Parallel Processing of Numerical Integrations on High-Performance Computing cluster

May Myint Thwe*, Pho Kaung**

Abstract

Numerical methods can be used to solve the increasing complexity of scientific problems. The solution of scientific and engineering problems sometimes requires integration of an expression. High-Performance Computing(HPC) uses a cluster of servers to split complex computations into smaller tasks that can be calculated in parallel on multiple servers. HPC Services for Excel include service-oriented architecture (SOA) clients and services that enable developers to quickly convert workbooks to run on a cluster. As computations and models of workbooks increase in size and complexity, the time required for calculation grows and workbook use becomes cumbersome due to long waits for calculation results. The HPC system is developed to provide users speed-up in Excel calculations on it by reducing in size, complexity and calculation time. This work describes how an HPC cluster can be utilized to run Excel calculations in parallel and generating Fresnel diffraction of arbitrary dimensions by using Fresnel integrals. Excel includes Visual Basic for Applications (VBA) language. VBA is the tool to develop programs that control Excel. The system is operated by Windows® HPC Server 2008 R2 and Office Excel 2010 process is used to calculate the workbook in the cluster.
Keywords: High Performance Computing, HPC cluster, Windows® HPC Server 2008 R2, SOA clients, Microsoft Office Excel 2010, Numerical Integrations

1. Introduction

Numerical methods are involved in almost every aspect of engineering. Multi-dimensional Integration needs extra computing power and therefore is a scientific application to be run in parallel on the High-Performance Computing (HPC) cluster. HPC uses supercomputers and computer clusters to solve advanced computation problems. The key to any HPC solution is parallelization.[12] Parallelization refers to splitting a complex computation into component parts that can be run simultaneously on multiple servers. Large problems are divided into smaller parts and distributed among the many computers. Clusters are just multiple computers connected together to solve a specific problem. [6, 7, 9]

To run an Office Excel 2010 workbook in a Windows HPC cluster, some part of the calculation must be able to run in parallel. Windows HPC Server 2008 R2 enables running multiple instances of Office Excel 2010 in a Windows HPC cluster, where each instance is running an independent calculation or iteration of the same workbook with a different dataset or parameters. Windows HPC Server 2008 R2 also supports one or more Windows Communication Foundation (WCF) Broker nodes, which enable running WCF services in the cluster. WCF enables and simplifies building service-oriented applications and Windows HPC Server 2008 R2 supports Service Oriented Architecture (SOA) model that enables running WCF services in a Windows HPC cluster. WCF and SOA provide a simple framework for building service applications to run on a cluster.[10]

In a SOA system, distinct computational functions are packaged as software modules called *services*. Developers can create cluster-SOA client applications to provide access to services that are deployed to a Windows HPC cluster. The client application submits a job that contains a **Service** task to the cluster, initiates a session with the broker node, and sends service requests and receives responses (calculation results). This work is parallel processing of numerical integrations for Fresnel transform for faster calculation time and better performance on HPC cluster with Excel VBA. This system addresses how the calculation

* Lecturer, Dr, Department of Computer Studies, University of Yadanabon
** Rector, University of Yangon

time required for long-running workbooks can be reduced to give users faster access to critical information.

2. Developing HPC Cluster for Excel

2.1 Hardware Development

HPC uses a cluster of servers to split complex computations into smaller tasks that can be calculated in parallel on multiple servers. A Windows HPC cluster consists of several networked servers.[4, 6]

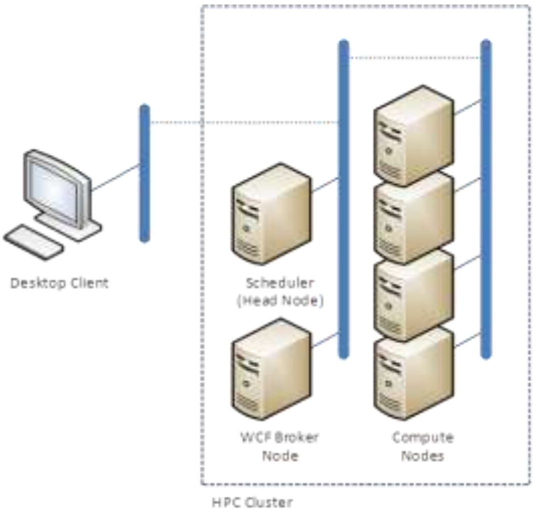


Figure 1. Windows HPC cluster topology

Each server in the cluster performs one or more specific roles. A Windows HPC cluster uses a Scheduler (hosted on a *Head node)* to manage computation tasks and Compute nodes (that is, servers) to perform the actual work. In a typical cluster calculation, a client computer—such as a desktop computer running the Windows operating system—connects to the Scheduler over the network to submit a job. The Scheduler handles the work of managing the Compute nodes, locating available resources, and running the computation. When the Compute nodes complete units of work, the Scheduler sends calculation results to the client computer.

2.2. Software Requirements

2.2.1. Windows® HPC Server 2008 R2

Windows® HPC Server 2008 R2 integrates with Microsoft® Excel® 2010 to help run workbooks and user-defined functions (UDFs) faster by offloading calculations to the cluster.[10] If a workbook contains independent units of calculation, multiple compute nodes can perform the calculations simultaneously. Parallel computation can significantly reduce workbook calculation time and make calculations across larger data sets more feasible. Many complex and long-running workbooks run iteratively—that is, they perform a single calculation many times over different sets of input data. These workbooks might contain complex Microsoft Visual Basic for Applications (VBA) functions or calculation intensive XLL Excel Library add-ins. This type of workbook is suitable for cluster acceleration.

