Regular Paper

An MEG Data Analysis System Using Grid Technology

Susumu Date,[†1] Yuko Mizuno-Matsumoto,[†1,†2,†3]
Youki Kadobayashi[†4] and Shinji Shimojo[†5]

Recently, the lack of computational power for analyzing scientific data, and the distribution of knowledge (by scientists) and technologies (for advanced scientific devices) are two major problems which are common in every scientific field. Emerging grid technology provides a new computational platform for a variety of scientific problems. MEG (magnetoencephalography) data analysis is an important research topic in brain science. In MEG data analysis, the lack of computational power and the distribution of knowledge and technologies lead to inefficient diagnoses and analyses of brain functions. In this research, we have built an MEG data analysis system on a grid environment, simulated with Globus grid toolkit, which is an implementation of grid technology. In this system, we attempted to reduce analysis time and to seamlessly integrate computational resources. Our evaluation results show that the system was highly efficient in reducing analysis time. Furthermore, we succeeded in integrating computers on our grid environment to seamlessly transfer data between each other. We believe that grid technology is effective and promising for real-life medical and scientific problems.

1. Introduction

Magnetoencephalography is a sophisticated measurement device which dynamically captures neural activities inside the brain. Electrical activity in the brain is caused by movements of ions inside and outside cellular membranes[1)], and these electrical currents give rise to magnetic fields. MEG has the ability to measure the strength of the magnetic field generated by these neural activities. The MEG measurement is performed at multiple points around a head.

The most prominent features of MEG over traditional measurement devices like electroencephalography (EEG) and electrocorticography (ECoG) are non-invasiveness and a high degree of measurement accuracy. Traditionally, these two features had a trade-off relationship. In fact, the measurement with ECoG is intrusive in spite of achieving a high degree of measurement accuracy. This point is one reason why MEG is promising in brain science.

However, the small number of available MEGs poses a serious problem. Presently, only tens of MEGs exist in the world in spite of their prominent capability. One of the main reasons for this lack of worldwide MEGs is the high cost for purchase and maintenance. Unfortunately, many brain scientists need to go to a remote hospital or scientific institute where an MEG is located.

The amount of MEG data poses another serious problem. In the case of a one hour measurement with a 64-sensor MEG, the data amount reaches 0.9 GB. For clinical purposes, an MEG measurement longer than an hour is often performed. As it is difficult for even a specialist to understand the meaning of complex MEG data at a glance, a variety of signal processing techniques are used for MEG data analysis. Such techniques are categorized as a type of computationally-intensive problem. A great amount of time is required to adequately analyze a large amount of MEG data with a computationally-intensive signal processing technique, despite the fact that early treatment is effective for brain diseases.

Consequently, in order to support advances in brain science, we have to solve the following two issues:

(1) Seamless integration of computational resources
(2) Fast analysis of MEG data

Today, a diversity of resources are connected to the Internet. However, in order to use these resources, we need to identify ourselves several times with credentials like passwords. This situation prevents seamless transfer of data, which results in inefficient diagnoses and analyses. In this research, we aim to realize seamless inte-

†1 Department of Information Systems Engineering, Graduate School of Engineering, Osaka University
†2 Osaka Jonan Women's University
†3 Division of Interdisciplinary Image Analysis, Osaka University Graduate School of Medicine
†4 Graduate School of Information Science, Nara Institute of Science and Technology
†5 Cybermedia Center, Osaka University

gration of the authentication process and the seamless transfer of data with a grid technology.

Furthermore, our aim is to gain enough computational power to analyze a large amount of MEG data with our analysis method, a *wavelet cross-correlation analysis*. In order to make full use of computational power, efficient parallel processing for wavelet cross-correlation analysis is also considered.

This paper is organized as follows. In Section 2, wavelet cross-correlation analysis is explained. In Section 3, the MEG data analysis system which has been built with grid technology is described. In Section 4, the system is evaluated. Section 5 concludes this paper.

2. Wavelet Cross-correlation Analysis

In this research, wavelet cross-correlation analysis[2] has been adopted as a signal processing technique. This analysis has the capability of investigating frequency components contained in MEG data without losing original time information. This feature is suitable for the analysis of non-stationary data like MEG data.

Wavelet cross-correlation analysis is composed of two types of analyses. One is wavelet analysis, and the other is cross-correlation analysis. First, wavelet analysis investigates frequency components contained in MEG data. Next, cross-correlation analysis quantifies the result of wavelet analysis.

