Running PIC code on a small HPC cluster

Lianxin Xin, Jian Zhao, Zachary Tyler, and Jun Ren

Delaware State University

Abstract

To improve the accuracies of computing real physical systems and extend the dimensions of simulation, large-scale computing systems with advanced infrastructure designs and continuing optimization of code algorithms are required. This presentation concentrates on improving parallel code performance by optimizing synchronous reading and writing I/O operations. An NVMe protocol is applied where Non-Volatile Memory express over Fabrics (NVMe-oF) is combined with the InfiniBand Remote Direct Memory Access (RDMA) technology. Further improvement is attained by replacing the Network file system (NFS) IO with the Lustre IO so that data can be provided/distributed faster. We run the Particle-in-Cell (PIC) code OSIRIS (A three-dimensional, massively parallel, fully relativistic PIC code) on our system. Code speed evaluation and benchmark test showed that the overall system performance is improved by nearly ten times. This newly upgraded system enables us to run 1D prototype simulations of light-plasma interactions with various complex boundary conditions and paves the way for collecting more sounding results from 2D and 3D runs when using larger computing facilities.

Introduction

Hardware in our HPC Cluster:

- Intel E5-2630 v4 X 16
- 64G ECC 2400 X 8
- Mellanox IS5023 40G Switch
- Mellanox ConnectX-3 VPI 40G X 8
- 1T SAS Hard Drive X 8
- NVME M.2 Samsung 980 Pro 1TB X 4

Software in our HPC Cluster :

- CentOS 7.3.1611 (core) Linux Platform
- Intel MPI 2021.03
- Lustre Cluster 2.10 File Storage Server
- NFS Server 4.1
- OSIRIS 4.0
- HDF5 1.8.11

Methods

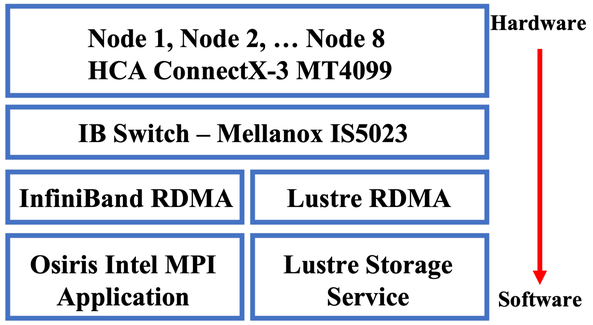

Figure 1. Lustre system architecture.

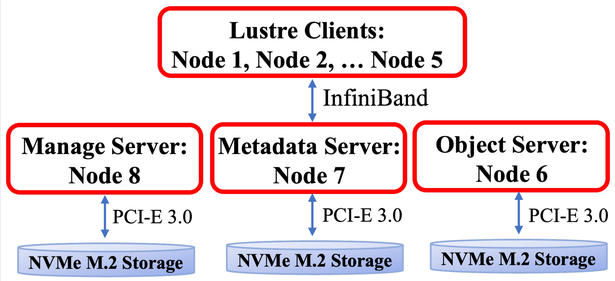

Figure 2. Lustre system architecture with NVMe protocol in our HPC cluster.

Results

Theory FIO Test (bandwidth) for NFS and Lustre Storage Server		
	NFS Storage Server	Lustre Storage Server
Random READ	2816MiB/s (2953MB/s)	4321MiB/s (4531MB/s)
Random WRITE	417MiB/s (437MB/s)	1511MiB/s (1585MB/s)
READ	3281MiB/s (3440MB/s)	4753MiB/s (4983MB/s)
WRITE	418MiB/s (439MB/s)	1752MiB/s (1837MB/s)

Table 1. The theory I/O bandwidth test results for NFS and Lustre storage server through the Flexible I/O tester (FIO).

Theory IOR Test for NFS and Lustre Storage Server			
		NFS Storage Server	Lustre Storage Server
WRITE (MiB)	Max	75.66	487.30
	Min	44.61	333.87
	Mean	67.39	436.34
READ (MiB)	Max	1123.36	1290.62
	Min	930.60	1122.03
	Mean	1047.99	1180.50

Table 2. The theory I/O performance test results for NFS and Lustre storage server through the Interleaved or Random (IOR)

