Parallel Performance Evaluation of MITgcm

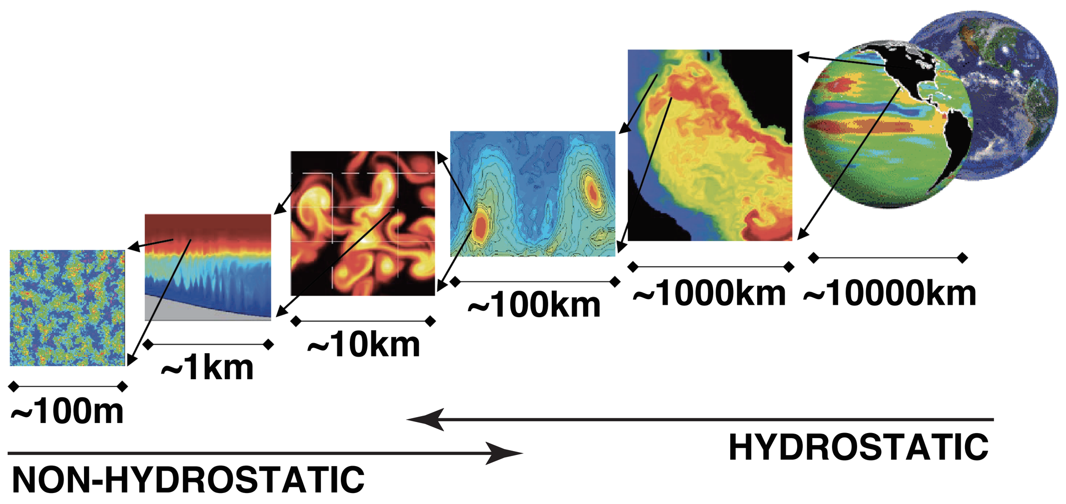
Rin Irie, Helen Stewart, Toru Nakamura, Masaki Hisada
Nippon Telegraph and Telephone Corporation

NTT

1 INTRODUCTION

What is MITgcm[1]?

- MITgcm is an ocean-atmosphere general circulation model developed by the Massachusetts Institute of Technology.
- It uses one hydrodynamical kernel including the non-hydrostatic capability to drive both atmospheric and ocean models.
- It can be used for both small-scale phenomena such as convection and large-scale phenomena such as global general circulation.
- It supports parallel computation using multiple processors by dividing the computation area in the horizontal direction.



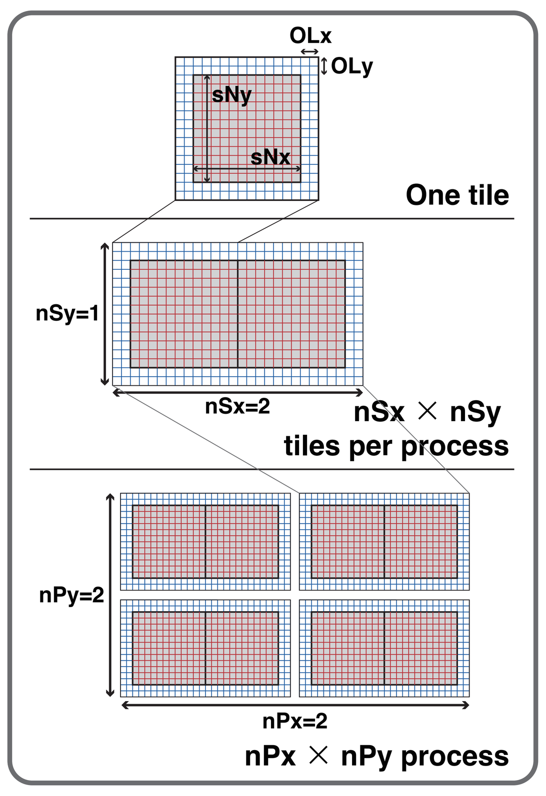
Goal

Evaluate the parallel performance (strong scaling) of MITgcm to get the computation time characteristics for large-scale computation.

