VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

INTEGRATION OF A BIG DATA EMERGING ON LARGE SPARSE SIMULATION AND ITS APPLICATION ON GREEN COMPUTING PLATFORM

Norma Alias[1] and Mohamad Hidayad Ahmad Kamal[2]
[1]Center for Sustainable Nanomaterials, Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia, Malaysia
[2]Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia, Malaysia
E-Mail: normaalias@utm.my

ABSTRACT

The process of analyzing large data and verifying a big data set are a challenge for understanding the fundamental concept behind it. Many big data analysis techniques suffer from the poor scalability, variation inequality, instability, lower convergence, and weak accuracy of the large-scale numerical algorithms. Due to these limitations, a wider opportunity for numerical analysts to develop the efficiency and novel parallel algorithms has emerged. Big data analytics plays an important role in the field of sciences and engineering for extracting patterns, trends, actionable information from large sets of data and improving strategies for making a decision. A large data set consists of a large-scale data collection via sensor network, transformation from signal to digital images, high resolution of a sensing system, industry forecasts, existing customer records to predict trends and prepare for new demand. This paper proposes three types of big data analytics in accordance to the analytics requirement involving a large-scale numerical simulation and mathematical modeling for solving a complex problem. First is a big data analytics for theory and fundamental of nanotechnology numerical simulation. Second, big data analytics for enhancing the digital images in 3D visualization, performance analysis of embedded system based on the large sparse data sets generated by the device. Lastly, extraction of patterns from the electroencephalogram (EEG) data set for detecting the horizontal-vertical eye movements. Thus, the process of examining a big data analytics is to investigate the behavior of hidden patterns, unknown correlations, identify anomalies, and discover structure inside unstructured data and extracting the essence, trend prediction, multi-dimensional visualization and real-time observation using the mathematical model. Parallel algorithms, mesh generation, domain-function decomposition approaches, inter-node communication design, mapping the subdomain, numerical analysis and parallel performance evaluations (PPE) are the processes of the big data analytics implementation. The superior of parallel numerical methods such as AGE, Brian and IADE were proven for solving a large sparse model on green computing by utilizing the obsolete computers, the old generation servers and outdated hardware, a distributed virtual memory and multi-processors. The integration of low-cost communication of message passing software and green computing platform is capable of increasing the PPE up to 60% when compared to the limited memory of a single processor. As a conclusion, large-scale numerical algorithms with great performance in scalability, equality, stability, convergence, and accuracy are important features in analyzing big data simulation.

Keywords: big data analytics, fine-grained, parallel numerical method, green computing.

INTRODUCTION

Introduction to modern methods of data analysis, spatial data handling, and visualization technologies gives high impact to big data analytics. The requirement of big data analytics refers to the real-time solution, the associated demand, imposes a unique solution for understanding the needs of actionable information from large sets of data and improving strategies for making the decision. Some examples of a complex problem involving large-scale data are complex modelling, grid generation; parallel algorithm, high speed simulation, numerical analysis and PPE were investigated. Furthermore, handling a high-level computational capacity of the large sparse simulation is supported through green computing technology of high-performance computing system (HPC). In this paper, the scalable computing, energy efficiency, sustainability, eco-friendly use of computers, intersection of data analytics pipelines on HPC systems are emerging green computing technology. The processes are meshed generation, domain-function decomposition technique, parallel algorithm development [1, 2], implicit and explicit strategies [3], inter-node communication design and mapping the subdomain. The two types of big data analytics are numerical analysis and parallel performance evaluations (PPE) are considered.

TYPES OF BIG DATA ANALYTICS

The three types of big data analytics exploit the data generated through mathematical modelling, general data extracted from EEG and data constructed based on digital image processing.

Theory and fundamental of parallel numerical simulation

The first type of big data analytics emphasizes theory and fundamental on numerical analysis. This type is the basic process for handling a high-level computational capacity of the large sparse simulation. Huge computation complexity cost was generated from arithmetic operations of the mathematical model. The theoretical

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

nanotechnology illustrates the fundamental nanoscale computation and produces resources in big data analytics. First and second order partial differential equations (PDE) with fine-grained feature were used to present the silicon nanowire growth in chemical vapour deposition (CVD) reactor. The discretization of PDE using finite difference method (FDM) increased the amount of data structure generated by numerical simulation of FDM [4].

The fundamental theory in silicon nanowire growth for the governing mathematical modelling of PDE-elliptic equation is self-consistently with the Poisson equation. Dealing with wave system, the governing equations reflect to some independent and dependent parameters such as energy, mass and space and time associated with initial and boundary conditions. The time-independent Schrodinger equation represents the properties of atomic systems in stationary conditions by considering the time-independent Schrodinger equation below:

$$\nabla^2\psi + \frac{2m}{\hbar^2}(E-V)\psi = 0 \tag{1}$$

Where

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} \tag{2}$$

m is the mass of the electron and

$$E = E_{kin} + V \tag{3}$$

is the total energy of the energy.

