Distributed Simulations

Tyler F. Cloutier

Goal

- Focus on optimizing distributed calculations
- Create an easily generalizable system that can be applied to many different models
- Learn how to optimize on an real system

Importance - Medicine

- Complex nature of biological macromolecules
- Difficult to study small and complex biological systems in lab
- Provide a new way to study enzyme activity
- If accurate can provide much more detailed information

Importance - Material

- Predict new material structures and properties
- Quickly assess a range of materials for a particular job (Ex. Batteries)
- Better understand the mechanics of phase transitions

Simulation Model

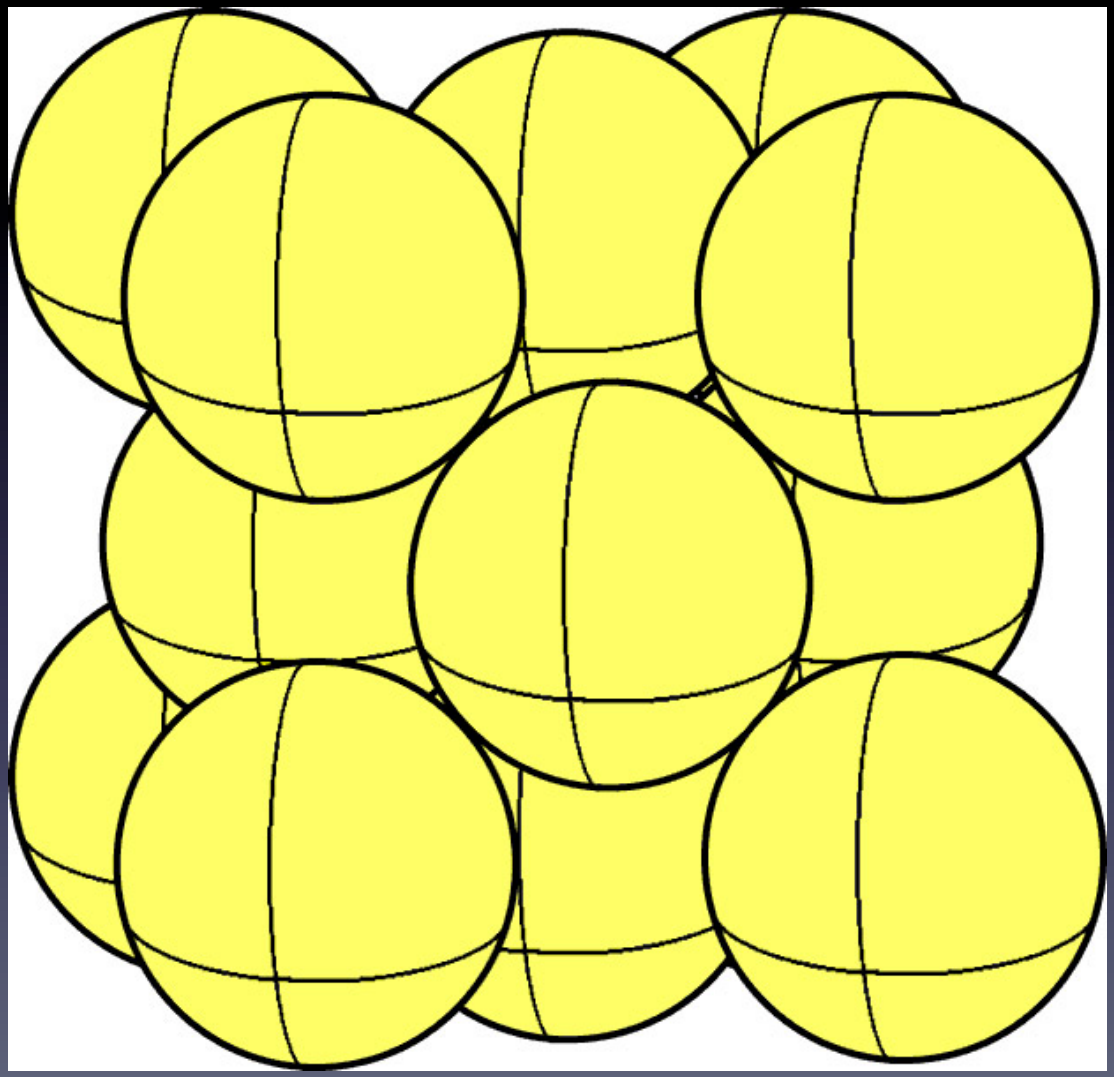
$$E_{ab} = \sum_{i}^{\text{on } a} \sum_{j}^{\text{on } b} \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{\text{OO}}^{12}} - \frac{B}{r_{\text{OO}}^{6}}$$

- Simple Point Charge Model
- Coulomb Potential
- Lennard-Jones Potential
- Other simulation models
  - More complex point charge
  - Monte Carlo and quantum mechanical

Simulation Model

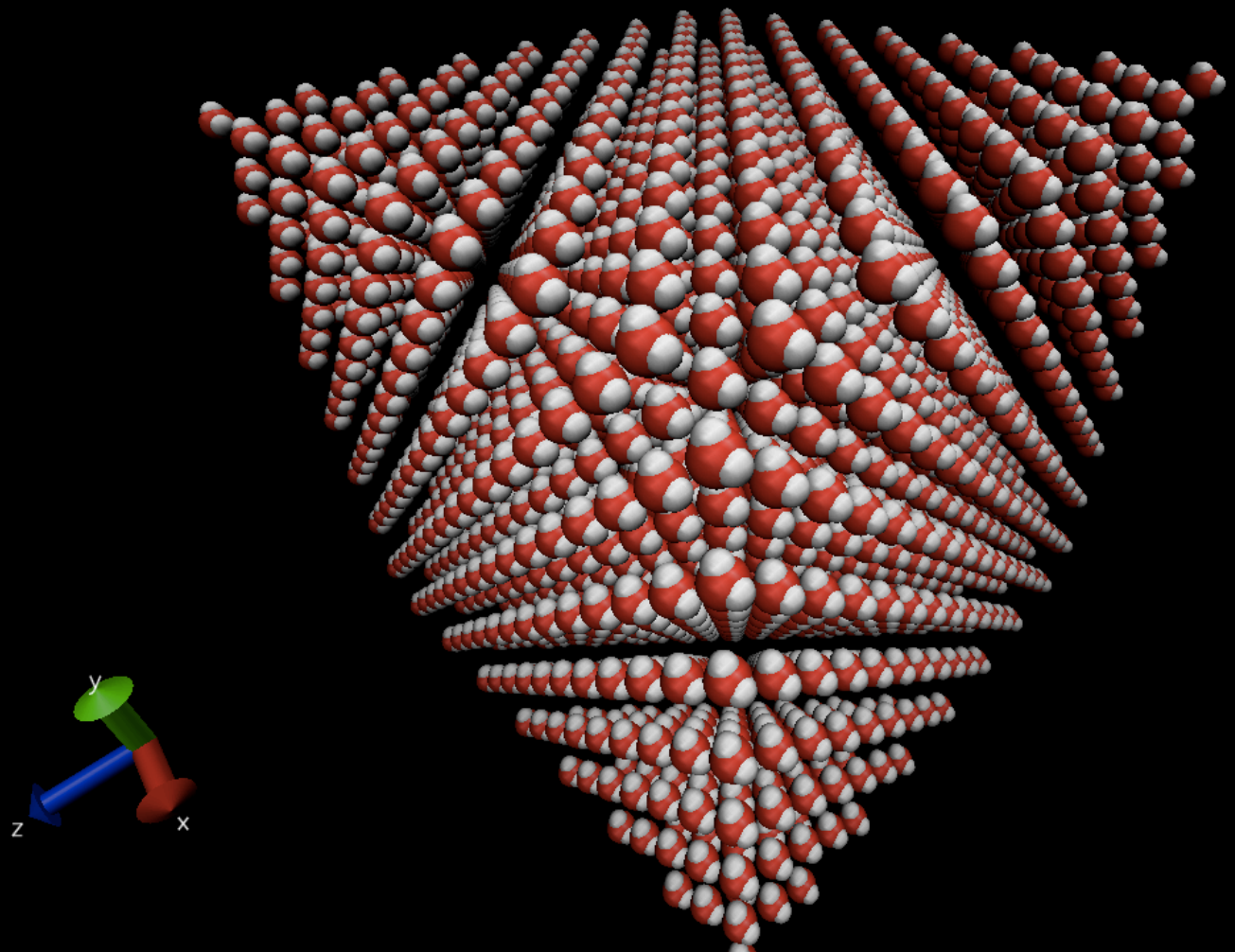
- Periodic Boundary Conditions
- Initial Molecule Placement
  - Face-Centered Cubic Close Packing of Spheres
  - Size of molecules are adjusted to fit desired reduced density

