© 2005 – ongoing JATIT & LLS

ISSN: 1992-8645

www.jatit.org



# NANOTECHNOLOGY THEORY USED FOR SIMULATION OF EMERGING BIG DATA SYSTEMS ON HIGH PERFORMANCE COMPUTING: A CONCEPTUAL FRAMEWORK

### <sup>1</sup>NORMA ALIAS, <sup>1</sup>MAI, 1HAFIZAH MUSA, <sup>1</sup>VICTOR RIYADH SERGEY, <sup>2</sup>NORHAFIZA HAMZAH <sup>1</sup>WALEED MUGAHED AL-RAHMI

<sup>1</sup>Ibnu Sina Institute for Scientific &Industrial Research (ISI-SIR), Faculty of Science, Universiti Teknologi

Malaysia, 81310, UTM Skudai, Johor, Malaysia

<sup>2</sup>Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu,

Sabah, Malaysia

E-mail: <sup>1</sup>norma@ibnusina.utm.my, <sup>2</sup>hafiza@ums.edu.my <sup>1</sup>waleed@ibnusina.utm.my,

### ABSTRACT

The Implications of big data analytics of current trends in nanotechnology theory, model and simulation are becoming impressive issues. The potential applications of nanotechnology in the industrial sector, identifying and prioritizing research across the emerging technology are motivating to perform the integration between the conceptual framework of nanotechnology and a big data system development. This paper presents six variations to meet the contexts of a conceptual framework for modeling the complex systems involve nanotechnology theory, modeling, large scale numerical simulation in the real world problem. Integrated mathematical modeling and large scale numerical simulations are the tools to solve the complex systems. The conceptual framework is a comprehensive concept in theory, ordinary differential equation (ODE) or partial differential equation (PDE) modeling and simulation based on high performance computing (HPC). The main objective is to improve the process of huge computation of the big data modeling and to increase the performance evaluation of parallel programming on HPC platform. The framework organizes the idea and step to be considered for solving the integrated theory, mathematical modeling with fast numerical simulation , specific parallel computing strategy , communication software and HPC hardware system which are applicable for solving large scale nanotechnology applications.

Keywords: Nanotechnology Theory, Big Data, High Performance Computing

#### 1. INTRODUCTION

High Performance Computing (HPC) serves as the main platform to support the visualization and animation; HPC of the communication model for solving the interdisciplinary problems above will be analyzed using a flow of the algorithm, numerical analysis and the comparison of parallel performance evaluations [1]. Furthermore, HPC is turning out to be a major source of hope for future applications development that require greater amounts of computing resources in various modern such Science domains as bioengineering, nanotechnology, and energy where HPC capabilities are mandatory in order to run simulations and perform visualization tasks [2]. Several numerical methods offer a straightforward implementation of converting the mathematical

model into a set of grid points. FDM, FEM and direct methods work as tools to discretize the mathematical modeling and to generate large sparse linear systems. Numerical simulations of explicit, implicit and Crank-Nicolson schemes fit well in solving the large sparse matrix of linear systems. Finer scale modeling, decomposition techniques, aggregation and mapping processes are able to obtain the parallel simulation of the large-scale modeling. Thus, a wide-range power simulation tools are required for handling a very large sparse matrices, huge memory allocation and high-speed processors of HPC systems [1]. The emergence of big data inevitably impacts many conventional computing paradigms such as high performance computing (HPC) and supercomputing that involve the most powerful machines worldwide [2]. Moreover, The limitation of nanotechnology that

<u>30<sup>th</sup> November 2017. Vol.95. No 22</u> © 2005 – ongoing JATIT & LLS

#### ISSN: 1992-8645

www.jatit.org

would present reasonable balances among the fundamental theory, principle of the foundational scientific law, fundamental molecular and the real phenomena at the nanoscale level. Thus, the theoretical principles are important to present the real phenomena in models by considering the independent and dependent parameters. In addition, а wide range of *mathematical* equations and theoretical principles have been used to obtain a good prediction and observation of the nanotechnology application. However, predictive modeling and simulation for large-scale and system of nanotechnology phenomena complex

dealing with nanoscale dimensions by scaling down their size. As an impact, the simulation requires a set of techniques with a new form of large size databases, fine granular simulation, parallel algorithm and computing technology, high speed processes and a large memory allocation of HPC platform. Thus, the motivation of this paper is to perform the conceptual framework for designing and emerging big data integration with the theoretical principles of nanotechnology, multiscale model of the complex system and large scale simulation on HPC platform.