And rewrite as a wave function with lower case explicitly as,

$$\psi(x,y,z,t) = \psi(x,y,z)\cdot e^{i\omega t} \tag{4}$$

Trends and opportunities in nanoscale computation and HPC system for nanotechnology-based data storage were utilized in this case study.

Pattern extraction from EEG graph

The second type of big data analytics emphasizes on the extraction of patterns from EEG readings for detecting problems in eye movement. The aim of this application is to detect patterns in the horizontal-vertical eye movements, predicting the attributes of a centrally generated motor pattern and monitoring the abnormal eye movements. Based on the Digitizer Module of MATLAB 2009a software, the EEG graph was converted into digital data based. Mathematical model generated a big data, with emphasis on ANN algorithms, data training and parallel implementation. Figure-1 shows the line mesh generation method for describing and applying EEG digital data based on eye movements.

Digital images in 3D visualization

The third type of big data analytics emphasizes on digital images in 3D visualization by applying advanced techniques such as spatial data handling and

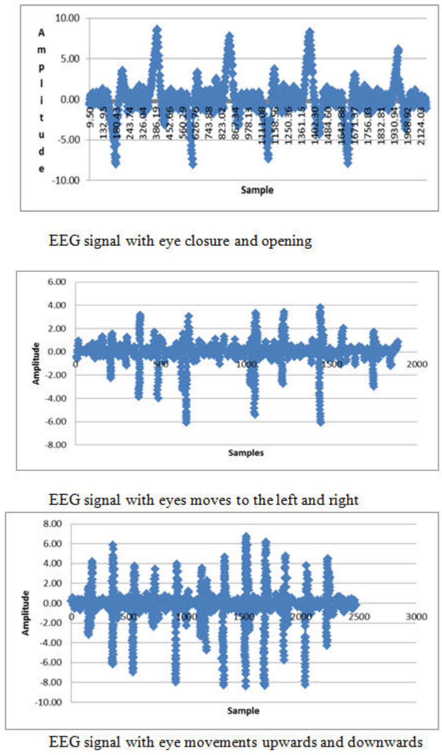

Figure-1. Mesh generation for EEG digital data based on multi-direction of eye movements.

data analysis. A series of MRI images of the head, taken from multiple different directions of the brain. The set of images was produced had huge digital image through the process of converting the signal image to digital form. A huge amount of digital images requires the process of analysing a comprehensive set of a big data in an effort to perform the 3D visualization.

Figure-2 shows the steps of digital image processing for contour lines projected onto the 2D image. The contour line with developments from the cropped image of the original image and considered as the approximate edge of the tumour growth. The processes of contour line development are based on image segmentation algorithms such as semi-implicit additive operator splitting (AOS) scheme and geodesic active contour model (GAC) with fine-grained discretization [2]. Repeat the contour line development process to different directions of a human tumour image such as front, side and top view as shown in Figure 3. High resolution of edge detection can be generated by fine-grained digital features referred to a specific interval size between meshes. The 3D construction is simply based on the VE method by utilizing the set of different directions of human tumour images.

ARPN Journal of Engineering and Applied Sciences

Figure-2. Processes for human tumor edge detection.

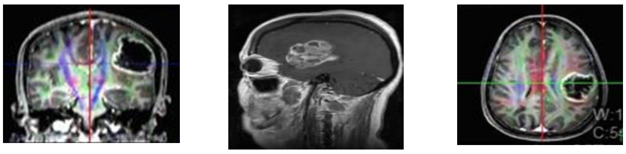

Figure-3. Different directions of human tumor images.

MODEL AND METHOD

Theory and fundamental of parallel numerical simulation

Partial differential equation (PDE) model of the Schrodinger equation was used to represent the growth process of silicon nanowires based on CVD method. The parameter of PDE describes the temperature behaviour, time, space and density. By FDM, the growth process of silicon nanowires can be presented in huge nanotechnology-based data storage. The mesh generation of FDM produces a huge amount of meshes by reducing the interval size of the grids and transforming the coarser mesh into finer mesh. The numerical simulation involving the iterative schemes such as Jacobi (JB), Gauss-Seidel (GS), Red Black Gauss-Seidel (RBGS) and Alternating Group Explicit (AGE). The graph of comparison among the methods can be visualized in Figure-4. The initial and boundary conditions of flow process are generated by the experimental data based on CVD method. Parallel implementation was used in the green computing platform integrated with the PVM communication software and C language. Graph visualization, table form, comparison results are the indicators to validate the experimental data, PDE model and numerical analysis.

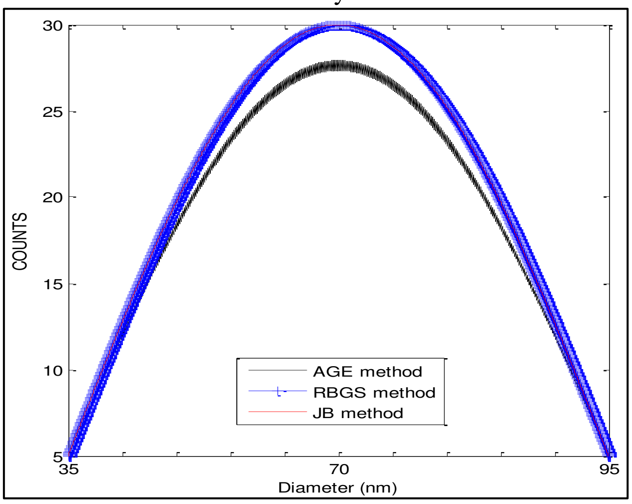

Figure-4. The visualization of silicon nanowires growth with different numerical scheme using diameter of silicon nanowire is 25 nm.

