIR filter based model (shown in Fig. 2), because only one filter is used (at two instants of time by switching the data signal using MUX and DEMUX) instead of two filters of the previous model. This block diagram also shows various input, filter enable, clock, reset and some output status signals. The control path provides the various control signals to different blocks of the data path. First-in last-out (FILO) block is used for storing the filter output and then reading in reverse order by another filter (replace the R-block as in Fig. 2). In this paper, the size of dual-port RAM (used in FILO) is equal to 2048×32 .

First of all, external filter input is applied to the I₀ input of MUX, whose output (Mux_out) is then supplied to lattice ANF filter input. The output of the lattice ANF filter is stored in FILO₁ through DEMUX Y₀ (FL1_in) output. After complete writing of filter output into FILO₁, it is read in the reverse direction from FILO₁ and supplied again to lattice ANF filter through MUX I₁ input. The filter output is again stored (through DEMUX Y₁ (FL2_in) output) in FILO₂, which is then read in the reverse direction for supplying the final external filter output (Filt_out).

3.2. Control path FSM of proposed BPN filter based hot-spots detection system

FSM (shown in Fig. 4) of the control path provides the various control signals (like select and enable signals) to MUX, DEMUX, lattice ANF filter and FILO. In this FSM diagram, FSM inputs are indicated on the transition arrow and outputs are printed along the arrow.

Table 2
Various proteins used for finding the consensus spectrum of protein families of SARS CoV-2 virus.

S. no.	Name of protein family	No. of proteins	PDB/Uni-prot IDs
1	Spike (S)	14	6LZG, 6M1V, 6VXX, 6W41, 6WPT, 6X6P, 6XDG, 6YOR, 6YZ7, 6Z2M, 6Z43, 7BYR, 7BZ5, 7CAN
2	RdRp	9	A0A2I4S557, A0A2P1E984, A0A2P1E991, A0A2R3SUZ4, A0A1W6S769, A0A2R3SUN8, A0A2R3SUU4, A0A1U9X1J7, A0A2D3HYN3
3	ACE2	7	Q5EGZ1, Q58DDO, Q9BYF1, Q56NL1, Q5RFN1, Q56H28, Q8R0I0
4	Membrane (M)	10	Q0Q472, A7J8L8, Q6SRM8, E0XIZ6, A0A4Y6GN58, R9QTR4, QLG76880, A0A6B9XUA0, A0A088DIE6, F1BYM2
5	Envelop (E)	8	B8Q8W2, U5WI28, E0XIZ5, R9QJTJ1, A0A1W5YKU8, Q6JH43, D2E2J8, PODTC4

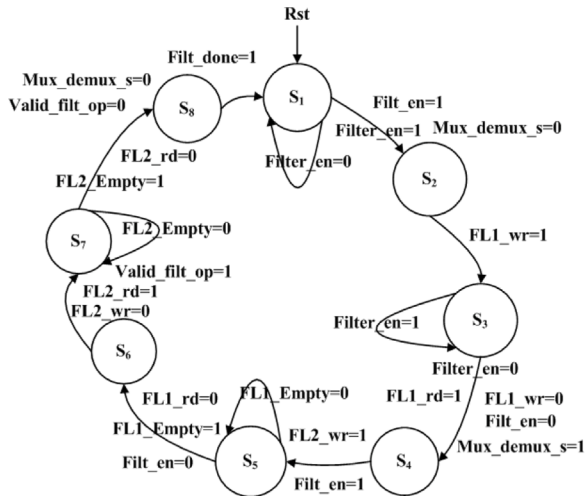


Fig. 4. FSM of control path of BPN lattice IIR digital filter system for hot-spots detection in SARS CoV2.

There are total 8-states in this FSM, out of which four states S_1 , S_3 , S_5 and S_7 are used for reset state, writing and reading from FILO, supplying the data to lattice ANF filter and providing select signals to the MUX and DEMUX. On the other hand, the other four extra states S_2 , S_4 , S_6 and S_8 are utilized for providing the synchronization and keeping the sufficient delay between various control signals. Mealy (i.e. output changes during the state transition) type FSM is used here in this control path.