Face Centered Cubic

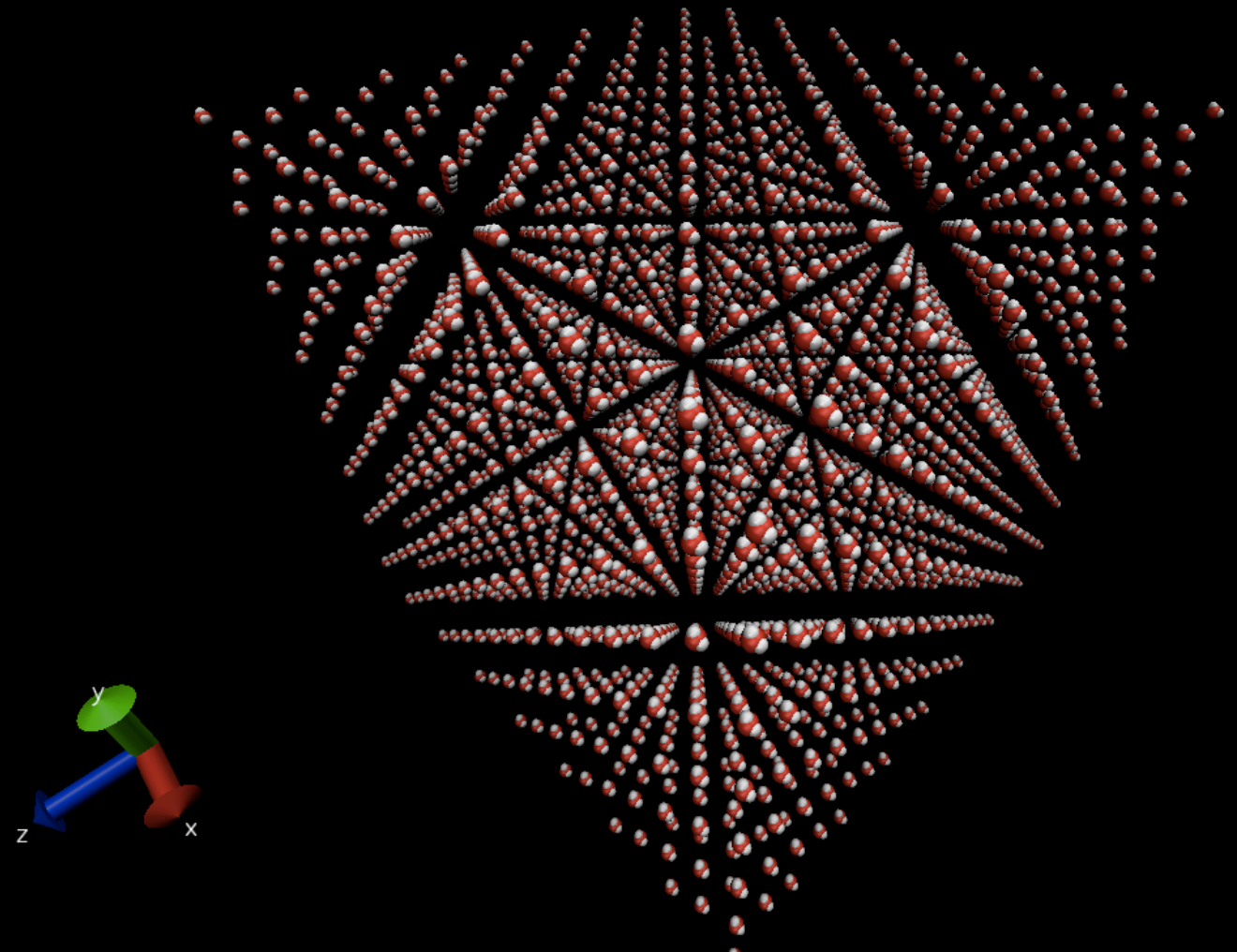


http://education.mrsec.wisc.edu/Edetc/SlideShow/images/unit_cells/face_centered_cubic.jpg

FCC Lattice, Reduced Density - 0.2



FCC Lattice, Reduced Density - 0.05



Tools

- Spread 4.2
- Visual Molecular Dynamics
- C programming language
- Everything else from scratch

Design

- How do we pass messages?
- What data do we send?
- Why did we choose this?
- What is the computation payload?
- What is the scale of computation vs messaging?

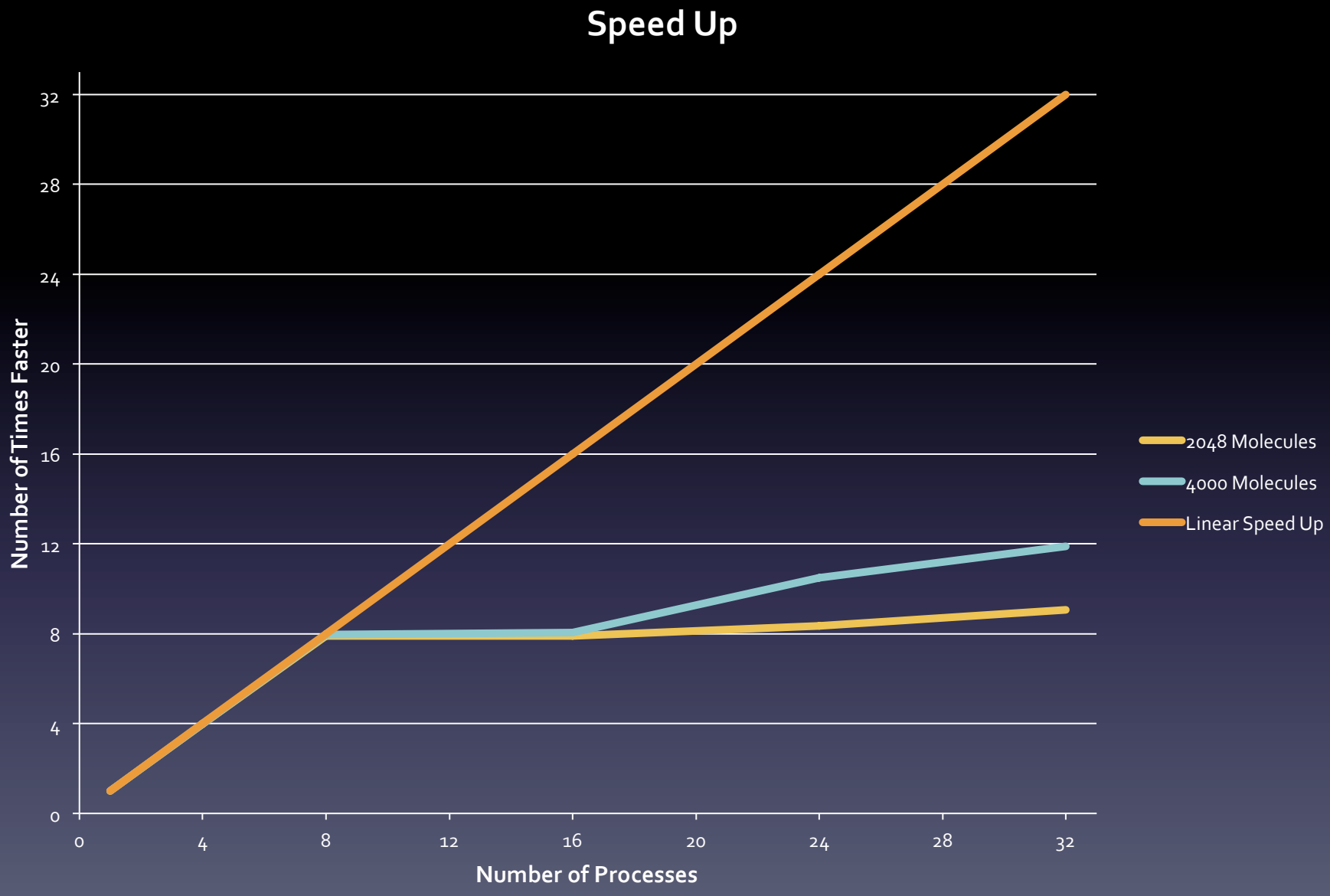
Program Structure

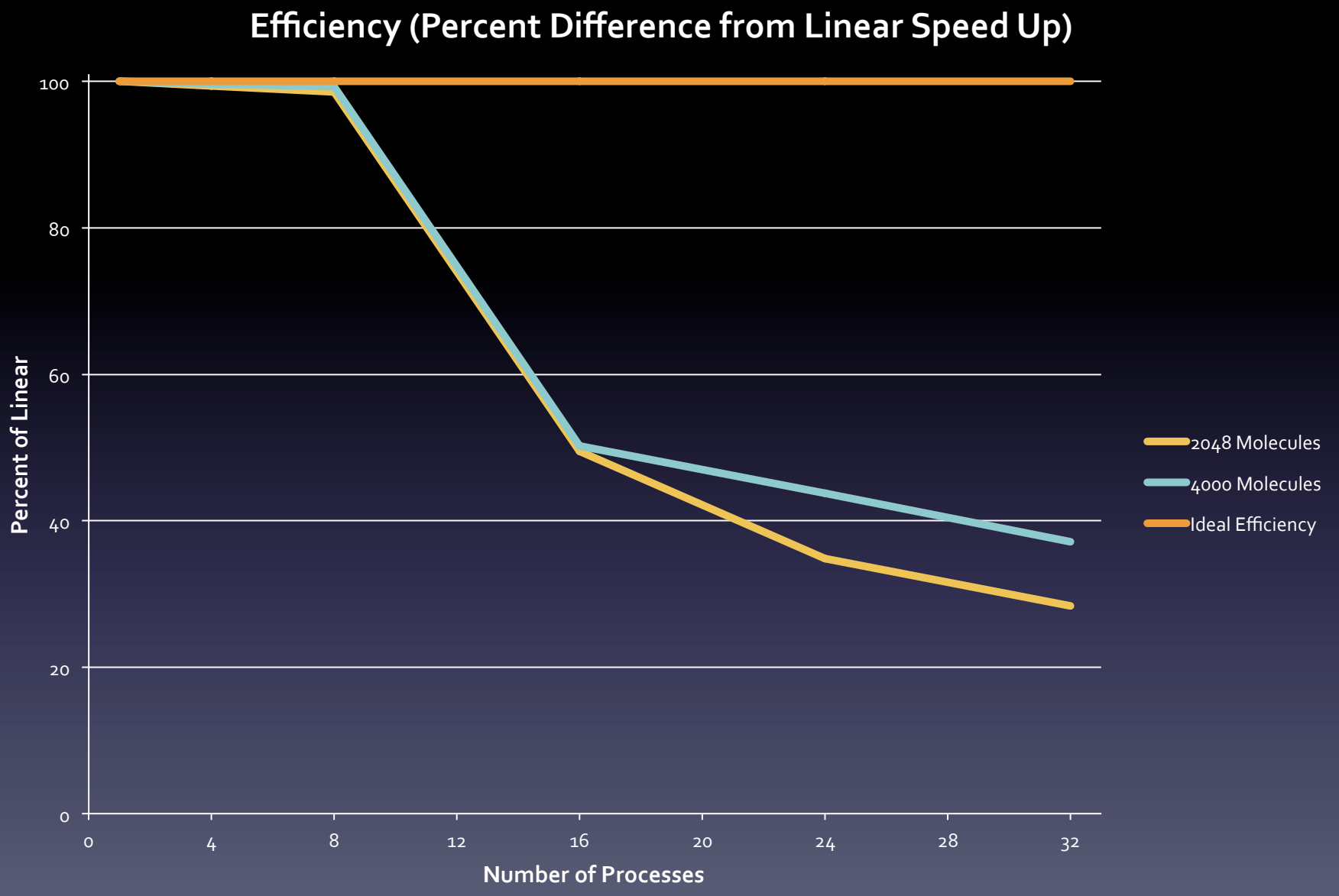
- 2 Parts
  - Computation Handler
  - Communication Handler

Results – Let's we how we did!

