A Performance Study and Linear Optimization Modeling of Collective Communication Algorithms

Mamadou, Hyacinthe Nzigou
Graduate school of Information Science and Electrical Engineering, Kyushu University

出版情報：SLRC プレゼンテーション, 2007-07-25. 九州大学システムLSI研究センター
バージョン：
権利関係：

A Performance Study and Linear Optimization Modeling of Collective Communication Algorithms

九州大学システム情報科学府
情報理学専攻 博士課程2年　村上研究室

ヤセント　ジグ　ママドゥ
Hyacinthe NZIGOU MAMADOU

2007 年7月25日

NGArch Forum 2007

Presentation Flow

- Motivation and Objective
- Background and Current System Limitation
- Performance Modeling Technique
  - Point-to-point performance model (P-LogP)
  - Alltoall Algorithm (Ring).
- Result Evaluations
- Extended Model Optimization
  - Linear regression method
- Conclusions and Future work

Motivation and Objective

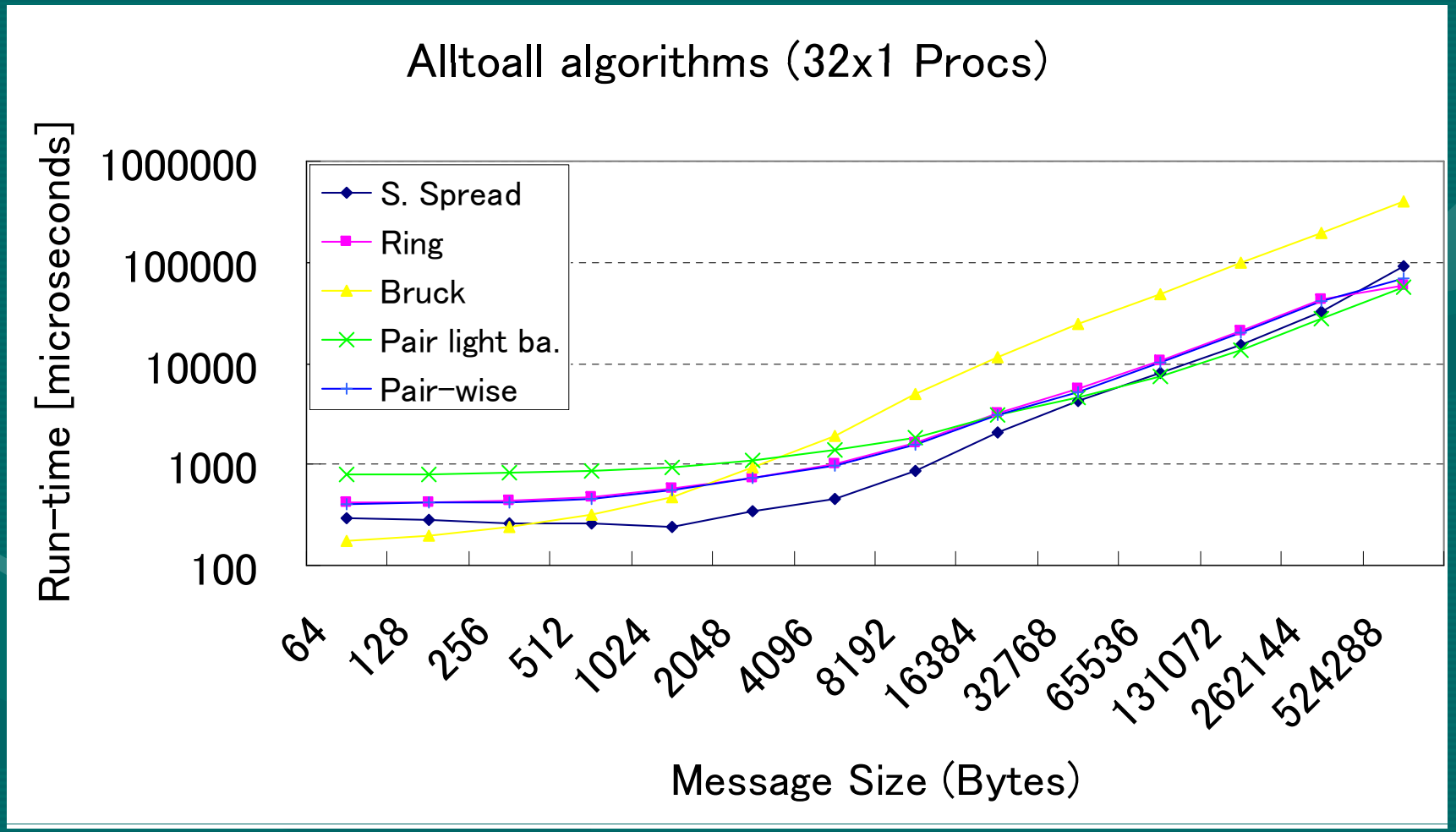
- Reduce the overall execution time of MPI application programs.
  - By reducing the run-time of collective communication operations.
- Achieve high performance of such application software.