Windows HPC Server 2008 R2 now enables running multiple instances of Office Excel 2010 in a Windows HPC cluster, where each instance is running an independent calculation or iteration of the same workbook with a different dataset. Many complex and long-running workbooks run *iteratively*—that is, they perform a single calculation many times over different sets of input data. These workbooks might include intensive mathematical calculations contained in multiple worksheets, or they might contain complex Microsoft VBA functions.[11]

When a workbook runs iteratively, the best option for parallelizing the calculation is to run the entire workbook in the cluster. In this model, individual calculations need not be split into component parts, but the overall calculation—generating the results from many individual calculations—can be run in parallel.

Every application that benefits from this solution has three parts: the *workbook,* a *service,* and a *client.* Office Excel 2010 and Windows HPC Server 2008 R2 must be installed on each cluster server. Microsoft Office Excel 2010 must be installed on the client computer. The client computer can run the Windows XP with Service Pack 3, Windows Vista®, or Windows 7 operating system.

2.2.2. The workbook

The workbook refers to a standard Excel workbook. This solution runs multiple instances of Office Excel 2010 on cluster servers, meaning that it supports workbooks that use VBA or XLL add-ins or an Excel Add-in (XLA) as well as external resources (provided these resources are accessible from the servers).

In some cases, workbooks may need to be modified to work with this solution. When Office Excel 2010 runs on the server, it does not support user interaction. Windows HPC Server 2008 R2 includes a comprehensive pop-up manager that can handle occasional dialog boxes and pop-up messages, but it is not designed to support interactive Office Excel 2010 features: Users cannot create a PivotTable when running on the server, for example, because doing so requires user interaction.

When a workbook is used in this scenario, it is important to identify the input values and the output or result of the calculation. The input values might be cells within a worksheet in which the user enters a value, or they might be parameters to a VBA function. The output might be a second set of cells within a worksheet or the result of a VBA function. Identifying the workbook input and output values are important when developing the service, which will run on the cluster servers and execute the workbook calculation.

2.2.3. The Service

The service is a WCF service that controls the execution of the Office Excel 2010 workbooks on the cluster servers. The service starts Office Excel 2010, calculates a workbook, and returns results.

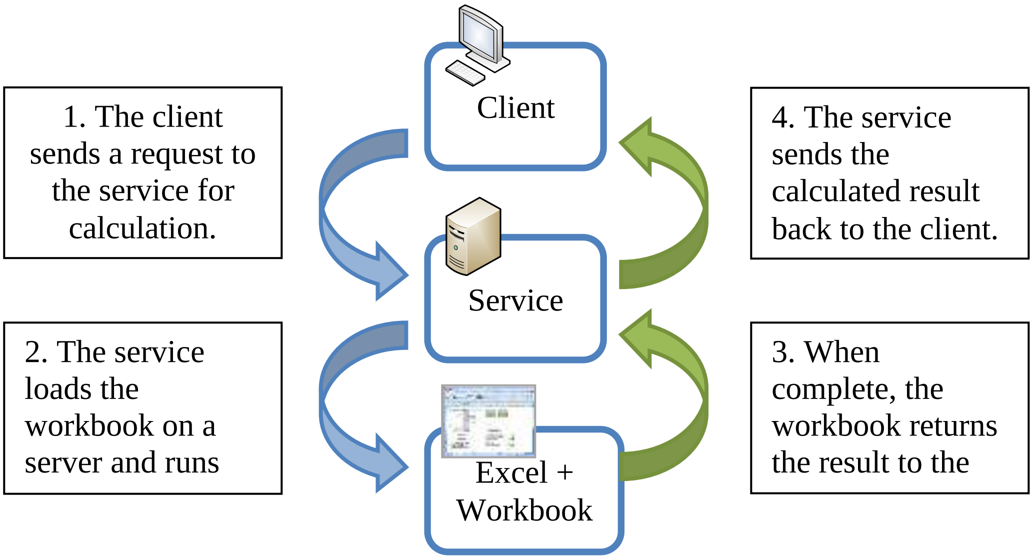
Windows HPC Server 2008 R2 offers an SOA model that enables running WCF services in the cluster. In a Windows HPC cluster, the WCF Broker node handles managing and hosting the service library. The Scheduler handles assigning and managing compute resources. From the standpoint of the developer, using WCF removes all the complexity of hosting and managing the service. The developer need only build the calculation functions.

Once the service has started Office Excel 2010 on a server, it can interact with the workbook in either of two ways to provides access to the Excel object model (similar to using Excel automation on a client computer): Using the Excel object model, the service can read from and write to the workbook, and it can trigger calculation of the workbook. For example, in a simple workbook, the service might write values into some input cells, recalculate the workbook, and then read the values of some output cells.

2.2.4. The Client

The client is a program that controls the overall calculation. It is designed to work with the WCF service, and it tells the service which workbook to calculate along with the parameters or options to use, and it receives results when the calculation is complete (see Figure 2). The client typically runs on a user's client computer. Client programs can be written in any language supported by Microsoft .NET or COM, and the client can be a Windows application or a command-line program.