2048 water molecules

1 process:	1218.3 seconds
4 processes:	306.40 seconds - 3.98 x faster
8 processes:	154.46 seconds – 7.89 x faster
16 processes:	153.92 seconds - 8.04 x faster
32 processes:	134.31 seconds - 9.07 x faster


Speed Up
Number of Times Faster
0
4
8
12
16
20
24
28
32
Number of Processes
0
4
8
12
16
20
24
28
32
2048 Molecules
4000 Molecules
Linear Speed Up


Efficiency (Percent Difference from Linear Speed Up)
Percent of Linear
100
80
60
40
20
0
0
4
8
12
16
20
24
28
32
Number of Processes
2048 Molecules
4000 Molecules
Ideal Efficiency

What went wrong?

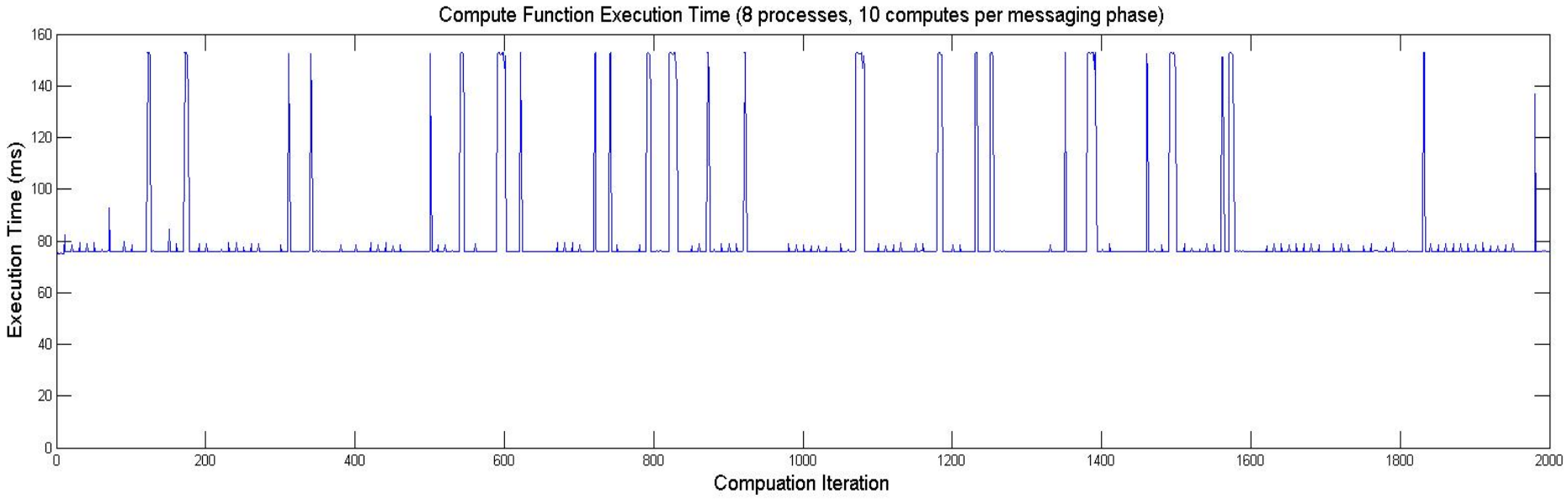
- 2 or more processes on the same machine
  - Processes Per Machine = PPM
- Problem persisted on 2 architectures
  - Clouds and Rains
- 2 serial programs run as expected

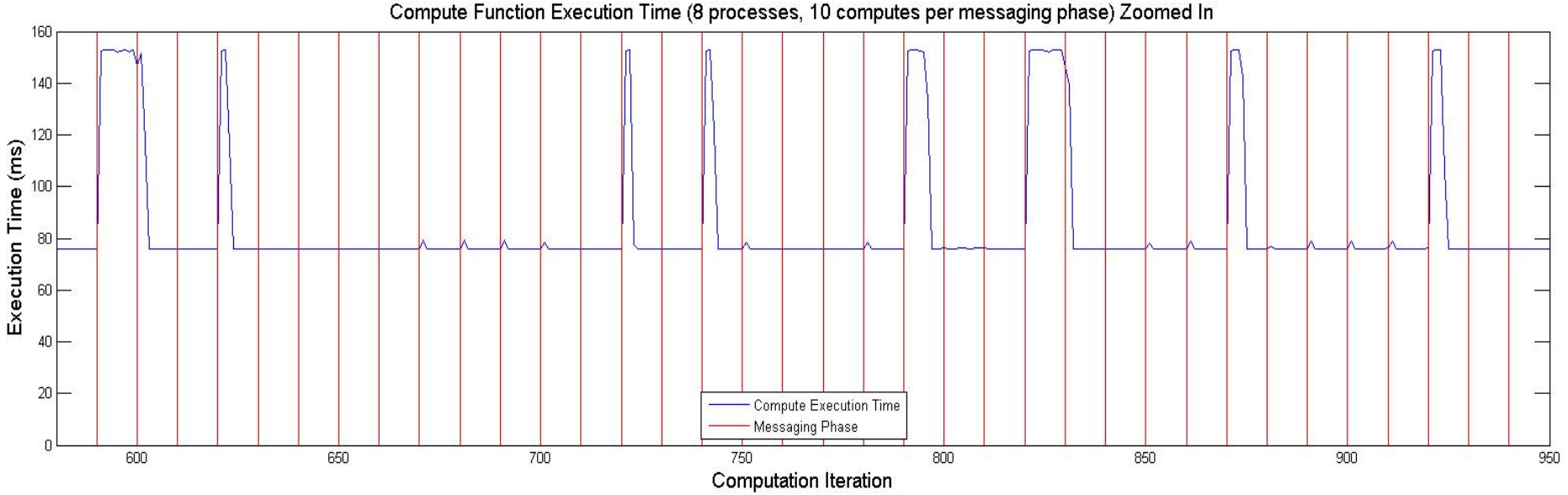
Testing

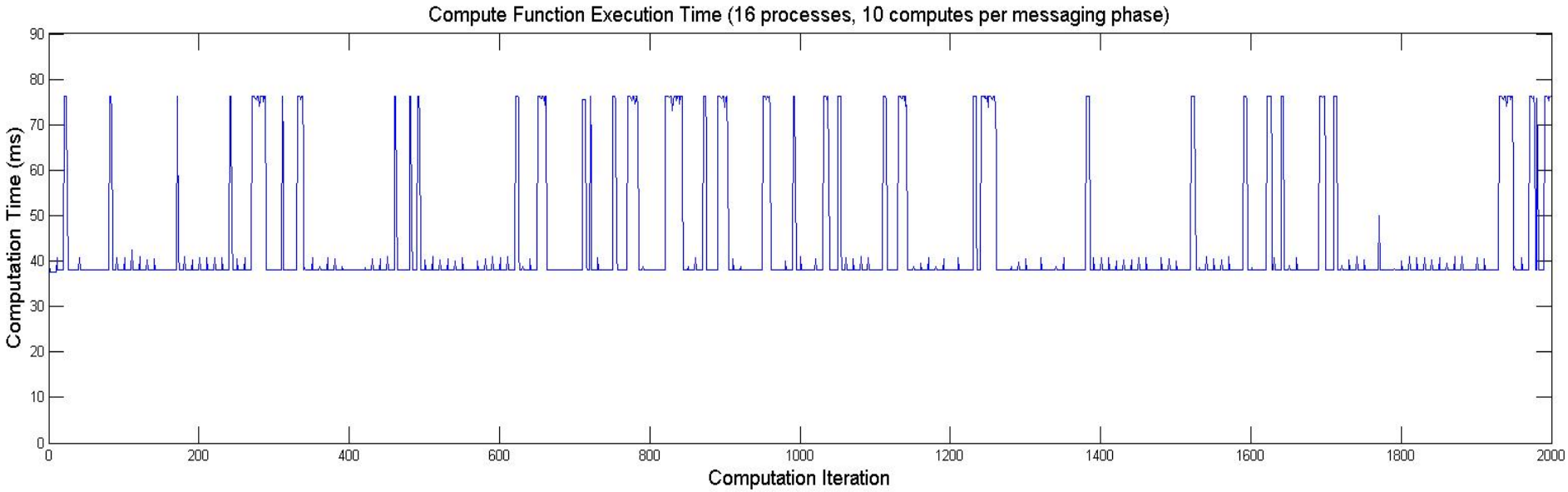
- CPU usage for 1 PPM is near 100%
- CPU usage for 2 PPM fluctuates – 50% - 100%
  - Processes on the same machine in tandem
- No more Mr. Nice Guy! (sudo nice --20)
- Need to time each phase of the simulation

Testing

- Most time-steps take almost 2x on 2 PPM
- Increase the ratio of calculation to messaging
- 10X Computation -> A very interesting pattern!


Compute Function Execution Time (8 processes, 10 computes per messaging phase)
Execution Time (ms)
Compuation Iteration
0
20
40
60
80
100
120
140
160
0
200
400
600
800
1000
1200
1400
1600
1800
2000


Compute Function Execution Time (8 processes, 10 computes per messaging phase) Zoomed In
Execution Time (ms)
Computation Iteration
Compute Execution Time
Messaging Phase
0
20
40
60
80
100
120
140
160
600
650
700
750
800
850
900
950


Compute Function Execution Time (16 processes, 10 computes per messaging phase)
Computation Time (ms)
Computation Iteration
0
10
20
30
40
50
60
70
80
90
0
200
400
600
800
1000
1200
1400
1600
1800
2000


Compute Function Execution Time (16 processes, 10 computes per messaging phase) Zoomed In
Computation Time (ms)
Computation Iteration
Compute Execution Time
Messaging Phase
0
10
20
30
40
50
60
70
80
90
600
650
700
750
800
850
900