3.3. Optimized realization of lattice architectures of ANF digital IIR filter

BPN (ANF) IIR filter can be realized either by direct form or lattice structures whose Hardware architectures are explored in our previous work [21,29,30], which were used for identification of exon regions in eukaryotic genes (different application). In this paper, lattice structures are followed due to the requirement of less logic hardware as compared to direct form structures, which are used for separate applications of hot-spots identification in SARS CoV2 proteins. In this paper four VLSI architectures of ANF lattice structures are considered: (i) original ANF lattice filter (Method-1(M1)) (ii) Retimed ANF lattice filter (Method-2(M2)) (iii) Pipelined ANF lattice filter (Method-3(M3)) (iv) Pipelined and retimed ANF lattice filter (Method-4(M4)).

After analyzing the critical path delay of various proposed optimized lattice architectures, it is concluded that retiming (M2) and pipelining (M3) alone have a higher delay. But if we apply the retiming on pipelined architecture (M4), then critical path delay is significantly reduced by one adder delay, which improves the performance (maximum clock frequency) of the BPN lattice IIR filter based hot-spots detection system.

In this paper real data type is used, because lattice filter coefficients (k_1 and k_2) and EIIP values of protein sequences are real numbers. There are two options for denoting the real numbers in digital form: floating-point and fixed-point. In this paper, the floating-point number system (32-bit single-precision IEEE standard [31]) is followed due to its higher range, precision and resolution as compared to fixed-point numbers.

It is clear from different lattice architectures that the main blocks of ANF (BPN) IIR filter are: Adder, multiplier and register. Hardware implementation of adders and multiplier [32,33] is executed using 32-bit single-precision floating-point arithmetic. Then, at last, these blocks are combined according to different optimized hardware architectures to form the ANF lattice IIR filter.

4. Results and discussion

4.1. Simulation results for determination of characteristics freq. in SARS CoV-2 proteins family

It is known that structural proteins (S, M, E, N), RdRp proteins of SARS CoV-2 virus and human ACE-2 proteins are very important for the drug designing purpose. So data sets of these five protein families (as per details in Table 2) are used for finding the characteristics frequency, which are downloaded either from PDB data bank [34] or Uni-Prot [35] using their IDs. Different numbers of proteins are used for finding the characteristics frequency of any SARS CoV-2 proteins, which are listed in Table 3. For example, 14 and 7 number of proteins are used for the determination of characteristics of Spike and ACE2 protein respectively.

A MATLAB program is written for finding the characteristics frequency of a particular protein family by checking the peaks in the consensus spectrum of a specific functional group. Consensus spectrum of ACE2 and RdRp proteins, are shown in Fig. 5(a) and (b) respectively. The consensus spectrum of these figures clearly shows the peaks in their respective plots, which indicate the characteristics frequency of various SARS CoV-2 proteins and are listed in Table 3.

Different identified characteristics frequencies of SARS CoV2 Spike (S), Membrane (M), Envelop (E), RdRp proteins and human ACE2 protein are 0.2738, 0.7333, 0.8378, 0.8194 and 0.4938 respectively. These characteristics frequencies will provide an insight for the interaction of SARS CoV-2 proteins with other living organisms.

4.2. Synthesis results of BPN lattice digital filter based hot spot-detection system

Xilinx ISE 14.4 version of VLSI EDA tool is used for the synthesis of lattice filter-based hot-spots detection system. In this manuscript, HDL (here VHDL) program is written for hardware implementation of lattice structures. The proposed design is implemented on Zynq-series (Zybo-board) FPGA with the actual device name 'xc7z010-3-clg400' for actual FPGA validation of the proposed lattice filter design. Summary of FPGA resources utilization and main FPGA blocks used by different