Pattern extraction from EEG graph

The concept of signal processing of EEG graph is adapted to identify the full range of negative and positive frequency components of the EEG spectral edge frequency record. This paper used signal-processing software called Digitizer Module of MATLAB 2009a to convert the analog signals into digital data based. The digital data are broken into a large sparse grid. A huge amount of grid is decomposed into blocks of a configured size for parallel implementation. The artificial neural network (ANN) is used to train the digital data of EEG. The aims of ANN are to detect patterns of horizontal-vertical eye movements. Dealing with a big data set, large training times and large networks are required. The sequential algorithm for block mode of back-propagation learning patterns was developed. However, sequential training of ANN is a time consuming job. Thus, parallel algorithm for ANN was implemented using domain decomposition techniques integrated between the threads and blocks. The process of the data flow can be shown in Figure-5. The parallel algorithm for training process was employed by decomposing the full neural network in multiple threads, the hierarchy of grids. The thread will be executed by block. The programming model provided by CUDA and makes the parallelism process straightforward.

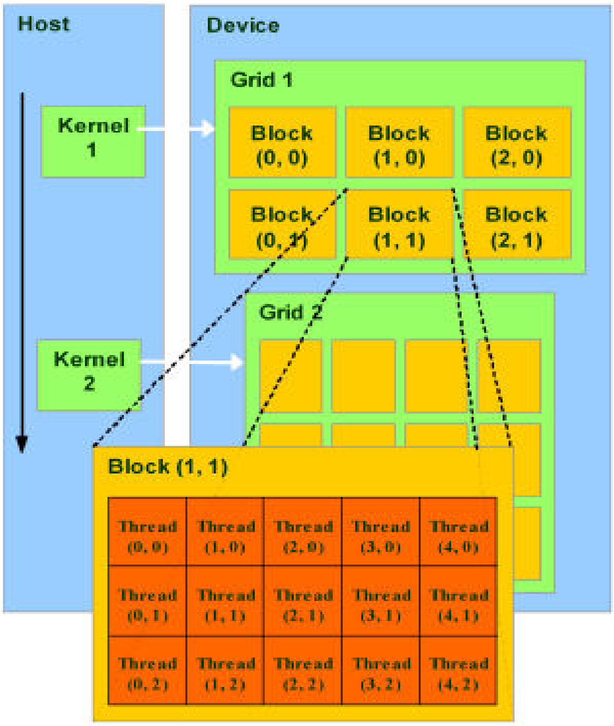

Figure-5. Threads and blocks definition.

Digital images in 3D visualization

There are many techniques to reconstruct a 3D object visualization of tumor growth from 2D MRI slices. Data dimensionality estimation method, image manifold method, volume estimation method and entropy estimation are methods to construct a 3D object. In this paper, data dimensionality estimation methods based on volume

ARPN Journal of Engineering and Applied Sciences

estimation are used for quantifying the tumour by external feature; length, height, and volume. Figure-6 shows the slices of the tumour were arranged in one direction by referring to the formula,

$V = t \times a$

Where:

V is the volume
t is the height and
a is the base area of object

Based on the plenary results, $V = t \times \Sigma a\ cm^3$, Σadenotes the section areas in cm^2 and t is the sectioning interval in cm for the consecutive sections. The route of converting based on the main following aspects as shown in Figure-6.

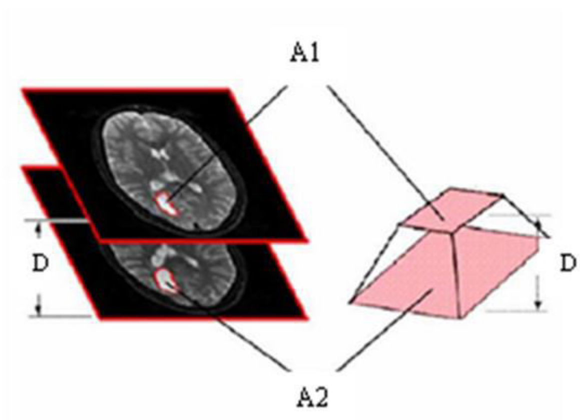

Figure-6. Process of reconstruction of 2D images to 3D visualization.

This method counters the density estimation of grids and the sufficient number of slices for smooth profiled subjects. The practical estimate volume from the 2D MRI image is based on the cross-sectional 2D images arrangement and combined with the point-counting estimator. In terms of clinical treatment, the 3D visualization through the estimate volume construction gives high significance for predicting and monitoring the tumor growth.

PARALLEL ALGORITHM

The parallel algorithm for large-scale numerical simulation of case study 1 and 3 refer to the Figure-7. All cases emphasized on a client-server model for implementing the data exchange among tasks.