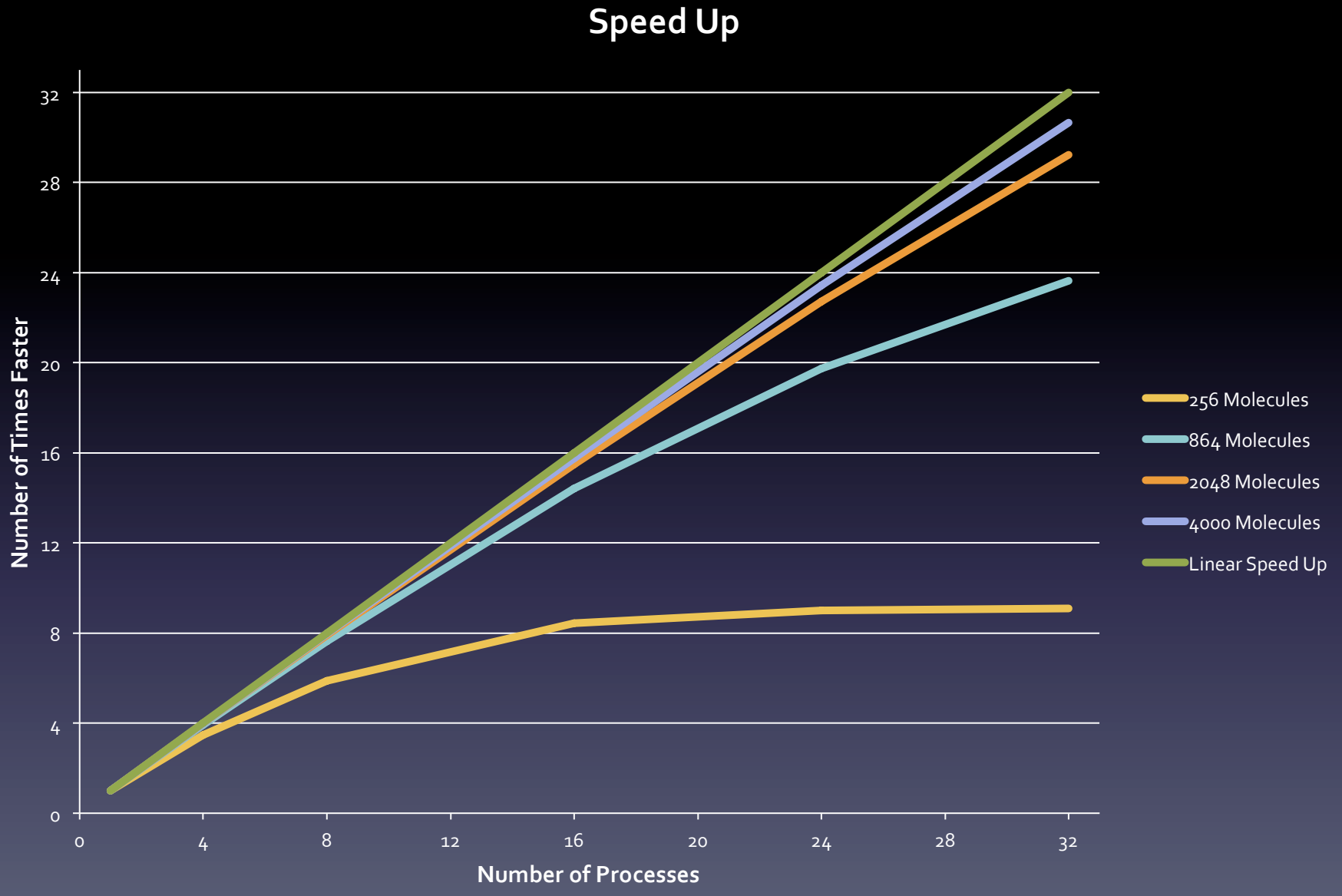
Solution

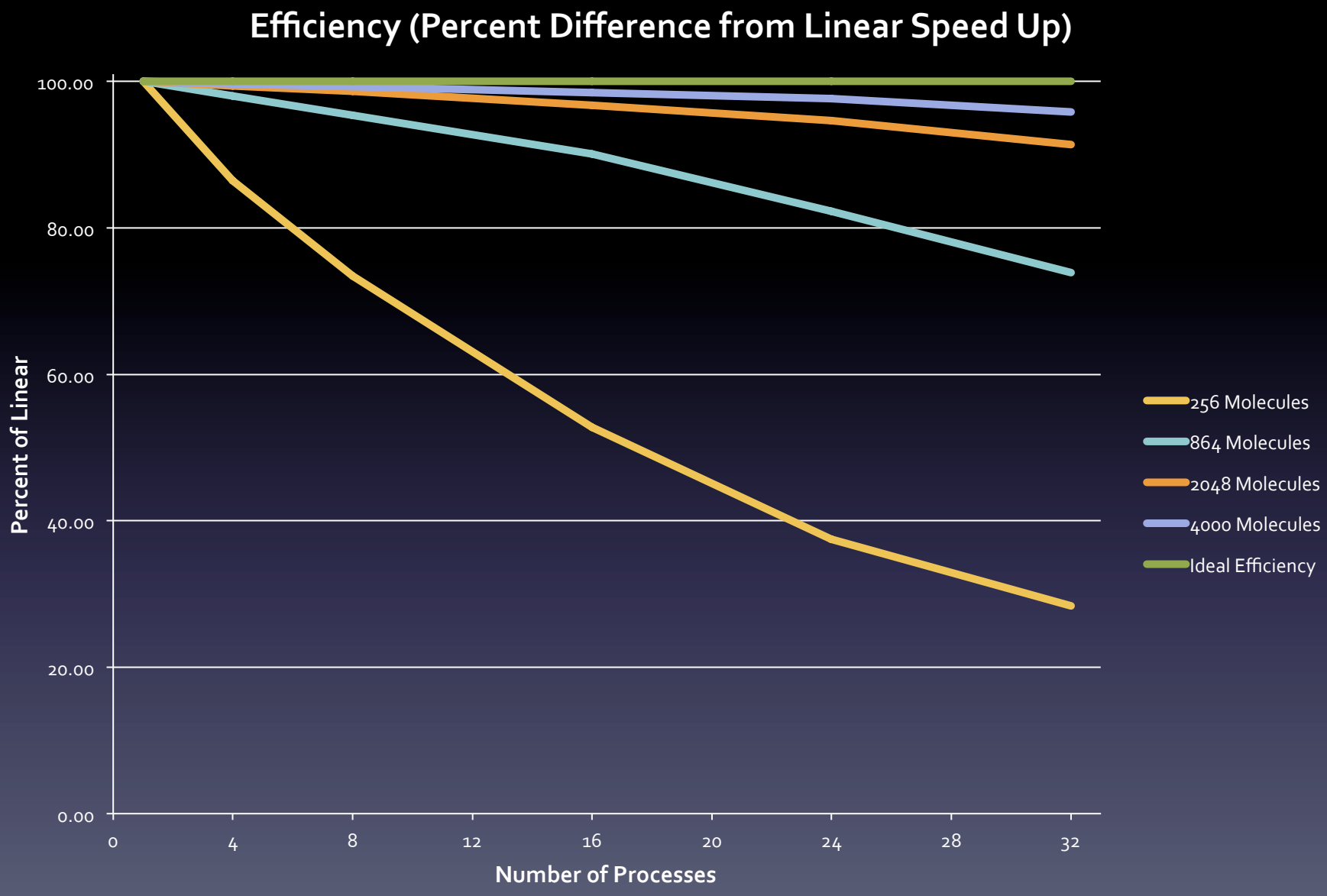
TURN OFF HYPER-THREADING

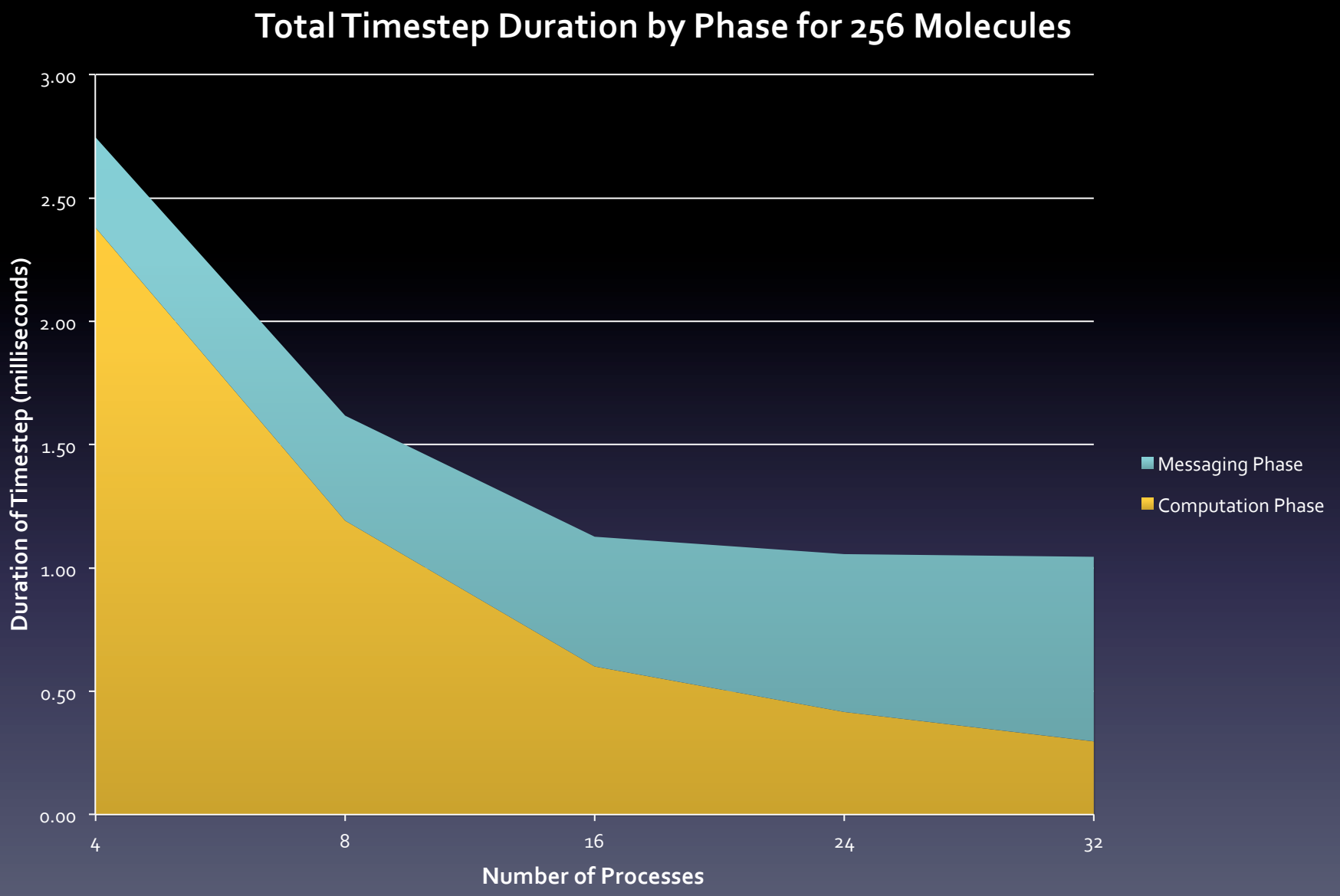
Better Results!