Figure 1 illustrates the overview of wavelet cross-correlation analysis. Wavelet analysis is performed for MEG data acquired from a single sensor. The wavelet analysis is performed based on the following Eqs. (1)–(3).

$$Wf(a,b) = \int_{-\infty}^{\infty} \overline{g_{a,b}(t)} f(t)dt \qquad (1)$$

$$g_{a,b}(t) = \frac{1}{\sqrt{a}} g\left(\frac{t-b}{a}\right) \qquad (2)$$

$$g(t) = e^{-\frac{t^2}{2}}\left(e^{j\Omega t} - e^{-\frac{\Omega^2}{2}}\right), \Omega = 2\pi \qquad (3)$$

The function $f(t)$ represents MEG data acquired from a single sensor. The function $g(t)$ is a Gaussian basis which is a kind of mother wavelets (analyzing wavelets). The upper images in Fig. 1 are the visualized results of wavelet analysis. These images allow doctors to intuitively understand a frequency distribution map of the corresponding MEG data.

Cross-correlation analysis is performed for each pair of the results of wavelet analysis. This analysis is performed based on the Eq. (4).

$$WC_{1,2}(a,\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \overline{Wf_1(b,a)} Wf_2(b+\tau,a)db \qquad (4)$$

Wf_1 and Wf_2 are the results of wavelet analysis. This analysis provides information on the correlation of frequency distribution between the corresponding two sensors. The bottom image in Fig. 1 is the visualized result of cross-correlation analysis. This image indicates that a brain signal with a frequency f' component

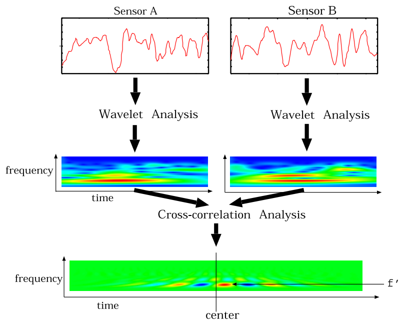

Fig. 1 Overview of wavelet cross-correlation analysis.

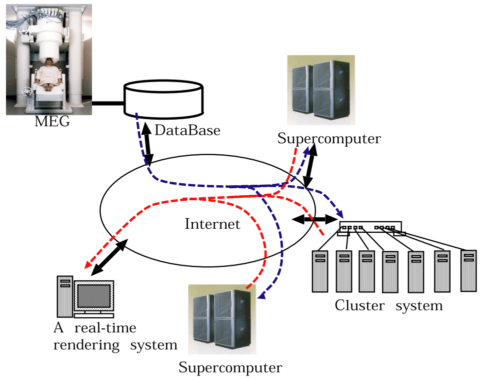

Fig. 2 The concept of MEG data analysis (The figure of MEG is cited from http://www.ctf.com.).

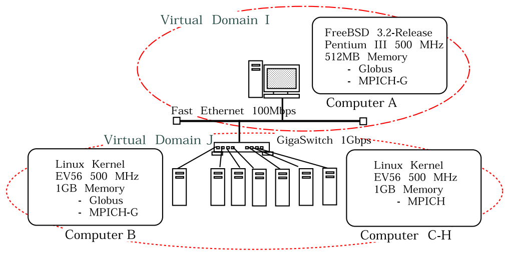

Fig. 3 Physical computational environment.

is detected in sensor B earlier than in sensor A. This is the reason why this analysis can localize the source of a brain signal.

However, the space of wavelet cross-correlation analysis becomes large because of the existence of many program parameters and the large amount of data. Accordingly, the wavelet cross-correlation analysis is used as follows. First, doctors select the temporal region of the MEG data which contains a brain signal of interest to them. Next, using wavelet analysis, doctors investigate if the MEG data contains the frequency component of interest. Next, using cross-correlation analysis, doctors investigate the correlations in frequency components contained in MEG data among all pairs of sensors. Finally, they investigate the time-lag of the signal's emergence and localize the source of the signal.

The computational workload for wavelet cross-correlation analysis increases intensively with the increase of MEG sensors. The use of a 64-sensor MEG is assumed in this research. In this case, then, cross-correlation analysis should be ideally performed 2,016 times. Recently, the number of MEG sensors tends to increase. The MEG with more than 200 sensors exists. In the case of a 200-sensor MEG, a cross-correlation analysis should be ideally performed 19,900 times. In this research, we distribute the computational workload for this analysis in order to reduce the analysis time.

