3 Study of exascale computing: Advancements, challenges, and future directions

Neha Sharma[a], Sadhana Tiwari, Mahendra Singh Thakur, Reena Disawal and Rupali Pathak

Prestige Institute of Engineering Management and Research, Indore, India

Abstract

Exascale computing is the high performance computing system that can measure quintillion calculations per second. It is capable to perform the calculations of 1018 floating point operations (FLOPS) per second. It is the term given to the next 50–100 times increased speed over very fast super computers used today. High performance computing application helps to simulate large scale application, machine learning, artificial intelligence, industrial IoT, weather forecasting, healthcare industries and many more. The increased computational power will enable researchers to tackle more complex problems, collects and analyze larger data sets, perform simulations with high accuracy and resolutions. Exascale computing has the power to transform scientific research, spur innovation, and tackle complex issues that were previously computationally impractical. This paper describes a brief description, architecture and various applications of exascale computing such as healthcare, microbiome analysis, etc. This paper also presents the future and research aspects of exascale computing.

Keywords: High performance computing, exascale computing, super computers, parallel processing, data analytics, computer architecture

I. Introduction

High performance computing (HPC) technology affects almost every sphere of our life that includes education, communication, entertainment, economy, engineering and science, etc. The next stage of HPC is known as exascale computing (EC), where computer systems is capable to perform the calculation at least 10^{18} floating point operations per second (1 exaFLOPS) or a billion (**i.e. a quintillion**) **calculations per second.**

In comparison to existing petascale systems, it represents a huge increase in computational capability. It is thousand times faster than petaflops machines (Huang et al., 2019; Matthew et al., 2020). EC will enable simulations and analyze previously unheard of complexity and scope, which will revolutionize scientific research, engineering, and data analytics (Fabrizio et al., 2019; Matthew et al., 2020).

EC is developed with the increasing demands of scientific and industrial applications. Because these applications requires large computational capability to solve complex problems in areas like astrophysics, materials science, energy research, climate modeling, and more. These applications generate large amount of data and requires complex simulation. These simulations demand extremely high level of processor power, memory capacity, and storage bandwidth.

To achieve exascale processor, significant advancement is required in computer architecture, system design, software development and energy management. Exascale systems are designed to get over the drawbacks and difficulties that current HPC systems experience, including high power usage, memory and storage bottlenecks, limited programmability, and scalability as shown in Table 3.1.

Exascale computer provides extra ordinary power and memory so that it can be applied in HPC areas like large scale simulation, machine learning, deep learning, and multi physics (Francis et al., 2020; Matthew et al., 2020; Choongseok et al., 2023). EC has done lots of improvement in scientific, medical, weather forecasting, and artificial intelligence (Francis et al., 2020; Yuhui et al., 2022).

Media governments such as India, US, EU, China, Japan, etc., and industries such as IBM, Intel, etc., together are putting their efforts to build exascale computers. There is competition between United States and China to become the first nation that has an exascale computer. The estimated cost of this exotic computer equipment will be in between $ 400 million and $ 600 million. Aurora 2021 (A21) (Matthew et al., 2020) is therefore first US exascale system.

Components of EC

a. **Processor:** Homogenous or heterogeneous platforms can be used for designing of exascale systems. In heterogeneous, exascale uses CPU and GPUs to improve the performance efficiently.
b. **Memory requirement:** In order to meet the performance requirement, exascale needs high bandwidth memory (HBM). HBM stack can contain

[a]nsharma@piemr.edu.in

Table 3.1 Technological overview of exascale system

Parameter	2009	2018 Swimlane 1 (extrapolation of multi-core design)	2018 Swimlane 2 (represent the GPU design point)
Power	6 MW	~20 MW	Same as SL1
Memory	0.3 PB	32–64 PB	Same as SL1
Node performance	125 GF	1.2 TF	10TF
Latency	1–5 µs	0.5–1 µs	Same
Memory Latency	150–250 clock cycles (~70–100 ns)	100–200 clock cycles (~50 ns)	Same
Node memory BW	25 GB/s	0.4 TB/s	4–5 TB/s
Storage	15 PB	500–1000 PB	Same
System size (nodes)	18,700	1 M	100,000
IO	0.2 TB	60 TB/s	Same as SL1

up to eight Dynamic Random Access Memory (DRAM) module which are connected through two channels per module. It includes silicon interposer base die with a memory controller and interconnected through-silicon via (TSVs) and microbumps. Double Data Rate (DDR) memories are generally off-chip dual-in line memory modules means they are separated from CPU die. HBM offers low latency and has high throughput as compared to DDR because it is close to the processor die.