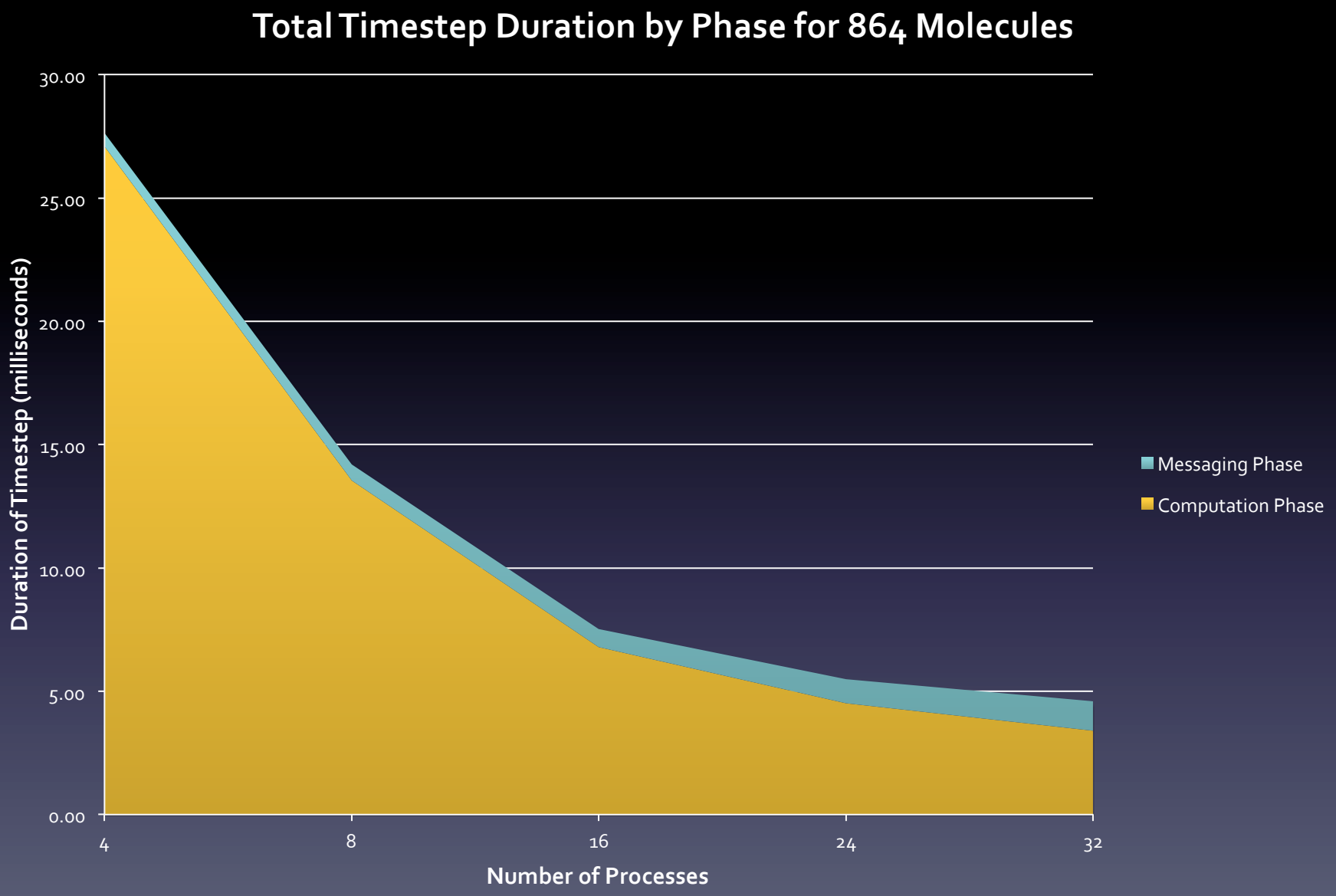
2048 water molecules

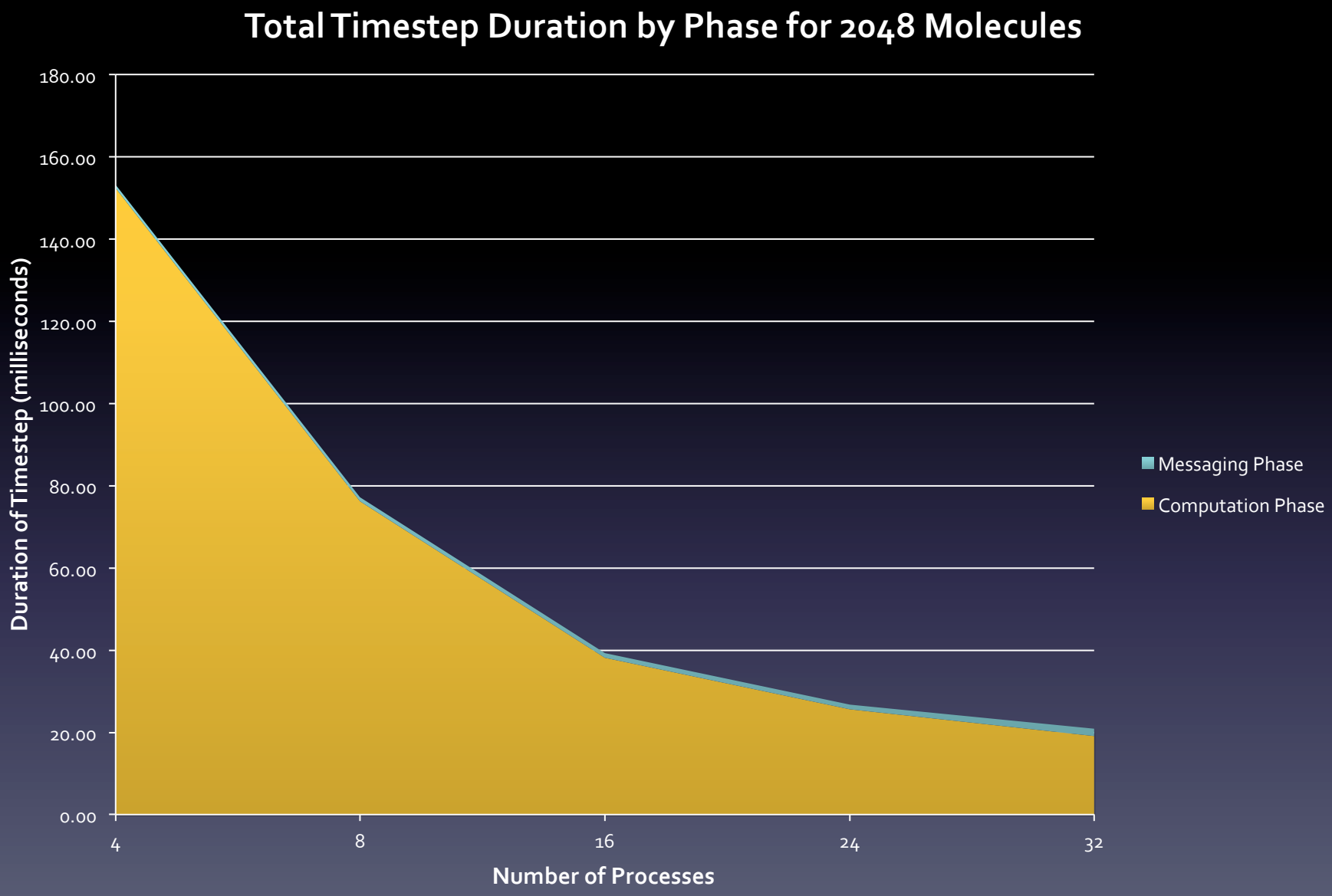
1 process:	1218.3 seconds
4 processes:	306.40 seconds - 3.98 x faster
8 processes:	154.40 seconds - 7.89 x faster
16 processes:	78.67 seconds - 15.49 x faster
32 processes:	41.66 seconds -29.24 x faster

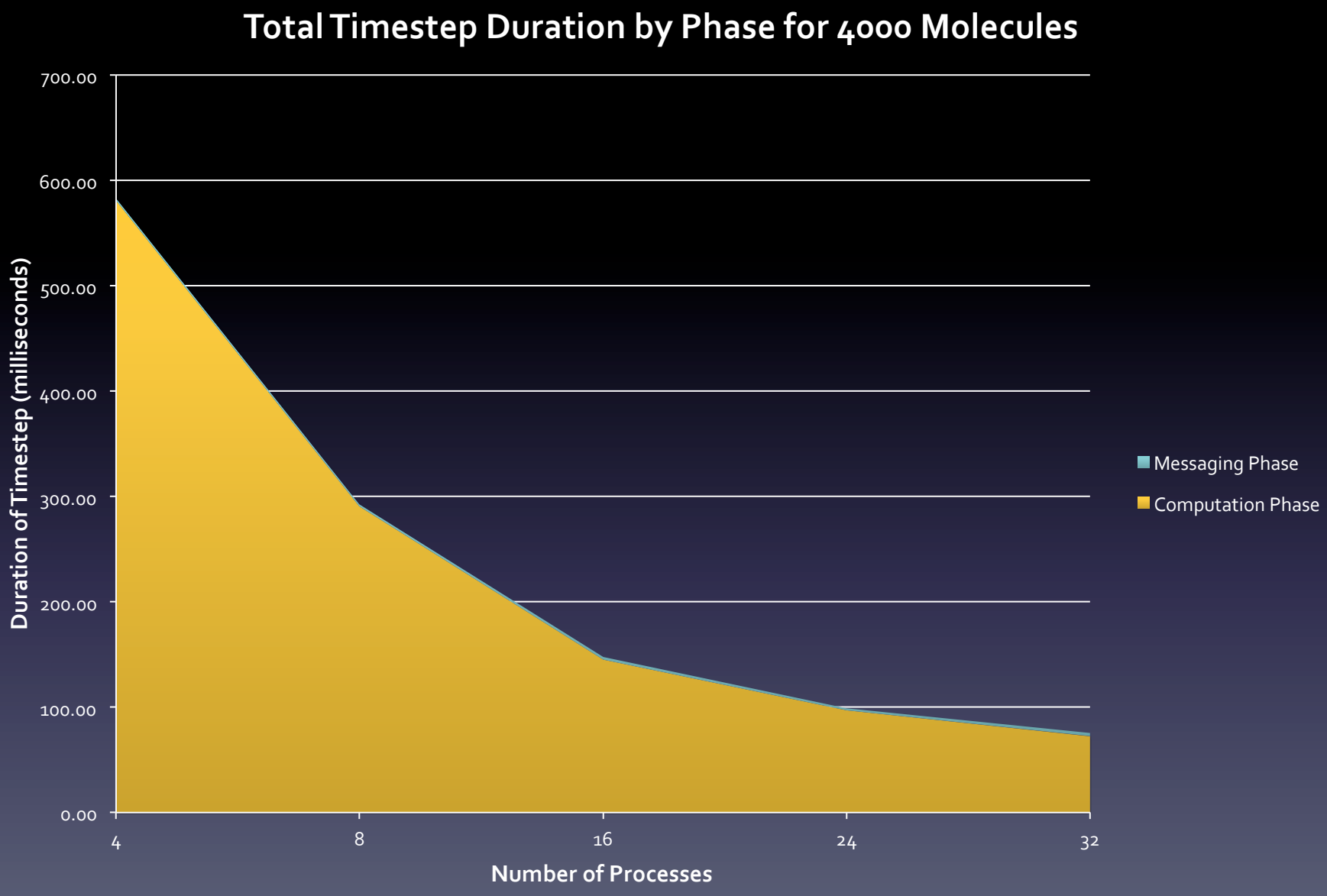

Speed Up
Number of Times Faster
0
4
8
12
16
20
24
28
32
Number of Processes
256 Molecules
864 Molecules
2048 Molecules
4000 Molecules
Linear Speed Up