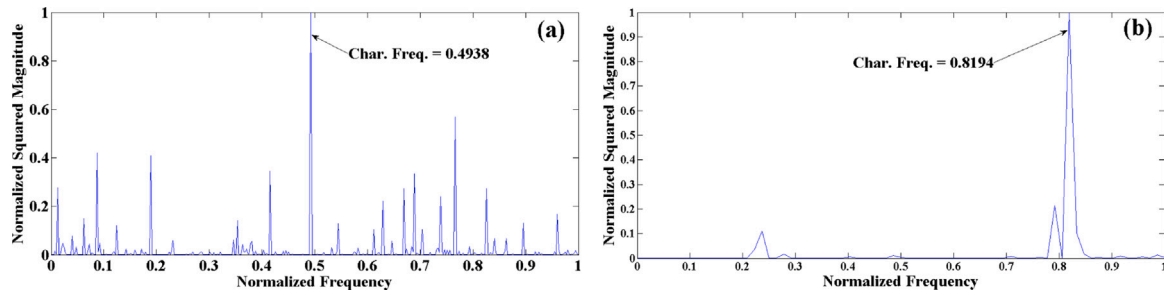


Fig. 5. Consensus spectrum of SARS CoV2 proteins like (a) ACE2 protein (b) RdRp protein.

Table 3

Characteristics frequency of protein families of SARS CoV-2 virus, detected by our proposed RRM model.

S. no.	Name of protein family	Sequence length	Char. Freq.	PDB/Uni-prot ID
1	Spike (S)	1281, 1247	0.2738	6VXX, 7BYR
2	RdRp	145	0.8194	A0A1W6S769
3	ACE2	805	0.4938	Q9BYF1
4	Membrane (M)	222	0.7333	QLG76880
5	Envelop (E)	75	0.8378	P0DTC4

Table 4

FPGA resource utilization.

Resource name	Available	Total used resources			
		M1	M2	M3	M4
Slice registers	35 200	232	228	295	302
Slice LUTs	17 600	3365	3344	3389	2918
Bonded IOBs	100	101	101	101	101
Block RAM/FIFO	60	1	1	1	1
BUFG/BUFGCTRLs	32	2	2	2	2
DSP48E1s	80	6	6	6	6

Table 5

Summary of FPGA hardware blocks.

Hardware block	M1	M2	M3	M4
Dual port RAM	2	2	2	2
Multipliers	3	3	3	3
Adders/Subtractors	9	9	9	9
Adders	13	13	13	13
Subtractors	17	17	17	17
Registers	16	16	18	18
Comparators	21	21	21	21
Multiplexers	312	312	312	312
Xors	10	10	10	10
FSMs	1	1	1	1

hardware architectures are indicated in Tables 4 and 5 respectively. It is clear from Table 4 that almost the same amount of FPGA resources are used for different lattice architectures, in-fact less number of slice LUTs are used for M4 structure. Table 5 also indicates the similar amount of FPGA Hardware blocks used by different architectures (except only two extra registers for pipelined (M3 & M4) lattice filter structures).

The timing summary of FPGA implementation of various proposed lattice filter architectures is also shown in Table 6, which reveals that the practical minimum period of the M4 structure has the lowest value (28.595 ns) compared to other lattice structures. Power analysis of the proposed lattice IIR filter is also carried out using the Xilinx XPower Analyzer tool without applying any constraints. Total power using this tool is calculated as 42 mW. Therefore, it is clear from Tables 4–6 that the performance of the BPN filter based hot-spots detection system has been improved using pipelined and retimed (M4) lattice architecture without much overhead of logic hardware.

Table 6

Timing summary.

Parameter	M1	M2	M3	M4
Min. Period (ns)	35.99	46.97	36.03	28.59
Max. Clk Freq. (MHz)	27.78	21.28	27.74	34.97
Min. I/P arrival time before clk (ns)	11.58	40.17	11.60	13.31
Max. O/P required time after clock (ns)	2.12	2.12	2.12	2.12

4.3. Simulation results of BPN lattice digital filter based hot spot-detection system

Protein data sets are downloaded into MATLAB through their UniProt/PDB IDs using the MATLAB sequence viewer toolbox. These protein character sequences are then mapped to numerical values using the EIIP method. These EIIP values are then utilized for finding the characteristics frequency in the SARS CoV2 protein family and hot-spots detection in those proteins. For validation (through FPGA implementation) of the lattice BPN filter-based hot-spots detection system, these numerical sequences (real values) are then converted to 32-bit single-precision floating-point numbers using the 'float2bin' MATLAB function and saved in an input text file. These floating-point values are then read into the Xilinx ISE tool through the VHDL testbench and supplied to the proposed hardware design under test (DUT). Floating-point output values of this lattice filter are written into the output text file, which is then read into MATLAB and again converted to real values using the 'bin2float' MATLAB function. The power spectrum of these real values of lattice BPN filter output is calculated and the graph is plotted for hot-spots detection in SARS CoV2 proteins.