Figure 2. Application flow for a single service call of running workbooks on a cluster



2.2.5. Service-Oriented Architecture

Service oriented architecture (SOA) is an approach to building distributed, loosely coupled systems. HPC Services for Excel uses a SOA infrastructure to run Excel jobs on a cluster. HPC Services for Excel includes ready-made SOA clients and services that enable developers to quickly convert workbooks to run on a cluster. The HPC Services for Excel features support a number of programming models that enable Excel workbooks to run in parallel on a cluster. The features include SOA clients and services that enable developers to quickly convert workbooks to run on a cluster. In a SOA system, services can be distributed across a network and accessed by other applications. A *client* application provides an interface for the end-user to access the functionality of one or more services. The client-side and developer features are available in the Microsoft® HPC Pack 2008 R2 client utilities and software development kit (SDK). The cluster-side features are available with the Enterprise Edition of HPC Pack 2008 R2.

3. Numerical Integration

Numerical methods are involved in almost every aspect of engineering. Scientific and engineering problems may be difficult or even impossible to obtain an expression for the integral of a particular function. But by using numerical methods, a value for the definite integral can always be obtained.[5, 8]

3.1. Numerical calculations for Fresnel transform

In linear optics, electromagnetic diffraction through a plane screen is described in good approximation with Fresnel transform.[1, 2, 3] Let $V_0(s, t)$ the electromagnetic field on a point (p, q, z= 0) on a plane z = 0 and V(x, y, z) the electromagnetic field on point (x, y, z) as shown in Figure 3.

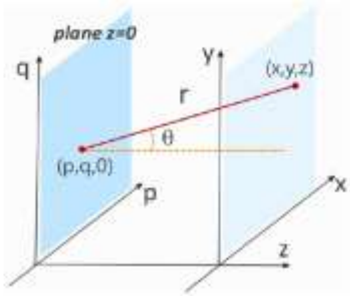


Figure 3. The electromagnetic field on point (x, y, z)

At distance z>> λ the electromagnetic field V(x, y, z) is described with good approximation through the Fresnel transform:

$$V(x,y,z) = \iint V_0\ (p,q) e^{ikr}\, e^{\frac{i\pi}{\lambda z}((x-p)^2+(y-q)^2)}\, dpdq \qquad (1)$$

where $r = [(x-p)^2 + (y-q)^2 + z^2]^{1/2}$; i is the imaginary unit.

By substitution of Euler's formula,

$$e^{ikr} = \cos kr + i\ \sin kr$$

In Equation 1, the electromagnetic pattern V(x, y, z) on a plan (x, y, z) is composed of real part and imaginary part:

$$V(x,y,z) = \iint V_0(p,q)(\cos kr + i \sin kr)\ [\ (\cos \frac{\pi}{\lambda z}[(x-p)^2 + (y-q)^2]) + i\left(\sin \frac{\pi}{\lambda z}[(x-p)^2 + (y-q)^2]\right)]dpdq \qquad (2)$$

Fresnel equation slit up in real and imaginary parts

$$Re(x,y,z) = \iint V_0(p,q)(\cos\ kr) \times (\cos\ \frac{\pi}{\lambda z}[(x-p)^2 + (y-q)^2])\ dpdq$$

$$Im(x,y,z) = i \iint V_0(p,q)(\sin\ kr) \times \left(\sin \frac{\pi}{\lambda z}[(x-p)^2 + (y-q)^2]\right)\ dpdq \quad (3)$$

To run a computer simulation, apply a sampling in plane (p, q, z = 0) and in plane (x, y, z). Let consider a segment L_p centered around the p axis, and L_q centered around the q axis, a segment X_m centered around the x axis, and Y_n centered around the y axis.

N_p is the number of samples in the p axis, N_q is the number of samples in the q axis, N_m is the number of samples in the x axis, and N_n is the number of samples in the y axis.

$$V(x,y,z) = \iint V_0(p,q)(\cos kr + i \sin kr)\ [\ (\cos \frac{\pi}{\lambda z}[(x-p)^2 + (y-q)^2]) + i\left(\sin \frac{\pi}{\lambda z}[(x-p)^2 + (y-q)^2]\right)]dpdq$$

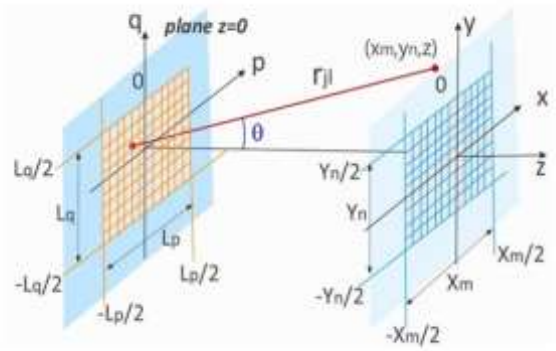


Figure 4. Sampling intervals on two planes

As shown in Figure 4, the sampling intervals and the relative ranges are as follow;

$$\Delta p = \frac{L_p}{N_p}\ ;\ \ j = -\frac{N_p}{2}\ , \dots, -1, 0, +1, \dots, \frac{N_p}{2}\ \ ; p_j = j\Delta p$$
$$\Delta q = \frac{L_q}{N_q}\ ;\ \ l = -\frac{N_q}{2}\ , \dots, -1, 0, +1, \dots, \frac{N_q}{2}\ \ ; q_l = l\Delta q$$
$$\Delta x = \frac{L_x}{N_x}\ ;\ \ m = -\frac{N_x}{2}\ , \dots, -1, 0, +1, \dots, \frac{N_x}{2}\ \ ; x_m = m\Delta x$$
$$\Delta y = \frac{L_y}{N_y}\ ;\ \ j = -\frac{N_y}{2}\ , \dots, -1, 0, +1, \dots, \frac{N_y}{2}\ \ ; y_n = n\Delta y \qquad (4)$$