Current System Limitation

- MPICH, Fujitsu MPI etc. software adaptability is based on
  - Message size
  - Number of processors involved in the communication

- Difficult to achieve high performance of such MPI implementations on different architectures.

Background : Performance of Different Algorithms



Final Goal

User application program

```
for (i=0; i<100; i++)
Alltoall (msg_sz )
128 B
Alltoall (msg_sz )
512 KB
Allreduce (msg_sz )
…
```

MY_Alltoall

1	Bruck
2	Simple spread
3	Pair-wise
4	Ring
…	…

1	Pair light barrier
2	Ring
3	Pair-wise
…	…

Develop a mechanism that can select the best performing algorithm based on performance prediction models.

Problem

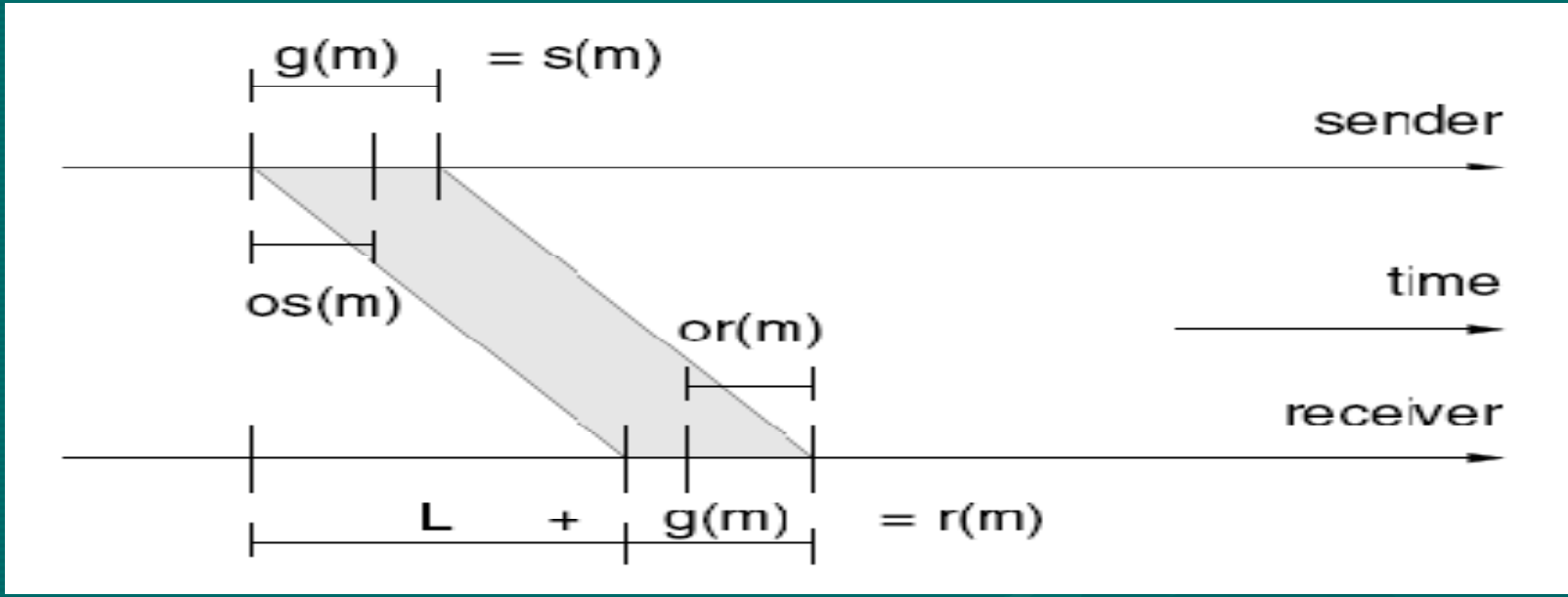
- No standard model exist to predict the performance of collective communication operations.
- How to develop collective communication routines that can achieve high performance computing ?