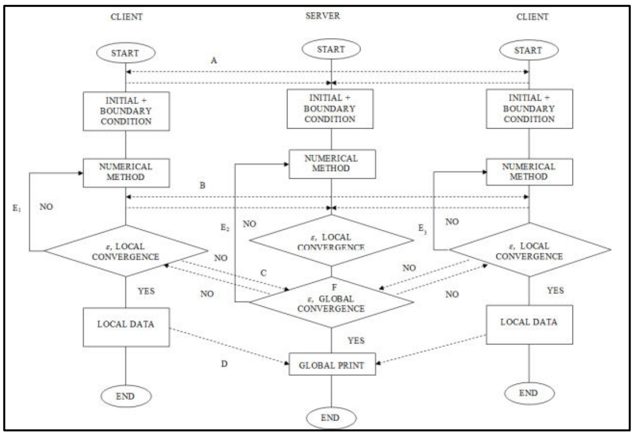

Figure-7. Parallel algorithms for case study 1 and 2.

In case study 1, server controls the functionality of client for subdomain execution in parallel. Server obtained the global convergence and clients performed the local convergence of the numerical methods. The relaxation process of the parallel numerical method requires the data exchanges among neighborhood clients. Each client sends its dependencies to its neighbors. *Thus,* PVM allows *a* client with different *data* representations to exchange by sending and receiving data via the function pvm_initsend creates and clears a buffer and returns a buffer identifier.

In case study 3, the parallel implementation of volume estimation refers to the domain decomposition technique due to the physical feature of the surface force of digital images. The integrated AOS and GAC for image segmentation process are generated by an explicit strategy. Using FDM to discretize second order PDEs of AOS and GAC, large sparse meshes are generated. Thus, mesh generation of FDM obtains a curve smoothing for tumor edge reconstruction. Domain decomposition corresponds to a number of meshes of image domain and number of blocks. Matching between a number of blocks or subdomain and tasks are done based on an intensity pattern depended on the features in images, local and global meshed, communication design between neighborhood meshes and mapping the subdomain into a number of tasks.

The parallel algorithm for case study 2 refers to the Figure-8. The parallel algorithm of ANN starts with generating the largest amount of grids as an initial weight process on the GPU. The parallel programming using CUDA with C language, demonstrating the flow for GPU implementation is shown in Figure-8. The process starts with generating the random number. Random number has been generated by using it for initial weight process. The first part is in CPU where all variables were defined. Next, GPU starts with the kernel. In order to allocate space for GPU device, the variables are defined using CUDA API, cudaMalloc. There were two types of weight that need to find in this study. There were the weights of hidden output and weights of input hidden. The data has been loaded from this network. In addition, the kernel function call grid and block variables and it's using three angular brackets

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

<<<grid, block>>>. After running the kernel, the result is copied back to host using CUDA API *cudaMemcpy*. The result will be displayed after completing the execution of all the iterations.

HARDWARE AND SOFTWARE OF GREEN COMPUTING

These environmentally sustainable computing platforms have tremendous potential to design and structure the complex software engineering in producing integrated HPC software and embedded system. As shown in Figure-1, this paper focuses on the low-cost multiprocessor systems of the DPCS laboratory that includes the Intel Xeon Series CPUs, Intel Xeon L5420 2.50 GHz CPU, and CPU-GPU with a NVIDIA GTX460 CUDA platform and Intel server. Based on the classification of single instruction multiple data stream (SIMD), DPCS reconfigurable for memory allocation and communication paradigm on a homogeneous or heterogeneous cluster of workstations.

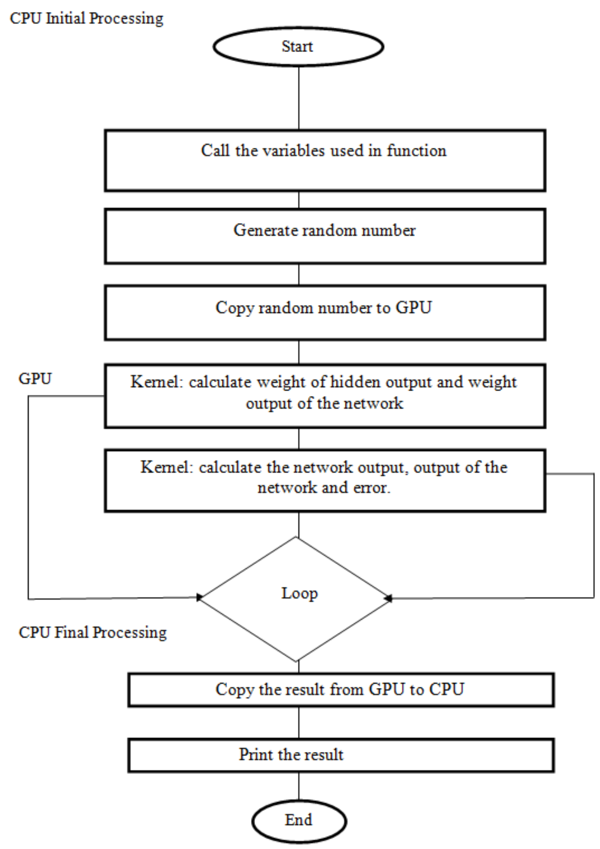

Figure-8. Parallel algorithm for case study 3.

