Integrating and introducing CERN and NCBI data science to understand quantum realm computations

Raghavendra Rao Sankaramanchi¹, V Kamakshi Prasad², Kumara Chandra Singarapu³, Tejaswini Thallapalli³, Sandeep Sagar², Shilpa Chakra Chidurala¹, Upender Gaddam⁴, and Shrawan Kumar⁵.

¹Research Associate and Independent Researcher, CSE, JNTUH, Hyderabad, India
²Professor, CSE, JNTUH, Hyderabad, India
³Software Analyst, Hexagon Capability Centre India Pvt.Ltd, Hyderabad, India
²Intern Student, CSE, JNTUH, Hyderabad, India
²Project Student, CSE, JNTUH, Hyderabad, India
³Assistant Professor, CNST, IST, JNTUH, Hyderabad, India
⁴Independent Researcher, JNTUH, Hyderabad, India
⁵Assistant Professor, CSE, Vardhaman College of Engineering, Hyderabad, India

E-Mail: raghavendra.sankaramanchi@gmail.com

Abstract. Big data in the coming days is going to become an integral part of life existing on this planet. Hence, introducing the early stage data analytics with the European Organization for Nuclear Research (CERN) and the National Center for Biotechnology Information (NCBI) would generate interest in the research community for the rapidly changing world of technology. This study is especially intended to integrate two organizations data for understanding the quantum realm computations. Generally, Quantum Computation (QC) creates the notion in the brain as Qu-bits. But here, QCs are the principles followed by the particles, atoms, molecules, and systems. Few of them are Pauli’s exclusion principle, superposition and entanglement. These principles played a vital role in the construction of proton to an atom and to higher systems. In this present work, we have collected homo-sapiens’ chromosomes data from NCBI and integrated it with CERN data to understand quantum computations. We have identified a couple of computations and communications taking place inside the genes by integrating the recent scientific results obtained at CERN. We have proposed a novel code-named peacock code (PCOC) and designed a prototype to develop a unique identity with blockchain technology. A typical correlation between CERN and NCBI datasets are integrated into this study statistically with R programming tool to understand supersymmetry and develop future computations with natural principles.

1. Introduction

European Organization for Nuclear Research (CERN) deals with particles behaviour/interaction. Physicists and engineers of the European Organization for Nuclear Research are probing into the fundamental structure of the Universe. To study the basic constituents of matter, CERN uses the world's largest and most complex scientific instruments. The protons/particles are made to collide with each other with a speed nearly equal to that of light. This paves the way for physicists and other relevant experts to understand how the particles interact and decay by constructing new particles with
fundamental laws of nature. In 2012, the CERN experiments confirmed the existence of the Higgs particle which was one of the greatest prediction ever made by Peter Higgs and his colleagues in 1960s. Recent results obtained at ATLAS and CMS detectors at the Large Hadron Collider (LHC) revealed how efficiently the Higgs boson interacts with the heaviest known elementary particle (a subatomic particle with no substructure, thus not composed of any other particles), the top quark. The main function of this organization is to study the sub-atomic particles’ functions and behaviours.

National Centre for Biotechnology Information (NCBI) provides access to biomedical and genomic information. It maintains more than 160,000 organisms’ molecular data in its databases (DB). Major databases include GenBank, PubMed, and NCBI Epigenomics. GenBank is for DNA sequences, PubMed is for biomedical literature and NCBI Epigenomics for epigenetic modifications. Gene and its expressions are confirmed with several trial and error experiments and stored at NCBI database center. This organization’s main function is to study the genomic data of all the organisms. We have integrated both the organizations data and designed a unique identity prototype with blockchain technology in the way the DNA designs the fingerprints and iris in living systems. Proton-Proton collision data is stored at CERN database center. CERN results have proved that all the particles in the standard model are generated by the Higgs boson. Agglomeration of these generated particles formed the initial atom i.e., Hydrogen. Later, the fusion process might have led to the generation of other higher elements. Further, the combination of diatoms and carbon with other elements led to the formation of first life with the construction of DNA which led to genomics. DNA succeeded in the race by evolving the life faster than the xenomics (XNA). Right now, we have DNA data available from NCBI. In this present study, we have specifically concentrated to integrate two organizations data for understanding the computations and communications of the quantum realm [1]. The data collected from CERN & NCBI are different from each other in their functions. The integration of these two organizations data paved a path for us to understand the quantum-computations taking place within the dark matter and light matter [2]. This process is named as "Peacock Code", Particle Communications and Optical Computing (PCOC). We have designed a prototype to develop a unique identity with blockchain technology by utilizing PCOC technology.