3. System

3.1 System Concept

Figure 2 illustrates the concept of the MEG data analysis system. This concept envisages the seamless integration of an MEG (brain database), high performance computers such as supercomputers and cluster systems, and a real-time rendering system on the Internet.

Our concept can integrate not only physical resources but also processes of MEG data analysis. MEG data analysis is composed of a data acquisition process, a data analysis process and an implementation process of analysis results. Presently, these processes are rarely performed in the same place. This situation leads to difficulties in obtaining the final results of analysis and diagnosis within a realistically efficient amount of time. This inefficiency may be fundamentally caused by the fact that MEG data is often used with a magnetic storage like an optical disk, hand to hand. In the concept, the seamless sharing of MEG data is achieved through the three processes of acquisition, analysis and implementation of analysis results.

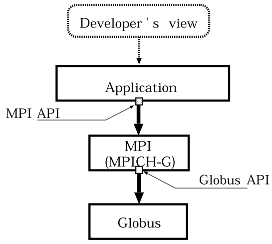

Fig. 4 The grid architecture in our system.

Furthermore, our concept has the potential to dramatically reduce the analysis of MEG data. We plan to integrate high-performance computers which are in low load status on the Internet. By making full use of such computers, we hope to gain enough computational power to analyze a large amount of MEG data.

In sum, our concept aims to achieve a seamless and efficient data mining method for MEG data analysis on the Internet. Our concept would dramatically improve the efficiency of MEG data analysis.

3.2 Implementation

In this paper, we describe the MEG data analysis system built on our concept. The system has been realized on the LAN at Cybermedia Center, in Osaka University. Our physical computational environment is shown in **Fig. 3**. In an initial stage of development, the number of computational resources available for our research was limited due to an administrative problem; hence, the simulation of a grid environment has been performed on eight computers.

The connection of MEG to a public network is, at present, difficult to realize in Japan. This difficulty involves a security problem regarding the patients' privacy. Also, certain Japanese conventions in the medical area may prevent this work. Thus, we have assumed that MEG data are already stored in the visualization computer.

A cluster system which is composed of seven computers B–H in Fig. 3. The cluster system has been used as a computational engine. In this research, the visualization computer and the cluster system are assumed to be in different organizations such as clinics, hospitals, and scientific institutions.

In the system, the Globus grid toolkit[3] has been adopted as a building block for constructing the grid environment. First, in **Fig. 4**, a grid architecture in our system is presented. Next, a task distribution mechanism of wavelet cross-correlation analysis on the grid architecture is explained.

3.2.1 Grid Architecture

The Globus grid toolkit is the middleware that embodies the concept of "grid" computing[3]. Generally, a grid refers to recent, advanced technology that allows scientists to easily integrate a diversity of computational resources geographically distributed on the Internet. These resources include supercomputers and scientific instruments such as the MEG. In other words, we can easily gain far more computational power through the use of grid technology than through the use of traditional parallel computing technology. Also, the remote-control of scientific instruments is possible using a grid.

In order to allow end-users without special knowledge on the Internet technology to easily use globus, globus provides multiple complex services essential for widely distributed parallel computing on the Internet. Examples of these services include communication, security, and information management, to name a few. These services are extremely difficult to implement in the Internet. This difficulty is easy to understand if you imagine there are many different administrative policies on the Internet. However, globus has a design strategy that offers users uniform and integrated methods to access these resources in an API (Application Programming Interface)-based manner. Accordingly, we can build a parallelized application over organizations such as clinics, hospitals and scientific institutions on the Internet with relative ease.

In the system, to develop MEG data analysis easily, MPICH-G[4] has been adopted. MPICH-G is a grid-enabled version of MPICH, which is an implementation of MPI (Message Passing Interface) specification for a standard library of message passing that was originally defined by the MPI Forum[5]. MPICH-G enables users to easily access globus services via MPI APIs to develop grid applications.

Figure 4 illustrates the grid architecture which our system has adopted. This grid architecture enables application developers to develop grid applications without globus APIs, by using familiar MPI APIs. Furthermore, the best feature of this architecture is that application developers can reuse existing MPI programs on grid environments without modifying them.

To simulate a grid environment on the LAN, we have installed the Globus grid toolkit and the MPICH-G into computer A and computer B. For computer B, MPICH also has been installed. For computers C–H, only MPICH has been installed. Under this environment, computer A and the group of computers B–H can be placed in different domains.