Figure-9. Green computing based on DPCS architecture system.

CUDA/C-based "green" technology is considered as fast analysis of nanoscale computation. Thus, the integrated CPU green computing system with the main unit of PC with general CPU-GPU in CUDA terminology is utilized. The GPU is graphics card or device host and the basic unit of execution in CUDA kernel with C routine. It is executed in parallel by CUDA threads. Implementation of several threads is grouped into larger groups called thread blocks. Figure-10 and Figure-11 illustrate the CPU-GPU in CUDA prospective.

In order to run a kernel in the device, memory must be allocated in device memory before invocation of kernel functions. CUDA program has four data processing steps. First of all, allocate memory to host and device separately. Next, if it required copying the data from a host to device. It is done using the CUDA API. After that, execution in parallel of a kernel function in every core process occurs. Then the data is copied back from the GPU device to the CPU.

The parallel algorithm of case study 2 was implemented on CPU-GPU platform. The PPE indicators stated the formulation of execution time, speedup, efficiency, effectiveness and temporal performance. Speedup describes the scalability of the system as the number of processors is increased. The speedup would be the ratio of the execution time on a single processor to a number of processors. Efficiency is the speedup divided by a number of processors. It is a value between zero and one. Efficiency is countered with how well-utilized the processors in terms of communication and synchronization features. The efficiency is defined as the average contribution of the processors towards the global computation. Effectiveness is used to calculate the speedup and the efficiency. Effectiveness is the speedup divided by a number of processor execution time. Temporal performance is a parameter to measure the performance of a parallel algorithm.

VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

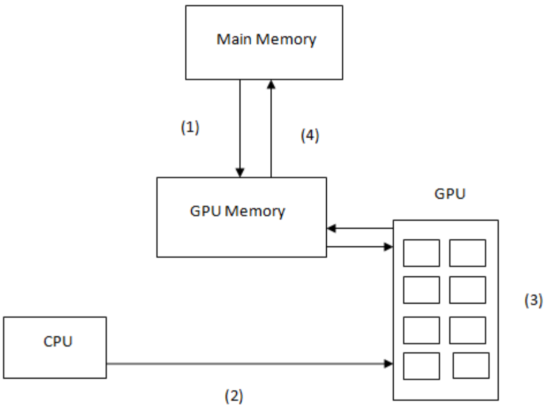

Figure-10. The process of data flow between CPU and GPU.

Figure-11. Integrated green computing with] CPU-GPU based.

The parallel algorithm of case study 2 was implemented on CPU-GPU platform. The PPE indicators stated the formulation of execution time, speedup, efficiency, effectiveness and temporal performance. Speedup describes the scalability of the system as the number of processors is increased. The speedup would be the ratio of the execution time on a single processor to a number of processors. Efficiency is the speedup divided by a number of processors. It is a value between zero and one. Efficiency is countered with how well-utilized the processors in terms of communication and synchronization features. The efficiency is defined as the average contribution of the processors towards the global computation. Effectiveness is used to calculate the speedup and the efficiency. Effectiveness is the speedup divided by a number of processor execution time. Temporal performance is a parameter to measure the performance of a parallel algorithm.

NUMERICAL RESULTS, PPE AND DISCUSSIONS

Theory and fundamental of parallel numerical simulation

The simulation graphs from the governing equation based on three different numerical methods employed are plotted as in Table-1. From the figure, it is clearly shown that AGE method gives a more accurate prediction with the experimental data of CVD compared to JB and RBGS method. In [5-6] were compared the three numerical methods for solving PDEs with parabolic and elliptic types and they have proven that AGE method is more stable and accurate compared to JB and RBGS methods. Table-1 show that AGE has the lowest iterations and the minimum run time compared to GS and JB. AGE achieves a faster convergence compared to other methods. Thus, this proves that AGE is the superior scheme in simulating the growth of silicon nanowire. Based on the root means square error (RMSE), the AGE algorithm is most accurate schemes compared to JB and GS.

Table-1. Numerical analysis for the model of silicon nanowire growth.

Numerical indicator	Method		
	AGE	GS	JB
$\Delta x, \Delta t$	2×10^{-2}	2×10^{-2}	2×10^{-2}
Matrix sizes, m	5000	5000	5000
Tolerance, ε	1.0×10^{-7}	1.0×10^{-7}	1.0×10^{-7}
Run time (sec)	4.103	6.083	6.839
Iterations number	51	69	82
Max error	5.925055×10^{-8}	9.605787×10^{-8}	7.916650×10^{-8}
RMSE	2.057652×10^{-14}	1.781009×10^{-10}	4.680225×10^{-10}

Based on 100% of simulation is contributed from previous time steps, JB performs the longest execution time and a larger number of iterations for solving the silicon nanowire growth model compared to the GS and AGE methods. Meanwhile, the GS scheme converges faster than JB scheme. However, the GS scheme converges slower than AGE scheme because 70% of the simulation is contributed from the next time steps and 30%

VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

of the simulations are collected at the previous time steps. The AGE scheme is more stable and accurate for solving a large sparse elliptic-Poisson and parabolic simulation compared to JB and GS schemes. Figure-12 shows the comparison of the parallel performance between the two methods: AGE and RBGS using 8 processors. The parallel performance of AGE is superior compared to RBGS in terms of execution time, speedup, temporal performance efficiency, and effectiveness.

