International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

High Performance Computing in Genomics and Space Science

Vedika Agarwal[1], Prof. Anitha Sandeep[2]

[1]*B.E Student, Dept. of Computer Science, R.V College of Engineering, Karnataka, India*
[2]*Assistant Professor, Dept. of Computer Science, R.V College of Engineering, Karnataka, India*

Abstract – *Earth and space science investigation (ESS) has one of the major goals as understanding, observation, prediction and modeling of effects on earth and in space. All of these stages require advanced computing. Starting from processing to archiving and then retrieving. For the extraction of meaningful information from all of these various observations, filtering this data is important and scientists are assisted by advanced algorithms for achieving the same. Also, as there is an increase in experiments involving research at the genomic level as well the advanced DNA sequencing technologies the complexity as well as the amount of biological data is increasing very rapidly. As the volume of data produced is too huge it is very difficult or better to say impossible to process such huge amount of data in standalone executions using an ordinary desktop machine. This is where the concept of High-Performance Computing in Genomics and Space Science comes into play. In this paper we tried discussing various applications of High-Performance Computing in various aspects of Genomics as well as Space Science. Furthermore, an effort has been made to give emphasis on the use of High-Performance Computing in visualizing one of the biggest Space related issues of the present times-the Space Debris.*

Key Words: **High Performance Computing, Genomics, Space Science, Space Debris**

1.INTRODUCTION

Genomics involves incorporating elements from genetics and studying the whole genome of organisms. It makes use of DNA sequencing methods as well as different combinations of recombinant DNA as well as Bioinformatics for assembling, sequencing, and analysing the function and structure of genomes. The genomic experiments taking place today has to process the so-called "biological big data" reaching the size of Petabytes and Terabytes now. If scientists use their own workstation then they may require weeks or even months for processing this huge amount of data. Parallelism techniques and high-performance computing environments can be applied to solve this problem. The various mysteries of Genetics can be unlocked by making use of High-Performance Computing. Furthermore, all the investigations happening in space science require advanced computing on all the stages. Be it because of thousands and thousands of satellites orbiting around the earth's orbit or be it because of numerous experiments taking place by astronauts and scientists in the International Space Station (ISS), all of these include a huge amount of data which needs to be properly processed in order to derive meaningful information. The search of life forms happening in Mars along with various other research needs the play of parallel computing in some way or the other.

In this work, we aim to provide an overview of the various applications of parallel and high-performance computing in various aspects of Genomics and Space Science. Additionally, we emphasis on one of the major applications of parallel computing in space debris visualization. Space debris also known as space junk refers to debris or pieces of machinery left in space by the humans. Space junk can also be hazardous to spacecraft and active satellites. If the collision risk grows to become too high then the Earth orbit could even become impassable. Last satellite that was destroyed due to collision by space junk was in 2009.

1.1 Parallel Computing in Genomics

Parallel Computing makes it possible to build and maintain deep, broad databases of real-world genetic information. Automating the large-scale analysis of genetic information with this level of big data available, can help in improving healthcare for whole populations. Since even a single genome or proteome contains an immense quality of data, performing even a simple analysis on genome-scale data quickly turns into a computationally difficult problem. Thus, High-Performance computing as well as its various related components in database systems, computer engineering and visualization is required by computational biology. The cost of sequencing human genomes has been drastically dropping over the years because of the usage of parallel computing. Unlocking more of the secrets held within the microscopic structure of DNA is an area of research which has huge potential benefits for all of us; everything from assisting us to understand, diagnose and treat diseases to boosting food production. For this to happen parallel computing has a very crucial role to play. Due to increase in the no. of experiments involving genomic research as well as advances of DNA sequencing technologies the amount and complexity of biological data is being increased. This is also directly affecting the performance achieved in the computational execution of bioinformatics experiments. As the volume of data produced is too huge it is very difficult or better to say impossible to process such huge amount of data in standalone executions using an ordinary desktop machine.