In short, we have built an MEG data analysis system based on this grid architecture. In the next section, we will describe a task distribution mechanism pertaining to wavelet cross-correlation analysis on our grid environment.

3.2.2 Task Distribution Mechanism

In order to consider an efficient parallel processing method for the wavelet cross-correlation analysis, investigating how the wavelet cross-correlation analysis is used and the most efficient way to use it is necessary. After understanding the computing requirements regarding wavelet cross-correlation analysis, we need to design a task distribution mechanism so as to explore the space of wavelet cross-correlation analysis.

In wavelet cross-correlation analysis, the wavelet analysis needs to be repeatedly performed. In general, brain signals that are of interest to doctors, such as epileptic wave signals, last for 4 to 10 seconds and such signals appear many times. Furthermore, the emergence of such signals is non-stationary. Doctors need to find the data regions which contain such signals in a large MEG data space. To do this, doctors often perform wavelet analysis many times, finding arbitrary temporal regions of MEG data. This work is time and labor - consuming. Thus, quickly finding the pertinent brain signals is an important key to success in diagnosing various brain diseases.

Doctors and scientists use this wavelet analysis to find brain signals by focusing on two different regions in MEG data space. The two regions are as follows (**Fig. 5**).

(1) The region along the temporal axis
(2) The region along the spatial axis

The first region is used where the source of the brain signal is, to some degree, localized by a medical test performed in advance. In this case, the doctors' main concern is how the frequency components in the MEG data vary over time, since doctors want to precisely localize

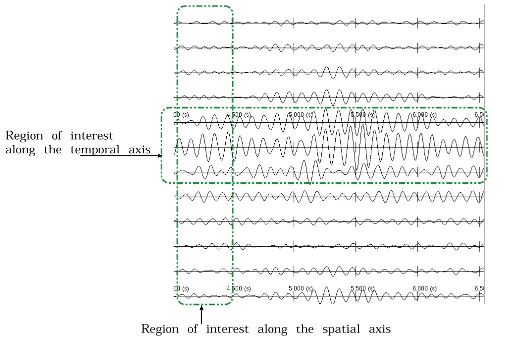

Fig. 5 Two regions of interest to doctors in the MEG data space.

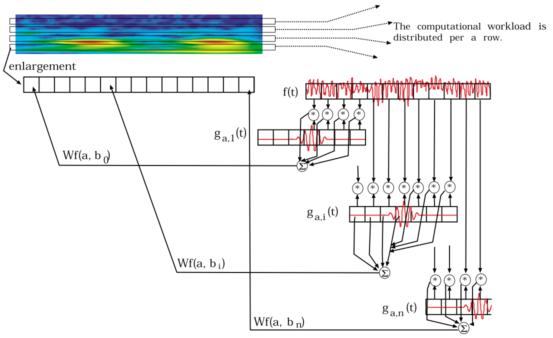

Fig. 6 A parallel processing method which focuses on the region along the temporal axis.

the source of brain signals by comparing them among sensors. Hence, long MEG data such as 60 second data, which are acquired from only a few sensors around the source are analyzed.

The second region is used when doctors have no information on the source of the brain signal which interests them. In this case, in order to observe which MEG data contains the frequency components of the brain signal, doctors need to investigate the MEG data from all sensors. Here, short MEG data ranging from 1 to 4 seconds acquired from all sensors are analyzed.

In our system, two types of parallel processing methods for wavelet analysis have been implemented on our grid environment, based on the previously described doctors' needs. For problems focusing on the region along the temporal axis in the MEG data space, the workload of wavelet analysis itself is distributed to computers B–H. In other words, the Eq. (1) is computed in parallel.

Figure 6 shows the parallel processing method which focuses on the region along the temporal axis. This wavelet analysis result provides a 2-dimensional time-frequency map for a single MEG data. Each row shows the result

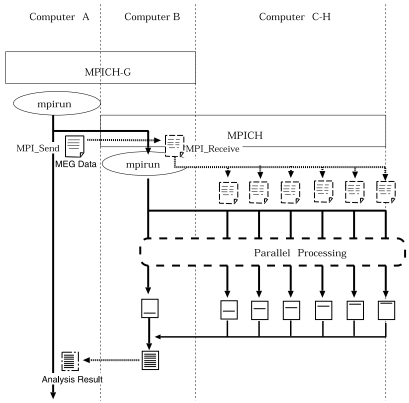

Fig. 7 Overview of task distribution mechanism.

of a convolution operation between MEG data $f(t)$ and a Gaussian basis $g_{a,b}(t)$ with a certain frequency response. The convolution operation for each row can be easily performed in parallel. Our system has utilized this computational locality to distribute the workload for a single wavelet analysis.