Verification of the proposed lattice filter design is carried out through its hardware simulation using Xilinx ISE 14.4 simulator. For this simulation, the k_2 filter coefficient is provided to the proposed design for various proteins according to the characteristics frequency of a particular SARS protein family. The simulation waveform of the proposed design indicates the different inputs, outputs and intermediate signals of the proposed design. Filtering starts after filter_en = 1 and it continue until filtering_done = 1 (i.e. filtering is complete). Valid filter output is available on output data bus for valid_filter_op = 1. Hence hardware simulation time (i.e. total computational time (CT)) is calculated from filter_en = 1 up to filtering_done = 1. For calculation of this CT, the total number of clock pulses (L) for this duration are

Table 7

Computational time comparison for MATLAB and hardware simulation of proposed BPN lattice IIR filter based hot-spot detection system.

S. no.	Protein name	MATLAB CT (μ s)	M	CT (μ s) of proposed hardware		Speed improvement	
				Without any optimization	With pipelining and retiming	Without any optimization	with pipelining and retiming
1	E protein	204	233	8.386	6.662	24	31
2	RdRp protein	367	443	15.945	12.667	23	29
3	M protein	574	674	24.259	19.272	24	30
4	ACE-2 protein	1117	2423	87.213	69.283	13	16
5	S protein	1571	3851	138.612	110.115	11	14

counted. After exploring the waveform it is found that external protein sequence data (of sequence length = N) is traced three times plus some extra 8-clock pulses are required for synchronization between different blocks. Hence, total number of clock pulses $L = 3 \times N + 8$. Then total CT is computed as $CT = L \times T$, where T is the minimum critical path delay (which is found from synthesis results).

The hardware simulation results of the proposed design are also validated through its software simulation, which are carried out using MATLAB 2014a. Computational time (CT) of software simulation is calculated using MATLAB 'tic' and 'toc' commands. For this calculation average CPU time of MATLAB 100-runs (for each data sample) is recorded.

CT of hardware and software (MATLAB) simulation of proposed BPN IIR filter for hot-spots detection in various SARS CoV2 proteins is compared in Table 7. This table signifies that computational time increases with the sequence length of different SARS-related proteins. CT of proposed lattice hardware BPN filter (without any optimization) varies from 8.386 to 138.612 μ s for various SARS proteins, which is further reduced to 6.662 to 110.115 μ s using pipelining and retiming techniques. This table also implies that the performance (speed) of the proposed hardware BPN filter-based hot-spots detections system is improved by 14 to 31 times for different SARS CoV2 protein data samples as compared to its MATLAB implementation.

4.4. Discussion of simulation results for hot-spots identification in SARS CoV-2 related proteins

After determining the consensus spectrum of SARS CoV-2 related protein families, the proposed tunable BPN IIR filter is tuned according to the characteristics frequencies of various proteins, which are then used for finding the hot-spots in those proteins. Both FPGA hardware implementation and MATLAB simulation are done for the proposed design and hot-spots in proteins are detected by peaks in the plot of the power spectrum of BPN lattice filter response. In our previous work [19,20], the proposed BPN Filter was verified using 11-protein sequence data-sets of easily available ASEdb database [36,37] (in which hot-spots were identified using laboratory-based wet-lab experiment ASM-method). Here for example result of one standard data-set of FGF protein is shown in Fig. 6, which indicates the similar identified hot-spots for both hardware implementation and MATLAB simulation of the proposed BPN filter. Same hot spot locations (24, 96 and 103) as of reported in ASEdb database [36,37] are detected. In paper [19,20], mostly hot-spots were identified (by our proposed BPN filter system) mentioned in the ASEdb database. More than 70% success rate was obtained by our proposed BPN filter for hot-spots identification. In this paper, the same BPN filter is used, therefore we can believe that the proposed BPN filter will also correctly identify the hot-spots in SARS CoV-2 related proteins.