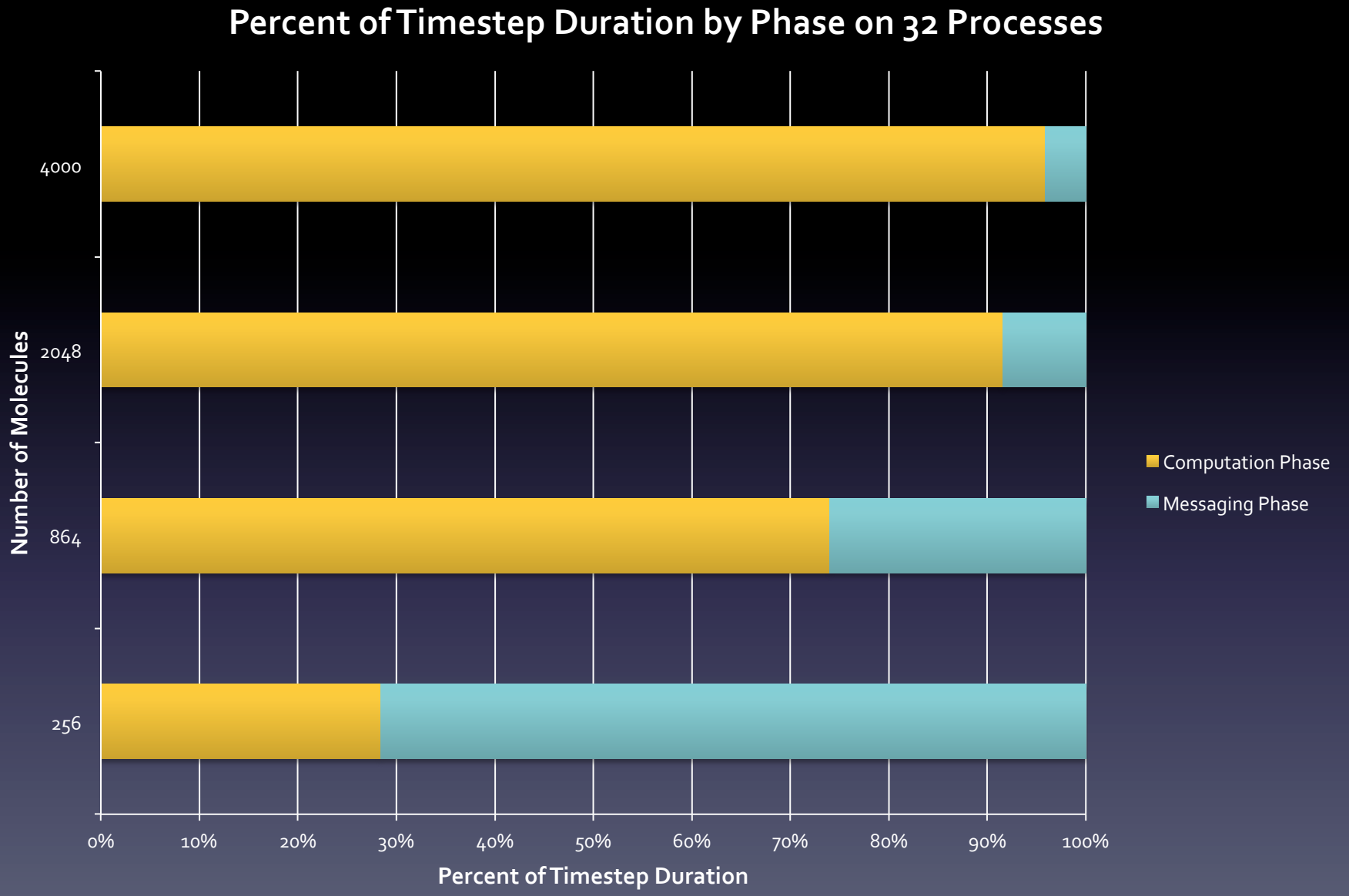

Efficiency (Percent Difference from Linear Speed Up)
100.00
80.00
60.00
40.00
20.00
0.00
Percent of Linear
0
4
8
12
16
20
24
28
32
Number of Processes
256 Molecules
864 Molecules
2048 Molecules
4000 Molecules
Ideal Efficiency

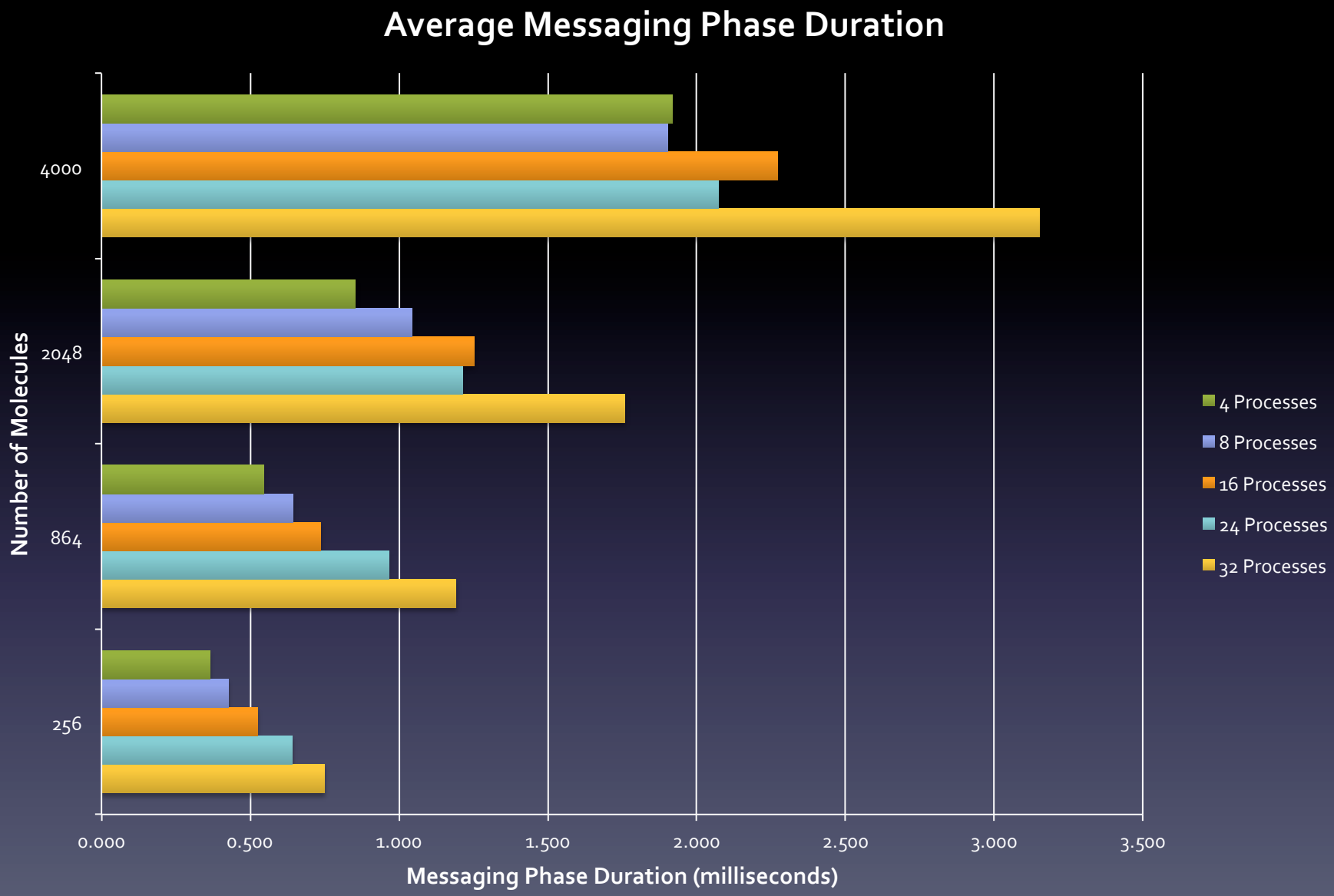

Total Timestep Duration by Phase for 256 Molecules
Duration of Timestep (milliseconds)
3.00
2.50
2.00
1.50
1.00
0.50
0.00
4
8
16
24
32
Number of Processes
Messaging Phase
Computation Phase