Proposal Solution

- Make efficient performance models for the prediction of collective communication algorithms.

- Use these models to select the best performing algorithm for a given situation
  - Network topology (Number of processors, Latency, Bandwidth etc.)
  - Message size
  - Load imbalance …

Point-to-Point Performance Model

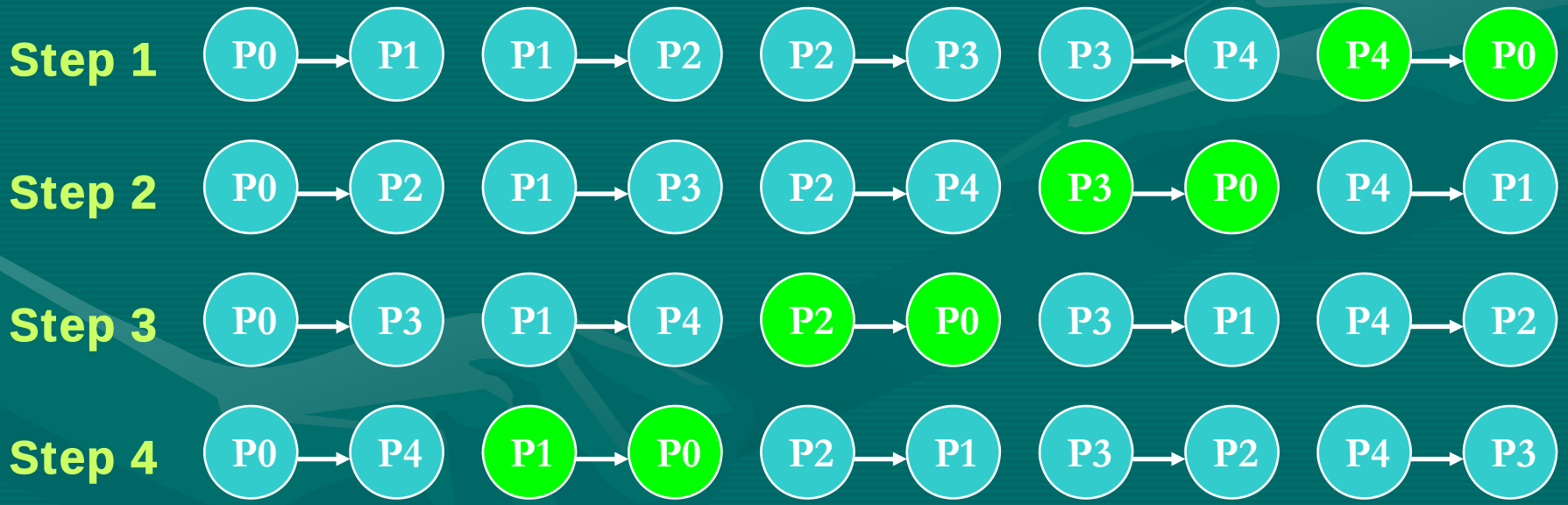


L : Communication latency
Os(m) : Overhead send of a message size m
Or(m) : Overhead receive of a message size m
g(m) : Time a message size m occupies a link.
P : Number of Processors