II. Related work

Enormous research is going on HPC technology to improve the performance of high speed application. In 2018, exascale system was introduced which performs calculation of 10^{18} FLOPS (Matthew et al., 2020). Exascale system helps to simulate high speed applications such as healthcare industry, industrial IoT, data analytics and many more (Levent Gurel et al., 2018; Tanmoy et al., 2019; Francis et al., 2020). Tanmoy et al. (2019) described that how artificial intelligence (AI), Big data and HPC helps to discover new drug with reduce cost and minimize development cost. Francis et al. (2020) explored the role of EC in different areas such as microbiome analysis, healthcare industry, chemistry and material applications, data analysis and optimization applications, energy application, earth and space science applications and many more. Tanmoy et al. (2019) explained how EC technique and AI helps to predict the cancer and tumor response in advance. Exascale computing enables engineers and researchers to design, optimize, and test new products and technologies more efficiently and quickly. In his paper, L. Gurel et al. (2018) reviewed that contribution of EC in autonomous driving and how EC reduces the software complexity with available hardware. EC has various technical challenges such as power consumption, memory management, parallelism, fault tolerance, and scalability (John et al., 2011; Pete et al., 2012; Judicael et al., 2015; Mahendra et al., 2020; Matthew et al. 2020). In literature, authors Matthew et al. (2020), Maxwell et al. (2021), Francis et al. (2020), John et al. (2011) have discussed various benefits, opportunities and challenges in EC. Fabrizio et al. (2019) reviewed the political and social aspects of exascale computing along with history of HPC architecture. Peter et al. (2013) and Martin et al. (2019) have explained the requirement analysis of exascale based on cases use. Author has described reference architecture and technology-based architecture of the process project in EC. Martin et al. (2019) have proposed novel hardware designs and architectures that can deliver exascale performance while maintaining energy efficiency and reliability. Peter et al. (2013), Martin et al. (2019) authors summarized the different challenges in operating system such as technical, business and social for exascale system. This includes research on resource management, job scheduling, power management, fault tolerance.

III. Architecture of EC

In view of the requirement of different industry, the architecture of exascale is divided in to three groups: virtualization, data and computing requirement (Peter et al., 2013; Martin et al., 2019). The exascale computing architecture is shown in figure 3.1.

In virtualization layer, virtualization requirements are taken directly from the application basis of our user communities-container support that provides lightweight virtualization method similar to app packages. Advantages of this technique are flexibility, reliability, ease of deployment and maintenance. User applications require to be distributed across a

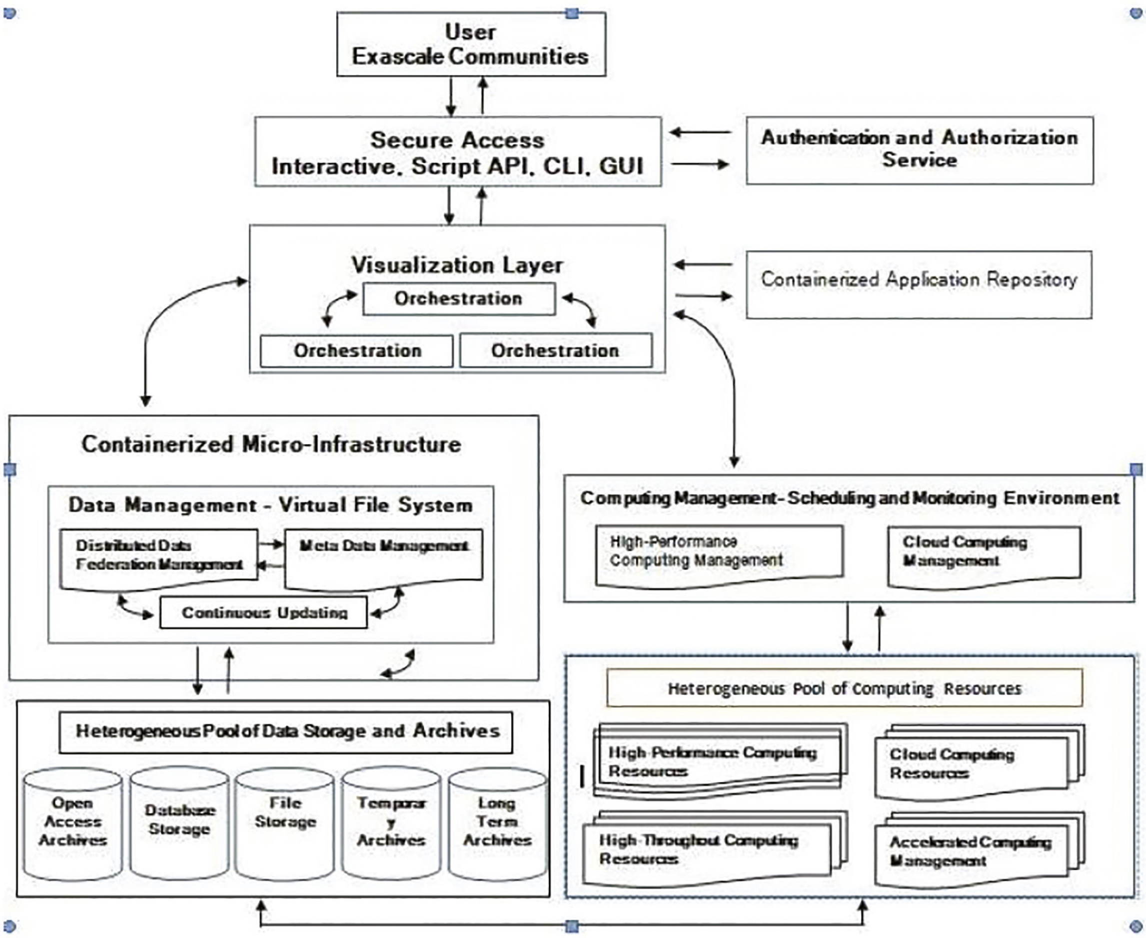

Figure 3.1 Exascale computing architecture

variety of computer infrastructure, portability and collaboration.

The primary requirement is to manage exascale data sets or excessive data flow, it is impossible to alter and manage at a single data center. It is also integrated with Meta data management. Depending on the data services, data connection or data transfer is big challenge. The exascale platform should support large data transfer in all infrastructures (Martin et al., 2019).

