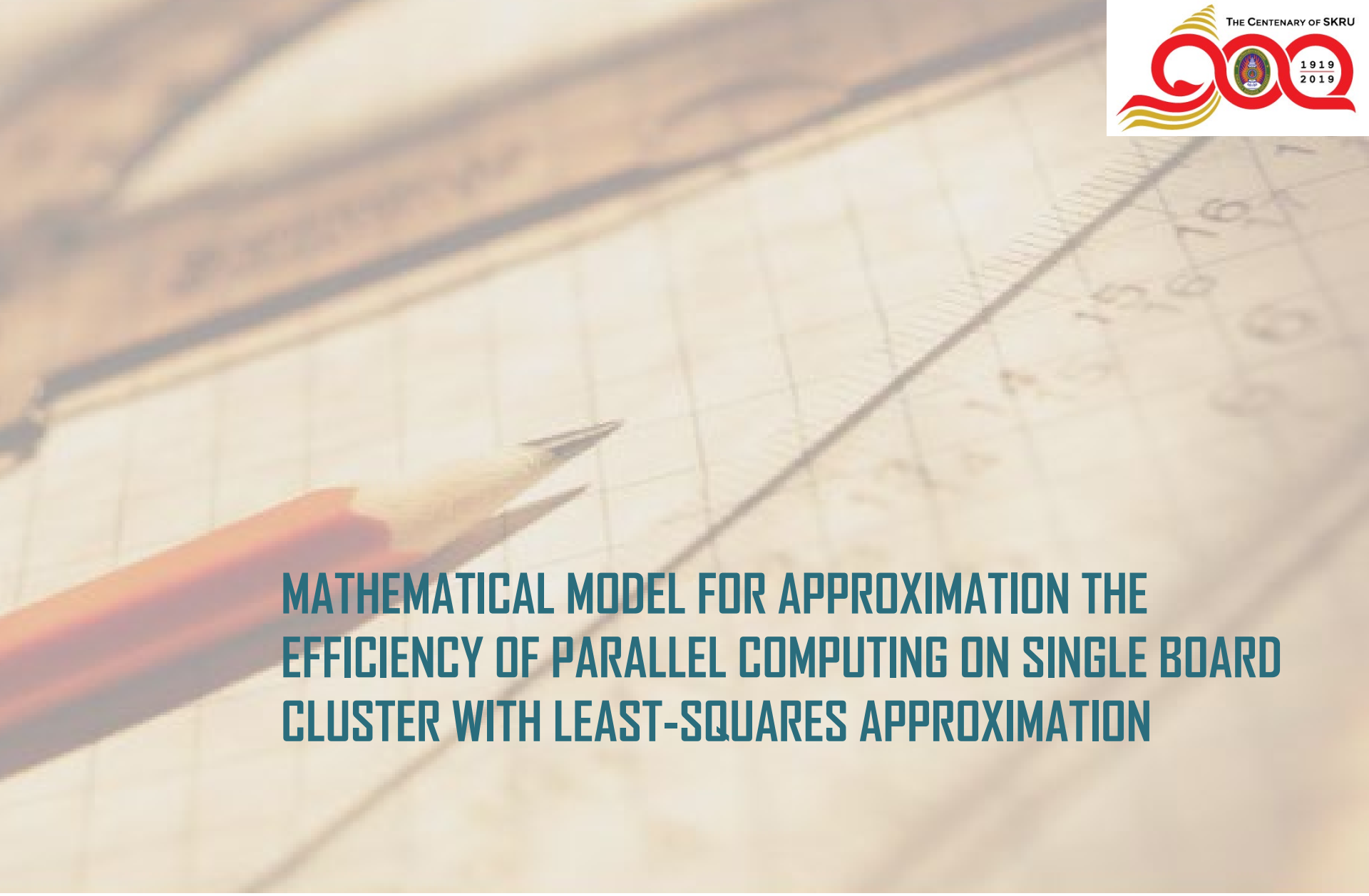

MATHEMATICAL MODEL FOR APPROXIMATION THE EFFICIENCY OF PARALLEL COMPUTING ON SINGLE BOARD CLUSTER WITH LEAST-SQUARES APPROXIMATION

Chotitham Thanrak and Sasalak Tongkaw

Agenda

1. Introduction
2. Objectives
3. Implementation
4. Results and Discussion
5. Conclusion

Introduction

- Computer systems in order to process or deal with big data and there is a tendency for demand to increase rapidly.
- 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.