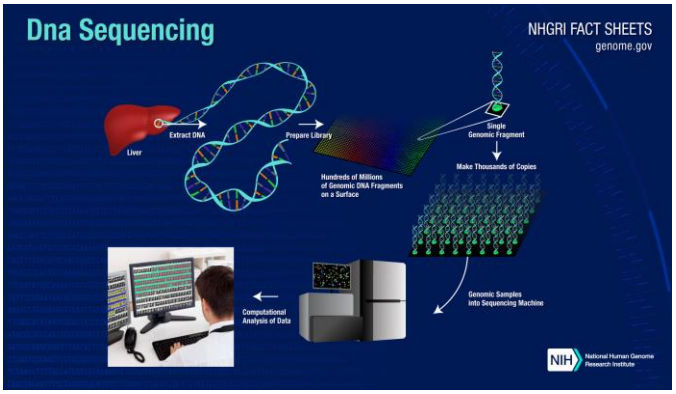

Fig-1: DNA Sequencing in National Human Genome Research Institute

1.2 Parallel Computing in Space Science

Problems associated with the catalog and improving catalog of earth orbiting satellites is achieved by Parallel and High-Performance Computing. Computations once taking many days with traditional mainframes are now being performed in only few hours. Uncorrelated target processing, conjunction prediction, and a new space catalog based on orbit determination and prediction with special perturbations method are some of the efforts underway for US Naval Space Command. Due to the chaotic nature of the orbits of the asteroids crossing the orbits of other planets as well as the orbit of the earth in order to study the probability collisions it is not possible to be computed in a deterministic way for a long-time span. Though studying the statistical behaviour over a long-time span for such large no. of orbits is still possible only provided if there are intelligent post processing software and enough computing resources available. The enormous power of High-Performance and parallel computing systems can be used to handle with the former problem. Parallel computing can help in space debris visualization which can lead to adaptation of various space debris mitigation methods.

Fig-2: High Performance Spaceflight Computing (HPSC)

2. What is Space Debris? Why is Space Debris a Problem?

Space debris also known as space junk refers to debris or pieces of machinery left in space by the humans. It may refer to comparatively small things such as paint flecks or bits of debris that might have fallen off a rocket. Or this may refer but is not limited to very big objects like dead satellites that might have been left in the orbit at the end of their missions or might have failed.

Fig-3: Space Debris or Space Junk in orbit around the earth

Space junk can also be hazardous to spacecraft and active satellites. If the collision risk grows to become too high then the Earth orbit could even become impassable. Last satellite that was destroyed due to collision by space junk was in 2009. The increase in the population of space junk is a matter of concern as this would lead to increase in danger to the spacecraft with humans aboard as well as to the International Space stations and all other space vehicles.

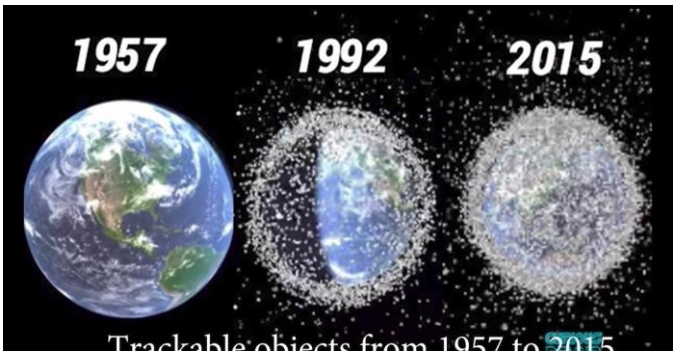

Fig-4: The rising population of space debris over the years

There is an estimated population of 500,000 for particles lying in the range of 1 to 10 cm in diameter. For the particles having their sizes larger than 1mm the count exceeds some 100 million particles. The amount of space junk or space debris orbiting around the Earth exceeded 8,000 metric tones as of January 1, 2020.

3. High Performance Computing for Space Debris Visualization

The 'Institute of Aerospace Systems', is currently developing a program that can produce a real-time graphical animated representation of simulated and actual space junk or space debris objects in the orbit of the Earth. It makes use of OpenCL as a graphics library and is written in C++. The image in **Fig 5** displays rocket bodies in yellow, satellites in red, simulated explosion fragments in blue and mission related objects in green.

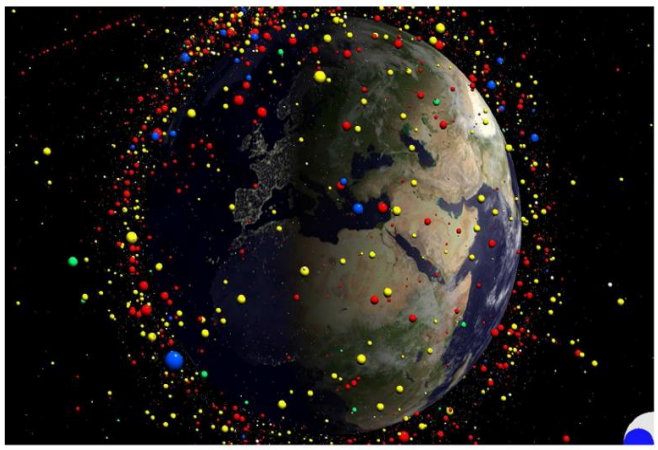

Fig-5: Simulation of Space Debris Environment with the help of Parallel Computing