The computing requirement is needed that can support all HPC, cloud computing and speed requirement. Aim of current scientific application is huge data distribution at all computer research centers or sites. As a result, degree of parallelism and concurrency is also increased (J. Singh et al., 2009; Jiangang et al., 2021). These requirements need to be fulfilled while designing computing requirement. In continuation, computing architecture is proposed based on modularity and scalability. These two approaches are useful in high degree parallelism and high distribution. It offers flexibility to be used small modules and method that exploits various sources of exascale systems efficiently (Martin et al., 2019).

IV. Key technology for EC

The essential technologies needed for EC are discussed in the following section.

A. HPC system

Since EC handles large amount of data (10^{18} FLOPS) and computational workloads (Matthew et al., 2020). Therefore, it includes numerous linked processing units, such as CPUs, GPUs, or specialized accelerators (Thiruvengadam et al., 2017).

B. Parallel processing

EC strongly relies on parallel processing, which involves running numerous computer processes concurrently to achieve high throughput. This entails decomposing complicated issues into simpler issues so that numerous processing units can handle them simultaneously (Matthew et al., 2020).

C. High speed processor

EC necessitates the creation of advanced processor that can supply the necessary levels of computational power. This can entail utilizing heterogeneous architecture, which pair conventional CPUs with specialized accelerators like GPUs or FPGAs (Thiruvengadam et al., 2017).

D. Memory
To manage the enormous amount of data required for EC, large-capacity, fast memory and storage devices are needed. Improvements in random access memory (RAM), high-speed cache, and storage technologies like solid-state drives (SSDs) or non-volatile memory are all included in this (Matthew et al., 2020).

E. Energy efficiency
EC systems use a lot of electricity, so energy efficiency is important. Sustainable energy source are essential for EC. To address the power and thermal concerns, this entails creating low-power CPUs, optimizing algorithms, and using cutting-edge cooling techniques.

F. Software model
It is essential to create software and programming models that effectively make use of the extreme parallelism and diverse architectures found in exascale systems. This includes providing tools for managing and debugging intricate software systems, as well as optimizing algorithms and constructing parallel programming frameworks.

G. Data management and analytics
EC generates large amount of data that need to be managed and analyzed. To get useful insight from the enormous datasets produced by exascale simulations and computations, effective data management approaches, including data storage, retrieval, analysis, and visualization, are required.

V. Emerging applications of EC

There are various applications of EC that is shown in Figure 3.2 and the detail descriptions are given below:

A. Advances in healthcare (accelerating drug discovery with AI, HPC and big data)
The current state of drug development is a long, expensive process and, to some extent, a shot in the dark. The cost of developing even a single drug is high. According to research by the Tufts Center, the cost of drug development was found to be more than $2.5 billion.

There are many different healthcare sectors, such as the pharmaceutical industry, and many more are struggling to develop new drugs, and patients are also waiting for new drugs to improve their medical condition. AI, cloud computing, IoT and Big data aim to shorten development time and reduce costs at every step of the new drug development chain, from initial research to clinical trials. Different emerging technologies help scientists do retrospective analysis on existing data analytics, also help find new drugs for disease. AI that runs through large amounts of genetic data to determine the correlation between a particular DNA sequence and a disease that will help identify potentially useful drugs. Once this process is complete, AI uses electronic media recording to identify potential drug for the target audience and enable the industry to develop setup and put drugs into trials (Tanmoy et al., 2019; Francis et al., 2020).

Traditionally, multiple clinical trial phases are required once the most promising drugs have been identified, which becomes time-consuming, demanding and costly. Data analytics, IoT and cloud computing already offer benefits here and promise to bring more in the future. Wearable and implantable IoT

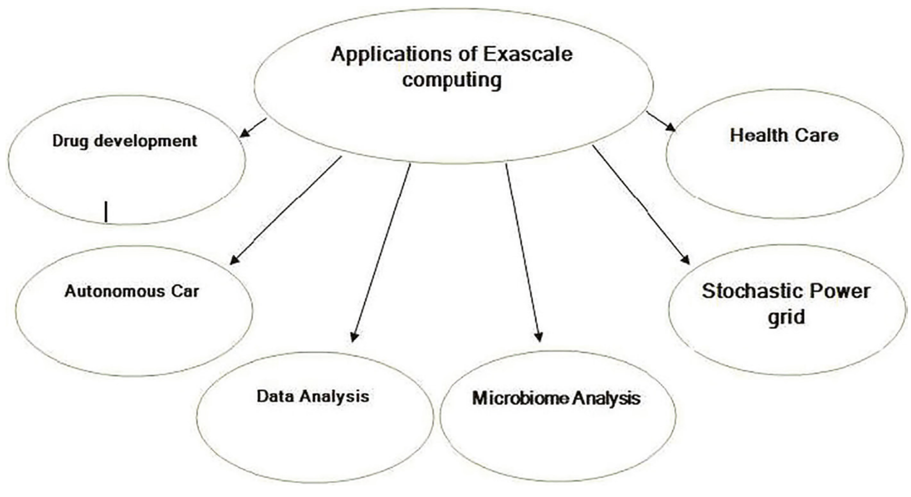

Figure 3.2 Application of exascale computing

devices collect enormous amounts of patient information from sensors and data storage in the cloud. The cloud provides large storage that is cheaper and requires high computing power that assists in the data analysis process. In short, accelerating drug discovery with AI, HPC and Big data (Francis et al., 2020):

- Current processes for drug discovery are time consuming and expensive.
- Cutting-edge technologies such as artificial intelligence, HPC and Big data will reshape method of drug discovery (Tanmoy et al., 2019).
- Requires high computing hardware power results for the ability to model further drug progress before moving on to clinical trials.