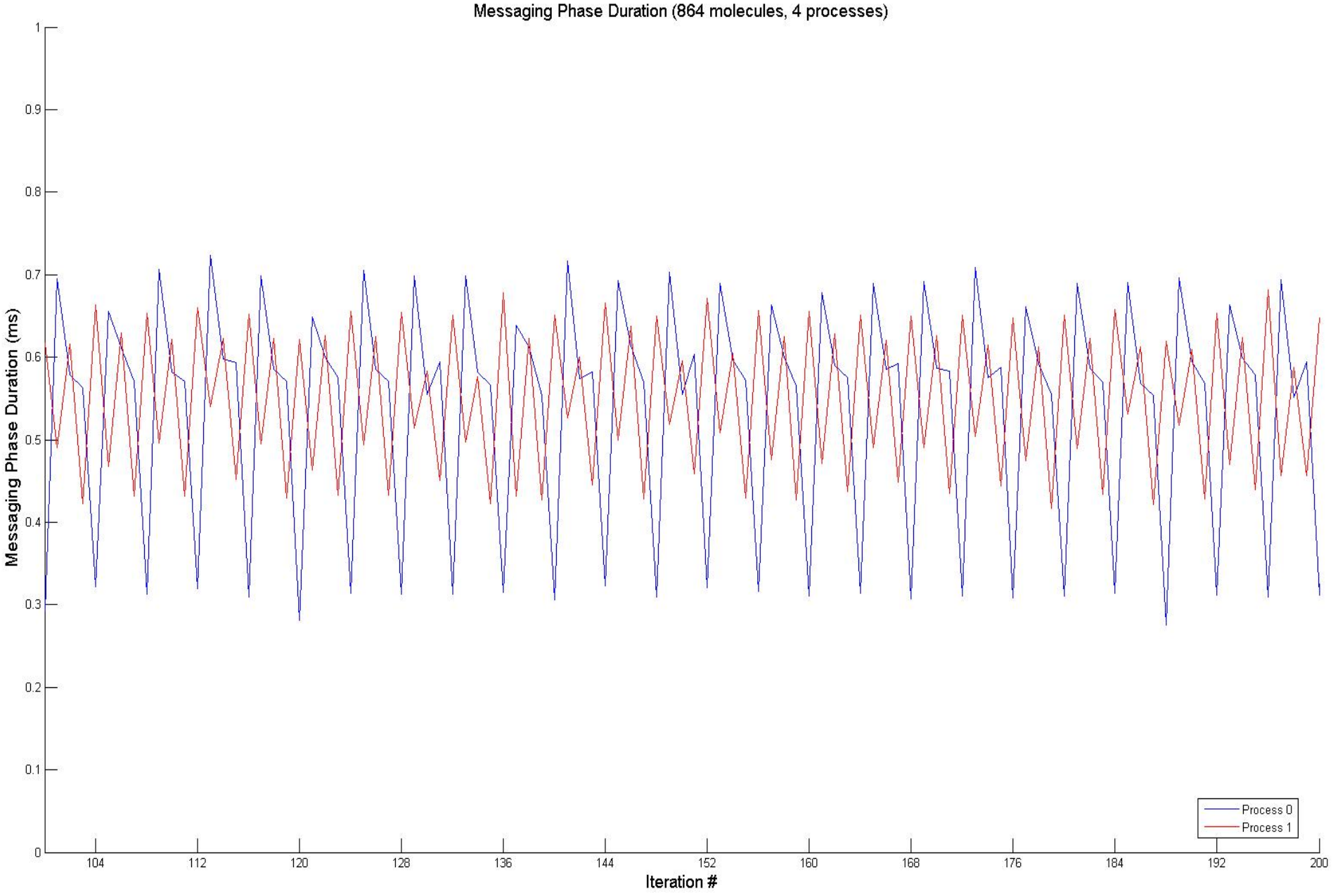

Total Timestep Duration by Phase for 864 Molecules
Duration of Timestep (milliseconds)
30.00
25.00
20.00
15.00
10.00
5.00
0.00
4
8
16
24
32
Number of Processes
Messaging Phase
Computation Phase

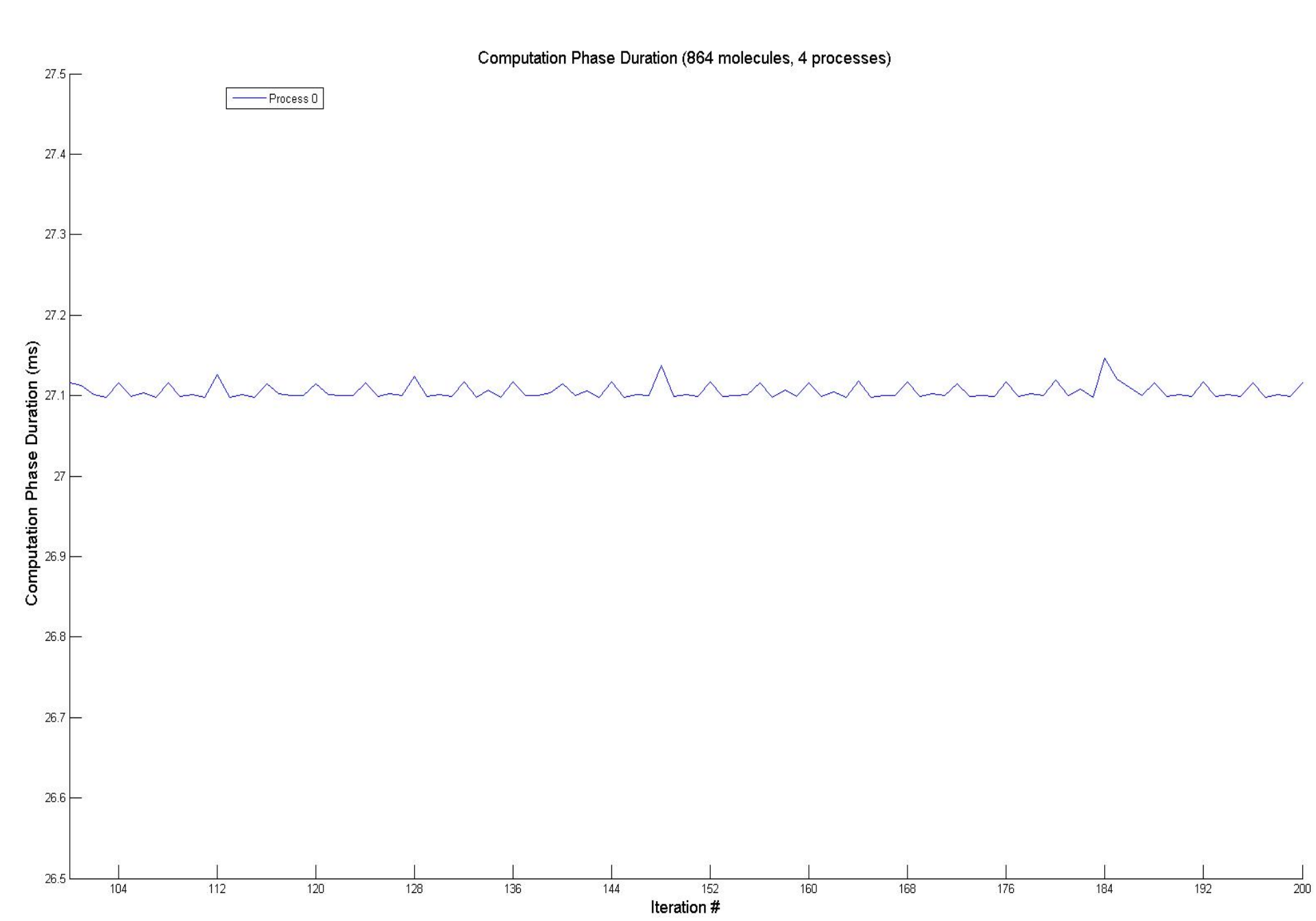

Total Timestep Duration by Phase for 2048 Molecules
Duration of Timestep (milliseconds)
180.00
160.00
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
4
8
16
24
32
Number of Processes
Messaging Phase
Computation Phase

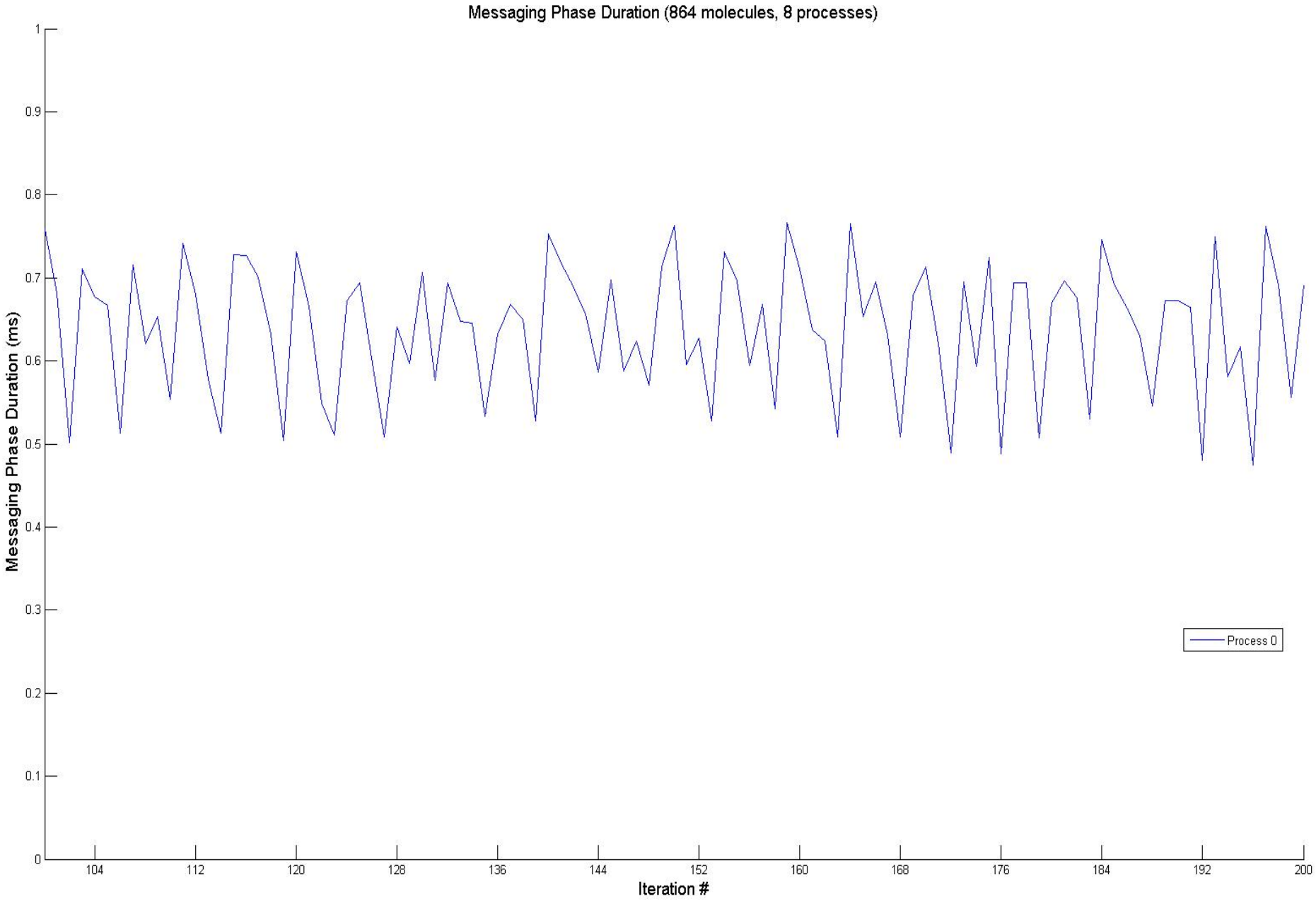

Total Timestep Duration by Phase for 4000 Molecules
700.00
600.00
500.00
400.00
300.00
200.00
100.00
0.00
Duration of Timestep (milliseconds)
4
8
16
24
32
Number of Processes
Messaging Phase
Computation Phase