Plots of Power spectrum vs amino acids locations for SARS CoV-2 envelop (E) protein is shown in Fig. 7. In this figure, (a), (b) and (c) parts respectively indicate the power spectrum plots of MATLAB implementation, FPGA hardware (without any optimization) and FPGA hardware (with pipelining and retiming technique) implementation. These same plots signify the same behavior of the proposed BPN lattice

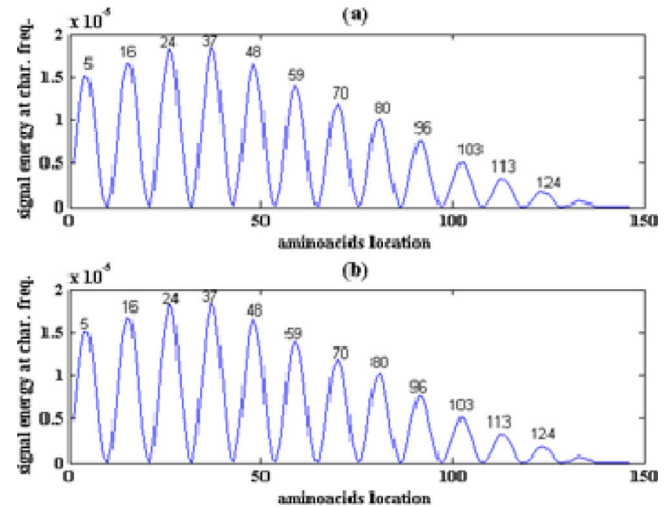


Fig. 6. Detection of hot-spot locations for standard data-set of FGF protein by BPN IIR filter using (a) MATLAB simulation (b) Hardware implementation.

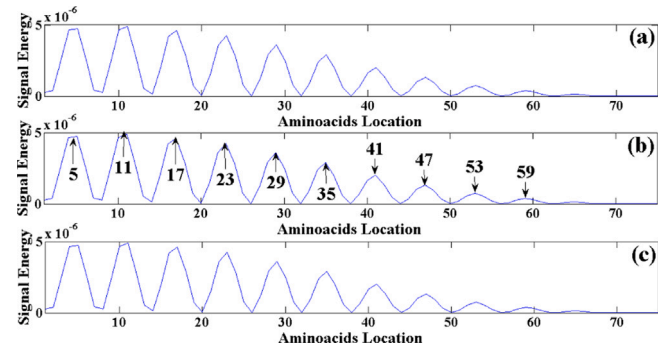


Fig. 7. Power spectrum plot for hot spot detection in SARS CoV2 E-protein by BPN lattice IIR filter using (a) MATLAB approach (b) FPGA hardware without retiming (c) FPGA hardware with retiming.

filter for hot-spots identification in SARS CoV2 proteins for three types of implementation.

Various hot-spots identified in other SARS CoV2 proteins such as Membrane (M) and RdRp proteins are shown in Fig. 8(a) and (b) respectively. The peaks in these figures denote the locations of hot-spots, which are clearly visible. Some of the hot-spots are denoted in these figures and other complete hot-spots (identified by our proposed hardware architecture) of E, M and RdRp proteins are shown in Table 8. This table reveals that 10 (at the amino acid location of 5, 11, 17, 23...), 24 (at locations 5, 10, 16, 21...) and 34 (at locations 5, 9, 12, 16...) numbers of hot-spots are identified by our proposed hardware BPN lattice filter architecture in E, RdRp and M proteins respectively.

ACE2 proteins are the primary target for SARS CoV-2 attack. Except for human this protein also exist in other organism species like rat,