2. Methodology
We have collected homo-sapiens total chromosomes sequence data from NCBI. In this present study, we have targeted insulin and glucagon gene data, for fast data processing with the available computation facilities. We have applied the top-down approach for understanding the quantum realm communications and computations. Initially, we have calculated the total number of nucleosides present in the insulin and glucagon hormones genes [3]. In the second step, we have calculated the number of Adenine, Guanine, Cytosine and Thymine molecules and their ratios present in the given gene sequence. In the third step, we have calculated the number of Carbon atoms, Nitrogen atoms, Oxygen atoms and Hydrogen atoms present in the given sequence. In the fourth step, we have estimated the probability of presence of different types of isotopes of Carbon, Nitrogen, Oxygen and Hydrogen in Adenine, Guanine, Cytosine and Thymine molecules. In the fifth step, we have integrated sub-atomic particles data, obtained from the open data portal of CERN organization, by statistically programming with R. In the sixth step, we have analyzed the statistical data obtained from the R tool with additional knowledge of particles behaviors at higher and lower levels of its own fabricated systems. In the seventh step, we have proposed the internal structures of fundamental particles and their interactions with photons and named it as PCOC (peacock code). In the eighth step, we have designed a Unique Identity (UID) prototype.
3. Workflow

The above figure (Fig.1) explains the methodology of this present work. It has four layers, namely Presentation tier, Application tier, Database tier and Infinite tier. All these four tiers are the mixtures of light matter, light energy, dark matter and dark energy. Unknown matter/data is more than that of the known. Dark energy impacts the Universe's expansion. We tried to decode the complete mystery with peacock code. Dark energy is roughly 68% of the universe while the dark matter is about 27%. Everything ever observed on this Earth and the cosmos with all of the high-end technical instruments including our sense organs, all normal matter - adds up to less than 5% of the universe. The infinite tier matter/data with various data types (p, q, r and s) is responsible for the formation of the rest of the tiers. The database tier forms the molecules and becomes application tier. DNA as application tier performs all the functions with matter/data in dominant and recessive manner to function at presentation tier [4]. Different systems in the presentation tier are formed from the DNA where the dominant and recessive genes show the differences among the systems. Eagle-view of the above blue print gives us the information about the data processing with the data types (waves with interacting and non-interacting behaviours) and datasets (particles) in infinite tier. Data types are the data carriers for the communications between the systems. Datasets are the physical matter, which are observable and interactable. Entropy in all the systems and the Universe expansion is happening to fill the datasets with data types and to communicate with rest of the data types to perform sophisticated functions. These communications and computations of quantum realm paved the way for us to understand several mysteries of the Universe and to design an UID in the same way the application tier designs the fingerprints and iris at presentation tier.
4. Results and Discussion

4.1 CERN Data Analysis

Higgs boson decay graphs

Fig. 2: (a): $\text{Em}_Hb$ and $\text{Im}_2\text{jets}$, (b): $\text{Em}_Hb$ and $\text{Im}_{Ht\_L}$, (c): $\text{Em}_Hb$ and $\text{Tm}_{te\_L}$, (d): $\text{Im}_{Ht\_L}$ and $\text{Im}_2\text{jets}$, (e): $\text{Tm}_{te\_L}$ and $\text{Im}_2\text{jets}$, (f): $\text{Tm}_{te\_L}$ and $\text{Im}_{Ht\_L}$

The above (Fig. 2) represents the results obtained by plotting the CERN data [5]. The Higgs Boson candidate ($\text{Em}_Hb$) is taken on X-axis and the invariant mass of two jets ($\text{Im}_2\text{jets}$), the invariant mass of hadronic tau and Lepton ($\text{Im}_{Ht\_L}$), the transverse mass between missing transverse energy and the lepton ($\text{Tm}_{te\_L}$) are taken on Y-axis correspondingly in Figs. 2 (a), (b) and (c) respectively. The invariant mass of Hadronic tau and Lepton ($\text{Im}_{Ht\_L}$) is taken on X-axis and the invariant mass of two jets ($\text{Im}_2\text{jets}$) is taken on Y-axis in Fig.2 (d). The transverse mass between missing transverse energy and the lepton ($\text{Tm}_{te\_L}$) is taken on X-axis and the invariant mass of two jets ($\text{Im}_2\text{jets}$), the invariant mass of hadronic tau and lepton($\text{Im}_{Ht\_L}$) are taken on the Y-axis respectively in Figs. 2 (e) and (f) respectively. These results indicate the computations and communications that are taking place at quantum realm.

4.1.1 Muon interaction graphs
Fig 3: (g): $Db_t$ and $E_{FD}$, (h): $Db_t$ and $Tm_{p2m}$, (i): $Db_t$ and $M_{P2m}$, (j): $E_{FD}$ and $Tm_{P2m}$, (k): $E_{FD}$ and $M_{P2m}$, (l): $Tm_{P2m}$ and $M_{P2m}$.