The distance r_{jl} between the sample (p_j, q_l, z = 0) and the sample (x_m, y_n, z) is shown in Equation 5:

$$r_{jl} = \left[\left(x_m - p_j\right)^2 + (y_n - q_l)^2 + z^2\right]^{1/2} \qquad (5)$$

The discrete expression for real and imaginary parts of electromagnetic field are shown in the Equation 6:

$$\mathrm{Re}(x_m, y_n, z) = \sum_{l=-\frac{N_q}{2}}^{\frac{N_q}{2}} \sum_{j=-\frac{N_p}{2}}^{\frac{N_p}{2}} V_0\left(p_j, q_l\right)\left(\cos k r_{jl}\right) \times \left(\cos\frac{\pi}{\lambda z}\left[\left((m\Delta x) - (j\Delta p)\right)^2 + \left((n\Delta y) - (l\Delta q)\right)^2\right]\right) \Delta p \Delta q$$

$$\mathrm{Im}(x_m, y_n, z) = \sum_{l=-\frac{N_q}{2}}^{\frac{N_q}{2}} \sum_{j=-\frac{N_p}{2}}^{\frac{N_p}{2}} V_0\left(p_j, q_l\right)\left(\sin k r_{jl}\right) \times \left(\sin\frac{\pi}{\lambda z}\left[\left((m\Delta x) - (j\Delta p)\right)^2 + \left((n\Delta y) - (l\Delta q)\right)^2\right]\right) \Delta p \Delta q \qquad (6)$$

The intensity I(x_m, y_n, z) of electromagnetic field is equal to the square of module of electromagnetic field V(x_m, y_n, z)

$$I(x_m,\ y_n,\ z) = \mathrm{Re}(x_m,\ y_n,\ z)^2 + \mathrm{Im}(x_m,\ y_n,\ z)^2 \qquad (7)$$

4. Results and Discussion

The test platform was built on a cluster with one head node, one broker node and three compute nodes. The system was operated by Windows® HPC Server 2008 R2 and Office Excel 2010 process was used to calculate the workbook in the cluster. The Excel workbook contains the four modules and reference to HPC library (Microsoft_Hpc_Excel) is added and it may be ready to run on Windows HPC cluster. Through the job template, the HPC administrator can specify which resources. To evaluate the performance of the Excel worksheet running many times the same discrete Fresnel transform with same simulation input parameters but with different number of HPC cores, it is created for this purpose by the HPC management console HPC job template with different number of usable cores. The results of Local and HPC with different cores for the number of samples of 64×64 are shown in Figures 5, 6, 7, 8, 9 and 10.

