Research on Computational Techniques for JMA's NWP Models

Kensuke Takenouchi, Shintaro Yokoi, Chiashi Muroi, Junichi Ishida and Kohei Aranami
Numerical Prediction Division, Japan Meteorological Agency,
1-3-4 Otemachi, Chiyoda-ku, Tokyo 100-8122, Japan
E-mail: k.takenouchi@met.kishou.go.jp

1. Introduction

The Japan Meteorological Agency (JMA) is currently implementing a comprehensive program of assessment and research on computational techniques.

Recent trends regarding HPCs (high-performance computers) indicate that cluster machines have accounted for a large share of the market in the last decade.

This report describes the performance of JMA's NWP models on two kinds of cluster machines and the effectiveness of computational techniques implemented on them.

2. TSUBAME and SCS8

Table 1 shows the specifications of TSUBAME (a computing system at the Tokyo Institute of Technology) and SCS8 (JMA's eighth operational supercomputer system). TSUBAME is composed of 655 nodes (Sun Fire X4600), and has more memory per node than ordinary PC clusters. SCS8 is composed of 210 nodes (Hitachi SR11000), and represents a typical SMP HPC.

Table 1. Machine specifications

Machine name	TSUBAME	SCS8
Total performance	85 TFlops	27.5 TFlops
Number of nodes	655 (16 CPU/node)	210 (16 CPU/node)
Amount of memory	32 GB/node	64 GB/node
Memory bandwidth	6.4 GB/s	15 GB/s
Speed between nodes	Two-way, 2 GB/s	Two-way, 16 GB/s
CPU	AMD Opteron 880 (2.4 GHz), 885 (2.6 GHz)	POWER5+ (1.9 GHz, 2.1 GHz)

3. Scalability and profiles

The scalability and profiles of JMA's Global Spectrum Model (GSM; for details, see http://www.jma.go.jp/jma/jma-eng/jma-center/nwp/nwp-top.htm) and its NonHydrostatic regional Model (NHM, Saito et al., 2007) were measured using TSUBAME and SCS8. Table 2 shows the settings of the experiments, while Figs. 1 and 2 show the scalability and profiles, respectively.

The scalability of both the GSM and the NHM on TSUBAME are close to the ideal lines, while the performance of the GSM on SCS8 hardly increases over about 32 processes. This does not necessarily signify that TSUBAME is better than SCS8. In fact, the total execution time with eight nodes on TSUBAME is five times longer for the GSM and ten times longer for the NHM than on SCS8, although the theoretical peak performance per node of TSUBAME is almost the same as that of SR11000. Considering that the memory bandwidth of TSUBAME is lower than that of SCS8, the better scalability on TSUBAME is considered to be due to the decrease in memory load per process resulting from the increased number of nodes.

From Fig. 2, it can be seen that the rates of communication time on the GSM and NHM increase with higher numbers of nodes, which indicates that measures to reduce communication time are necessary for both models.

Table 2. Experimental settings of models

Model	GSM	NHM
Resolution	TL319	2 km
Number of grid points	640 x 320 x 60	800 x 550 x 60
Number of calculation steps (forecast period)	72 (one day)	60 (ten minutes)
Number of processes	1 per node	1 per node
Number of threads	16 per process	16 per process
SMP	OpenMP	OpenMP

4. Computational technique

(i) Parallelization of calculation and communication

The parallelization of calculation and communication was examined as a way of reducing the total execution times of the models.

Two cases of such parallelization are shown in Fig 3. One is parallelization in one process, and the other is parallelization with threads. Research has shown that these parallelization techniques generally reduce the total execution time to some extent. However, the amount of reduction depends on the kind of machine, and the effects in total models are unclear. Additionally, extensive modification of model codes is necessary to enable these techniques to be applied to JMA's NWP models. As a result, these parallelization methods have not been applied, and remain under research at JMA.

(ii) kij-ordering

In JMA's NWP models (such as the NHM), variable arrays are set sequentially in the order of x, y, z (ijk-ordering: a(i, j, k)). However, setting them sequentially in the order of z, x, y (kij-ordering: a(k, i, j)) works more effectively to increase speed in terms of cache tuning, for example, in the case of calculation in the Z direction (M. Ashworth et al., 2001; J. Michalakes et al., 2001). The effectiveness of kij-ordering is to be investigated in the development of a new dynamical core for nonhydrostatic model.

Reference

Saito, K., J. Ishida, K. Aranami, T. Hara, T. Segawa, M. Narita, and Y. Honda, 2007: Nonhydrostatic Atmospheric Models and Operational Development at JMA. J. Meteor. Soc. Japan, 85B, 271–304.

