INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING VOL. 12 NO. 4 (2020) 181-186

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Mathematical Model for Approximation the Efficiency of Parallel Computing on Single Board Cluster with Least-squares Approximation

Chotitham Thanrak[1], Sasalak Tongkaw[2]

[1] Faculty of Science and Technology,
Songkhla Rajabhat University, Songkhla, THAILAND

[2] Faculty of Science and Technology,
Songkhla Rajabhat University, Songkhla, THAILAND

DOI: https://doi.org/10.30880/ijie.2020.12.07.020
Received 3 September 2020; Accepted 12 October 2020; Available online 31 October 2020

Abstract: This research aims to study the relationship between parallel processing efficiency and several nodes on a single board cluster using a mathematical model, approximating least squares. This research tested on the Raspberry Pi single-board in the form of a high-performance computing system. It divided the tasks that need to be processed in each particular part and sent it to each unit to process simultaneously via the MPI (Messaging Passing Interface). This process is the standard division of work with communication between processors in the form of messages on the cluster system. It consists of eight nodes of Raspberry Pi. It measures the instruction set's ability to perform decimal operations per second or Floating-point Operation Per Second (FLOPS) with High-Performance Linpack Benchmarks (HPL). As a result, the efficiency of the ability to process instruction set in decimal per second increases the performance continuously when increasing the number of the node on the cluster. Which corresponds to the mathematical model obtained f(x) = 1.0684x^(0.8256).It shows a relationship between parallel processing performance values and the number of nodes on the cluster and can be estimated with the mathematical model above.

Keywords: Cluster, least-squares approximation, parallel computing

1. Introduction

Here nowadays, we use computer systems in order to process or deal with big data. There is a tendency for demand to increase rapidly. As a result, computer systems need to have higher processing efficiency to support extensive data and sophisticated data analysis or solve mathematical equations. It requires a highly efficient computer system and has a high price to procure.

However, there is a Single Board Computers (SBC) use to run mainstream operating systems such as Raspberry Pi, which creates a range of SBCs with flat cost, high performance, and expandable features for clustering. SBC cluster can also integrate with the full range of IoT and Smart City systems and the game application with sufficient power. [1] This compute application can provide extensive data source computation. It can reduce bandwidth across the network and latency. [2]

There are many types of SBC; however, this research use SBC with Model B + of Raspberry Pi 3. As for the Raspberry Pi 3 Model B +, the Broadcom BCM2837B0 has quad-core processor speeds up to 1.4GHz from 1.2GHz, adding a dual-band wireless connection (2.4GHz and 5GHz), adding Bluetooth 4.2 chips from previously only Bluetooth Low Energy. It also supports Gigabit Ethernet via USB 2.0, speeds up to 300 Mbps, three times more than before, and comes with built-in Power over Ethernet (PoE) to enable power through the LAN port.

In the newest update, A single Raspberry Pi board contains four processing cores. [3] Each unit processes simultaneously via Messaging Passing Interface (MPI) between the processors in the form of messages on a cluster

*Corresponding author: helmy@uthm.edu.my

system. It consists of 8 nodes of Raspberry Pi and measures the instruction set's performance in decimal per second or Floating-point Operation Per Second (FLOPS) with High-Performance Linpack (HPL) program. The prestigious rankings of supercomputer rankings that many people know well refer to ranking. An analysis of Linpack and power performances of the world's TOP500 supercomputers is an example of Machines that are further capable of measuring how hardy or frail the computer is. [4] This is done by running a benchmark program called HPL; the performance is evaluated in terms of data transfer; if the process is faster than others, it would be considered more efficient. Each Raspberry Pi model has different features; Notwithstanding the following age of the Internet of Things and cluster technology. The SBC cluster, like Pi Stack, will use to construct for computing manner in employment circumstances different than education. The study presented a Pi Stack, which reduces the computing power and low cost to support edge computing applications, and it also has individual heartbeat monitoring in Pi Stack. [5] HPL is a benchmark that runs to find solutions to systems of linear equations $Ax = b$ when A is a dense matrix or matrix where most of its members are not 0, with the time complexity of working as $O(n^3)$ when using HPL to run on the device will get one FLOPS number. HPL is considered a suitable measure method since most scientific work presently implied a lot of linear system equations and measurement by floating-point numbers. That can be processed in one second reflects the ability of that parallel computer very well.