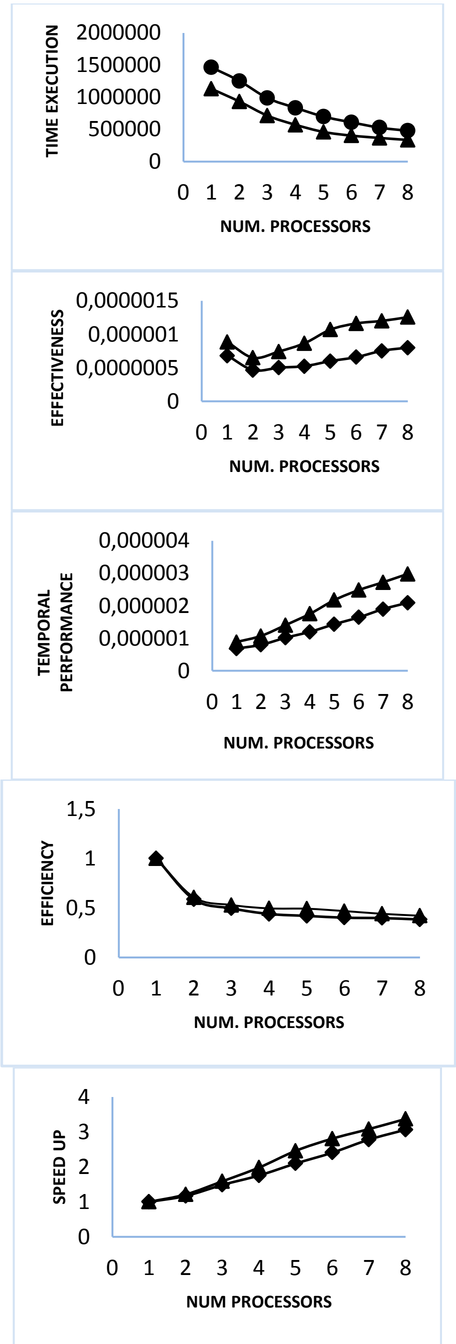

Figure-12. Execution time, speedup, temporal performance efficiency, and effectiveness versus number of processors for parallel AGE (▲) and RBGS (■).

Pattern extraction from EEG graph

Figure-13 shows the graph of the actual data or digital EEG data and the prediction of r eyes open and close condition. The graph can be described having spike patterns for predicted value. The total mean square error is 0.10 to represent the accuracy of the ANN. As a conclusion, both patterns of the graph have spike pattern and if for the next number of samples can predict that the pattern will also go up and down following the digital EEG data. This means that for an eye condition, the blink of eyes increases as the amplitude increase in EEG graph.

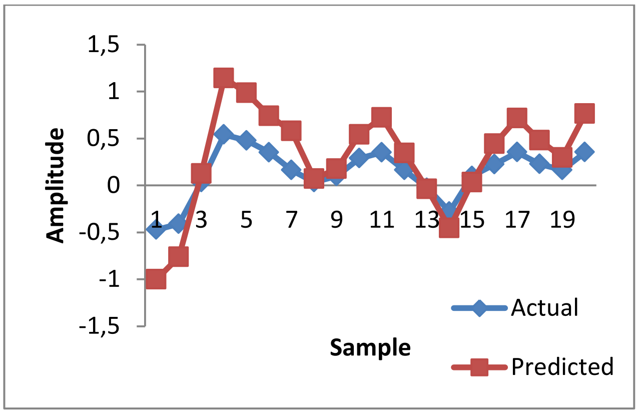

Figure-13. The graph of eyes open and close.

Figure-14 indicates the two conditions of execution time for left-right and close-open eye movement. The execution time for CPU-GPU decreases if the number of blocks is increasing for both conditions. The speedup increases as the amount of blocks increase. This is because the high number of blocks contributes to speed the execution. Basically, the efficiency is decreased when the amount of blocks increase. In comparison, the efficiency of left-right eye movement is higher than the efficiency of close-open eye movement.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