2 METHODS

Parallel Computation

- The simulation area is horizontally partitioned into grid tiles.
- MITgcm supports running multiple processes (MPI[4]) and multiple threads (OpenMP[3]) in parallel.
- The number of tiles and threads to run in a single process must be set specified by the user.



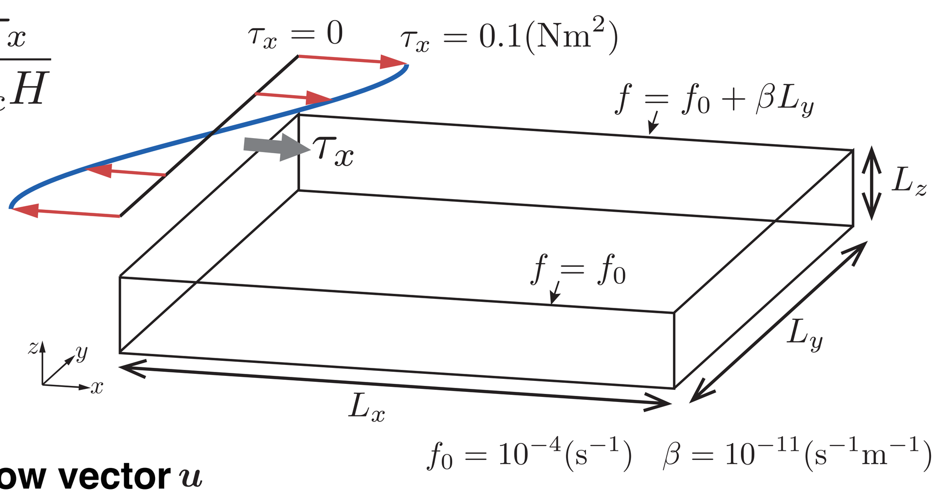
Benchmark Conditions

- Physics Model: Barotropic ocean gyre (MITgcm's Manual 4.1[1])

$$\frac{Du}{Dt} - fv + g\frac{\partial \eta}{\partial x} - A_h \nabla_h^2 u = \frac{\tau_x}{\rho_c H}$$

$$\frac{Dv}{Dt} + fu + g\frac{\partial \eta}{\partial y} - A_h \nabla_h^2 v = 0$$

$$\frac{\partial \eta}{\partial t} + \nabla_h \cdot (H\mathbf{u}) = 0$$



u, v : x and y components of the flow vector $\mathbf{u}$
η : free surface height
A_h : horizontal Laplacian viscosity
ρ_c : fluid density
g : acceleration due to gravity

- Discretization Condition:

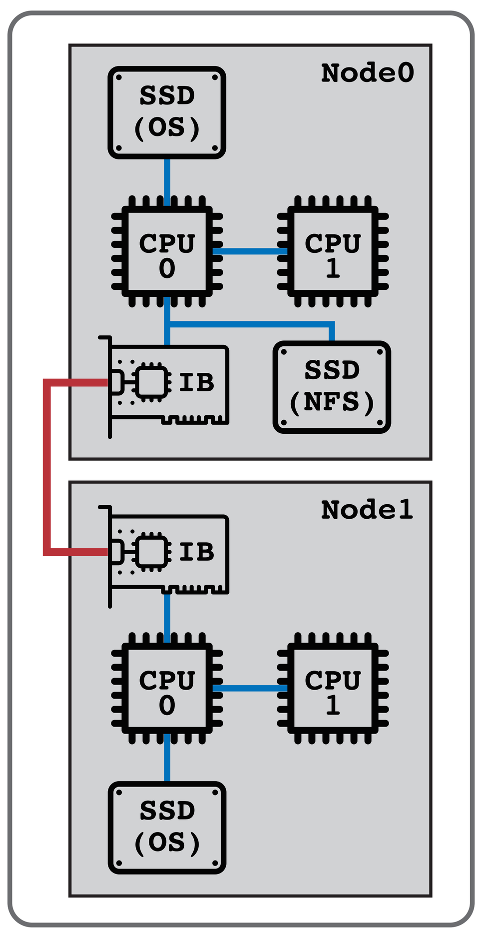
Grid spacing	$\Delta x = \Delta y = 2.0$ km, $\Delta z = 5.0$ km
Grid points	$N_x = N_y = \{672, 1008, 1344\}$, $N_z = 1$
Time step	$\Delta t = 5.0$ minutes
Integration Time	$N_t = 104832$ steps $(= 1.0$ year$)$