The advantage of Raspberry Pi is a wide variety from home automation, image processing, IoT, navigating robots, and also for supporting machine learning in terms of solving mathematical problems by distributed architecture in many kinds of applications such as to serve tourism applications. [6] The test of Machine Learning performance implemented in Raspberry Pi 4 with some libraries, such as DeepLearning4J and TensorFlow. Although the test represents the cluster's performance of training and executing, the performance scale also required additional nodes in the cluster. Another project uses Raspberry Pi to calculate the prime number, it presents the timing records of testing results with nine node Raspberry Pi 2. [7]

Therefore, this research presents a study of the relationship between efficiency, parallel computing, and the number of nodes on a single board cluster with an approximation of least squares by testing on the Raspberry Pi single-board in the form of high-performance computing clusters.

2. Objectives

This research aims to study the relationship between parallel processing efficiency and several nodes on a single board cluster using a mathematical model, an approximation of least squares.

3. Implementation

This research tested the high-performance of the single board on the Raspberry Pi in a parallel computing system. The testing process divides each particular part's tasks and sends it to each unit to process simultaneously via the MPI (Messaging Passing Interface). The MPI exists a standard archive for a piece concerning information for passing a piece of a message in a parallel computer. Many vendors provide MPI libraries, for example, MPICP, which frame the primary of a movable parallel programming environment. [8] The MPICH used pair MPI mpptest, programs, and goptest. Both programs afford true analyses regarding execution. Another one is MVAPICH MPI implementation on a Linux cluster of 512 nodes. However, today there is much research also test on the performance of Hybrid MPI and Open MP on a parallel application, for example, CHAMELEON. [9]

There is much research use SBC for classroom training or developing an IoT project. Some research conducts general practice during the implementation of the system. [10] However, in some cases, SBC is used for the computing section; to reduce computation power volume for a portable cluster. [5] For example, the single board can also use for clustering or test application similar to supercomputing techniques alike Open Multi-Processing (OpenMP) and Message Passing Interface (MPI). [11]

This research will implement the SBC testing operation and configuration as follow.

3.1 Testing Operation

Testing is done on a simulated network and a sole board, the Raspberry pi 3 Model B+, which consists of 2 parts, which run OS Raspbian Buster kernel 4.19. Linux based operating systems like Raspbian presents a high ability to manage Message Pass Interface (MPI). It can be sent tasks across nodes in parallel processing. A VPN server is developed in this study for clustering the computer at various places and data mining. [12] Originally, the Raspberry Pi3 Model B+ model has featured in Table 1.

Table 1 - Raspberry Pi 3 Model B+ features

Features	Raspberry Pi 3 Model B+
Processor cores	4
Processor speed (GHz)	1.4
RAM (GB)	1
Network speed (Mbit/s)	1000
Network connection	USB2
Storage	Micro-SD
Operating System	Raspbian stretch lite

Figure 1 shows that the Raspberry pi machine works as a controller node for a single node. While the second part, the Raspberry pi machine works as a node with seven nodes. The HPL package stores proper testing and gflops timing period to quantify the obtained solution's precision and the period it used to calculate it. Next, we test the execution of those two parts with the High-Performance Linpack program, HPLinpack 2.3, shown in Figure 1.

