# Determining the MPP LS-DYNA Communication and Computation Costs with the 3-Vehicle Collision Model and the Infiniband Interconnect

Yih-Yih Lin
*Hewlett-Packard Company*
*MR01-3*
*200 Forest Street*
*Marlborough, MA 01752*
*yih-yih.lin@hp.com*

## Abstract

*The least square error approach applied to LS-DYNA communication and computation costs for the Neon model was shown previously by this author to be useful in predicting performance on a given interconnect of known ping-pong latency and bandwidth, and both Gigabit Ethernet and HyperFabric 2 results were presented. In this paper, this prediction method is applied to a much larger public-domain crash model, the 3-vehicle collision model, to determine communication and computation costs for models representative of the most demanding requirements. Furthermore, the result is verified against the new, high-speed low- latency Infiniband interconnect. Users of this method may perform trade-off analysis for optimum hardware configuration decisions without the need for extensive benchmark testing.*

## Introduction

In a previous paper [1], the author presented the least square error approach to determine the MPP LS-DYNA communication and computation costs for the Neon Model. As described in that paper, the elapsed time of an MPP LS-DYNA job is comprised of two parts, the computation and the communication costs:

$$T_{elapsed} = T_{comput} + T_{comm} \quad (1)$$

The communication cost can be further approximated by the following formula:

$$T_{comm} = M(\alpha t^{lan} + \beta t^{bw} s / t^{bw}) \quad (2)$$

where M is the average number of messages per processors, s is the average message size, $t^{lan}$ and $t^{bw}$ are the ping-pong latency and bandwidth, $\alpha$ is the latency constant (so called because $\alpha t^{lan}$ represents the aggregate latency), and $\beta$ is the bandwidth constant (so called because $t^{bw}/\beta$ represents the aggregate bandwidth). For a given MPP LS-DYNA job, all those quantities, except the latency and bandwidth constants $\alpha$ and $\beta$, can be measured and assumed known. To determine the quantities $\alpha$ and $\beta$, the least square error approach is proposed. In the approach, two clusters, with interconnects *a* and *b*, are used, and a set of elapsed times for the two clusters, $T^a_{elapsed}$ and $T^b_{elapsed}$, are measured with various numbers of processors. Then, from formulas (1) and (2), the quantities $\alpha$ and $\beta$ can be determined as the quantities that minimize the following sum of squares of errors:

$$S = \sum_i [(M_{n,i}(t^{lan}_a - t^{lan}_b)\alpha + M_{n,i}s(1/t^{bw}_a - 1/t^{bw}_b)\beta) - (T^a_{elapsed,i} - T^b_{elapsed,i})]^2 \quad (3)$$

where i ranges with the varying numbers of processors. Once the two quantities are determined, one can then apply formulas (1) and (2) to predict the elapsed time, or the performance, of MPP LS-DYNA for a given model with an interconnect with known ping-pong latency and bandwidth.

In this paper, the same least square error approach is applied to a much larger public domain model, the 3-vehicle collision model, to determine its communication and computation costs and, thus, to provide a quantitative relationship between the performance of this currently representative model of the most demanding requirements and an interconnect. Furthermore, the quantitative relationship is verified against the new, high-speed low-latency Infiniband interconnect.

## Model, Machine, Interconnects, Measured Data

**Model, Machine, and OS**

In this paper, the 3-vehicle collision model from NCAC, whose website is http://www.ncac.gwu.edu, of 971 thousand elements and with simulation time of 150 milliseconds, is used; furthermore, decompositions used are the same one as described in the website http://www.topcrunch.org. The single-precision 970.3858 version of MPP LS-DYNA is used. A 32-processor cluster, consisted of 16 machines of HP's 1.5 GHz Rx2600 with HP-UX 11.23, is used. The Rx2600 is a 2-CPU Itanium2 machine.

**Interconnects and Their Characteristics**

Two interconnects are used to determine the communication and computation costs: the Gigabit Ethernet (GigE) and HP's HyperFabric 2 (HF2); their ping-pong latencies and bandwidths have been measured and are shown in Table 1. Once the communication and computation costs are determined, the performance with a different interconnect can be predicted. The new, high-speed low-latency Infiniband interconnect, whose ping-pong latency and bandwidth has been measured and is also shown in Table 1, is used to verify the accuracy of the resulted prediction and thus validate the approach.

**Elapsed times**

Table 2 and Figure 1 show elapsed times, actually measured, for jobs with numbers of processors 2, 4, 8, 12, 16, 24, and 32; and each with the three interconnects: GigE, HF2, and Infiniband.