- Hardware Specifications:

CPU	Intel Xeon Gold 6258R (2.7GHz, 28c) x2
Memory	768GB (DDR4-2933 ECC 64GB x12)
Storage	Crucial MX500 2TB (SATA SSD)
Interconnect	Mellanox ConnectX-6 (200Gb-HCA)[2]

- Software Specifications:

OS	Ubuntu 22.04 LTS
Compiler	gfortran v11.3.0
MPI	OpenMPI v4.1.2
mpirun opt	--mca pml ucx
MITgcm	checkpoint68i (Mid 2022 version)



3 RESULTS

Fig. 1 Measurements of computation times as the number of processes is increased for grids of 672^2, 10082 and 1344^2.

- The results are taken as the average of 3 measurements.
- The optimal computation time for strong scaling should be 1/n of the computation time for one process, when n is the number of processes (shown by 'Slope-2').

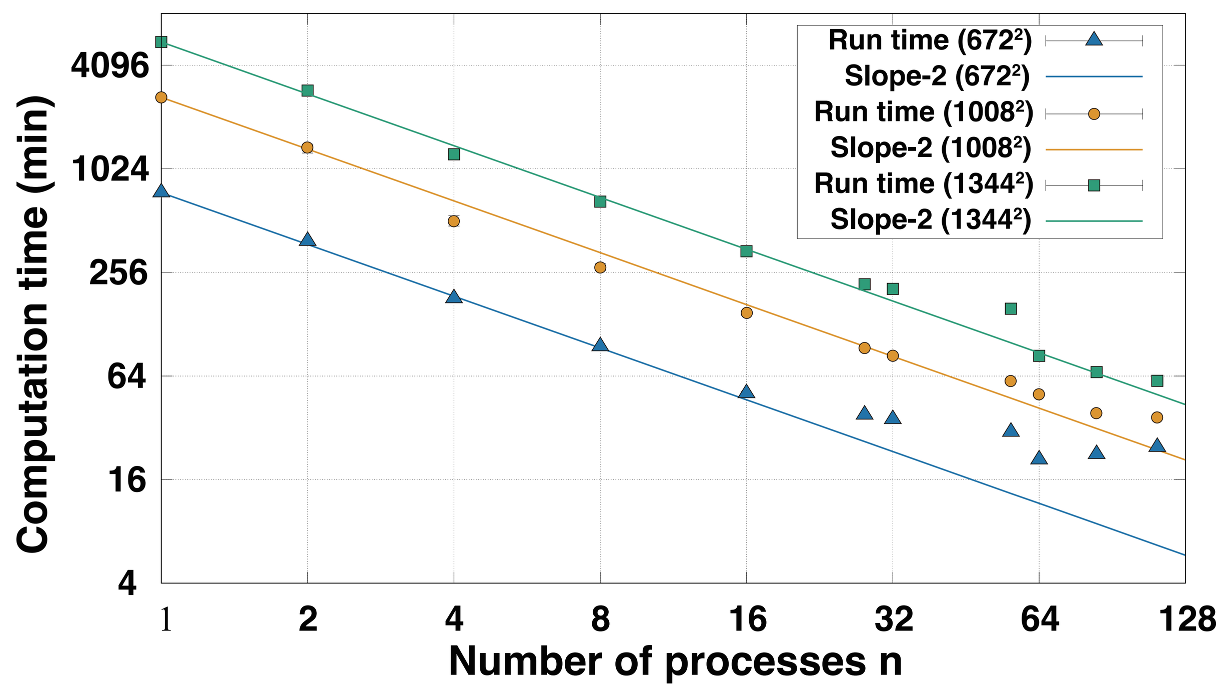
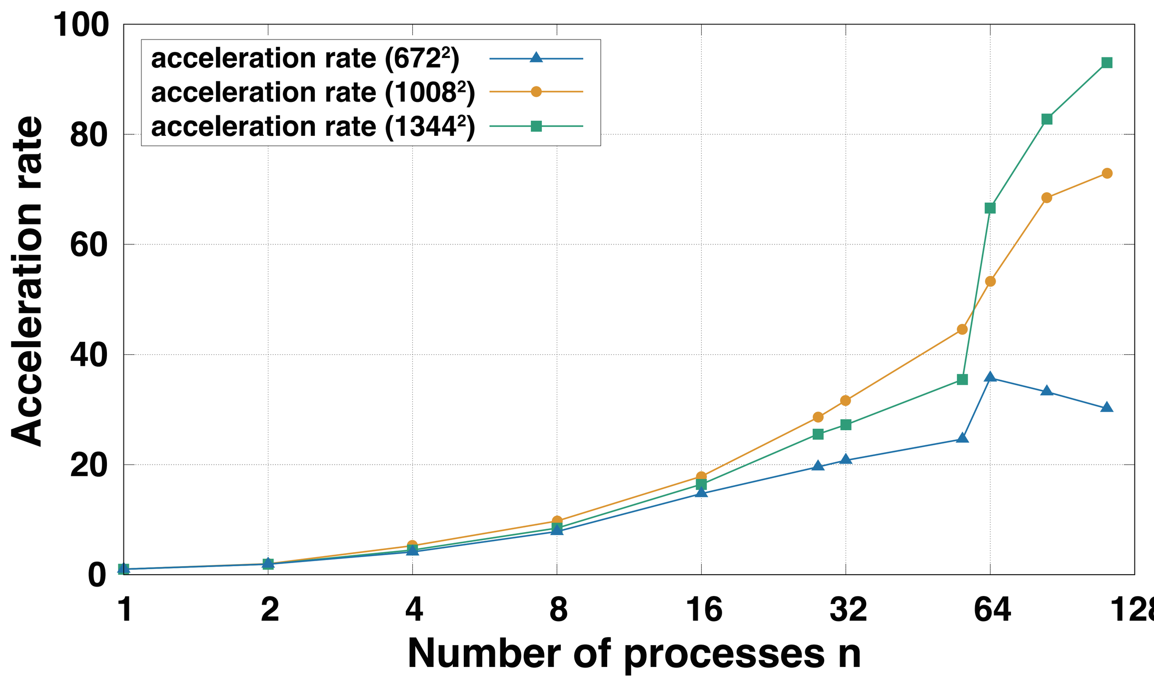