Figure 1: Flow Diagram Of The Conceptual Framework

This paper proposes seven processes in the Vcycle diagram to overcome the limitation above as shows in Fig 1. The first process, complex system of the theoretical principle of nanotechnology law is presented in mathematical modeling. Next, the large scale numerical simulation is governed by using discretization strategy such as FDM, FEM or FVM for very fine computational mesh [3]. The purposed of this process is to perform an effectiveness of the high resolution and accurate prediction. For solving high speed simulation of fine granular and large sparse complex system, the requirements for the technology of parallel algorithm and computing are needed. Some technology of domain or function decomposition, load balancing, concurrent and synchronization communication processes, protocol, task control are essential for parallel performance indicator (PPI) on HPC systems. Then, a comparison of some numerical methods is used to validate an analytical solution of the model and to verify an industrial data set. Last, the prediction of the real phenomena in models based on the independent and dependent parameters are presented by conducting high-resolution visualization and numerical analysis indicators [4].

Therefore. Conceptualizing sustainable nanotechnology within the TBL framework can be utilized to help solve nanotechnology problems that have many variables and which call for detailed analysis. A Multi Criteria Decision Analysis (MCDA) model is one method of systematic topdown decision making appropriate for risk-based fields such as nanotechnology [5]. The literature review and model results call for a detailed and pragmatic articulation of the concept of sustainable nanotechnology, to better bridge the divergence in knowledge and stakeholder values. This is particularly important for manufacturers and regulators who are shaping the trajectory of ENM development with their decisions about product and process design and nanotechnology regulation. A benchmarking exercise in US educational institutions reveals that sustainability is being increasingly incorporated engineering into education [6].

# 2. INTEGRATING NANOTECHNOLOGY THEORY AND MODEL

In nanotechnology, the theory of physics, chemistry and engineering are assessed. Some fundamental theory in nanotechnology is the

	S 2005 - Oligoling SATT & LLS	JATIT
ISSN: 1992-8645	www.jatit.org	E-ISSN: 1817-3195

manipulation of matter on atomic and molecular features [3] such as Newton's law, Darcy's law [7], be

extended in terms of Nano scale level different properties. Thus, the enhancing model, a wide range of mathematical equations and theoretical principles, simulation and computational platform from small data volume to big data have high impact on data accuracy, precise, high resolution and visualization of the nanotechnology application.

#### 2.1 Transition from small data to big data

Figure 2 shows the requirement of the transition from small data to a big data and its impact to the development of data science at a nano-scale level. The transition is based on a simple model, small simulation system for analyzing, observing and predicting the enhanced model and complex model with large sparse simulation. Fig 2 shows the flow of the transition in terms of sequential and parallel processes, structured or unstructured data modeling, communication protocol and software, single and multi-processor architectures and small local memory and distributed-shared memory allocation on HPC. The impacts of the transition from small model to complex and enhancing model are to obtain high resolution in multi-dimensional space, accurate prediction and observation at nanoscale level. There has been substantial discussion of the convergence of big data analytics, simulations and HPC [9, 10]. In several papers [11] looked at the model in big data problems and studied the model performance on both cloud and HPC systems. In thermodynamics law [8], Fourier's law, Bernoulli's principle, Poiseuille's Law, Danish's law can

[11], look at performance in great detail showing excellent data analytics speed up on an Infiniband connected HPC cluster using MPI. Deep Learning [12] has clearly shown importance of HPC and uses many ideas originally developed for simulations. High Performance Computing enhanced Apache Big Data Stack as many critical core components of the commodity stack (Spark and Hbase) come from open source projects while HPC is needed to bring performance and other parallel computing capabilities [13]. In [4], New Iterative Alternating Group Explicit (NAGE) was introduced which is a powerful parallel numerical algorithm for multidimensional temperature prediction. The discretization was based on finite difference method of Partial Differential Equation (PDE) with parabolic type. The critical 3-Dimensional temperature visualization involves large scale of computational complexity. This computational challenge inspired the authors to utilize the power advanced high performance computing of resources. Incomplete blow-up is a condition under the quasilinear heat equation [14]. The Porous Medium Equation (PME) with power source are admitting incomplete blow-up. It is used as one of the filtration process in the industry. Authors proposed a new variance of the Alternating Group Explicit Scheme (AGE) algorithms to solve incomplete blow-up problem through High Performance Computing (HPC).