For the region along the spatial axis, the distribution method is simple. In this case, as MEG data acquired from all sensors are analyzed, we distribute these multiple tasks of wavelet analysis to computers B–H. In our system, 64 data at most are simultaneously investigated with wavelet analysis.

After wavelet analysis, cross-correlation analysis is performed for each pair of the wavelet analysis results. When a 64-sensor MEG is used, the number of such pairs reaches 2,016. In our system, this computational workload for 2,016 pairs of MEG data at most is distributed between computers B–H.

These three different parallel processing methods for wavelet cross-correlation analysis have been realized in an identical task distribution mechanism. **Figure 7** diagrams the task distribution mechanism in our MEG data analysis system.

The idea is simple and easy to understand. First, MPICH-G processes are generated by a user on computer A. Next, the user is required to identify him or herself with his or her passphrase to the Grid Security Infrastructure (GSI)[6),7)], which the Globus grid toolkit provides. We have utilized GSI to realize seamless integration of the authentication process in making simultaneous use of multiple resources. After the authentication, the MEG data on computer A is transferred to computer B with

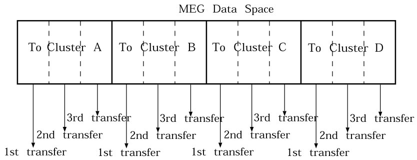

Fig. 8 Data parallelism among MPICH-G processes.

MPI_Send() and MPI_Recv(), and then computer B distributes the MEG data to computers C–H. The MEG data that will be wavelet-transformed is transferred from computer A to computer B. This transfer of data is performed using globus services. After receiving MEG data, a MPICH-G process on computer B invokes MPICH processes on computers B–H. These MPICH processes perform a computation of wavelet and cross-correlation analysis in harness, according to the situation. Each analysis which results on computer B–H is gathered with an MPI_Gather() and is then transferred to the visualization machine, or to computer A. This task distribution mechanism is applicable to the computational environment which is composed of a visualization machine and multiple high-performance computers such as our cluster system, by simply creating different groups of MPICH processes on such computers in the same way.

At the time of writing this paper, in wavelet cross-correlation analysis, the task distribution mechanism based on MPICH-G has been mainly used for data parallelism. The MEG data space is divided into the same number of regions as the number of available clusters along the temporal axis and each region is analyzed on the corresponding cluster system.

Figure 8 shows that four clusters of computers are assumed to be used. The data transfer to the corresponding cluster system is performed at one time. Only MEG data parts which is necessary for the one-time wavelet analysis is transfered. After computation, the results of the analysis are transferred back to the visualization computer. This process is repeated until the analysis of the assigned data is finished.

In addition, we plan to perform parameter parallelism among MPICH-G processes in the near future. Currently, for example, wavelet cross-correlation analysis changes parameter "a" in the Eq. (2) in 51 different ways. This set of parameters "a" has been optimized for brain data analysis in advanced research and controls the frequency response of mother wavelet $g(t)$ from 1.0 to 50.0 Hz. Nonetheless, an analysis with more detailed resolution regarding parameter "a" is necessary for advanced brain science.

Brain data analysis requires the detailed investigation of analysis space. Data space and parameter space make analysis space large. In the future, we will also simultaneously perform not only wavelet cross-correlation analysis but also multiple brain data analysis methods like ICA (Independent Component Analysis) using grid technology.

4. Evaluation

As described in the Section 2, wavelet analysis is performed for analysis of frequency changes in time scale. On the other hand, cross-correlation analysis is used for the quantification of the wavelet analysis results.

Based on the concept introduced in the section 3, we will implement the MEG data analysis system on an actual wide-area network like the Internet. **Figure 9** illustrates how a doctor can analyze MEG data on the MEG data analysis system. In the blueprint, a doctor can obtain the result of wavelet analysis in a user-interactive manner, by using a cluster of high performance computers on the fast network which connects his or her visualization computer with the cluster. After that, if the doctor would like to know the results in detail, the doctor can use the cross-correlation analysis to quantify the result in a batch manner. For the cross-correlation analysis, the use of a number of high performance computers including a supercomputer is assumed in the blueprint.

In this section, we evaluate the MEG data analysis system from a comprehensive standpoint, based on the viewpoint as described now. For the purpose, the following two items are investigated:

(1) Efficiency of the parallel computing using grid technology.