4.1. Run Time on HPC cluster with different number of cores

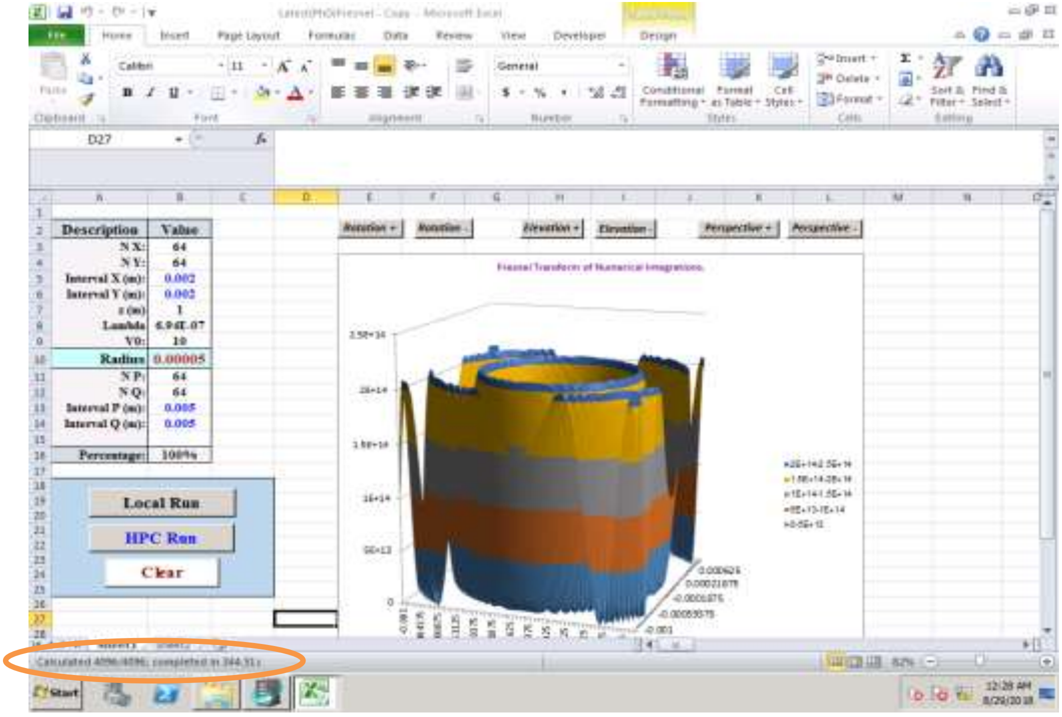


Figure 5. Run time of Local Run taking 344.51 s

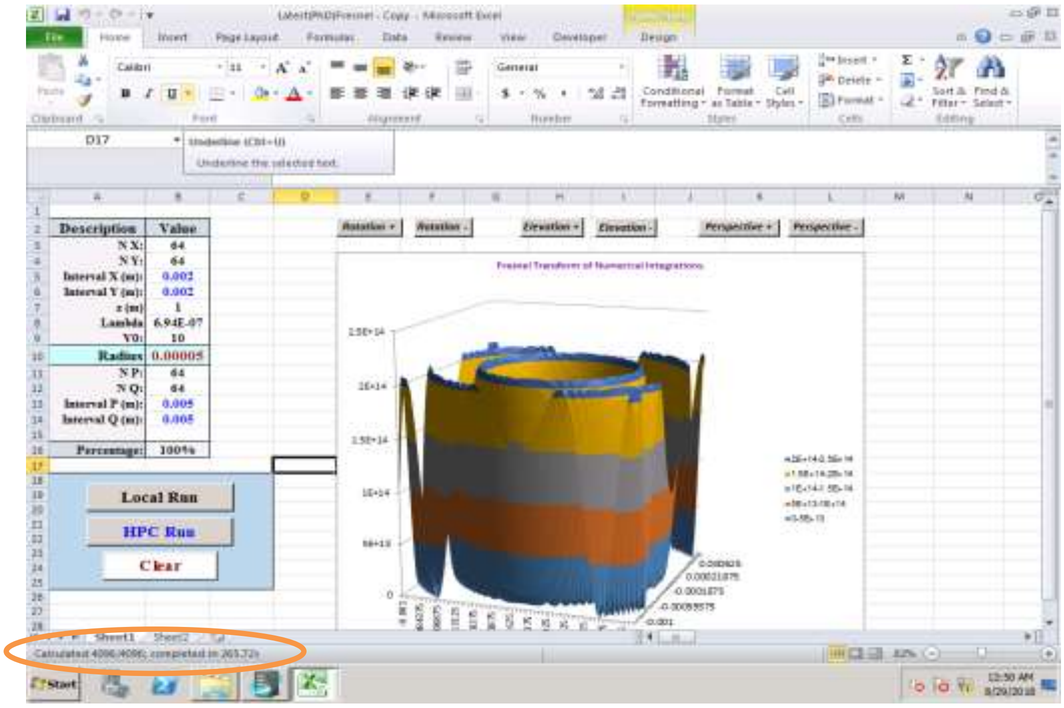


Figure 6. Run time of HPC with 2Cores taking 265.72 s

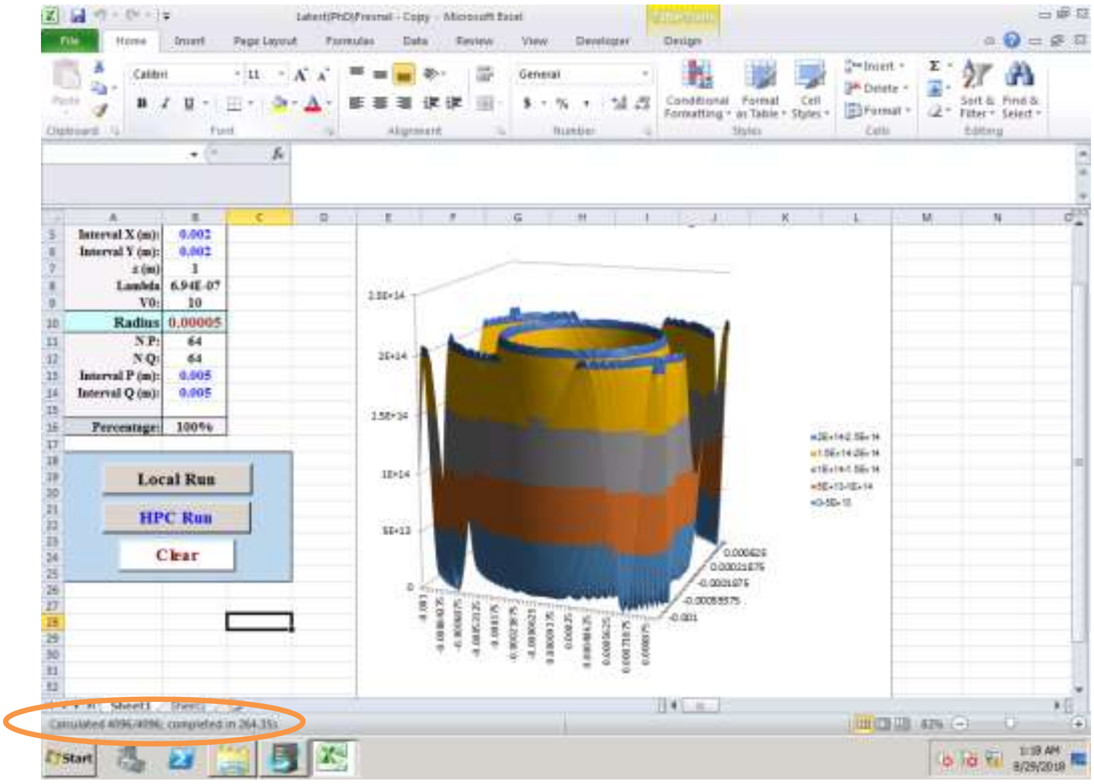


Figure 7. Run time of HPC with 4Cores taking 264.35 s

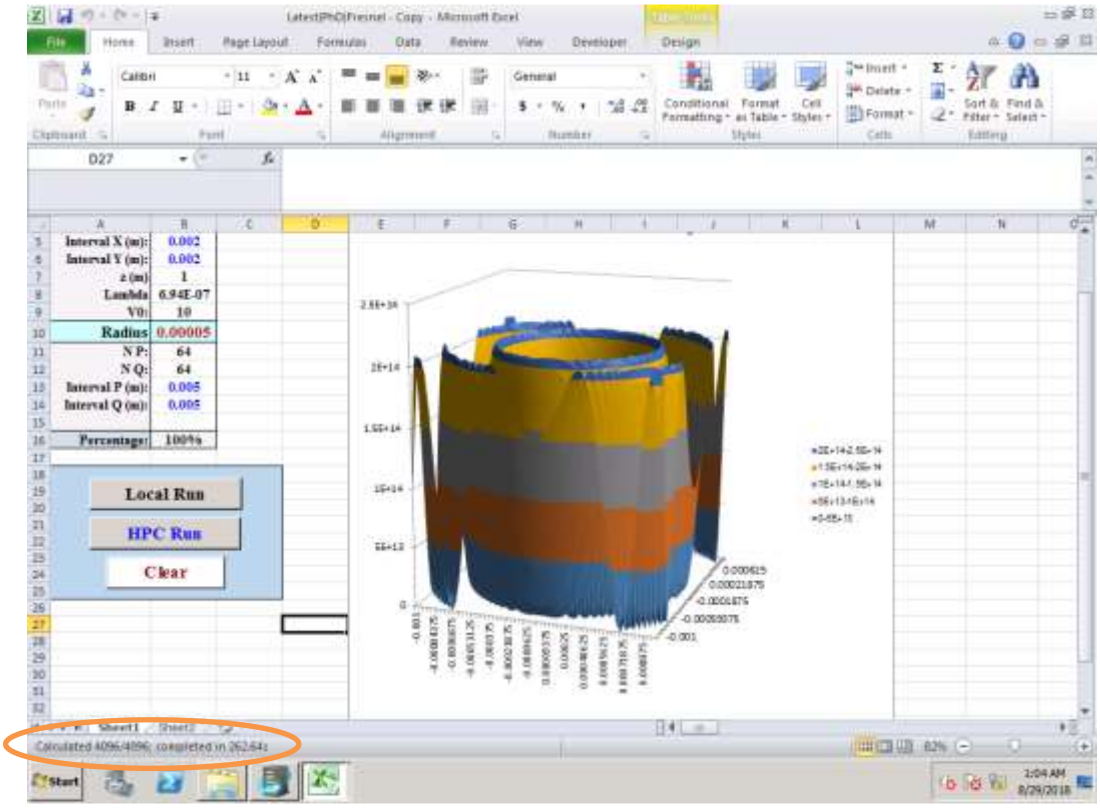


Figure 8. Run time of HPC with 8Cores taking 262.64s