#### Impact



Big Data Complex system/ multi-model/ Sensor data/device data/image processing resolution/GIS/ grid generation/multi-D

Large/shared/distributed memory Parallel computing Communication protocol : TCP/IP as the basis for communication

 MDCS
 client-server model

 GRIDMATHEMATICA
 client-server

 model
 client-server

 CUDA
 inter-GPU communication

 LUNA
 comp-comm model

 PVM
 master-slave

 MPI
 point to point comm

 P-Performance Envaluation + Numerical

 Anavsis

High resolution Visualization in multi-D

Accurate prediction classification

Real time observation Monitoring

Visual inspection data represent

Precise negotiation between the practical and the **prediction** system



#### Small data

- Not complex/Small scale
- model/simulation
  Small size local memory
- Sequential computing
- Small da
- No Communication
- C
- C++ • VB
- JAVA
- Simple analysis

Figure 2: Enhancing Model, Simulation And Platform From Small Data To Big Data



#### www.jatit.org



#### 2.2 Large Scale Numerical method and simulation

Based the Fick's first on and second laws describe the diffusion equation of PDE, [15] first, proposed the sequential algorithm and some numerical methods for predicting the temperature behavior of curing process of thermoset nanocomposite material. Fourth-order Runge-Kutta scheme is superior to obtain the degree of cure for curing process. Second, extended the sequential algorithm to parallel algorithm integrated with a big data analytics approach for solving the large sparse simulation of the PDE using parallel Virtual Machine (PVM) communication software with master-slave protocol integrated with C programming on distributed parallel computing system (DPCS). Numerical analysis and PPI show that parallel AGE method performs high accurate prediction and high resolution visualization of the temperature behavior of curing process. Simulation results and information of drying kinetics of fruit material such as time temperature-moisture content distributions, as well as theoretical approaches to moisture movement, is very essential for the prevention of quality degradation and for the achievement of fast and effective drying. Such information will be very useful to optimize production processes of tropical fruits dried. Hence, the authors' contribution of the study [16] is successful modified the mathematical simulation in representing the actual process of dehydration in commercial foodstuff industry in terms of heat and mass transfer inside tropical fruits material. The development of machine learning has enhanced the growth and importance of data science. The methodology to scale the modeling

into a big data technology involving the repre-processing oganizing and the data. Development and implementation of algorithms for the efficient determination of computational solutions to mathematical problems. Efficiency numerical approximation can depending on the context, mean algorithmic accuracy, stability, consistency, convergence. Thus, the development of machine learning has enhanced the growth and importance of data science. In additional, proposed the parallel artificial neural network (PANN) method to validate the real data set from eye therapy industry, IC Herbz Sdn Bhd to predict the eve movement behavior during driving. PANN is the alternative machine learning strategy dealing classification with the big data of electroencephalogram (EEG) signals. In detail, refer to the contribution of the literature review and Fig 1 above, this paper highlights the data source from the experiment or industrial can be the initial and boundary condition of the mathematical modeling. In addition, the ODE and PDE are the methodology to obtain the set of equations of some parameters by considering the integrated with *scientific* principles for domain-specific applications. The outcome of the modeling is considered as decision science and the visualizing models with interactions, including plot, contour, graph and table are the relationship between the *outcome* and analysis [15]. Furthermore, the numerical analysis and the validation can be the decision support for the nanotechnology application [16]. The flow of diagram can be presented in Figure 3.

Data engineering  $\rightarrow \rightarrow$ Data science $\rightarrow \rightarrow$ Decision science $\rightarrow \rightarrow$ Decision supportData sourcemodelingvisualizationanalysis

Figure 3: The Flow Of Decision Support For Data Science Based On V-Cycle Framework

Moreover, [15] developed high performance nanotechnology software (HPNS) by following the conceptual framework in Fig 1 and 3 for predicting, visualizing and observing the temperature behavior and some parameters characteristics for the multilayer nanochip system and nanowire fabrication. The implementations of parallel algorithms for solving the heat transfer, wave motion, pressure, density, space, time control based on parabolic elliptic types of PDE. and Furthermore. The HPNS supports the supercomputing of numerical simulations and its repository on distributed memory architecture.

#### 3. PARALLEL DESIGN AND GRID STRUCTURED

The transition of sequential algorithm from small scale model and course granular to the parallel algorithm for large scale model involves the fine granular discretization of FDM, FEM or FVM. For low grain-size ratios between dependent parameters and time-space, an increase in the fine fraction up to 50%, lead to re-design the grid structure of the discretization. Thus, the parallel strategy for fine grain decomposition starts by partitioning the domain into subdomain or tasks. Next, design the communication activities among the

<u>30<sup>th</sup> November 2017. Vol.95. No 22</u> © 2005 – ongoing JATIT & LLS



<u>www.jatit.org</u>



E-ISSN: 1817-3195

neighborhood tasks to be interactive. Then, choose the agglomeration strategy, indicating global communication protocol model such as masterslave, client-server or fragmented messaging transport protocol. Lastly, mapping the block of tasks into a number of processes on HPC systems. Figure 4 shows the domain decomposition process of fine granular discretization of FDM, FEM or FVM.



Figure 4: Domain Decomposition Process

Besides, the fine granular discretization can be obtained by increasing the space dimensional of the PDE in 1D, 2D and 3D models by increasing the order of Taylor's series of the FDM discretization in higher order (Hamid DH et al 2014), increasing nodes per element of the FEM or decrease the volume in terms of length, area and volume of FVM. For example, the grid generation of 1D, 2D & 3D modeling can be presented in Fig 5 below. The computational molecule of the multidimensional numerical method called AGE is the fine subdomain presented and the communication activities among the tasks.