Objectives

- This research is 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.

Implementation

- This research tested the high-performance of the single board on the Raspberry Pi in a parallel computing system.
- The testing process divides the tasks in each particular part and sent it to each unit to process simultaneously via the MPI (Messaging Passing Interface).

Implementation

- Testing Operation

  Simulated network and a single board, the Raspberry pi 3 Model B+ , which consists of 2 parts, Frist Raspberry pi machine works as a controller node for one node and second Raspberry pi machine works as a node with seven nodes.

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

Implementation

- Testing Operation

  Each node will work to create SSH-Keys distributed on each node in place of users and passwords, enabling communication with MPI

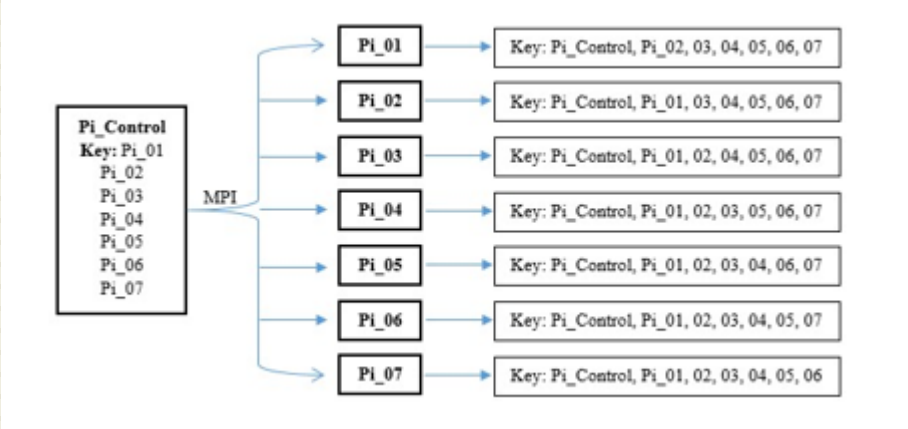

Fig. 1 - Network model for research operations

Implementation

- Configure Parameters

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

  Gflops: Rate of execution for solving the linear system

Implementation

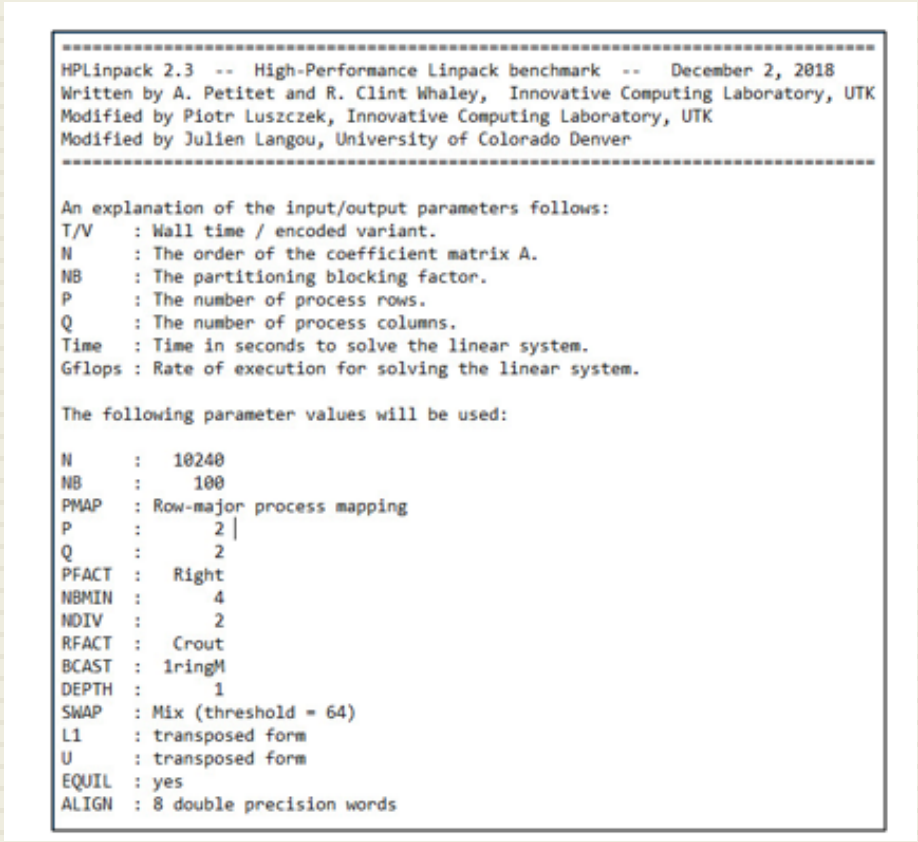

Fig. 2 - Network model for research operations

Implementation

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