(2) The data transfer problems and issues.

Finally, we evaluate the effectiveness of our system with grid technology from the viewpoint of practical diagnosis.

4.1 A Performance Evaluation for Wavelet Analysis

In the MEG data analysis system, two types of parallel processing methods for wavelet analysis have been implemented. These parallel processing methods are different in terms of the number of MEG data that is analyzed. However, the purposes are the same, namely, the

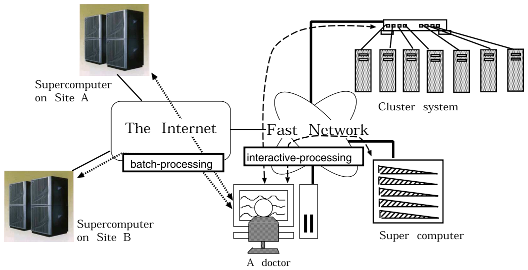

Fig. 9 The blueprint for practical diagnosis: a doctor can simultaneously perform wavelet analysis and cross-correlation analysis in a batch-processing manner and in a user-interactive manner, respectively.

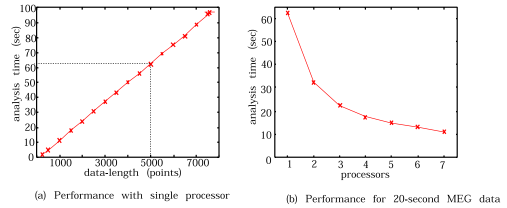

Fig. 10 The effect of our parallel processing method which focuses on the region along the temporal axis.

quick provision of the analysis results, so that doctors can analyze frequency changes in time scale in a user-interactive manner.

4.1.1 A Parallel Processing Method Which Focuses on the Region along Temporal Axis

Figure 10 shows the analysis time which varies with the increase of MEG data length. The data length is defined as the number of sampling points in the figure. Graph (a) shows the analysis time measured in the traditional processing manner on computer B. The results show that analysis time increases in proportion to data length.

Graph (b) shows the effects of our parallel processing method which focuses on the region along the temporal axis in the MEG data space. The analysis time in the graph was measured when wavelet analysis was performed for MEG data with 5,000 sampling points (20-second MEG data) acquired from two sensors. The analysis time is the time from submitting a computational request to computer A to obtaining the results on computer A from computers B–H. The graph shows that the analysis time decreases with the increase of processors. In particular, the analysis time was reduced to 11.25 seconds when we used seven computers

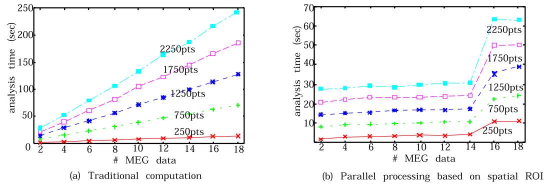

Fig. 11 The effect of our parallel processing method which focuses on the region along the spatial axis.

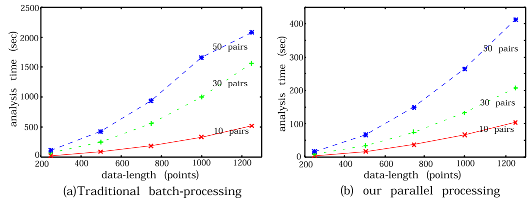

Fig. 12 The effect of our parallel processing method for cross-correlation analysis.

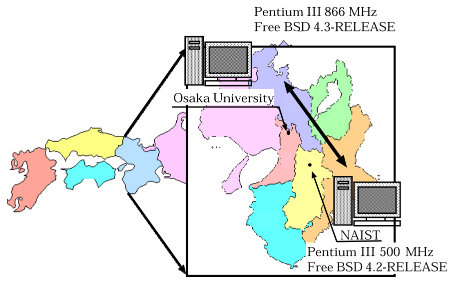

Fig. 13 The data transfer experiment over the network between NAIST and Osaka University.

B–H from computer A for this parallel processing method, by using the grid mechanism shown in the Fig. 4 and the Fig. 7, compared with the more than 60 seconds necessary when only a single computer B was used. These results mean that an approximately 5.5-fold analysis was achieved with 7 computers as remote computational resources and that our system can realize near real-time analysis.

4.1.2 A Parallel Processing Method Which Focuses on the Region along the Spatial Axis

A parallel processing method focusing on the region along the spatial axis in the MEG data space distributes the workload for multiple tasks of wavelet analysis. In order to evaluate the effect of this processing method, we compared the analysis time in the traditional batch-processing way with that in our parallel processing method.