B. Dynamic stochastic power grid

ExaSGD application is used to preserve the integrity of power grids and address load imbalances. With the help of this programe, the grid's real-time response optimization against probable disruption occurrences is created using models and algorithms. ExaSGD serves power grid operators and planners and is based on exascale computing.

Power grids keep the supply and demand for electricity in balance. Attacks on the grid, whether physical or digital, can result in costly power grid components being permanently damaged or experiencing large-scale blackouts. Load shedding is utilized to prevent generation-load imbalance and maintain the functionality of the power grid (Francis et al., 2020).

Cyber-enabled control and sensing, plug-in storage devices, censored elements, and smart meters managed automatically and remotely can all have an impact on how the electrical grid behaves. To avoid generation and load shedding at the moment, load shedding is employed. Using simulations, the ExaSGD tool offers additional ideal configurations for resolving generation-load imbalance. This method enhances the electricity grid's ability to recover from various risks (Francis et al., 2020).

C. Deep learning (DL) enabled the precise cure for cancer

Project "CANDLE application" was started by the DOE and NCI (National Cancer Institute) of the NIH (National Institutes of Health). The goal of this project is to develop CANDLE (Cancer Learning Area), an amazing and in-depth learning environment for exascale programs. Three key challenges are being addressed by the CANDLE programe (Tanmoy et al., 2019; Francis et al., 2020):

1. Find a solution for the drug response issue and create models for predicted drug responses. These models can be expanded upon in order to enhance pre-clinical drug testing and accelerate cancer patients' access to drug-based therapies.
2. RAS (Rat sarcoma virus) pathway issue 2.
3. Planning for a treatment approach.

In order to predict treatment response, complicated, indirect interactions between drug structures and tumor structures are captured using supervised mechanical learning techniques to address drug responses. Using the history of past simulations, the RAS technique uses multi-tasking to search a large-scale space to define the scope of a series of simulations. Machine learning (ML) models are used to automatically read and compile millions of clinical records in order to deal with the treatment approach. Direct conclusions about are provided by ML models. Every issue calls for a distinct approach for to integrate the learning, yet they are all supported by the same CANDLE environment.

Python library, the runtime manager, and a set of deep neural networks are all included in the CANDLE package. Tensor Flow, PyTorch, and deep neural networks that download and represent three issues are employed for exascale computing, with a runtime supervisor organizing the distribution of work throughout the HPC system. Performance features include semi-automated uncertainty quantification, large-scale search for hyper parameters, and automatic search for best model performance.

Exascale challenges are represented in the urgent requirement to train many related models. Each test application's demand results in cutting-edge models that span the speculative space (which is not specific to the idea of an accurate medicine).

D. Microbiome analysis

Microbial species are important part of our ecosystem. They are influencing various domains such as agricultural production, pharmaceutical and also used to make oils, medicines and other products. To study and gather information about the microbe's genome, sequence methods are used. In genome sequencing, Metagenomics data are larger and more plentiful results in increased cost of computation. As a solution, the ExaBiome application develops data integration tools with high computing power (Francis et al., 2020).

Metagenomics is a domain that explores functional and structural details of the microbiome. Metagenome integration, protein synthesis and signature-based methods are three major computational problems faced in bioinformatics domain. ExaBiome attempts to provide measurable tools for above stated problems. Metagenome integration means capturing raw data sequences and generates long gene sequences and signature-based methods enable comparable and effective metagenome analysis (Francis et al., 2020).

MetaHipMer, a well-known metagenome compiler created by the ExaBiome team, scales thousands of computers in contemporary petascale-class architecture. Additionally, a sizable ecological database has been created. To take advantage of the chance for enhanced node compatibility with memory structures, including GPUs, work is being done on measurable upgrades across nodes and node level improvements. With other collaborators, MetaHipMer exhibits competitiveness. The second long-term compiler is also being developed and has a significantly larger computer density, making it well suited to exascale systems even though MetaHipMer is made for short reading data (Illumina) and is meant for long-term data. HipMCL, the second code from ExaBiome, offers a way to measure proteins. The structure of protein families in the billions of proteins may be seen thanks to HipMCL, which has thousands of nodes. These codes are based on typical compound patterns with flexible character unit (DNA or protein) algorithm alignment, minimal layout, calculation, and analysis of fixed-length strands, as well as a range of graphs and small matrix techniques. Metagenome integration is core of the ExaBiome complicated challenge, but that capability will make it simpler for new bioinformatics problems to emerge (Francis et al., 2020).

E. Analysis of data for free electron laser

X-ray diffraction is used by the Linac Coherent Light Source (LCLS) at the Stanford Linear Accelerator Centre (SLAC) to model individual atoms and molecules for crucial scientific activities. The representation of molecular structure revealed by X-ray fragmentation in close to real time will need for previously unheard-of computer compression scales and bandwidth data techniques. Data detector measurements in light sources have substantially increased; after LCLS-II-HE development is complete, LCLS will grow its data by three orders in magnitude by 2025. The ExaFEL programme uses exascale computation to accelerate the process of reconstructing molecular structures from X-ray diffraction data from weeks to minutes (Francis et al., 2020).

Users of LCLS demand an integrated approach to data processing and scientific interpretation which calls for in-depth computer analysis. Exascale processing capacity will be needed to meet demand for real-time analysis of the data explosion which will take about 10 minutes (Francis et al., 2020).