Ashworth M., R. Proctor, J. T. Holt, J. I. Allen, and J. C. Blackford, 2001: Coupled marine ecosystem modelling on high performance computers. Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the Use of High Performance Computing in Meteorology, W. Zwieflhofer and N. Kreitz, Eds., World Scientific, 150–163.

Michalakes, J., R. Loft, A. Bourgeois, 2001 : "Performance-Portability and the Weather Research and Forecast Model" in on-line proceedings of the HPC Asia 2001 conference, Gold Coast, Queensland, Australia, September 24-28.

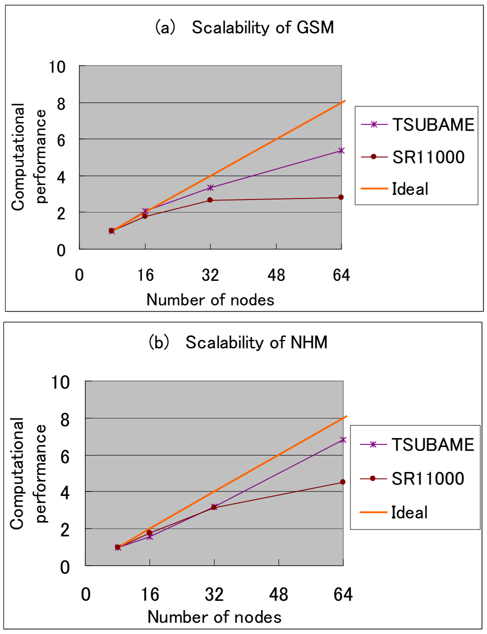

Fig. 1. Scalability of JMA's NWP models based on eight-node performance. (a) GSM, (b) NHM (1 process/node, 16 threads/process)

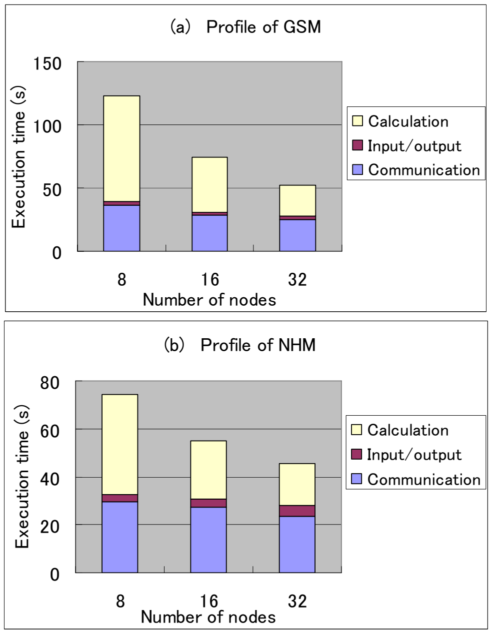

Fig. 2. Profiles of JMA's NWP models on SCS8. (a) GSM, (b) NHM (1 process/node, 16 threads/process)

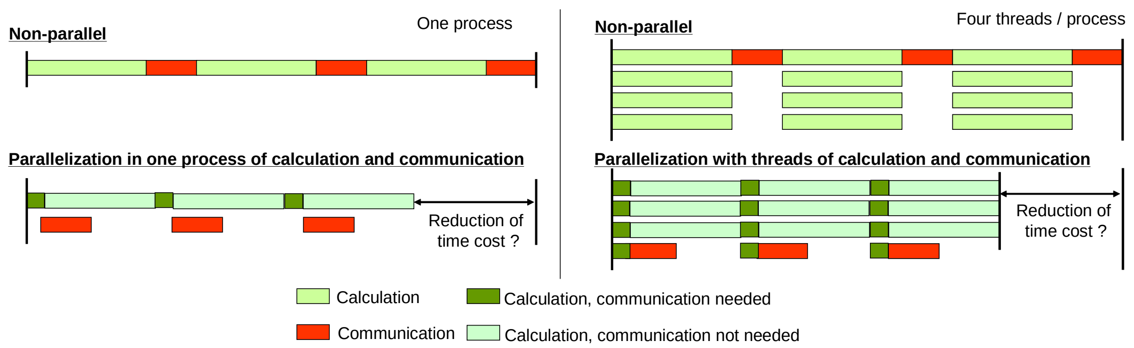

Fig. 3. Examples showing parallelization of calculation and communication. Left: parallelization in one process. Right: parallelization with four threads.