#### 3.1 Software systems, languages, and libraries

Most applications are computationally intensive. Scientists have traditionally attempted to parallelize their algorithms across HPC infrastructures. However, this task requires significant work and learning parallel programming. effort for Developing parallel libraries and high-level and easy to use languages are required to hide parallelization complexity of programs. [18], in their paper Implicit parallelism on 2D meshes using SkelGIS, tackle the issue of overcoming restrictions in parallelization of scientific simulations because of the complexity of functional concepts and specific features. Parallelization of scientific simulations requires a lot of efforts and fieldspecific knowledge to produce efficient parallel programs. For this reason, the authors introduce SkelGIS as a solution for abstracted and implicit parallelism. They apply SkelGIS to solve heat equations and shallow-water equations and compare both the SkelGIS performance and the SkelGIS programming effort with Message Passing Interface (MPI) solutions counterparts. Another important issue that can influence HPC systems performance is vectorisation. Many scientific codes have vectorisation potential that cannot be exploited due to an algorithm-driven choice of data layouts. In their article Data Layout Inference for Code Vectorisation, [19] propose an interesting approach for automatically generating efficient code for vectorisation by mainly focusing on the evaluation of a family of data layout transformations. See figure 5.



ISSN: 1992-8645

www.jatit.org



Figure 5: Grid Structured, Communication Activity Between Tasks For 1d, 2d And 3d Age Method

### 4. PARALLEL IMPLEMENTATION, COMMUNICATION SOFTWARE AND PROTOCOL ON HPC

The first step for parallel implementation is rewriting the serial algorithm of numerical method in single processing executing too concurrently or in parallel algorithm. There are many ways to implement the parallel algorithms. The methodologies of domain or function decomposition and the communication protocol influence the process of parallel programming. The familiar communication software Parallel Virtual Machine (PVM), Message Passing Interface (MPI), LuNA, Matlab Distributed Computing System (MDCS), OpenMP, Compute Unified Device Architecture (CUDA), COMSOL, LINDA and integrated with C++, C, FORTRAN, embedded system. Communication protocol for programming the parallel algorithm, PVM used master-slave model, MPI used message-passing model, and client-worker MDCS used model. Web technologies and protocols built around the clientserver model. Job manager is a central part of MDCE to coordinate the execution of jobs. Moreover, LINDA is a model of coordination and operating as a generative communication. Fig 6 shows the flow chart of message passing model among the clients and server. The latest technologies are CUDA used multi-threaded cores model and LuNA used data fragmentation model. CPU – GPU with CUDA computing consist of thousands of smaller cores for handling multiple tasks in SIMT (Single Instruction, Multiple Thread) protocol. [15], proof that PVM is well suited for solving large sparse numerical simulation compared to MDCS in terms PPI because of time consuming on embedded system operations with high range of rating levels for calling the library. CUDA is superior programming for CPU-GPU architecture by integrating shared and distributed memory architectures. generative Based on the communication, Gaussian integration with LINDA programming language is efficient for predicting chemical properties of benzene on Multicores processor, 2X3TB, WD Black Caviar, 16GB RAM [1]. Dealing with the integrated data and



ISSN: 1992-8645

www.jatit.org

E-ISSN: 1817-3195

computation fragments, LuNA is highly practical l to support a very large numerical simulation up to terabyte range. Figure 7 shows the LuNA fragmentation model by differentiate the data fragmented and the computation fragments for designing the parallel algorithm.



Figure 6: The Transition Of Sequential Algorithm To Parallel Algorithm



Figure 7: The Integrated Data And Computation Fragments In Luna

The communication design of the numerical algorithm, concurrent access and synchronization process for sharing the block of simulation, the feature of load balance to determine the overall performance and testing for level of *granularity*. The finer the granularity, the greater the potential

<u>30<sup>th</sup> November 2017. Vol.95. No 22</u> © 2005 – ongoing JATIT & LLS

ISSN: 1992-8645

www.jatit.org



E-ISSN: 1817-3195

for parallelism and speedup. However, its contribute the greater the overheads of

synchronization and communication. See Fig 8 below shows message passing paradigm.



Figure 8: The Message Passing Paradigm On Distributed Memory Architecture

The HPC architecture for emerging big data simulation can be categorized as shared, distributed and hybrid computing system. Shared memory contains multiple processors can operate independently but share the same memory resources. Data sharing between tasks is both fast and uniform due to proximity of memory to CPUs. Distributed Memory consists of several units of independent processors connected by a network to communicate between processors. Increase the number of processors and the size of memory increases proportionately. Rapid access to memory for each processor. Granularity is the ratio of computation and communication. The finer the granularity, the greater the potential for parallelism and speed-up, but the greater the overheads of synchronization and communication. Other PPI are as the following table.