Because of its high repetition rate and brightness, LCLS can map individual molecules' inherent fluctuation in relation to flexibility and ascertain their composition. The cornerstone for comprehending engineering structures, materials, and energy science is structural strengths and heterogeneities, or conformational mutations in macromolecules. Single-particle imaging (SPI) and X-ray scattering variation, which are non-crystalline based diffractive imaging techniques, may see and analyses these structural heterogeneity and variations. This characteristic encourages interest in the creation of X-ray free-electron lasers. Effective data processing, fragmentation patterns, and reconstruction of 3D electron cones, however, enable the visualization of structural changes over time (Francis et al., 2020).

The problem with ExaFEL is to devise an automatic analysis pipeline for single-part imaging using different techniques. This requires the reconstruction of a 3D cell structure from 2D separating images. This conversion is done by new Multi-Tiered Iterative Phasing (M-TIP) algorithm.

Diffraction images from distinct particles are gathered in SPI. The production of molecules (or atoms) and cohesive areas (or comparable particles) under specific operating circumstances is also assessed using these diffraction images. Since the shapes and conditions of the particles in the image are unknown and heavily contaminated by sound, determining properties using the SPI test is challenging. Additionally, the quantity of accessible particles typically places a cap on the number of viable images. To determine the form, areas, and molecular structure from a single particle's data obtained utilizing structural barriers simultaneously, the M-TIP algorithm uses a duplicate guessing framework. Additionally, it aids in the comprehensive information extraction from single-particle diffraction.

A quick response is necessary to direct the test, ensure that enough data is gathered, and modify the sample concentration to obtain a single particle rate. Together, exascale computing power and HPC processes can handle the analysis of the expanding data explosion. As a result, researchers will be able to analyses data quickly, respond quickly to test-quality data, and simultaneously decide on a three-dimensional sample design.

F. Autonomous car

Self-driving vehicle will generate and use a variety of data to analyze various parameters such as location, road condition, and passenger safety. To manage all the data, you need HPC (Levent Gurel et al., 2018).

The car is equipped with sensors, embedded computers, cameras, high-precision GPS and satellite, wireless network, 5G connectors to connect to the internet. Autonomous car will exchange data with the management and control system and will sync with a large database that continuously provides real-time information such as weather, traffic conditions, emergency alerts, etc.

Autonomous car will generate a large amount of data and will send more than four terabytes of data per hour to the cloud. Exascale high performance computing and Big data are therefore capable of delivering the computing power required to use predictive decision support systems to evaluate large amounts of data.

VI. Benefits of EC

The speed of EC is 50 to 100 times faster than latest supercomputer. Therefore, this kind of HPC application helps to simulate large scale application, ML and AI, etc. (Matthew et al., 2020; Maxwell et al., 2021). It is fast, and cost effective. As a result, intelligent storage capacity, computing power can be applied in industries like health care, chemical, National Security, reducing pollution, and many more (Francis et al., 2020). EC helps to minimize health issues, and proves the better quality of life by optimizing the transportation facilities. In short, EC has a number of advantages that could revolutionize fields including engineering, society, and scientific study is shown in Figure 3.3. Some advantages of exascale computing are:

A. Scientific discovery
EC enables scientists and researchers to run simulations and models at a scale and resolution that have never been possible before. This may result in fresh scientific understandings, discoveries, and a better comprehension of intricate processes. Exascale simulations can facilitate discoveries and speed up scientific development in areas including climate modeling, astronomy, materials research, and computational biology (Francis et al., 2020).

B. Accelerated innovation
EC enables engineers and scientists to swiftly and efficiently build, optimize, and test new products and technologies. It enables rapid innovation in fields including aerospace, automobile design, energy systems, and material research by allowing for the study of a broad design space. EC aids in the identification of optimal designs, resulting in improved products and solutions, by modeling and analyzing complicated systems.

C. Advances in data analytics and AI
EC enables the processing and analysis of enormous datasets in real-time, opening up new opportunities in data analytics and AI. It makes possible for DL and machine learning models to be more accurate and effective, which advances fields like genomics, personalized medicine, social network analysis, autonomous systems, and recommendation systems (Tanmoy et al., 2019; Francis et al., 2020). EC facilitates the extraction of useful insights from enormous amounts of data, fostering innovation and decision-making.

D. Cross-disciplinary collaboration
EC fosters cross-disciplinary cooperation among scholars. Exascale systems' computational capacity and resources can be used by scientists, engineers, and subject-matter specialists to tackle challenging issues that call for interdisciplinary solutions (R. Arya et al., 2021). Through information exchange and integrated problem-solving, this partnership may result in advances in areas like fusion energy, drug development, urban design, and computational social sciences.

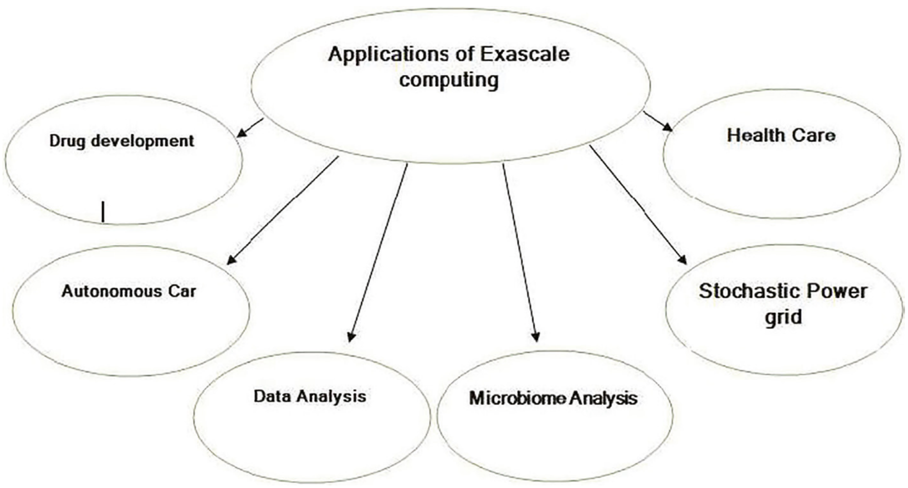

Figure 3.3 Benefits of exascale computing

E. Precision and realism

EC allows for simulations and modeling with a level of accuracy and realism never before possible. Exascale simulations deliver more precise results by including complex interconnections and finer-grained details. This improves decision-making processes, which helps in better forecasts, and encourages the creation of trustworthy and durable systems and technologies.