The graph (a) in **Fig. 11** shows the result of analysis time in the traditional computational method. The analysis time in the graph was the time required from submitting a computational request to computer A to obtaining the results. The horizontal axis indicates the number of MEG data which should be simultaneously analyzed. The graph shows that the analysis time doubles regardless of data length when the number of MEG data simultaneously analyzed doubles. It takes 56.45 seconds to analyze a set of MEG data with 1,250 sampling points acquired from eight sensors. From this fact, we can guess that it would take approximately 450 seconds (7.5 minutes) if 5-second MEG data acquired from sixty-four sensors are simultaneously analyzed.

Graph (b) shows that our parallel processing method has the capability of analyzing multiple MEG data in almost the same time as the analysis time for a set of MEG data from two sensors. The sudden jump-up when the number of MEG data reaches 16 can be explained by the implementational reason that the computational workload is distributed to seven computers per pair of MEG data. The two graphs (a) and (b) show that we can obtain the analysis results of fourteen 5-second MEG data in 17.08 seconds, while 99.39 seconds is necessary to analyze the MEG data with a single computer.

4.2 A Performance Evaluation for Cross-correlation Analysis

Cross-correlation analysis, performed after wavelet analysis is computationally-intensive. In order to improve efficiency in the entire wavelet cross-correlation analysis, quick provision of cross-correlation analysis is the key to success. In this research, multiple tasks of cross-correlation analysis are distributed to multiple computers.

The left graph (a) in **Fig. 12** shows analysis time using traditional batch processing on a single processor basis. The analysis time increases on the order of n^2 when the length of MEG data is n. Also, as expected, the analysis time increases in proportion to the number of MEG data pairs to be analyzed.

In contrast, the right graph (b) shows analysis time from submitting a computational request to computer A to obtaining the result from computers B–H when our parallel processing method was used on seven computers B–H as remote computational resources. Here, we obtained the analysis results of fifty cross-correlation analyses for 250 sampling points, or 1-second MEG data in 413.86 seconds (approximately 7 minutes), while 2,078.42 seconds are required in the traditional batch processing manner. We gained approximately a 5-fold quicker analysis in comparison with traditional batch-processing. It would be possible to apply our method to a wide-area network environment.

4.3 Data Transfer Issues

In the preceding Sections 4.1 and 4.2, the performance of the MEG data analysis system which consists of eight computers using the grid architecture as shown in Fig. 4 was evaluated. The result showed that parallel computing using the remote cluster system is highly efficient. However, before actually managing the system on a wide-area network, we need to consider several issues of data transfer.

Two major problems need to be considered in building the MEG data analysis system on an actual wide-area network such as the Internet:

(1) network parameters

(2) communication topology

Network parameters such as latency and bandwidth become important factors for efficient data analysis. In general, latency and bandwidth in the wide-area network are larger and narrower than in the LAN environment. Unfortunately, these parameters are also difficult to predict.

Currently, our system transfers MEG data to the cluster system in a peer-to-peer manner, using a mechanism shown in Fig. 8. The amount of data which the system transfers to and from the cluster system at one time is small.

For example, for the parallel processing method which focuses on the region along the temporal axis, the amount of input data reaches nearly 120 KB at most when a set of MEG data with 15,000 sampling points (60-second MEG data) acquired from two sensors is transferred. On the other hand, for the parallel processing method which focuses on the region along the spatial axis, the amount of input data reaches nearly 320 KB when a set of MEG data with 1,250 sampling points acquired from 64 sensors is transferred.

The amount of output data which is transferred from the cluster system varies widely according to the type of parallel processing and to the doctor's request for analysis. If a set of MEG data with 1,250 sampling points acquired from eight sensors are analyzed with the wavelet analysis on the cluster system at one time as performed in the measurement test, and doctors want to see all of the results, then the total amount of output data reaches nearly 510 KB. However, if a doctor wants to see only the final result of the wavelet cross-correlation analysis for a set of MEG data acquired from two sensors, the amount of data is just 408 B. This number obviously increases with the doctors' demand for data. However, in general and under realistic conditions, only a 510 KB data output at most is required.

Approximately 510 KB of data transfer imposes little overhead on the system's performance under a fast LAN environment. In our evaluation test, the 510 KB data transfer was accomplished within nearly 60.0 milliseconds on the LAN environment. This data transfer time was much lower in comparison with the computation time of 56.45 seconds as shown in the previous Section 4.1.2.