PPE	Formula	Definition
Speedup	$S_{p} = \frac{T_{1}}{T_{p}}$	Time taken for a program to execute in one processor divided by the time taken to execute in many processors.
Efficiency	$C_{p} = \frac{S_{p}}{p}$	A measure of processor utilization.
Effectiveness	$F_p = \frac{S_p}{p \times T_p}$	Whenever the speedup increases, the effectiveness will also increase.
Temporal performance	$L_{p} = \frac{1}{T_{p}}$	Maximum value of $L_p$ shows better performance of parallel algorithm.
Granularity	$G = \frac{T_{comp}}{T_{comm}}$	Gives a good indication of the feasibility of parallelization. Ratio of the amount of computation to the amount of communication within a parallel algorithm implementation

Table 1: The PPI, Formula And Definition

Some example of the HPC under consideration; NVIDIA, Multicore on open operating system, the familiar communication software PVM, MPI, LuNA, MDCS, OpenMP, CUDA, COMSOL, LINDA and integrated with C++, C, Fortran, embedded system, the hardware DPCS computational toois MDCS R2011a. The MDCS contains with Intel®  $Core^{TM}$  Duo Processor Architecture under Fedora 8 featuring a 2.6.23 via Ethernet network on Linux kernel operating system,

www.jatit.org



E-ISSN: 1817-3195

connected with internal network 10/100/1000 NIC. Each core can execute a sequential thread, but the cores execute in what NVIDIA calls SIMT

ISSN: 1992-8645

(Single Instruction, Multiple Thread) protocol using CUDA.



Figure 9: HPC Platform: DPCS And Hybrid Architecture

### 4.1 Scientific software development

Recently, the large sparse simulation of the mathematical model based on the nanotechnology theory and law can be extended into a graphical user interface (GUI) applications. The GUI is the interface complexity of such programs would exceed the visualization of the dependent and

independent parameter change in the modeling, support the generic solution and provide the optimum results. [3] Developed the GUI for the embedded system of the digital temperature sensor (DTS). The functions of DTS are data collecting of temperature behavior and heat flow profile requirements for the *electronic* component manufacturing.



Figure 10: Conceptual Model Of HPC Software Development [17]

<u>30<sup>th</sup> November 2017. Vol.95. No 22</u> © 2005 – ongoing JATIT & LLS



#### ISSN: 1992-8645

www.jatit.org

The service Oriented Architecture (SOA) framework and MDCS Software version R2011a is used for the DTS software development. Webbased user interface components provide coarsely web user and the integrated programs to support data input, execution process, and output. For example, Hamid et al 2014 used UCLA Grid Portal open source software with parallel computing system has been implemented to design the efficient authoring infrastructure and to implement an online graphics, animation, audio and video technology. The GUI recognized the dependencies, requests the discriminative input data elements from the user, and then processes the input request for the next interaction cycle accordingly. There are many scientific software developments to complete and to support the process of modeling, discretization and simulation of the GUI on HPC platform. [17], developed HPNS software for characterizing the parameters to manipulate the behavior of nanowire fabrication and nanochip system model. HPNS is an enhancement of SCLP software for thermal process control of nanowire fabrication and multilayer nanochip system [8]. The conceptual diagram of HPNS software development consists the interaction process among the users, algorithm provider and user interface. The software engineering flow stated three different modules according to their functionality. The modules are authoring, publishing and supporting. Figure 16 shows there are several levels to access the code, view tutorials and explore the code. The functions of scientific librarian are to approve the status of a contribution of algorithm provider and publish the contribution code. Thus, a huge virtual memory and high-speed processors support the repository and the numerical library of HPNS software [17]. In addition, the scientific software development is based on the Parallel Virtual Machine (PVM) and C language on the Linux operating system serviceoriented architecture (SOA) style and Java for web services development. In this research, Web Services Definition Language (WSDL) is being selected as web services; XML standardized technology and developed using NetBean IDE tools. Glassfish 2.0 is used as an application server for this architecture. This platform development will accelerate the speeds of execution and scaledout across a virtualized grid. The distributed processors and large memory size are involved in this platform to support a large computational cost with inexpensive hardware specifications and configured with purely open source software.