**Message Patterns**

The numbers of processors used in the least square error approach are 4, 8, 12, 16, 24, and 32. Table 3 shows the average numbers of messages and average message sizes per processor for the MPP LS-DYNA jobs with this set of numbers of processors.

| | GigE | HF2 | Infiniband |
|---|---|---|---|
| **Latency** | 43 μsec | 22 μsec | 6.5 μsec |
| **Bandwidth** | 112 MB | 216 MB | 780 MB |

Table 1. Ping-pong latencies and bandwidths of Gigabit Ethernet, HF2, and Infiniband

| Number of Processors | 2 | 4 | 8 | 12 | 16 | 24 | 32 |
|---|---|---|---|---|---|---|---|
| Infiniband | 197422 | 100938 | 51250 | 35872 | 26778 | 18210 | 14182 |
| HF | 197422 | 100941 | 51429 | 35943 | 27461 | 19010 | 15211 |
| GigE | 197422 | 101257 | 52921 | 37811 | 28076 | 20795 | 17100 |

Table 2. Measured elapsed times in seconds

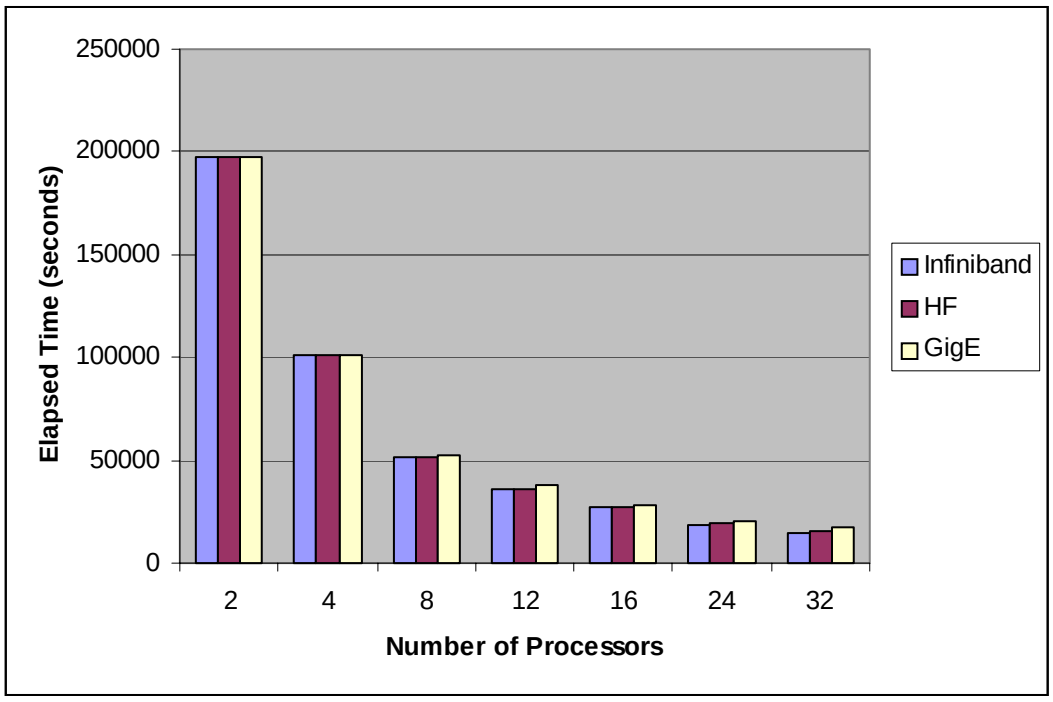


Figure 1. Graph for table 2

| Number of Processors | 4 | 8 | 12 | 16 | 24 | 32 |
|---|---|---|---|---|---|---|
| Ave. No. of Messages | 9924591 | 12787074 | 17521630 | 19141076 | 23910066 | 29620774 |
| Ave. Message Size in Bytes | 3177 | 2862 | 2338 | 2095 | 1830 | 1438 |

Table 3. Average number pf messages per processor and average messages sizes

## Estimation of Communication Costs

### Aggregated Latency and Aggregated Bandwidth

To estimate $\alpha$ and $\beta$, call the cluster with GigE as cluster *a* and the cluster with HF2 as cluster *b*. Then the sum of the squares of the 6 errors, with the numbers of processors being 4, 8, 12, 16, 24, and 32, as in formula (3), can then be obtained with the ping-pong latency and bandwidth in Table 1, the elapsed time data in Table 2, and the message data in Table 3. The sum of squares of these 6 errors is a quadratic function of $\alpha$ and $\beta$. The minimum of the quadratic function occurs when its partial derivatives with respect to $\alpha$ and $\beta$ are equal to zero, which, in turn, forms two

linear equations of the two unknowns $\alpha$ and $\beta$. A computer program based on this approach has been written to obtain

$$\alpha = 2.17 \text{ and } \beta = 2.89$$

This means that, for the 3-vehicle collision model, the aggregated latency of a given interconnect is 2.17 times its ping-pong latency, and its aggregated bandwidth is 0.346, or 1/2.89, times its ping-pong bandwidth.