F. Economic and social impact

EC holds the promise of fostering both societal and economic improvement. By quickening the pace of product development cycles, enhancing efficiency, and cutting costs, it encourages innovation and supports industries. By offering strong tools for modeling, analysis, and optimization, EC also helps to address major issues like climate change, healthcare, and sustainable energy.

G. Advances in computation

EC promotes improvements in computational methods and algorithms. To efficiently utilize the processing capacity of exascale computers, researchers investigate novel algorithms, optimization techniques, and parallel programming paradigms. Beyond exascale computing, these developments help other computer platforms and allow for further development in HPC.

VII. Challenges in EC

Exascale are facing different technical and social challenges (John et al., 2011; Pete et al., 2012; Judicael et al., 2015; Mahendra et al., 2020; Matthew et al., 2020) are shown in Figure 3.4. These challenges are as discussed in the following sections.

A. Technical challenges

Exascale system has identified four key challenges: increased number of faults, power requirement minimization, memory management and parallelism at node level. These challenges are directly related to exascale OS/R (operating system and runtime software) layer. Hardware complexity, resources challenges within OS, programming model, design issues are few more to handle.

i) **Resilience**
As the numbers of components are increasing on chip, the numbers of faults are also increases. These faults cannot be protected by other error detection and correction technique. Timely propagation fault notification across large network in limited bandwidth scenario is very difficult.

ii) **Power management**
It is one of the critical challenges of exascale system. It requires 20–30 MW to run any application. Resources can be change at any time to adopt power requirement.

iii) **Memory hierarchy**
New memory technology emphasizes on reducing the power cost while data are transferring between different nodes. In case of exascale system, the OS provide more support to runtime and application management and as the complexity reduces the OS overheads.

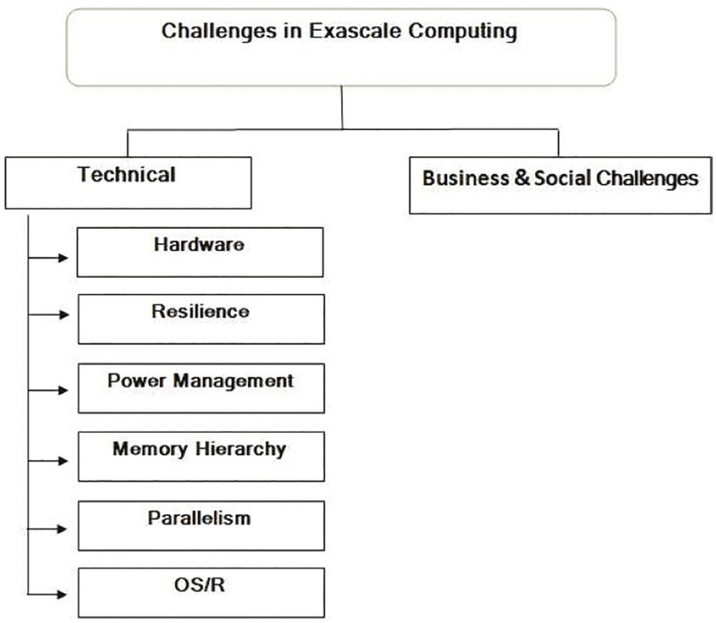

Figure 3.4 Challenges of exascale computing

iv) **Parallelism**
Exascale system performs the calculation of 10^{18} FLOPS. In order to achieve these calculations, application performs billions of calculation with in a second. This situation is significant challenge for developers. Performance cannot be increased through additional clock scaling but additional parallelism is required in order to support next generation of systems. But OS again faces different challenges like efficient and scalable synchronization, scheduling, scalable resource management, global consistency, coordination, and control.

v) **Hardware-related challenges**
The hardware resources are heterogeneous in nature. It includes multiple types of memory and different processing element. Different type of component may have different performance characteristics though they have capabilities to perform same function. Therefore, allocation of resources becomes complicated to perform the calculation. Handler establishment process for hardware event is more difficult in order to support responsiveness to faults, application monitoring system, and energy management and so on.

vi) **OS/R structural challenges**
The new operating system for exascale system is facing different challenges like misalignment of requirements, user-space resource management and parallel OS services.

a. **Misalignment of requirements:**
The interference of OS should be minimized while at the same time OS should provide necessary support to all other application.

b. **User-space resource management**
Generally OS directly manages and controls all resources to different application software. However different programming models and application requires different resources results in inefficiency. In the future application software and runtime will require increased control over resources like core, memory, power and so on.

c. **Parallel OS services**
OS performs parallel processing; effective support and development for this interface is difficult.

vii) **Legacy OS/R issues**

- The node operating and runtime systems for an exascale system will need to be highly parallel, with minimal synchronization. Legacy operating and runtime systems tend to be monolithic, frequently assuming mutual exclusion for large portions of the code.
- The node operating and runtime systems for an exascale system will need to support tight interaction across sets of nodes (enclaves). Legacy operating and runtime systems have been designed to impose strict barriers between nodes.
- The node operating and runtime systems for an exascale system will need to be right-sized in order to meet the needs of the application while minimizing overheads. Legacy operating and runtime systems do not emphasize restructuring to match the needs of an application.