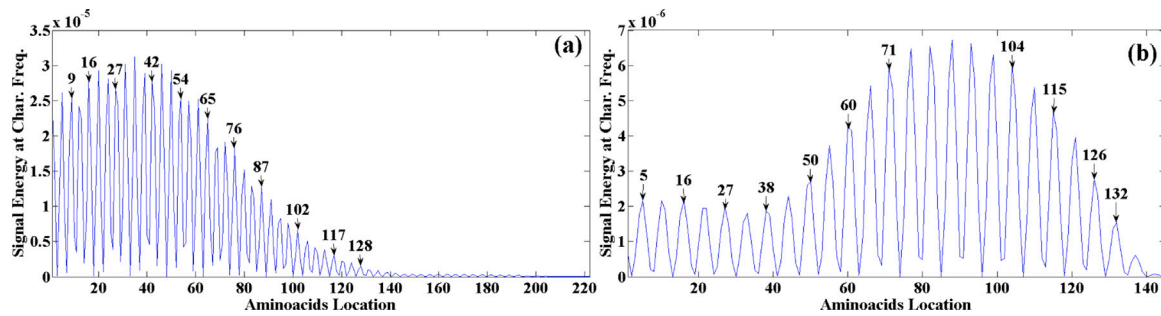


Fig. 8. Identification of hot-spots in SARS CoV2 proteins like (a) Membrane Glyco-protein (b) RdRp protein.

Table 8

Hot-spots identified by proposed BPN lattice IIR filter based hot-spot detection system.

S. no.	Protein name	Identified hot-spots
1	E protein	5, 11, 17, 23, 29, 35, 41, 47, 53, 59
2	RdRp protein	5, 10, 16, 21, 27, 33, 38, 44, 50, 55, 60, 66, 71, 77, 82, 88, 93, 99, 104, 110, 115, 121, 126, 132
3	M protein	5, 9, 12, 16, 20, 24, 27, 31, 35, 39, 42, 46, 50, 54, 57, 61, 65, 69, 72, 76, 80, 83, 87, 91, 95, 98, 102, 106, 109, 113, 117, 120, 124, 128

mouse, bovin, etc. So ACE2 proteins of these species are also considered for calculating the characteristics frequency through its consensus spectrum. It is well known that for developing any vaccine or drug for any disease like COVID-19, it is required to test the vaccine on other organisms like rats, mouse and monkeys before applying the vaccine on humans. So for finding the interaction of the SARS CoV2 virus with these organisms it is desired to get the knowledge of hot-spots, which will be very helpful for testing the vaccine or drug of COVID-19 on other species like rats, mouse before actual applying on human-being. The identified hot-spots in the SARS CoV-2 virus can be used as a useful resource for designing of new antiviral drugs and discovering the vaccine for SARS CoV-2 (COVID-19) disease. Hot-spots identification using the proposed hardware architecture of the BPN lattice IIR filter-based hot-spot detection system can speed up the drug design process due to its speed improvement of 14 to 31 times as compared to MATLAB implementation.

5. Conclusion

The paper presents the role of various proteins in the SARS CoV-2 virus. The RRM model is further used for finding the characteristics frequency of five protein families S, M, E, RdRp and ACE2 protein as 0.2738, 0.7333, 0.8378, 0.8194 and 0.4938 respectively. Various VLSI hardware architectures of the BPN lattice IIR filter-based hot-spot detection system are also explored. Then, the BPN lattice filter is tuned (according to characteristics frequency) for the identification of hot-spots in S, M, E, ACE2 and RdRp proteins. Performance (speed) of these hardware architectures is improved using different signal processing techniques like retiming, pipelining, etc. It is found that retiming along with pipelining decreases the critical path delay, which in turn increases the maximum clock frequency (hence performance is improved) of the BPN lattice IIR filter. It is observed that optimized hardware architecture (through its FPGA implementation) of BPN lattice IIR filter improves the speed by 14 to 31-times for hot-spots identification in various SARS CoV2 proteins as compared to its software (MATLAB) implementation with similar behavior. It is found that 34, 24 and 10 numbers of hot-spots are clearly detected in SARS CoV2 related M, RdRp and E proteins using the proposed hardware architecture of the BPN lattice IIR filter. Identification of characteristics frequency and

hot-spots of these proteins are helpful for designing of new antiviral drugs and discovering the vaccine for new variants of SARS CoV-2 (COVID-19) disease.