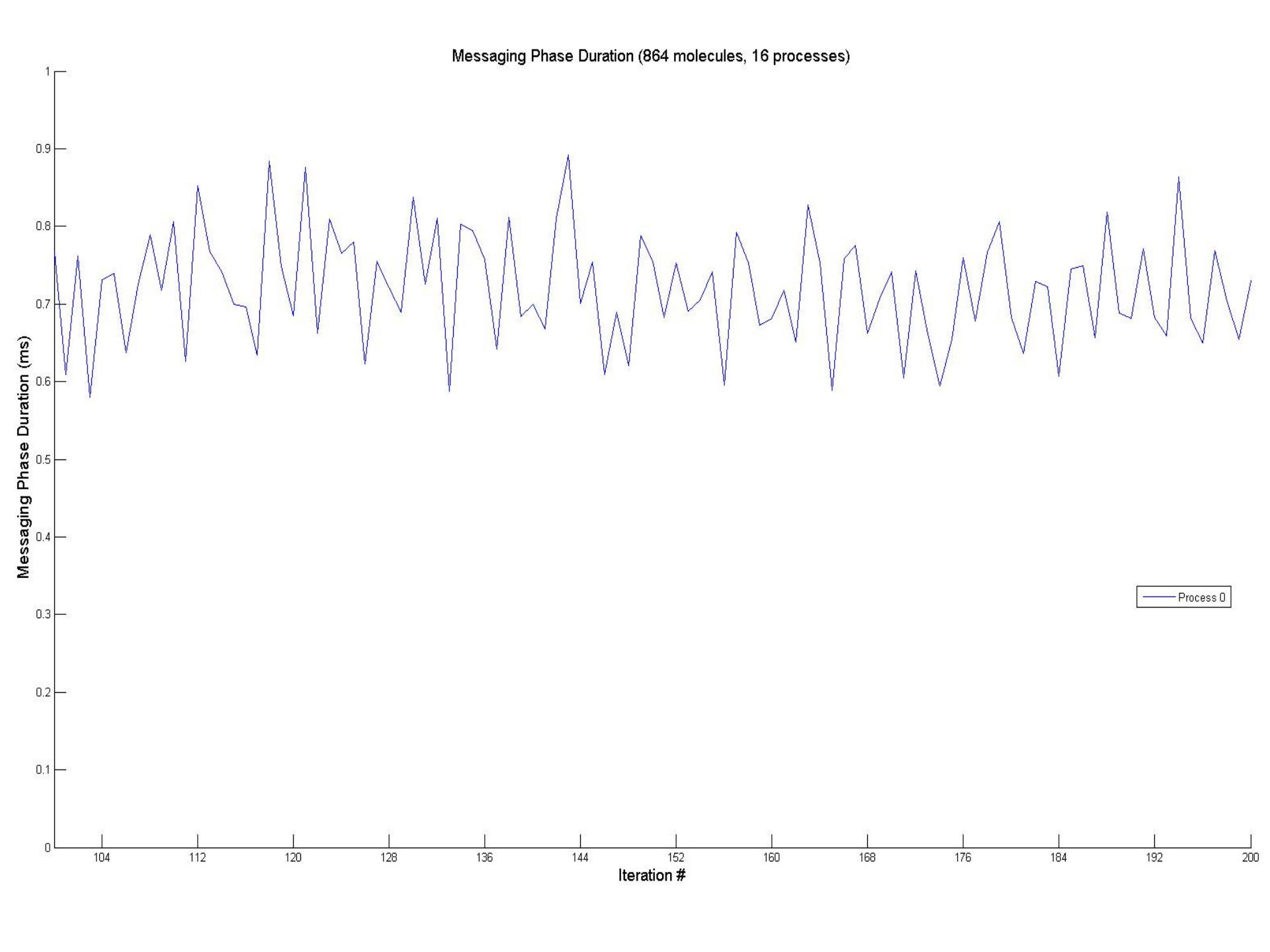

Percent of Timestep Duration by Phase on 32 Processes
Number of Molecules
4000
2048
864
256
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Percent of Timestep Duration
Computation Phase
Messaging Phase

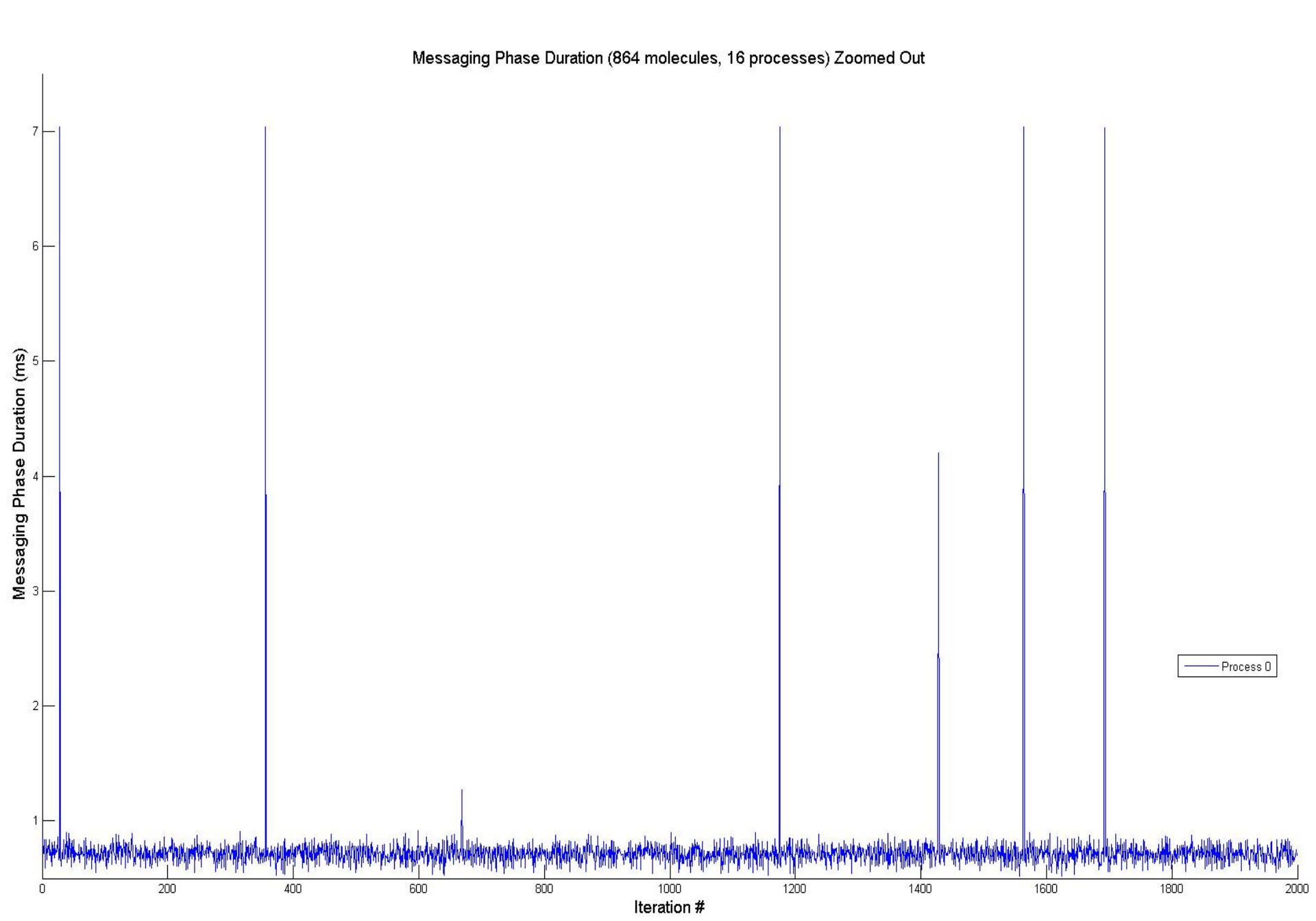

Average Messaging Phase Duration
Number of Molecules
4000
2048
864
256
0.000
0.500
1.000
1.500
2.000
2.500
3.000
3.500
Messaging Phase Duration (milliseconds)
4 Processes
8 Processes
16 Processes
24 Processes
32 Processes


Messaging Phase Duration (864 molecules, 4 processes)
Messaging Phase Duration (ms)
Iteration #
Process 0
Process 1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
104
112
120
128
136
144
152
160
168
176
184
192
200


Computation Phase Duration (864 molecules, 4 processes)
Process 0
Computation Phase Duration (ms)
27.5
27.4
27.3
27.2
27.1
27
26.9
26.8
26.7
26.6
26.5
104
112
120
128
136
144
152
160
168
176
184
192
200
Iteration #


Messaging Phase Duration (864 molecules, 8 processes)
Messaging Phase Duration (ms)
Iteration #
Process 0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
104
112
120
128
136
144
152
160
168
176
184
192
200


Messaging Phase Duration (864 molecules, 16 processes)
Messaging Phase Duration (ms)
Iteration #
Process 0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
104
112
120
128
136
144
152
160
168
176
184
192
200


Messaging Phase Duration (864 molecules, 16 processes) Zoomed Out
Messaging Phase Duration (ms)
Iteration #
Process 0
1
2
3
4
5
6
7
0
200
400
600
800
1000
1200
1400
1600
1800
2000

Future - Improvements

- Rotations
- Ewald summations - reciprocal space
- Electrostatic shielding and cut off distances
- Extend model to include many different atoms and molecules

Future - Performance Improvements

- Barnes-Hut Algorithm (Octree)
  - Approximations of point charges for groups
- Interleaving messaging and computation
- Graphics Processing Units

Questions?

THANK YOU!