Results and Discussion

- The experimental results will collect from Gflops value in three testing times in each node

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

- 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

Results and Discussion

- testing with the High-Performance Linpack program, HPLinpack 2.3, on the controller node, grouped in tests by the number of nodes 1, 2, 4, 6, and 8 nodes

Table 3 - 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

Results and Discussion

- Chose the least-squares method of estimation using transform least squares. The exponential equation

$$y = ax^k$$

- And adjusted to linear proportions by taking the logarithmic function ln on both sides of the equation

$$\ln y = k \ln x + \ln a$$

- Calculated using the least-squares method k = 0.8256 and a = 1.0684
- Substituting values in a mathematical model

$$f(x) = 1.0684x^{0.8256}$$

Results and Discussion

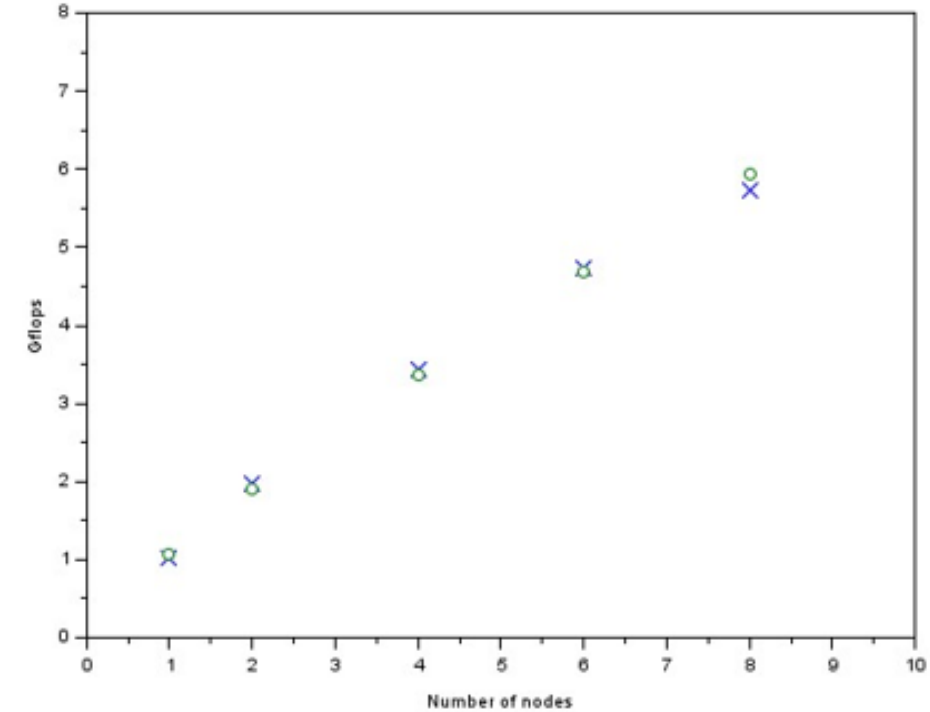

Fig. 4 - The relationship between number nodes and Gflops values shows x = test values, o = values obtained from mathematical models.

Conclusion

- The value Floating-point Operation Per Second tends to increase markedly 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.

$$f(x) = 1.0684x^{0.8256}$$

Q & A