## Accuracy of the Approach

In the above section, the latency $\alpha$ and the bandwidth constant $\beta$ are determined only with the knowledge of the elapsed times with the GigE and the HF2 interconnects, but not the Infiniband connect. With these two interconnect constants determined, for a given number of processors, we can use formula (2) to estimate the communication cost with the GigE, the HF2, or the Infiniband interconnect. The computation cost, which is independent of the interconnect, can be simply estimated to be the difference between the elapsed time and the communication cost with the GigE or the HF2, as indicated by formula (1). And the estimated, or predicted, elapsed time with the Infiniband interconnect is just the sum of the estimated communication and computations. Shown in Table 4 are comparisons between such predicted and measured elapsed times with the Infiniband interconnect. It is shown that the maximal percent error is just 3 percent.

| Number of Processors | 4 | 8 | 12 | 16 | 24 | 32 |
|---|---|---|---|---|---|---|
| Estimated Elapsed Time | 100039 | 50873 | 35239 | 26053 | 17860 | 13810 |
| Measured Elapsed Time | 100938 | 51250 | 35872 | 26778 | 18210 | 14182 |
| Percent Error | 1% | 1% | 2% | 3% | 2% | 3% |

Table 4. Comparison between predicted and measured elapsed times with the Infiniband interconnect

## Cost Percentages

As mentioned in the previous section, the approach allows one to quantify the two components, communication and computation costs, of the elapsed time and the two components, latency and bandwidth costs, of the communication cost. Table 5 shows the percentages of communication and computation costs in the elapsed times with Infiniband and the GigE interconnects. It is worthwhile to note that the slower GigE decreases its efficiency much faster than the fast Infiniband as the number of processors increases. Table 6 shows the percentages of latency and bandwidth costs in the communication costs with the same two interconnects. It is also worthwhile to note that the percentages of latency and bandwidth costs with the faster Infiniband and the slower GigE are, given a number of processors, almost the same. The conclusion that the Infiniband meets the processor speed of the current 1.5 GHz Itanium2 Rx2600 can be drawn immediately from these results.

| Number of Processors | 4 | 8 | 12 | 16 | 24 | 32 |
|---|---|---|---|---|---|---|
| GigE | | | | | | |
| Communication Cost | 1% | 4% | 8% | 9% | 17% | 23% |
| Computation Cost | 99% | 96% | 92% | 91% | 83% | 77% |
| Infiniband | | | | | | |
| Communication Cost | 0% | 1% | 1% | 2% | 3% | 4% |
| Computation Cost | 100% | 99% | 99% | 98% | 97% | 96% |

Table 5. Percentages of communication and computation costs in the elapsed times with the GigE and the Infiniband interconnects

| Number of Processors | 4 | 8 | 12 | 16 | 24 | 32 |
|---|---|---|---|---|---|---|
| GigE | | | | | | |
| Latency Cost | 53% | 56% | 61% | 63% | 66% | 72% |
| Bandwidth Cost | 47% | 44% | 39% | 37% | 34% | 28% |
| Infiniband | | | | | | |
| Latency Cost | 55% | 57% | 62% | 64% | 68% | 73% |
| Bandwidth Cost | 45% | 43% | 38% | 36% | 32% | 27% |

Table 6. Percentages of latency and bandwidth costs in communication costs with the GigE and the Infiniband interconnects

## Summary

In this paper, a previously proposed least square error approach for determining the MPP LS-DYNA communication and computation costs is applied to the very large public domain crash model of 3 vehicles, which is the currently representative model of the most demanding requirements. The result is verified against the measured elapsed times with the new, high-speed low-latency Infiniband interconnect and is shown to be within 3 percent error. This approach allows the quantitative breakdown of the MPP LS-DYNA simulation cost and thus can be used to perform trade-off analysis for optimum hardware configuration decisions without the need for extensive benchmark testing.

### References

Yih-Yih Lin, “A Quantitative Approach for Determining the Communication and Computation Costs in MPP LS-DYNA Simulations,” FEA News, February 2004.