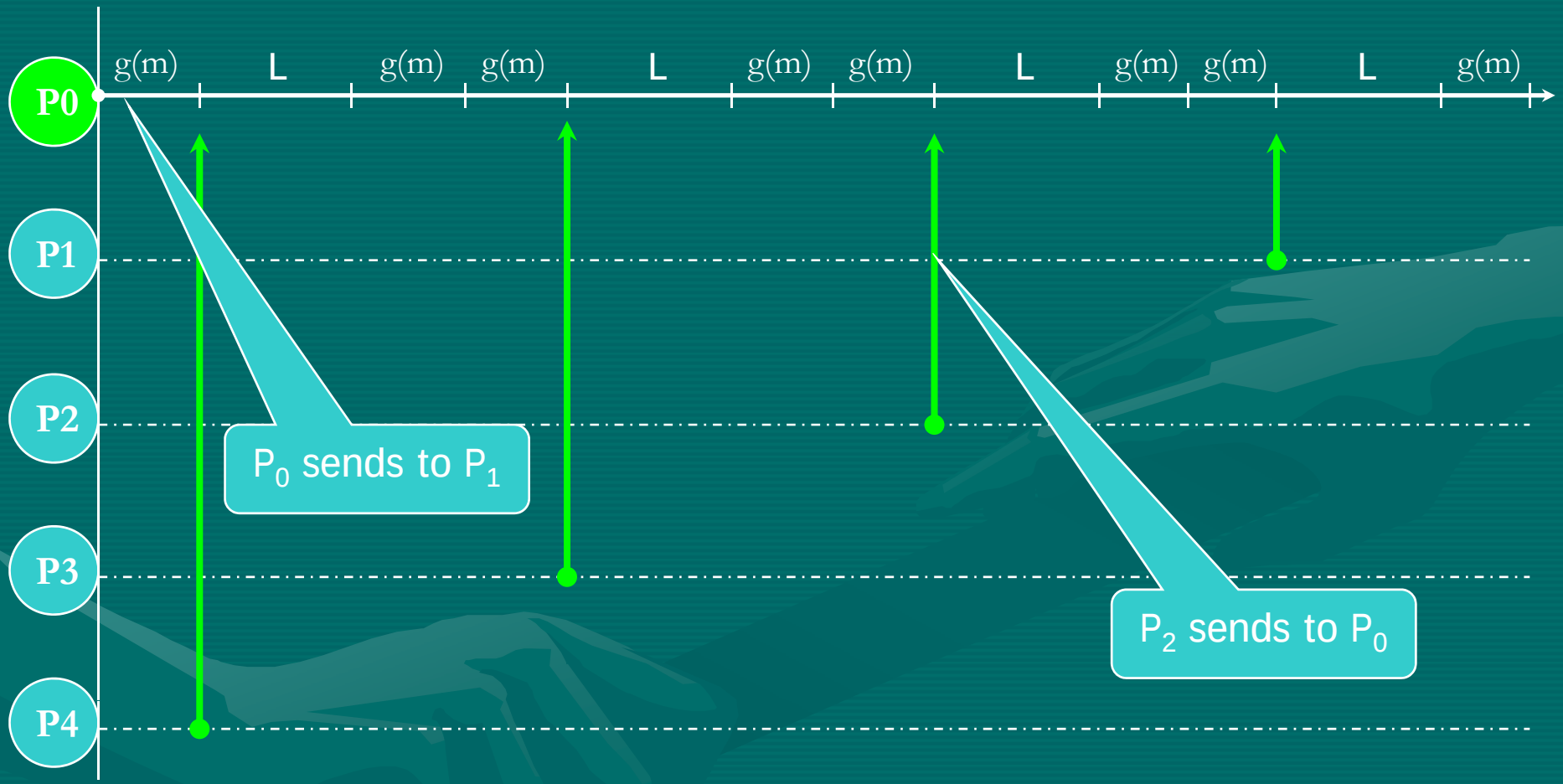
P-LogP standard model (Thilo Kielmann et al. 2000)

Alltoall Ring Algorithm

- Partition overall communication into (p – 1) steps to achieve MPI Alltoall.
- At step i, **(j – i) mod p → node j → (j + i) mod p.**