Fig. 2 Acceleration rates n, are defined as the ratio of computation time to the computation time when using one process (n = 1).



Discussion

- Up to 28 processes (one socket),
  The computation time are almost inline with 'Slope-2'.
- For 56 processes (one node),
  The computation time is slowing down. This is due to interrupted processing in the OS, MPI, etc. as all the cores in one node are used in the computation.
- 64~112 processes (2 nodes),
  - For the 672^2 grid, the acceleration rate slows down above 64 processes, for which there are 7056 grid cells per process.
  - For the 1008^2 grid, the acceleration rate starts to slow down above 112 processes, for which there are 9072 grid cells per process.

➡ The limit of strong scaling is about 7000 grid cells per process.

References

1. A. Adcroft et al.: MITgcm's User Manual. https://mitgcm.readthedocs.io/en/latest/
2. Infiniband Trade Association: Supplement to InfiniBand Architecture Specication, Vol 1.2.1 (2010)
3. OpenMP Architecture Review Board: OpenMP Application Programming Interface, Ver 5.0 (2018).
4. Message Passing Interface Forum: MPI: A Message Passing Interface. In Proceedings of the 1993 ACM/IEEE Conference on Supercomputing. 878–883 (1993).
5. Fluid Numerics Journal: MITgcm Benchmarks on Google Cloud Platform. https://journal.fluidnumerics.com/mitgcm-benchmarks-on-google-cloud-platform