The biggest advantage is it is possible to process them independently and individually due to the nature of these objects. This led to the implementation of an analytical orbit propagation algorithm to be run on GPU. A huge performance increase could be achieved by making use of its very efficient processing power compared to its CPU-based counterpart. Other than computer graphics even for other applications since many years there are several efforts being made to harness this computing power. With the increase in the acceptance of parallelism techniques software tools required for space debris simulation are among those that could really benefit and profit. OpenCL as well as other software development tools are emerging which is making it possible and aiding to transfer newer algorithms used in visualization even outside the field of computer graphics like one solid example being the simulation of the space debris environment. This way it is making use of Multi-Core -CPUs and GPUs as well as other parallel hardware for achieving faster computation.

4. High Performance Computing in Genomics and Space Science

Due to increase in the amount of data generated every single second for numerous research happening in the field of genomics and space science it is impossible to use standalone computers to derive meaningful information out of them. This is where Parallel and High-Performance Computing comes into play and becomes an utmost necessity. Genomics and Space science are two fields which hold the power to revolutionize the whole existence not only of human species but of various other species in existence, not only in earth but discovering the presence of life forms in other planetary bodies as well. For example, the search of life forms in Mars is going to be one of the biggest revolutions in the coming decades. In all these various research experiments going on in genomics as well as in space-science high performance and parallel computing has and will have a very crucial role to play. The combination of science and technology is all set to create history discovering and inventing the possibilities one might have never even imagined.

Fig-6: NASA uses TACC Super-Computers to Prepare for Next Gen Mars Rovers

5. Conclusions

To process huge amount of data generated every second due to various experiments going on in the domain of Genomics and Space Science, scientists may require weeks or even months if they use their own workstations. Parallel and high-performance computing is necessary at every step of research taking place both in earth for various life forms involving Genomics as well as in space. It has also aided in the research going on in the search of life forms at MARS. High-Performance computing has led to observation, understanding, modelling and prediction effects on earth and in space. Parallel computing can also help in solving the space debris problem with the help of space debris visualization techniques and can lead to various space debris mitigation methods.

In this work, we proposed the necessity of parallel computing in various applications of Genomics and Space Science. Both of these domains have many mysteries folded which if un-folded can lead to new evolutions for not only mankind but different life forms not only living in earth but discovering the life forms if existed in other planets as well. We also tried to discuss the space debris visualization methods and how High-Performance Computing can be a blessing in that. In Future this work will be further extended

to discover various other applications and discoveries in the field of Genomics and Space Science taking into the advantages of parallel and high-performance computing.

REFERENCES

[1] Kary Ocana and Daniel de Oliveira; Parallel Computing in Genomic Research: Advances and Applications, Nov 2015, Research Gate.

[2] Marek Moeckel; C. Wiedemann; Trappe, Using Parallel Computing for the display and Simulation of the space Debris Environment, July 2011, Research Gate.

[3] Ming Shen; Xiaozhong Guo; Pengqi Gao; Datao Yang; You Zhao; Zhongwei Fan; Application of Parallel Computing for Space Debris Close Approach Analysis, Dec 2012, IEEE Explorer.

[4] Miguel A. Vega-Rodriguez & Jose M. Granado-Criado; Parallel Programming in Bioinformatics: Some Interesting Approaches, Oct 2018, International Journal of Parallel Programming.

[5] George Lentaris; Konstantinos Maragos; Ioannis Stratakos; Lazaros Papadopoulos; High-Performance Embedded Computing in Space: Evaluation of Platforms for Vision Based Navigation, Feb 2018, Journal of Aerospace Information-Systems.

BIOGRAPHIES

Vedika Agarwal, B.E. (RVCE, VTU), Department of Computer Science & Engineering. She is a passionate student exploring the mysteries of Genomics and Space Science with the use of advances in technology.

Prof. Anitha Sandeep, Assistant Professor, Department of CSE, RV College of Engineering, Bangalore. She has 16 years of experience and has guided many UG and PG students. She has a no. of different National and International Publications.