Applications

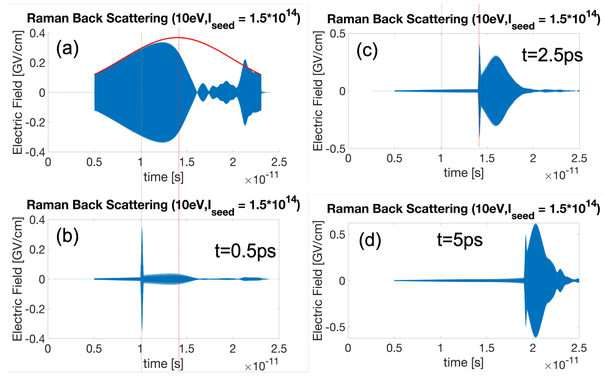

Figure 3. Energy transfer from the pump to the backscatter pulse. Pump: 800nm, 10ps, $1.8\times10^{14}\ W/cm^2$. Seed: 870nm, 125fs, $1.5\times10^{14}\ W/cm^2$. Plasma: 1.5mm, $1.13\times10^{19}\ /cm^3$.

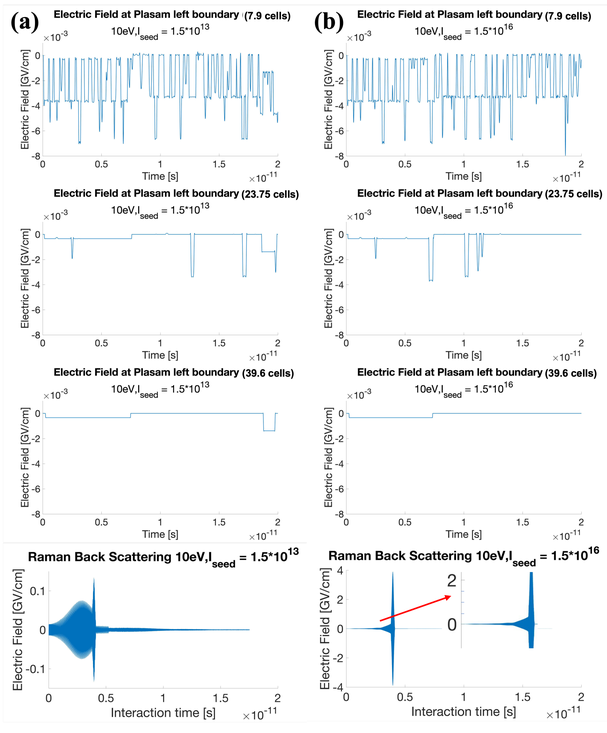

Figure 4. Static E-field accumulated at the plasma left boundary (Seed entrance). Pump: 800nm, 5ps, $1.8\times10^{14}\ W/cm^2$. Plasma: 4mm, $1.13\times10^{19}\ /cm^3$. Seed: 870nm, 125fs, (a) $1.5\times10^{13}\ \frac{W}{cm^2}$, (b) $1.5\times10^{16}\ \frac{W}{cm^2}$.

Condition	Total Time (s)	I/O Time (s)	Remaining Time
2 nodes 20 threads	4753	3134.5 (65.95%)	1618.5 (34.05%)
3 nodes 20 threads	5320	3993.15 (75.06%)	1326.85 (24.94%)
4 nodes 20 threads	4866	3727.42 (76.60%)	1138.58 (23.40%)

Table 3. NFS Simulation Time (s) and Proportion (%) Overview

Condition	Total Time (s)	I/O Time (s)	Remaining Time (s)
1 nodes 20 threads	2710	426 (15.72%)	2284 (84.28%)
2 nodes 20 threads	2698.82	1159.95 (42.98%)	1538.87 (57.02%)
3 nodes 20 threads	3181.82	1823.22 (57.30%)	1358.6 (42.70%)
4 nodes 20 threads	3167.79	2017.37 (63.68%)	1150.42 (36.32%)
5 nodes 20 threads	2800.48	1958.74 (69.94%)	841.74 (30.06%)

Table 4. Lustre Simulation Time (s) and Proportion (%) Overview

Summary

A Non-Volatile Memory express over Fabrics (NVMe-oF) technology with Remote Direct Memory Access (RDMA) is combined with upgrading the Network file system (NFS) IO to the Lustre IO. The system performance is significant improved. This scheme is particularly helpful for small system. With the improvement, we were able to study in detail the physical process of 3-waves interaction in plasma to generate extreme light and prove potency and effectiveness from 2D and 3D runs upon using larger computing facilities.

References

1. Lustre System: https://www.lustre.org/
2. Vieux, G., Cipiccia, S., Grant, D. W., Lemos, N., Grant, P., Ciocarlan, C., ... & Jaroszynski, D. A. (2017). An ultra-high gain and efficient amplifier based on Raman amplification in plasma. *Scientific reports*, *7*(1), 1-10.

Acknowledgement

Funding support from NSF #1833015 and NSF #1901397