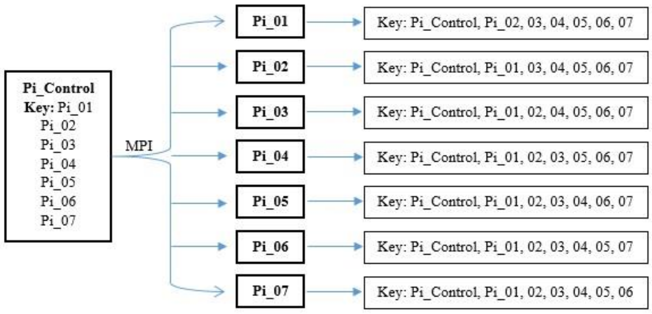


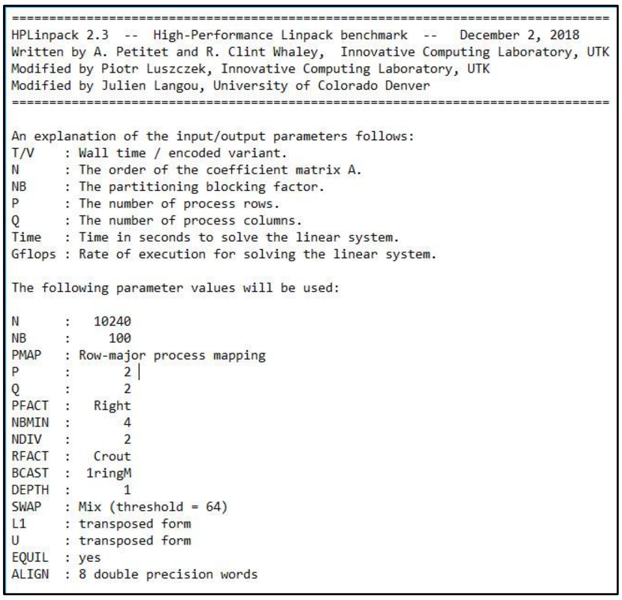
Fig. 1 - Network model for research operations

From Figure 1, each node will work to create SSH-Keys distributed on each node in place of users and passwords, enabling communication with MPI without the need to check permissions on each node.

3.2 Configure Parameters

The following parameter values of HPLinpack 2.3 were used and explain the parameters of input/output as follows:

T/V: Wall time / encoded variant.
N: The order of the coefficient matrix A.
NB: The partitioning blocking factor.
P: The number of process rows.
Q: The number of process columns.
Time: Time in seconds to solve the linear systems.



```
================================================================================
HPLinpack 2.3  --  High-Performance Linpack benchmark  --   December 2, 2018
Written by A. Petitet and R. Clint Whaley,  Innovative Computing Laboratory, UTK
Modified by Piotr Luszczek, Innovative Computing Laboratory, UTK
Modified by Julien Langou, University of Colorado Denver
================================================================================

An explanation of the input/output parameters follows:
T/V    : Wall time / encoded variant.
N      : The order of the coefficient matrix A.
NB     : The partitioning blocking factor.
P      : The number of process rows.
Q      : The number of process columns.
Time   : Time in seconds to solve the linear system.
Gflops : Rate of execution for solving the linear system.

The following parameter values will be used:

N      :   10240
NB     :     100
PMAP   : Row-major process mapping
P      :       2
Q      :       2
PFACT  :   Right
NBMIN  :       4
NDIV   :       2
RFACT  :   Crout
BCAST  :  1ringM
DEPTH  :       1
SWAP   : Mix (threshold = 64)
L1     : transposed form
U      : transposed form
EQUIL  : yes
ALIGN  : 8 double precision words
```

Gflops: Rate of execution for solving the linear system.

Fig. 2 - Network model for research operations

The parameter values used in the test is set as follow:

N: 10240
NB: 100
PMAP: Row-major process mapping
P: 2
Q: 2
PFACT: Right

NBMIN: 4
NDIV: 2
RFACT : Crout
BCAST : 1ringM
DEPTH: 1
SWAP: Mix (threshold =64)
L1: transposed form
U: transposed form
EQUIL: yes
ALIGN: 8 double precision words

For testing in each node process, this research used eight nodes, different processes, and different N. Some configurations showed in Table 2 below.

Table 2 - Some configurations on each node

Number of Nodes	Process	N	NB	P	Q
1	4	10240	100	2	2
2	8	14336	100	2	4
4	16	20480	100	4	4
6	24	25200	100	4	6
8	32	29184	100	4	8

4. Results and Discussion

This research paper will explain the experimental results and mathematical modeling results.