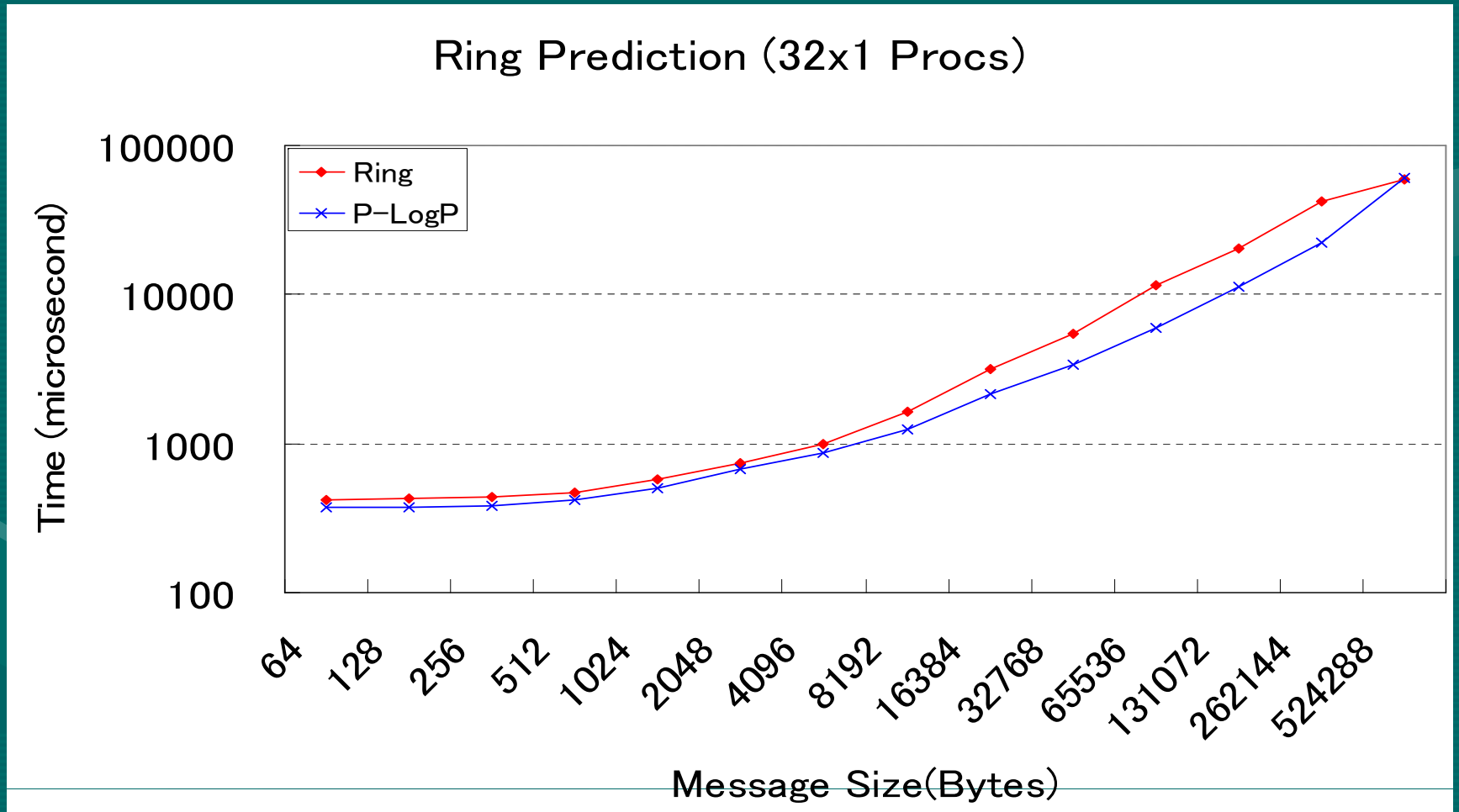


Performance Modeling (Ring)



Total time = 4 L + 8 g(m) = 4 (L + 2 g(m)) ⇒ (P-1) (L + 2 g(m)).