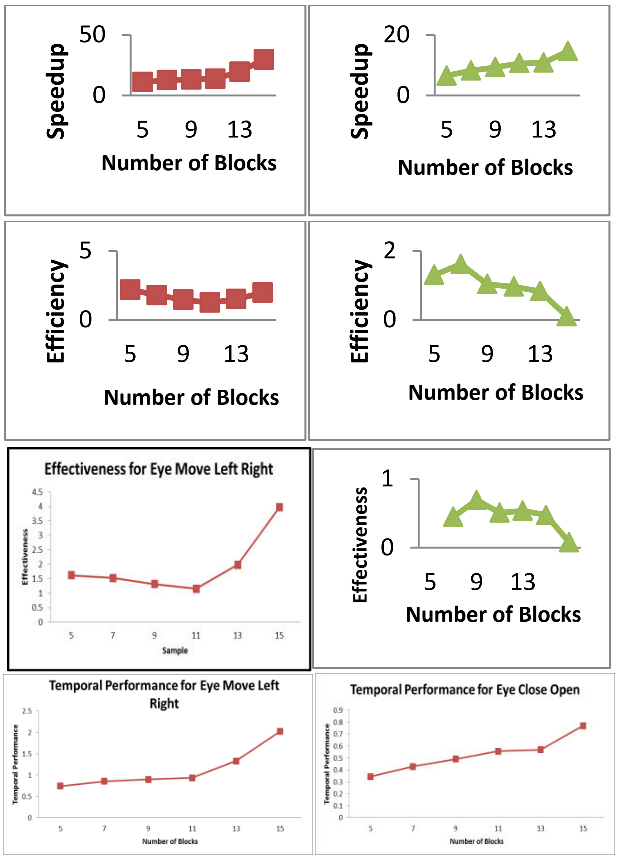

Figure-14. PPE for horizontal-vertical eye movements, left-righ (left) and close-open (right).

The effectiveness increases as the number of blocks increase as shown in Figure-14. However, the effectiveness of close-open eye movement is decreasing after 9 blocks. This could be due to CPU-GPU load balancing, waiting time for transferring data imbalance and message latency between CPU and GPU. Temporal performance increases with a number of blocks for both conditions. The optimum PPE for left-right and close-open eye movement conditions is 15 blocks. For both cases, the parallel algorithm is much better than sequential algorithm due to its PPE.

Digital images in 3D visualization

Figure-15 shows the 3D visualization of tumour growth based on VE on the green computing platform. According to the physical requirement, VE is the alternative method to reconstruct 3D visualization of tumour growth. The restructuring the tumour contours from the 2D surface of digital images to physical features of 3D representation is the simplification method to be considered.

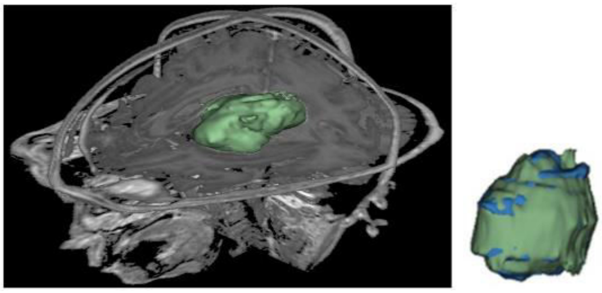

Figure-15. 3D visualization tumour growth based on VE.

Figure-16 demonstrates the domain decomposition strategy subdivide the geometric domain into subdomain s_i and mapping the s_i to a number of tasks t_i. In terms of minimizing the inter-processor communication, the static mapping is used to minimize the interactions in a wide *range* of algorithms and each *task* performs the same size of s_i.

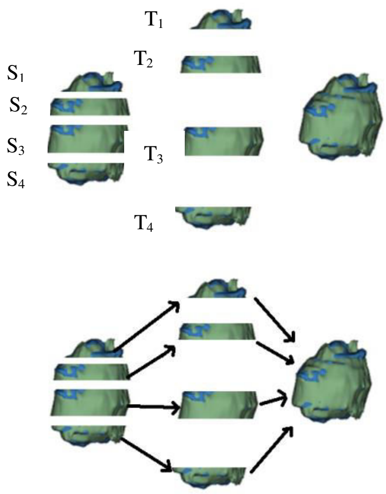

Figure-16. From domain decomposition techniques to task mapping process.

Based on PPE, its formulation and Table-2 shows the time of execution decreases when the number of processors increases. This is due to the distribution of a number of tasks from the server to clients. According to Amdahl's law, the speedup increases if the number of processors increases up to the optimum level. It is due to the effect of communication cost and message latency among the processors.

VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-2. PPE for AGEB, AGED, RBGS and JB.

AGEB				
Runtime	**Speedup**	**Efficiency**	**Effective**	**Temporal perform**
5.310174	1	1	0.188318	0.188318
4.31451	1.23077	0.41026	0.09509	0.23178
3.789342	1.401345	0.350336	0.092453	0.263898
3.239087	1.639405	0.327881	0.101226	0.308729
3.00531	1.766931	0.294488	0.097989	0.332744
2.578605	2.05932	0.294189	0.114088	0.387807
2.185948	2.429232	0.303654	0.138912	0.457467
AGED				
Runtime	Speedup	Efficiency	Effective	Temporal perform
5.960406	1	1	0.167774	0.167774
5.35987	1.11204	0.37068	0.06916	0.18657
4.42726	1.346297	0.336574	0.076023	0.225873
3.390985	1.757721	0.351544	0.10367	0.2949
3.176066	1.876663	0.312777	0.098479	0.314855
2.766507	2.154488	0.307784	0.111254	0.361467
2.310174	2.580068	0.322508	0.139604	0.432868
RBGS				
Runtime	Speedup	Efficiency	Effective	Temporal perform
6.558732	1	1	0.152468	0.152468
5.53471	1.18502	0.39501	0.07137	0.18068
4.777107	1.372951	0.343238	0.071851	0.209332
4.124067	1.590355	0.318071	0.077126	0.242479
3.536049	1.854819	0.309137	0.087424	0.282802
3.462952	1.893971	0.270567	0.078132	0.288771
3.043975	2.15466	0.269333	0.088481	0.328518
AGEB				
Runtime	Speedup	Efficiency	Effective	Temporal perform
11.80103	1	1	0.084738	0.084738
11.6227	1.01535	0.33845	0.02912	0.08604
11.00043	1.072779	0.268195	0.02438	0.090906
10.89907	1.082756	0.216551	0.019869	0.091751
10.59234	1.11411	0.185685	0.01753	0.094408
10.18893	1.158221	0.16546	0.016239	0.098146
9.080128	1.299654	0.162457	0.017891	0.110131