B. Business and social challenges
In spite of having technical excellence, exascale system is facing different challenges in business like lack of transparency from vendors, sustainability and portability, preservation of existing code base and so on.

VIII. Future and research aspects of EC

EC has the ability to significantly advance technological innovation, scientific discoveries, and societal concerns. The following are some crucial elements of EC's future and research potential (Tanmoy et al., 2019; Francis et al., 2020; Maxwell et al., 2021; Yuhui Don et al., 2022). The market growth of EC is represented in the table 3.2.

A. Simulation and modeling

EC will make it possible for scientists and researchers to run simulations and models at previously unheard-of scales and resolutions. This encompasses disciplines like computational biology, astrophysics, materials science, and quantum mechanics, among others. New scientific insights and discoveries will be made as a result of the ability to mimic and examine complicated phenomena in greater detail.

B. Data analytics and AI
EC will revolutionize these fields by making it possible to handle and analyses enormous datasets in real time. Applications in fields such as genetics, personalized medicine, social network analysis, intelligent cities, and autonomous systems are included in this. Exascale computing and AI techniques have the potential to revolutionize industries and spur innovation.

C. Multi-disciplinary research collaboration
EC will promote cross-disciplinary research cooperation. EC systems will enable scientists, engineers, and subject matter experts to solve complicated issues that call for interdisciplinary solutions. This partnership may result in innovations in industries like fusion energy, medication development, materials design, and urban planning.

D. Machine learning and DL
Exascale computing will make it possible to train and use deep learning and machine learning models

Table 3.2 Represents the market growth of EC

Criteria	Details
Reference year	2020
Forecast period	2021–2022
Revenue forecast in year 2028	USD 50.3 billion
Growth rate	CAGR of 6.3% for the year 2021–2028
Regions covered	North America, Europe, South America, Asia Pacific, Middle East and Africa
Profiles of significant players	Advanced micro devices (US), Intel (US), HPE (US), IBM (US) , Lenovo (China), Nvidia's (Japan), NEC Corporation
Portion covered	Through computation, devices, type, deployment, size, price, organization and server

that are more sophisticated and complicated. This will open up new opportunities in fields including speech and image recognition, natural language processing, robotics, autonomous cars, and recommendation engines. Researchers will investigate new architectures and algorithms to take use of exascale capabilities for more precise and effective machine learning.

E. Computational fluid dynamics

Engineers will be able to model and optimize fluid flow in unprecedented detail thanks to exascale computing's enormous impact on computational fluid dynamics (CFD) simulations. This has uses in environmental engineering, energy systems, automotive design, and aerospace. Higher resolution and more accurate simulation and analysis of complicated flow dynamics can result in better designs and more effective systems.

F. Quantum computing

EC has the potential to be extremely important for the growth and development of quantum computing. Exascale systems can be a great resource for expediting quantum research and applications because they can provide the enormous processing capacity that quantum simulators and quantum algorithms demand. This includes creating quantum-enabled algorithms for application in real-world situations, optimizing quantum algorithms, and simulating quantum systems.

G. Hardware and software innovation

Ongoing research and development of new hardware architectures, memory technologies, interconnects, and software frameworks will be necessary for exascale computing in the future. To fully utilize the capabilities of exascale systems, research will concentrate on enhancing energy efficiency, fault tolerance, scalability, and programmability.

IX. Market analysis of EC

Compound annual growth of EC is going to 6.3% throughout the course of the forecast and it is expected that market growth will reach USD 50.3 billion by 2028. This growth is driven by the increasing demand for HPC across industries such as healthcare, finance, energy, weather forecasting, and scientific research. EC is being actively embraced by numerous sectors to solve challenging computational issues and gain a competitive edge. For the instance, it provides sophisticated simulations for drug discovery, genomics, and personalized treatment in the healthcare industry. It supports high-frequency trading, risk modeling, and portfolio optimization in the financial sector. EC is also used by energy corporations for seismic imaging, reservoir modeling, and energy production optimization. EC is strategically important, and governments around the world are actively promoting its development. To speed up exascale computing research and deployment, numerous nations, including the United States, China, Japan, and European nations, have started national initiatives and funding programmes. These programmes seek to promote governmental, academic, and commercial cooperation in order to progress technology and preserve competitiveness. Several companies and organizations are leading the exascale computing such as Hewlett Packard Enterprise, IBM, Intel, NVIDIA, AMD, and Cray. Universities, research organizations, and national laboratories all contribute significantly to the development of exascale computer systems.

X. Conclusion

EC is the new frontier of HPC technique. It has capability to achieve performance of ExaFLOPS in terms of power and cost constraint. High computational capability is able to tackle various challenges such as scientific, medical, various social aspects and engineering. It includes various research areas such as system architecture, software and programming models, performance optimization, energy efficiency, resilience and fault tolerance, Big data analytics, and application-specific research. To fully utilize the capabilities of exascale systems, researchers are concentrating on creating innovative hardware architectures, implementing efficient algorithms, investigating new programming paradigms, and optimizing system performance.EC has a bright future ahead of it. It will promote multidisciplinary research collaboration, improve scientific discovery, enable advances in AI and data analytics, revolutionize fields like computational fluid dynamics and quantum computing, and ignite innovation across sectors. However, there are obstacles in the way of EC full potential. Power consumption, memory constraints, communication bottlenecks, and the complexity of programming for massively parallel systems are issues that researchers and engineers must overcome. Moreover, to overcome technical obstacles and assure the successful deployment of these potent systems, the development of EC necessitates close cooperation between academics, industry, and government.