CRedit authorship contribution statement

Vikas Pathak: Conception & design of study, Acquisition of data, Software, Methodology, Writing – original draft. **Satyasai Jagannath Nanda:** Conception & design of study, Analysis and/or interpretation of data, Supervision, Writing – review & editing. **Amit Mahesh Joshi:** CConception & design of study, Analysis and/or interpretation of data, Supervision, Writing – review & editing. **Sitanshu Sekhar Sahu:** Analysis and/or interpretation of data, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, et al., Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharm. Sin. B* 10 (5) (2020) 766–788.
- [2] A.K. Tripathy, A.G. Mohapatra, S.P. Mohanty, E. Kougianos, A.M. Joshi, G. Das, asyBand: A Wearable for safety-aware mobility during pandemic outbreak, *IEEE Consum. Electron. Mag.* 9 (5) (2020) 57–61.
- [3] A.M. Joshi, U.P. Shukla, S.P. Mohanty, Smart healthcare for diabetes: A COVID-19 perspective, 2020, pp. 1–18, arXiv preprint arXiv:2008.11153.
- [4] S.B. Rikan, A.S. Azar, A. Ghafari, J.B. Mohasefi, H. Pirnejad, COVID-19 diagnosis from routine blood tests using artificial intelligence techniques, *Biomed. Signal Process. Control* 72 (2021) 103263.
- [5] S. Thakur, A. Kumar, X-ray and CT-scan-based automated detection and classification of covid-19 using convolutional neural networks (CNN), *Biomed. Signal Process. Control* 69 (2021) 102920.
- [6] W. website, Coronavirus cases and deaths due to COVID-19 Pandemic, 2020, <https://www.worldometers.info/coronavirus/>, Accessed: 31-October, 2020.
- [7] T.S. Fung, D.X. Liu, Human coronavirus: host-pathogen interaction, *Annu. Rev. Microbiol.* 73 (2019) 223–227.
- [8] M.D.S. Demirci, A. Adan, Computational analysis of microRNA-mediated interactions in SARS-CoV-2 infection, *PeerJ* 8 (2020) 1–17.
- [9] Y.-J. Tan, S.G. Lim, W. Hong, Characterization of viral proteins encoded by the SARS-coronavirus genome, *Antivir. Res.* 65 (2) (2005) 69–78.
- [10] F.K. Yoshimoto, The proteins of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2 or n-COV19), the cause of COVID-19, *Protein J.* 39 (3) (2020) 198–216.
- [11] M. Notes, Structure and genome of SARS-CoV-2 (COVID-19) with diagram, 2020, <https://microbenotes.com/structure-and-genome-of-sars-cov-2/>, Accessed: September, 2020.
- [12] M. Bianchi, D. Benvenuto, M. Giovanetti, S. Angeletti, M. Ciccozzi, S. Pascarella, SARS-CoV-2 envelope and membrane proteins: Structural differences linked to virus characteristics? *BioMed Res. Int.* 2020 (2020) 1–6.
- [13] A.C. Walls, Y.-J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veessler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, *Cell* 181 (2) (2020) 281–292.