# 5. VALIDATION, VERIFICATION AND COMPARISON

The validation of the large sparse simulation is conducted because the complex system has multiple outputs with different data types. Numerical stability, analyses, such as consistency, convergence, accuracy, error estimation are the indicators to validate the numerical simulation. [15] used the table of numerical indicators to provide validation evidence whether the models are all rightly concerned for predicting the behavior. For example for user friendly interface, the numerical simulation is standardizing into software development. The function is to support embedded system for the specific equipment. Mathematical modeling and computer-based simulation are routine tools in most scientific disciplines, thus scientific programs as executable implementations of models represent an important part of the "output" by concentrating all the computational resources of the scientific software. The verification of the software is to prove the correctness of a scientific software development to support the large sparse simulation. Standardized user interface, User-friendly GUI, HPC-based platform and model testing [20] are the key components for verifying effective tools, the process of checking documents, data aliases, design, code, exclusion of recursion and program dynamically. The comparison is considered the indicator analysis for benchmarking the superior numerical method, optimum results and prediction, high resolution of the visualization, thus, the comparison focuses on numerical analysis among different numerical schemes and methods, dimension of the visualization. PPI provide comparative indicators on the size of sparse simulations and performance behavior across different communication protocol. Fig 10 shows the example of the numerical results for some numerical methods for solving 2D fundamental problem in heat conduction and diffusion process based on PDE with parabolic type as stated in [17]. The analysis of sequential algorithm of GSRB and AGE method for solving 2D matrix simulation of size  $(600 \times 600)$  and  $(1000 \times 1000)$ . Then, the large matrices size 104 or higher are obtained and analyzed by increasing the fine granularity in terms of time and space discretization of the 2D parabolic equation.

30th November 2017. Vol.95. No 22 © 2005 - ongoing JATIT & LLS

www.jatit.org



m	600  imes 600		1000  imes 1000	
Method	AGE_DOUGLAS	GSRB	AGE_DOUGLAS	GSRB
Run time (s)	35.6894	40.1592	143.2267	157.6732
Iteration	150	300	230	622
MSE	3.07398E <sup>-12</sup>	3.07461E <sup>-12</sup>	2.4416E <sup>-12</sup>	2.4482E <sup>-1</sup>
RMSE	1.0116E <sup>-11</sup>	$1.0151E^{-11}$	8.0597E <sup>-12</sup>	2.4482E <sup>-1</sup>
r	0.95	-	1.6	-
$\Delta x$	1.67E <sup>-3</sup>	1.67E <sup>-3</sup>	1.00E <sup>-3</sup>	1.00E <sup>-3</sup>
$\Delta v$	1.67E <sup>-3</sup>	1.67E <sup>-3</sup>	1.00E <sup>-3</sup>	1.00E <sup>-3</sup>

MSE=mean square error, RMSE = root mean square error, r=acceleration parameter for AGE method,  $\Delta x =$  grid meshes at x-axis and  $\Delta y = \text{grid}$  meshes at y-axis

Based on the number of iterations for both sizes of matrixes, AGE Douglas method 2 to 3 times faster convergence and 70-80% faster in execution time speed compared to GSRB. Fig 11 shows the

ISSN: 1992-8645

example of the visualization for multi-dimensional temperature behavior and power distribution for multilayer full-chip architecture in 2D and 3D heat equations stated in Alias et al (2015).



Figure 11: Visualization For Multi-Dimensional Model For Temperature Behaviour

The comparison of the performance behavior across different communication software and its protocol were stated in [17]. The comparison between MDCS and OpenMP on a multicore platform has been shown in Fig 12 to solve the large scale simulation of brain tumor modeling. An arallel performance comparison (a) PVM and MDC (b) OpenMP and MDCS (c) MPI and MPI+ OpenMP (d) Number of fragments in LuNA.



Figure 12: The PPI Comparison Between Open MP On Multicore Processors And MDCS

#### Journal of Theoretical and Applied Information Technology 30th November 2017. Vol.95. No 22

© 2005 – ongoing JATIT & LLS



www.jatit.org

The PPI of Open MP on multicore processors is higher than MDCS. It is because Open MP emphasizes on multi-threading protocol with minimum communication time, however MDCS is based on a standard programming interface, efficiently in a *distributed* environment with traditional client-server protocol. Fig 13 shows the PPI comparison between PVM and MDCS. The PPI of PVM is higher than MDCS. It is because PVM is more portable message passing library, however MDCS needs to use the MPI library to support the message passing process for fair comparison. The achievement of PVM is two times better performance compared to MDCS. PVM is superior features to *improve* throughput and latency of the message passing paradigm. Furthermore, for predicting the chemical properties of benzene using the MP4 method of Gaussian software, LINDA parallel programming was used to investigate the PPI of the multicore platform. The speedup and effectiveness of LINDA increase linearly respect to a number of processors as shown in *Figures* 13 and 14.