4.1 Experimental Results

After configuring parameter values of HPLinpack 2.3, the experimental results will collect from Gflops value in three testing times in each node, as shown in Figure 3 as follow.

```
--------------------------------------------------------------------------------

- The matrix A is randomly generated for each test.
- The following scaled residual check will be computed:
      ||Ax-b||_oo / ( eps * ( || x ||_oo * || A ||_oo + || b ||_oo ) * N )
- The relative machine precision (eps) is taken to be           1.110223e-16
- Computational tests pass if scaled residuals are less than            16.0

================================================================================
T/V                N    NB     P     Q               Time                 Gflops
--------------------------------------------------------------------------------
WR11C2R4       10240   100     2     2             695.93             1.0288e+00
HPL_pdgesv() start time Mon Aug  5 11:30:08 2019

HPL_pdgesv() end time   Mon Aug  5 11:41:44 2019

--------------------------------------------------------------------------------
||Ax-b||_oo/(eps*(||A||_oo*||x||_oo+||b||_oo)*N)=   1.29167729e-03 ...... PASSED
================================================================================

Finished      1 tests with the following results:
              1 tests completed and passed residual checks,
              0 tests completed and failed residual checks,
              0 tests skipped because of illegal input values.
--------------------------------------------------------------------------------

End of Tests.
================================================================================
```

Fig. 3 - Results of each node per test

The performance testing with the High-Performance Linpack program, HPLinpack 2.3, on the controller node, is grouped in tests by the number of nodes 1, 2, 4, 6, and 8 nodes with the same number NB (The partitioning blocking factor). Every node has the size of the equation given in the solution (matrix of the variable size) equal to 100*100, as shown in Table 3 as follows.

Table 2 - Floating-point Operation Per Second (FLOPS) value

Number of Nodes	Process	Gflops No. 1	2	3	Average
1	4	1.0260	1.0288	1.0302	1.0283
2	8	1.9686	1.9817	1.9873	1.9792
4	16	3.4237	3.4381	3.4353	3.4324
6	24	4.9422	4.6690	4.6043	4.7385
8	32	5.8639	5.5697	5.7378	5.7238

The data obtained from the number of nodes is averaged and compared. Table 3 in the average column shows that the number of nodes equal to 1 the average Gflops is 1.0283, 2 nodes are 1.9792, 4 nodes are 3.4324, 6 nodes are 4.7385, and 8 nodes is 5.7238, respectively. The efficiency of the ability to process instruction set in floating-point operation per second increases the performance continuously when increasing the number of the node on the cluster.

4.2 Mathematical Results

The experimental data is projected on a graphical scale for visualization. It can be seen that Floating-point Operation per second with the number of nodes inclined to correlate in a power function curve. So, this paper chose the least-squares method of estimation using transform least squares. The exponential equation $y = ax^k$ is adjusted to linear proportions by taking the logarithmic function ln on both sides of the equation. $ln\, y = ln(ax^k)$ Using logarithmic function adjust the equation to $ln\, y = k\, ln\, x + ln\, a$ And from the linear equation. $Y = AX + B$ Then we get $Y = ln\, y,\ A = k,\ X = ln\, x,$ and $B = ln\, a$. The results from Table 3 are calculated using the least-squares method. k = 0.8256 and a = 1.0684 substituting values in a mathematical model show the relationship between values Floating-point operation per second with number of nodes transforming to $f(x) = 1.0684x^{0.8256}$.

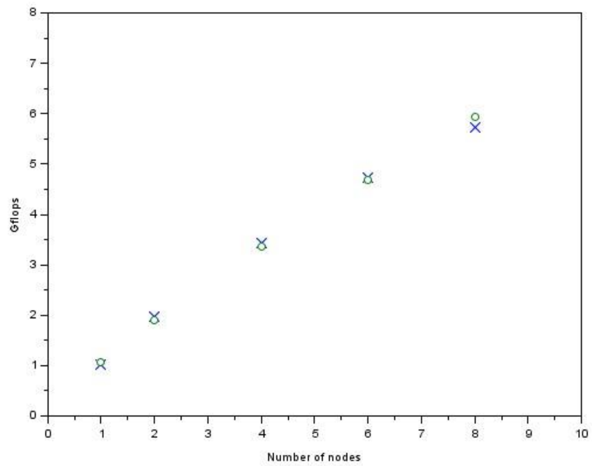


Fig. 4 - The relationship between number nodes and Gflops values

Figure 4 shows x = test values, o = values obtained from mathematical models.