The above (Fig. 3) represents the results obtained by plotting the kaggle Higgs Boson Machine Learning Challenge data. (Data Source: Training.csv file at [https://www.kaggle.com/c/flavours-of-physics/data](https://www.kaggle.com/c/flavours-of-physics/data)). The distance between tau and primary vertex ($Db_t$) is taken on X-axis and error on flight distance ($E_{FD}$), transfer momentum of p2 muon ($Tm_{p2m}$), momentum of p2 muon ($M_{P2m}$) are taken on corresponding Y-axis in Figs. (g), (h) and (i) respectively. Error on flight distance ($E_{FD}$) is taken on X-axis and transfer momentum of p2 muon ($Tm_{p2m}$), momentum of p2 muon ($M_{P2m}$) are taken on Y-axis correspondingly in Figs. (j) and (k) respectively. In Fig. (l), transfer momentum of p2 muon ($Tm_{P2m}$) is taken on X-axis and the momentum of p2 muon ($M_{P2m}$) is taken on Y-axis. These results indicate that the computations and communications are taking place at quantum realm.

Discussion: The ATLAS and ALICE detectors consider the particles such as electron, photon, muon etc. These detectors classify the type of the particle, its energy, and the 3D direction of travel of the particle. Based on these properties, LHC has provided huge amount of data in its open data portal. We have taken ATLAS data from opendata.cern database. The data source for ALICE is from the Higgs Boson Machine Learning Challenge by kaggle team. These graphs are only giving a path to understanding the super-symmetry principle. To explain super-symmetry, the fundamental particles' internal structures are hypothesized to explain the quantum computations that are taking place inside the quantum world and the regulatory process of the upper hierarchy.

4.2 CERN AND NCBI integrated data analysis

Fig 4: Prototype interface
Internal constituents of DNA can be viewed in the above front-end of the prototype. This was possible with the integration of CERN data into NCBI. The above Fig. 4 is the interface of prototype implemented to design the unique identity with blockchain and peacock technology. The homo-sapiens haploid chromosome data is taken from NCBI data centre. The selection of individual or specific gene can be done by the end-user/researcher to estimate the number of possible components inside a single gene up to the sub-atomic level. This gives a better insight to any research community to understand how a gene attains a specific function at the application tier and regulate at the presentation tier.

5. Conclusion
A typical correlation between CERN and NCBI is established and integration of big data to develop a prototype and design a UID has been implemented in this present work. The obtained results have paved a path for us to understand super-symmetry and quantum realm computations. The implemented prototype enhances the future research community to understand the communications and computations of quantum realm and so that they could make their desired material acquire the required function or property. In the future, Binary code may be replaced with Peacock code.

6. Future Scope
This integration will act as a gateway for many natural computations which can be solved by natural intelligence.

Acknowledgments
This research was supported by Mr. & Mrs. Eshwar Rao, Mr. & Mrs. G.L.N. Sharma, Mr. & Mrs. Gudichuttu Sathyanarayana and Mr. & Mrs. Thallapalli Srinivas. We thank all our professors of the University campus and the reputed intuitions of India, colleagues and co-researchers from all the departments, Jawaharlal Nehru Technological University Hyderabad and the other institutions who provided insight and expertise that greatly helped our research.

References
[1] Richard, P. (1982). Feynman: Simulating physics with computers. International Journal of Theoretical Physics, 21(6/7), 476-488.
[2] C. Turkoglu,* G. Angloher, P. Bauer, A. Bento, C. Bucci, L. Canonica, X. Defay, A. Erb, F. V. Feilitzsch, N. Ferreiro Iachellini, P. Gorla, A. Gütlein, D. Hauff, J. Jochum, M. Kiefer, H. Kluck, H. Kraus, J. C. Lanfranchi, A. Langenkämper, J. Loebell, M. Mancuso, E. Mondragon, A. Münster, C.E. Pagliarone, F. Petricca, W. Potzel, F. Pröbst, R. Puig, F. Reindl, J. Rothe, K. Schäffner, J. Schieck, S. Schöner, W. Seidel, M. Stahlsberg, C. Strandhagen, R. Strauss, A. Tanzke, H. H. Trinh Thi, M. Uffinger, A. Ulrich, I. Ushero, S. Wawoczny, M. Willers, M. Wüstrich, DOI: https://doi.org/10.22323/1.314.0086 PoS EPS-HEP2017 (2017) 086.
[3] Sanger F. 1980. Frederick Sanger — Biographical. (URL http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1980/sanger-bio.html).
[4] Watson J., Crick F. Molecular structure of nucleic acids - Nature. 1953; 171: 709–756. (URL http://www.nature.com/physics/looking-back/crick/)
[5] ATLAS collaboration (2014), ATLAS Data Access Policy, CERN Open DataPortal. http://opendata.cern.ch/record/328/files/atlas-higgs-challenge-2014-v2.csv.gz