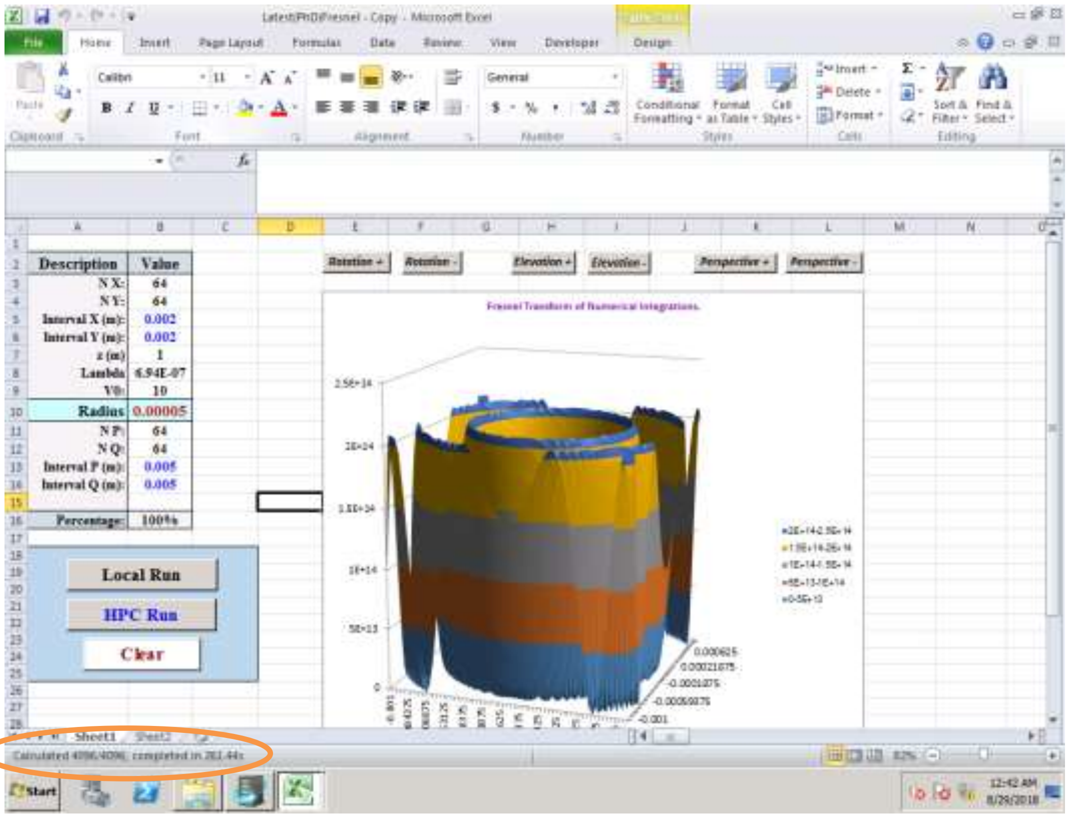


Figure 9. Run time of HPC with 16Cores taking 261.44 s

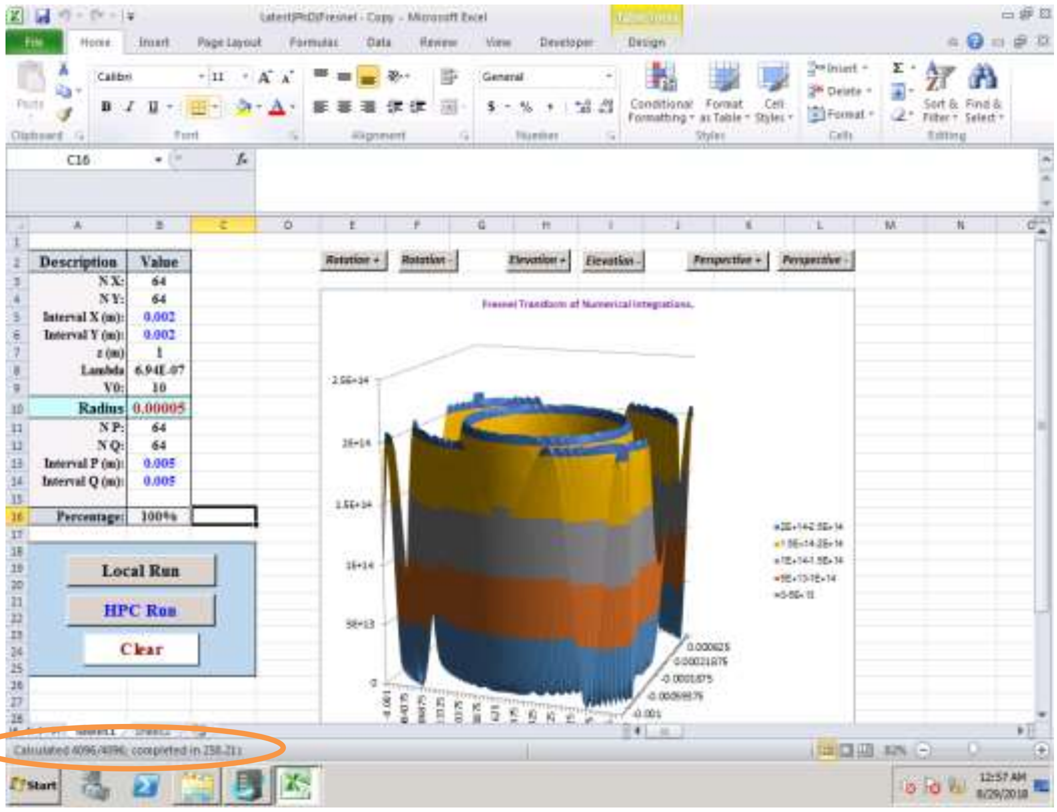


Figure 10. Run time of HPC with 32Cores taking 258.21 s

HPC with 2Cores is faster than Local by 78.28s.

HPC with 4Cores is faster than HPC with 2Cores by 1.37s and faster than Local by 79.65s.

HPC with 8Cores is faster than HPC with 4Cores by 1.71s, faster than HPC with 2Cores by 3.08s and faster than Local by 81.36s.

HPC with 16Cores is faster than HPC with 8Cores by 1.2s, faster than HPC with 4Cores by 2.91s, faster than HPC with 2Cores by 4.28s and faster than Local by 82.56s.

HPC with 32 Cores is faster than HPC with 16 Cores by 3.23s, HPC with 8 Cores by 4.43s faster than HPC with 4Cores by 6.14s, HPC with 2Cores by 7.51s and Local by 85.79s.

Figure 11 shows run time (seconds) taken by Local and HPC with the number of samples of 64 X 64.