From the graph showing the relationship above, it can be seen that the value obtained from the test is close to the value obtained from the mathematical model. It can be used to estimate the number of nodes used to design high-performing computer systems on a single board cluster with efficiency and cost-effectiveness or resource management.

5. Conclusion

In the performance testing, the value Floating-point Operation Per Second tends to increase markedly via growing the number of nodes during the initial phase, and the incline slowly subsided. It shows that increasing or decreasing the number of nodes on a single board cluster will affect the parallel computing efficiency, which is related and consistent with the mathematical model obtained. $f(x) = 1.0684x^{0.8256}$. However, HPL is starting to have features that are no longer similar to today's applications and should be far away. HPL focuses on measuring the power of pure calculations. In another way, the supercomputers that are good at computing have many other factors, such as the efficiency of network traffic speed in accessing memory (Main memory) and in other areas of the machine and network architecture. For this reason, HPL is no longer considered an accurate measurement procedure. Because a machine that has many FLOPS from HPL may be able to process many new programs not as good as the results show, therefore, the future work, this research may consider changing the test of software from HPL to HPCG testing on single board of Raspberry Pi.

Acknowledgment

The Information Technology Laboratory supports this paper at the Faculty of Science and Technology, Songkhla Rajabhat University.

References

[1] W. Anwaar and M. A. Shah, 'Energy efficient computing: a comparison of Raspberry PI with modern devices', Energy, vol. 4, no. 2, pp. 410–413, 2015
[2] S. J. Johnston et al., 'Commodity single board computer clusters and their applications', Future Generation Computer Systems, vol. 89, pp. 201–212, Dec. 2018
[3] D. Marković, D. Vujičić, D. Mitrović, and S. Ranđić, 'Image Processing on Raspberry Pi Cluster', International Journal of Electrical Engineering and Computing, vol. 2, no. 2, pp. 83–90, Dec. 2018
[4] Y. Deng, P. Zhang, C. Marques, R. Powell, and L. Zhang, 'Analysis of Linpack and power efficiencies of the world's TOP500 supercomputers', Parallel Computing, vol. 39, no. 6, pp. 271–279, Jun. 2013
[5] P. J. Basford et al., 'Performance analysis of single board computer clusters', Future Generation Computer Systems, vol. 102, pp. 278–291, Jan. 2020
[6] A. Komninos, I. Simou, N. Gkorgkolis, and J. D. Garofalakis, 'Performance of Raspberry Pi microclusters for Edge Machine Learning in Tourism', in AmI, 2019
[7] E. Ağbahca and A. A. Altun, 'Performance Test of MPI on Raspberry Pi 2 Beowulf Cluster', International Journa of Computing, vol. 5, no. 4, pp. 233–238, 2016
[8] W. Gropp, E. Lusk, N. Doss, and A. Skjellum, 'A high-performance, portable implementation of the MPI message passing interface standard', Parallel Computing, vol. 22, no. 6, pp. 789–828, Sep. 1996 S. Tongkaw et al., Int. J. of Integrated Engineering Vol. 12 No. 7 (2020) p. 181-187
[9] J. Klinkenberg, P. Samfass, M. Bader, C. Terboven, and M. S. Müller, 'CHAMELEON: Reactive Load Balancing for Hybrid MPI+OpenMP Task-Parallel Applications', Journal of Parallel and Distributed Computing, vol. 138, pp. 55–64, Apr. 2020
[10] S. Ray and A. Al Dhaheri, 'Using Single Board Computers in University Education: A Case Study', in Recent Advances in Information Systems and Technologies, Cham, 2017, pp. 371–377
[11] T. Klinger and C. Madritsch, 'Parallel computing for education using single-board computers', in 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), May 2018, pp. 0616–0619
[12] D. V. Diwedi and S. J. Sharma, 'Development of a Low Cost Cluster Computer Using Raspberry Pi', in 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), Nov. 2018, pp. 11–15