Figure 13: PPI Of LINDA Parallel Programming On Multicore Systems



Figure 14: PPI Comparison Between PVM And MDCS

30th November 2017. Vol.95. No 22 © 2005 – ongoing JATIT & LLS



www.jatit.org

E-ISSN: 1817-3195

PPI of CUDA parallel programming for solving the nanoparticle drug delivery modeling on CPU-GPU platforms was visualized in Fig 15. Based on the multi-threaded cores model, execution time, speedup and temporal performance was optimized at 64 numbers of threads of the CPU-GPU platform. The execution time for multiple threads on GPU lowers than *multiple cores* on CPU. By investigating the two different numerical methods, GD was achieved the maximum speed up and maximum temporal performance compared to BD for simulating for fine-grained parallelism, (24004x24004) size of matrix. For all indicators of PPI, GD is a superior method compared to BD.



Figure 15: The Optimum Execution Time, The Comparison Between CPU And GPU And The PPI Comparison Between BD And GD Methods

The performance characteristics of LuNA parallel programming on multicore platforms depend on size of matrixes, number of cluster nodes, and the percentage of fragmentation capability due to the overhead problem. For the comparison among three size of matrixes, two type of LuNa characteristics, two units of execution, threads and processes, 16 nodes is the minimum cores to execute the fine grained simulation as stated in Fig 16. Furthermore, Fig 16 shows the example of PPI on size of sparse simulations stated by [1]. The set of concepts contains the revise standard for guiding the scientific algorithm provider to model, discrete and simulate the integration of big data emerging based on parallel computing strategies. The six variations are identifying the parameter and governing the mathematical modeling for large scale complex system, transferring the continuous model, function and equation into discrete counterparts, developing the parallel algorithm and communication model, implementing the parallel program on HPC platforms, converting the discrete solution into

#### ISSN: 1992-8645

www.jatit.org



continuous feature's existence of exact solutions in any natural phenomenon of complex systems.



Figure 16: The Luna Execution Time For Three Different Matrixes, Two Different Characteristics On Multicore Platform

Last but not least is verifying and validating the numerical approximation and parallel performance evaluations. The matching skills between the numerical simulation of complex system and the computational tool such as communication model, message passing protocol, parallel programming language, and memory and processor performance on HPC platform were framed for data scientist and scientific algorithm provider. This framework employs some case study approaches to provide an adaptation of the conceptual framework and satisfy the variation requirements. The examples for variations are presented, the performance evaluations are analyzed and the predictions of the large scale complex system are provided. As a conclusion, the conceptual framework performs more comprehensive perspective based on future software system design dealing with the nextgeneration of HPC architecture system and assessing the environmental and societal implications of this big data emerging technology and enabling decision support for data science.

#### 6. CONCLUSION

A conceptual framework based on V-Cycle process is the alternative procedure for integrating the fundamental theory of nanotechnology application, governing mathematical model and its discretization. The large sparse simulation for emerging a big data assessed by the HPC platform and configured by communication software with specific message passing protocol. The big data analytic solve the grand challenge application involving huge data and large sparse matrix of the complex system modelling and nano-scale computation and large sparse simulation. Thus, the parallel computing and methodology is the tool to support the big data emerging on nanotechnology theory, model and simulation. Parallel performance analysis and numerical results are significant indicators for big data analytic in terms of time execution, speedup, efficiency, effectiveness, temporal performance, and granularity. In addition, model validation, scientific software verification and performance comparison in numerical schemes and methods, architecture of HPC, size of matrixes, number of processors, cores or multi-thread and its communication software for solving multidimension model are analyzed. The PPI comparison of communication paradigm, PVM, MDCE, Open MP, CUDA, Lu NA and LINDA enables to support high computational complexity problem on shared, distributed or hybrid architecture platforms. The conceptual framework in V-Cycle processes was well suited for designing the nanotechnology solver.

#### ACKNOWLEDGMENT

We would like to thank the Research Management Centre (RMC) at Universiti Teknologi Malaysia (UTM) for funding this project under grant number PY/2015/05399 / R.J130000.7826.4F795. We also would like to thank the Ibnu Sina Institute for Scientific and Industrial Research (ISI-SIR), Faculty of Science, Universiti Teknologi Malaysia. Moreover, we would like to acknowledge financial support by Ministry of Higher Education, Fundamental



www.jatit.org



[9] D.A. Reed, and J. Dongarra, "Exascale computing and big data," *Communications of the ACM* 58(7), 2015, pp. 56–68.

REFERENCES

2014 UMS, Malaysia.

[1] N. Alias, R. Sahnoun, and V. Malyshkin, "High-Performance Computing and Communication Models for Solving the Complex Interdisciplinary Problems On Dpcs", ARPN Journal of Engineering and Applied Sciences, 12, 2 2017, pp. 356- 364.