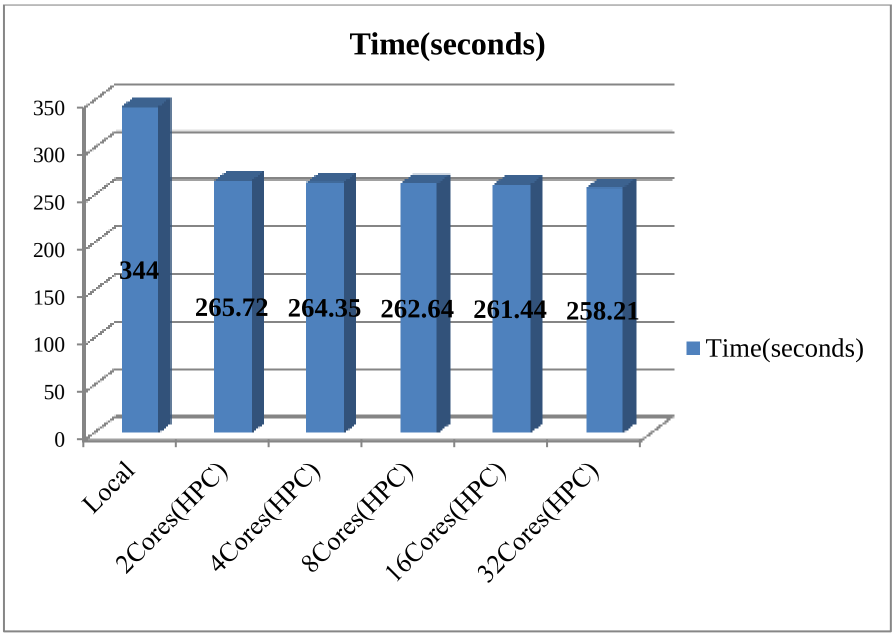


Figure 11. Comparison among run times taken by Local and HPC with different Cores for 64×64

It can be seen that HPC with 32 Cores can make Excel workbook run **1.3 times faster** than Local Run. Therefore, HPC Services for Excel with a Windows HPC cluster can improve calculation performance.

5. Conclusion

The basic objective of this system is to reduce the calculation time required for long-running Spread-sheet Calculations model in a Windows HPC cluster. Windows HPC Server 2008 R2 enables multiple instances of Office Excel 2010 run in a Windows HPC cluster. Faster execution of highly compute intensive tasks of multidimensional integration for advanced computation may be achieved by using HPC cluster. By using Windows HPC Server 2008 R2, the set-up of an HPC cluster can reduce calculation times for Excel 2010 workbooks. This system is able to remove all the complexity of hosting and managing the service and quickly make workbooks to be run on a cluster by applying WCF service and SOA model on a Windows HPC cluster. By using High-performance computing (HPC) solutions and installed facilities, the needs of scientific computing have been addressed. The increasing complexity of scientific problems can be solved by numerical method which is implemented well using Excel/VBA. This research also perform to reduce the calculation time required for long-running Spread-sheet Calculations and to get better performance for Fresnel transform of Numerical Integration by using HPC cluster. Therefore, this system is able to obtain improved calculation performance and faster calculation time for numerical integration by applying the Excel VBA on HPC cluster.

Acknowledgments

First of all, I would like to express my sincere gratitude to Dr Pho Kaung, Rector, University of Yangon for his enthusiastic guidance and various suggestions. I wish to express my sincere thanks to Dr Maung Maung Naing(Rector), Dr Si Si Khin(Pro-Rector) and Dr Tint Moe Thuzar (Pro-Rector), University of Yadanabon, for their kind permissions for the submission of this research paper. I also wish to thank Daw Wai Wai Myint (Lecturer and Head of Department of Computer Studies, University of Yadanabon) for her encouragement and support.

References

1. Aime C., Aristidi E. and Rabbia Y., (2012), *The Fresnel Diffraction: A Story of Light and Darkness*, EAS Publications Series.
2. Billo, E. J., (2007), *Numerical Techniques for Fresnel Diffraction in Computational Holography 1*.
3. Bradley, B. K., Beylkinz, G., Brandon A. J., and Axelradx, P., *A New Numerical Integration Technique in Astrodynamics*, AAS 12-216.
4. Buyya, R., (1999), *High Performance Cluster Computing*, Prentice Hall.
5. David, G. L., (2010), *Some Useful Excel VBA Codes for Numerical Methods in Engineering*, AIAA 2010-1360.
6. Eadline, D., (2009), *High Performance Computing For Dummies*, Indianapolis, Indiana Sun and AMD Special Edition, Wiley.
7. Grama, A. G., Karypis, G., and Kumer, V., (2003), Introduction *to Parallel Computing*, NewYork, Benjamin, Cummings.
8. Hoffman, (2001), *Numerical Methods for Engineers & Scientists*.
9. Karniadakis, G.E., and Kirby, R. M., (2008), *Parallel Scientific Computing in C++ and MPI*, Cambridge: CUP.
10. Microsoft Corporation, (2010), *THEWinServer2008_R2StepByStep*.
11. Microsoft Corporation, (2010), *Accelerating Excel 2010 with Windows® HPC Server 2008 R2: Building VBA*.
12. Middleton, A. M., Ph.D., (2011), *Introduction to HPCC High-Performance Computing Cluster*.