- [14] F. Li, Structure, function, and evolution of coronavirus spike proteins, *Annu. Rev. Virol.* 3 (2016) 237–261.
- [15] A.M. Joshi, U.P. Shukla, S.P. Mohanty, Smart healthcare for diabetes during COVID-19, *IEEE Consum. Electron. Mag.* (2020) 1–5.
- [16] Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, et al., Structure of the RNA-dependent RNA polymerase from COVID-19 virus, *Science* 368 (6492) (2020) 779–782.
- [17] A.A. Elfiky, SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: An in silico perspective, *J. Biomol. Struct. Dyn.* (2020) 1–9.
- [18] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, et al., Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV, *Nature Commun.* 11 (1) (2020) 1–12.
- [19] V. Pathak, S.J. Nanda, A.M. Joshi, S.S. Sahu, VLSI implementation of Tunable Band-Pass Notch IIR filter for localization of hot spots in proteins, in: *Proceedings of the Fourth International Conference on Microelectronics, Computing and Communication Systems*, Ranchi, India, Vol. 673, Springer, Singapore, Sept, 2020, pp. 563–575.
- [20] V. Pathak, S.J. Nanda, A.M. Joshi, S.S. Sahu, FPGA implementation of high-speed tunable IIR band pass notch filter for identification of hot-spots in protein, *Int. J. Circuit Theory Appl.* (2021).
- [21] V. Pathak, S.J. Nanda, A.M. Joshi, S.S. Sahu, VLSI implementation of anti-notch lattice structure for identification of exon regions in Eukaryotic genes, *IET Comput. Digit. Tech.* 15 (4) (2020) 217–229.
- [22] V. Pathak, S.J. Nanda, A.M. Joshi, S.S. Sahu, FPGA implementation of high speed anti-notch lattice filter for exon region identification in eukaryotic genes, in: *2021 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, IEEE, 2021, pp. 418–421.
- [23] P. Ramachandran, A. Antoniou, Identification of hot-spot locations in proteins using digital filters, *IEEE J. Sel. Top. Sign. Proces.* 2 (3) (2008) 378–389.
- [24] P. Ramachandran, A. Antoniou, Localization of hot spots in proteins using digital filters, in: *International Symposium on Signal Processing and Information Technology*, IEEE, Vancouver, BC, Canada, 2006, pp. 926–931.
- [25] P. Ramachandran, W.-S. Lu, A. Antoniou, Improved hot-spot location technique for proteins using a bandpass notch digital filter, in: *International Symposium on Circuits and Systems*, IEEE, Seattle, WA, USA, June, 2008, pp. 2673–2676.
- [26] P. Ramachandran, W.-S. Lu, A. Antoniou, Tuning technique for the location of hot spots in proteins using a bandpass notch digital filter, in: *International Workshop on Genomic Signal Processing and Statistics*, IEEE, Minneapolis, MN, USA, July, 2009, pp. 1–4.
- [27] P.P. Vaidyanathan, B.J. Yoon, Digital filters for gene prediction applications, in: *Proc. Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, Vol. 1, IEEE, Pacific Grove, CA, USA, Nov, 2002, pp. 306–310.
- [28] P.A. Regalia, S.K. Mitra, P. Vaidyanathan, The digital all-pass filter: A versatile signal processing building block, *Proc. IEEE* 76 (1) (1988) 19–37.
- [29] V. Pathak, S. Jagannath Nanda, A. Mahesh Joshi, S. Sekhar Sahu, Hardware implementation of infinite impulse response anti-notch filter for exon region identification in eukaryotic genes, *Int. J. Circuit Theory Appl.* (2020) 1–17.
- [30] V. Pathak, S.J. Nanda, A.M. Joshi, S.S. Sahu, High speed implementation of Notch/Anti-notch IIR filter on FPGA, in: *15th IEEE India Council International Conference (INDICON)*, IEEE, 2018, pp. 1–6.
- [31] W. Kahan, IEEE standard 754 for binary floating-point arithmetic, *Lect. Notes Status IEEE 754 (94720–1776)* (1996) 11.
- [32] P. Bharade, Y. Joshi, R. Manthalkar, Design and implementation of IIR lattice filter using floating point arithmetic in FPGA, in: *Proc. Conference on Advances in Signal Processing (CASP)*, IEEE, Pune, India, June 2016, pp. 321–326.
- [33] G. Marcus, P. Hinojosa, A. Avila, J. Nolasco-Flores, A fully synthesizable single-precision, floating-point adder/subtractor and multiplier in VHDL for general and educational use, in: *Proc. Fifth International Caracas Conference on Devices, Circuits and Systems*, Vol. 1, IEEE, Punta Cana, Dominican Republic, November 2004, pp. 319–323.
- [34] S.C.P. Database, National center for Biotechnology information (NCBI), national institutes of health (NIH), national library of medicine (NLM), 2020, <https://www.ncbi.nlm.nih.gov/sars-cov-2/>, Accessed: September, 2020.
- [35] S.C.U.P. Database, Uni-prot protein knowledge database. Swiss inst. Bioinformatics (SIB), 2020, <https://www.uniprot.org/uniprot/?query=sars+cov2+virus&sort=score>, Accessed: September, 2020.
- [36] K.S. Thorn, A.A. Bogan, ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions, *Bioinformatics* 17 (3) (2001) 284–285.
- [37] A. Database, Alanine scanning energetics database (ASEdb), 2020, <http://nic.ucsf.edu/asedb/index.php>, Accessed: August, 2020.