Research Grant Number (RACE) RACE0017-SG-

- [2] W. W. Smari, M. Bakhouya, S. Fiore, and G. Aloisio, "New advances in High Performance Computing and simulation: parallel and distributed systems, algorithms, and applications, "Concurrency and Computation: Practice and Experience, 2016.
- [3] N. Alias, H. Mokhtar, H. San, Y. Saipol, H.F. Mustaffa, M. N. Said, and N.M. Nor, "High speed computation of muscle stress problem using crack propagation modelling on DPCs platform," *Advanced Science Letters*, 20, 10-12, 2014, pp.1961-1966.
- [4] N. Alias, M. S. Masseri, M. R. Islam, and S. N. Khalid, "The visualization of three dimensional brain tumors growth on distributed parallel computer systems," *Journal Applied Science*, 9:2009, pp. 505-512.
- [5] I. Linkov, E. Anklam, and Z. Collier, "Environment, Systems, and Decisions, in: Risk-Based Standards: Integrating Top-Down and Bottom-Up Approaches,"http://dx.doi.org/10.1007/s10669-014-9488-3. 2014.
- [6] C. F. Murphy, D. Allen, B. Allenby, J. Crittenden, C.I. Davidson, C. Hendrickson, and H.S. Matthews, "Sustainability in engineering education and research at U.S. universities," *Environmental Science and Technology*, 43(15), 2009, pp. 5558-5564. DOI: 10.1021/es900170m
- [7] N. K. Kasmin-Bajuri, and N. Alias, "Implementation of finite difference method based on explicit and Crank-Nicolson method to the thermomechanical ice sheet model," *International Symposium on Mathematical Sciences and Computing Research*, Ismsc, 2015, pp. 419-425.
- [8] L.M. Yusuf, N. Alias, M.S. Ghouse, S.M. Shamsuddin, and S. M. Othman, "The parallel AGE variances method for temperature prediction on laser glass interaction, *ISCI 2011 -2011 IEEE Symposium on Computers and Informatics*, 5958927, 2011, pp. 279-283.

- [10] C. Baru, and T. Rabl, "Tutorial 4 Big Data Benchmarking," *IEEE International Conference on Big Data*, 2014.
- [11] B. Zhang, Y. Ruan, and J. Qiu, "Harp: Collective communication on hadoop," *IEEE International Conference on Cloud Engineering* (*IC2E*) conference, 2014.
- [12] A. Coates, B. Huval, T. Wang, B. WuD, Catanzaro, N. Andrew, "Deep learning with COTS HPC systems," *Proceedings of the 30th international conference on machine learning*, 2013, pp.1337–1345.
- [13] G.C. Fox, J. Qiu, S. Kamburugamuve, S. Jha, and A. Luckow, "HPC-ABDS High Performance Computing Enhanced Apache Big Data Stack," *In: Cluster, Cloud and Grid Computing (CCGrid), 2015 15th IEEE/ACM International Symposium,* 1057–1066. IEEE, 2015.
- [14] N. Alias, Z.S. Abdal-Ghaffar, N. Satam, R. Darwis, N. Hamzah, and R. Islam, "Some parallel numerical methods in solving parallel differential equations." *Proceedings of the 2nd International Conference on Computer Engineering and Technology*, 16-18, Chengdu, Sichuan, China, 2010, pp. 391-396.
- [15] N. Alias, N. N. Suhari, N. Y. Saipol, H.F. Dahawi, A.A. Saidi, and M.M. Hamlan, "Parallel computing of numerical schemes and big data analytic for solving real life applications," *Jurnal Teknologi*, 78, 8-2, 2016, pp. 151-162.
- [16] N. Alias, C.R. Teh, M. Berahim, Z.S.A. Ghaffar, M.R. Islam, and N. Hamzah, "Numerical methods for solving temperature and mass transfer simulation on dehydration process of tropical fruits," *Proceedings of the Simposium Kebangsaan Sains Matematik Ke-17 (SKSM17)*, 15-17, Melaka, Malaysia, 2009, pp. 928-934.
- [17] N. Alias, N. Satam, M.S. Othman, C.R. The, M.N. Mustaffa, and H.F. Saipol, "High performance nanotechnology software (HPNS) for parameter characterization of nanowire fabrication and nanochip system," *Communications in Computer and Information Science*, 513, 2015, pp. 251-268.
- [18] H. Coullon, M.H. Le, and S. Limet, "The SIPSim implicit parallelism model and the SkelGIS library," *Concurrency and*

© 2005 – ongoing JATIT & LLS



ISSN: 1992-8645

www.jatit.org

*Computation: Practice and Experience*, 28(7):2016, pp. 2120–2144.

- [19] A. Sinkarovs, and S.B. Scholz, "Type-driven data layouts for improved vectorisation," *Concurrency and Computation: Practice and Experience*, 28(7):2016, pp. 2092–2119.
- [20] N. Shah, S. Sara, A.K. Sher, N. Alias, and S. Anwar, "A Novel Rules Based Approach for Estimating Software Birthmark," *The Scientific World Journal*, ID 579390, 2015, pp. 1-8.