The efficiency decreases when the number of processors increases. The degradation of efficiency is due to the total overhead, concurrency and communication cost, idle time, start-up time and waiting time. The effectiveness increases with the increasing number of processors and the increasing of speedup. However, the

ARPN Journal of Engineering and Applied Sciences

effectiveness decreases when the number of processors exceeds 12. Table-2 shows the overall performance of the parallel programming based on execution time, speedup, efficiency, effectiveness and temporal performance.

CONCLUSIONS

According to the requirements of big data analytics for a large-scale numerical solution, three grand challenge applications were investigated. The requirements dealing with nanoscale computation for theoretical nanotechnology application is big data orientation for pattern extraction of eye-movement and 3D image construction. All the case studies were dealing with mathematical modelling of PDEs, AOS, GAC, ANN. The FDM is selected to discretize the domain and to obtain the numerical methods for sequential and parallel computation. The result of parallelization of ANN might reduce run time as well as training process is not time-consuming. In most cases, AGE is a superior method in terms of both performance indicators. Results obtained by AGE shows reasonably good agreement with experimental data. Therefore, the AGE scheme is suggested to be used for solving the multidimensional problem, although its governing equation is considered more complex compared to JB and RBGS methods. Thus, AGE initially predicating to a large extent on secondary data and influencing the big data analytics. Parallel implementation, domain decomposition technique and the corresponding block matrixes of the totally disconnected meshes were discussed. The requirement is based on the domain decomposition and partitions the domain into a number of subdomains. The fine-grained authorization and mesh generation were generated to a large extent on secondary data and to obtain the big data analytics involving a comprehensive analysis the numerical analysis, and PPE. For all case studies, the numerical analysis and PPE indicators were indicated the big data analytics in terms of numerical results, accuracy properties, error estimation, speedup, efficiency, effectiveness, temporal performance. As a conclusion, large-scale numerical algorithms with performance in numerical analysis, convergence, accuracy and PPE have been important features to analyse the big data simulation of 3 grant challenge applications.

ACKNOWLEDGEMENT

The author acknowledges the continuous financial support from Universiti Teknologi Malaysia (UTM) and funding from the Ministry of Science, Technology and Innovation (MOSTI under the Fundamental Research Grant Scheme (FRGS).

REFERENCES

[1] Alias, N., Mustaffa, M. N., Saipol, H. F. S., & Ghani, A. C. A. "High performance large sparse PDEs with parabolic and elliptic types using AGE method on DPCS." Advanced Science Letters, vol 20, no. 10-1, pp. 1956-1960, 2014.

[2] Alias, N., Satam, N., Othman, M. S., Teh, C. R. C., Mustaffa, M. N., & Saipol, H. F. S. "SCLP Software for Thermal Process Control of Nanowire Fabrication and Multilayer Nanochip System." So Me T. 2014.

[3] Alias, N., Anwar, R., Teh, C.R.C., Satam, N., Hamzah, N., Ghaffar, Z.S.A., Darwis, R. and Islam, M.R. "Performance evaluation of multidimensional parabolic type problems on distributed computing systems." Computers and Communications (ISCC), 2011 IEEE Symposium on. IEEE, 2011.

[4] Alias, N., Satam, N., Othman, M. S., Teh, C. R. C., Mustaffa, M. N., & Saipol, H. F.. High Performance Nanotechnology Software (HPNS) for Parameter Characterization of Nanowire Fabrication and Nanochip System. In International Conference on Intelligent Software Methodologies, Tools, and Techniques. Springer International Publishing, no. 513, pp. 251-268, 2014, September.

[5] Sahimi, M.S., Alias, N. and Sundararajan, E. "The AGEB algorithm for solving the heat equation in three space dimensions and its parallelization using PVM." International Conference on Computational Science. Springer Berlin Heidelberg, 2001.

[6] Alias, N., Sahimi, M.S. and Abdullah, A.R. "The AGEB algorithm for solving the heat equation in two space dimensions and its parallelization on a distributed memory machine." European Parallel Virtual Machine/Message Passing Interface Users' Group Meeting. Springer Berlin Heidelberg, 2003.