Ring Experimental Results



Linear Regression Modeling

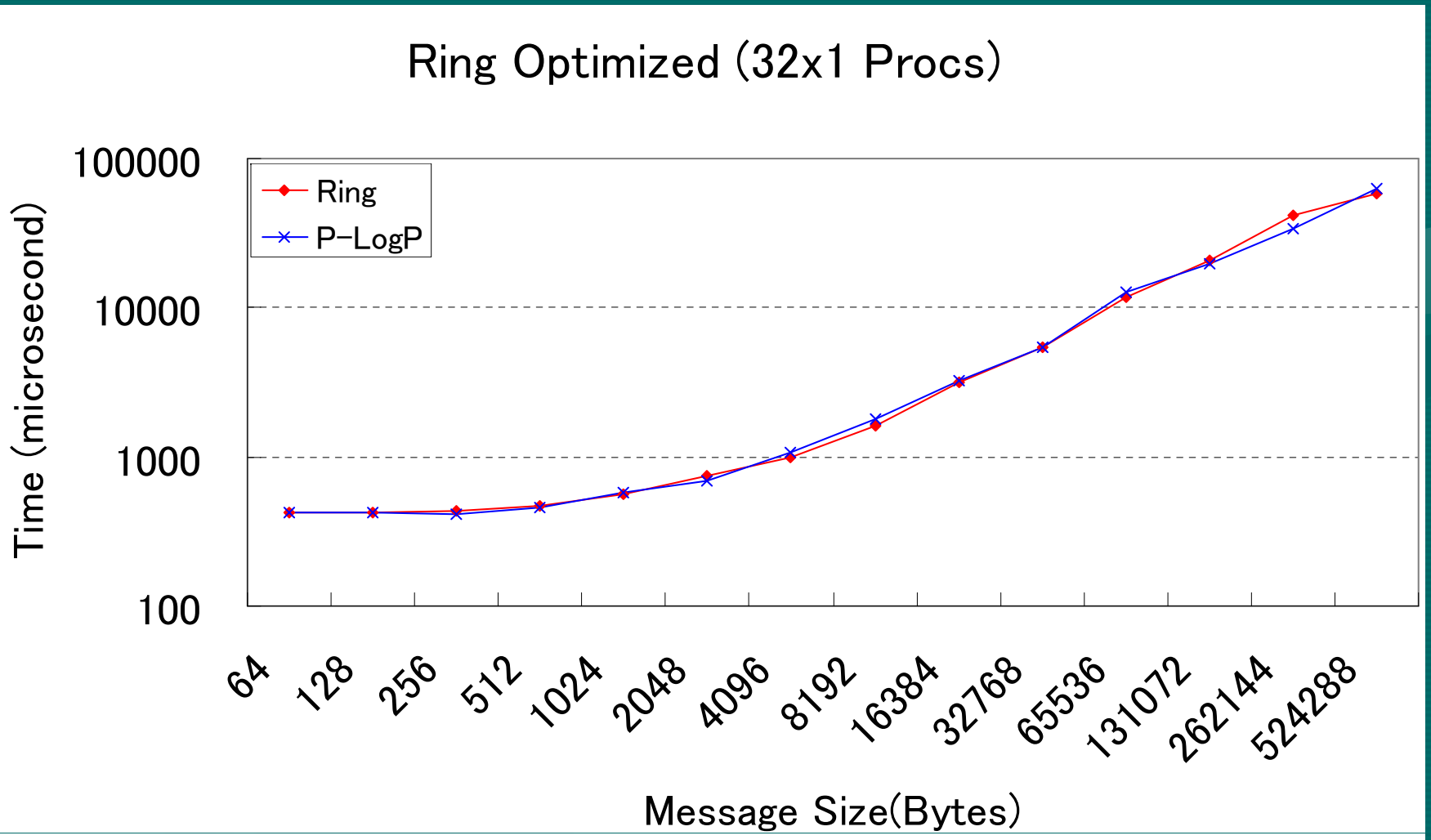
- Linear modeling
  - Process to determine the linear equation that is the best fit to a set of data points in terms of minimizing the sum of the squared distances between the line and the data points $(x_1, y_1), \ldots, (x_k, y_k)$.
  - **Y = A * x + B**.

$$A = \frac{\sum_{k=1}^{n} x_k \bullet \sum_{k=1}^{n} y_k - n \bullet \sum_{k=1}^{n} x_k y_k}{(\sum_{k=1}^{n} x_k)^2 - n \bullet \sum_{k=1}^{n} x_k^2} \qquad B = \frac{\sum_{k=1}^{n} x_k \bullet \sum_{k=1}^{n} x_k y_k - \sum_{k=1}^{n} x_k^2 \bullet \sum_{k=1}^{n} y_k}{(\sum_{k=1}^{n} x_k)^2 - n \bullet \sum_{k=1}^{n} x_k^2}$$

  - The value of **A** and **B** determined in this way represent the straight line with the minimum sum of squared distances.

Thus, the optimized model of ***Ring*** = **(P-1)(L + 2g(m)) + (A*m + B)**

Ring Experimental Results



Conclusions and Future Work

- Using this linear modeling technique, we try to handle the entire behavior of the system communication
  - Simple model with high accuracy
- Very Good results for all size messages, can achieve around 5% of the relative gap in most cases.
- Other collective communication performance analysis
  - Broadcast, Allreduce, Allgather …
- Theoretical performance improvement study.
- Dynamic implementation

ご静聴ありがとうございました。

Any Questions ?