References

Huang, H., Li-Qian, Z., YuTong, L. et al. (2019). An efficient real-time data collection framework on petascale systems. *Neurocomput.*, 361(7), 100–107. https://doi.org/10.1016/j.neucom.2019.06.039.

Matthew, N. O., Sadiku, Awada, E., and Musain, S. M. (2020). Exascale computing (Supercomputers): An overview of challenges and benefits. *J. Engg. Appl. Sci.*, 15(9), 2094–2096. DOI: 10.36478/jeasci.2020.2094.2096.

Gagliardi, F., Moreto, M., and Mateo Valero, M. O. (2019). The international race towards Exascale in Europe. *CCF Trans. HPC*, 1, 3–13. https://doi.org/10.1007/s42514-019-00002-y.

Kogge, P. and Shalf, J. (2013). Exascale computing trends: Adjusting to the new normal or computer architecture. *Comput. Sci. Engg.*, 15(6), 16–26. DOI: 10.1109/MCSE.2013.95.

Bobák, M., Hluchy, L., Belloum, A.S., Cushing, R., Meizner, J., Nowakowski, P., Tran, V., Habala, O., Maassen, J., Somosköi, B. and Graziani, M. (2019). Reference exascale architecture. In 2019 15th International Conference on eScience (eScience) 479–487. IEEE. doi: 10.1109/eScience.2019.00063.

Alexander, Francis, Ann Almgren, John Bell, Amitava Bhattacharjee, Jacqueline Chen, Phil Colella, David Daniel et al. (2020). Exascale applications: skin in the game. *Philosophical Transactions of the Royal Society* A 378(2166): 20190056. 1–31 doi: https://doi.org/10.1098/rsta.2019.0056.

Bhattacharya, Tanmoy, Thomas Brettin, James H. Doroshow, Yvonne A. Evrard, Emily J. Greenspan, Amy L. Gryshuk, Thuc T. Hoang et al. (2019). AI meets exascale computing: advancing cancer research with large-scale high performance computing. *Frontiers in oncology*, 9, 984. DOI: 10.3389/fonc.2019.00984.

Gürel, Levent (2018). Towards Exascale Computing for Autonomous Driving. *In 2018 International Workshop on Computing, Electromagnetics, and Machine Intelligence (CEMi)*, 17–18. IEEE, 2018. doi: 10.1109/CEMI.2018.8610529.

Beckman, P., Brightwell, R., de Supinski, B. R. et al. (2012). Exascale operating systems and runtime software report. US Department of Engineering. https://doi.org/10.2172/1471119.

Zounmevo, Judicael A., Swann Perarnau, Kamil Iskra, Kazutomo Yoshii, Roberto Gioiosa, Brian C. Van Essen, Maya B. Gokhale, and Edgar A. Leon. (2015). A container-based approach to OS specialization for exascale computing. *In 2015 IEEE International Conference on Cloud Engineering*, 359–364. IEEE, 2015. doi: 10.1109/IC2E.2015.78.

Gao, Jiangang, Fang Zheng, Fengbin Qi, Yajun Ding, Hongliang Li, Hongsheng Lu, Wangquan He et al. (2021). Sunway supercomputer architecture towards exascale computing: analysis and practice. *Science China Information Sciences*, 64(4): 141101. doi: https://doi.org/10.1007/s11432-020-3104-7

Arya, R., Singh, J., and Kumar, A. (2021). A survey of multidisciplinary domains contributing to affective computing. *Comp. Sci. Rev.*, 1–9. https://doi.org/10.1016/j.cosrev.2021.100399

Chang, C. et al. (2023). Simulations in the era of exascale computing. *Nat. Rev. Mat.*, 8, 309–313. https://doi.org/10.1038/s41578-023-00540-6.

Vijayaraghavan, Thiruvengadam, Yasuko Eckert, Gabriel H. Loh, Michael J. Schulte, Mike Ignatowski, Bradford M. Beckmann, William C. Brantley et al. (2017). Design and Analysis of an APU for Exascale Computing. *In 2017 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 85–96. IEEE, 2017. doi: 10.1109/HPCA.2017.42.

Shalf, J., Dosanjh, S., and Morrison, J. (2011). Exascale computing technology challenges. *High Perform. Comp. Comput. Sci. – VECPAR 2010: 9th Int. Conf.,* 6449, 1–25. https://link.springer.com/chapter/10.1007/978-3-642-19328-6_1.

Singh, Jaiteg, and Kawaljeet Singh. (2009). Statistically Analyzing the Impact of AutomatedETL Testing on the Data Quality of a DataWarehouse. *International Journal of Computer and electrical engineering.* 1(4) 488–495. DOI:10.7763/IJCEE.2009.V1.74

Don, Y. et al. (2022). Exascale image processing for next-generation beamlines in advanced light sources. *Nat. Rev. Phy.*, 4, 427–428. https://www.nature.com/articles/s42254-022-00465-z.

Verma, Mahendra K., Roshan Samuel, Soumyadeep Chatterjee, Shashwat Bhattacharya, and Ali Asad. (2020). Challenges in fluid flow simulations using exascale computing. *SN Computer Science.*, 1(3): 178 pp. 1-14. doi: https://doi.org/10.1007/s42979-020-00184-1

Zimmerman, M. I. et al. (2021). SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.*, 13, 651–659. https://www.nature.com/articles/s41557-021-00707-0.