In addition, we estimated the possibility of building the MEG data analysis system on an actual wide-area network by measuring the data transfer timebetween the computer on the LAN at Osaka University and the computer on the LAN at the Nara Institute of Science and Technology (NAIST) (**Fig. 13**). The slant distance between NAIST and Osaka University reaches approximately 40 Kilometers. For the test, a Pentium III 866 MHz computer at NAIST and a Pentium III 500 MHz computer at Osaka University were used. These two computers have been deployed with the grid architecture shown in the Fig. 4. In the measurement test, two MPICH-G processes exchange a single 520 KB message among two institutions one after the other. More specifically, the MPICH-G process on the LAN at the Cybermedia Center at Osaka University transfers the message with MPI_Send. The other process receives the message and then transfer the same message back immediately with MPI_Recv. This test was repeated 100 times.

The result of the transfer test showed that a round-trip of 520 KB, which is a typical model of the data communication required for the wavelet cross-correlation analysis, takes an average of 3.2 seconds. Also, the result indicates that the round-trip time of 520 KB data is small in comparison with the computation time of 56.45 seconds. However, we need to note that the transfer time varied over time and that the maximum time of each message data was 8.7 seconds. Although a long transfer time rarely occurred, a long transfer time may result in the lowering of a system's performance. For this problem, we will integrate the information service provided by the Globus grid toolkit into the system to predict the status of the network.

Finally, a communication topology needs to be considered when building the system into the Internet. As illustrated in the Fig. 9, we will combine multiple high-performance computers with the visualization computer. This communication form affects the high workload on the visualization computer because the visualization system needs to receive the result data from multiple high-performance computers. For this problem, we will introduce a data distribution method which takes the status of the network into consideration by integrating the information service provided by the Globus grid toolkit.

4.4 The Effectiveness of Grid Technology

So far, we have evaluated the performance of grid computing on a grid simulated on the LAN environment. We have also considered data transfer problems and related issues. In this section, we evaluate the effectiveness of the system from the viewpoint of a practical analysis based on the evaluation results.

The evaluation in the preceding sections has shown that the system achieving an approximately 5-fold quicker analysis on the grid environment in all cases of parallel processing methods introduced in this paper. This improvement in performance is small due to the number of computational resources available for this measurement. Importantly, however, in the results of the measurement, analysis time is slashed by approximately one-fifth through the use of only 7 computers as remote computational resources. In addition, the system has the potential to gain higher performance if we can make use of more computational resources.

MEG data analysis has the characteristic that the amount of data to be transferred as input and output is small and the amount of computation time is large. This characteristic allows doctors to analyze their patients' data with high efficiency. For example, a doctor can perform the cross-correlation analysis of a patients' MEG data in a batch manner, while the doctor can perform the wavelet analysis of other patients' MEG data in a user-interactive manner. In short, with multiple high-performance computers integrated with grid technology, the doctor can efficiently analyze his or her patients' MEG data without being aware of the existence of remote computers.

For wavelet analysis, if four more clusters of computers such as computers B–H connected to the visualization computer with the fast network are available, we would obtain the results of wavelet analysis for a set of 5-second MEG data acquired from 70 sensors within approximately 20 seconds. Also, it is possible to perform the wavelet analysis of the entire patient's MEG data in a batch manner and then observe all of the results after the computation. For example, it would take 14,400 seconds (4 hours) for a doctor to analyze a set of 1-hour MEG data acquired from 70 sensors in a batch-processing manner, compared with 71,560 seconds (20 days) in a traditional computing way.

Likewise, for cross-correlation analysis, if nineteen more clusters like our cluster system are available on the Internet with the Globus grid toolkit, approximately 420 seconds (7 minutes) would be required to perform 2,000 pairs of cross-correlation analysis for 5-second MEG data. This computation time of 420 seconds is much faster, when compared with the computation time of approximately 83,140 seconds (23 hours) performed in a traditional computational way.

In the current stage of development, it is impossible to realize the practical diagnosis and analysis with the wavelet cross-correlation analysis due to the limited number of available computers. Again, however, grid technology has been developed in order to integrate a diversity of heterogeneous high-performance computers which are in low load status on the Internet. Therefore, we can realize our blueprint in the near future.

5. Conclusion

In this paper, the application of emerging grid technology to MEG data analysis was described. Specifically, a MEG data analysis system has been built on a grid environment which was simulated on the LAN at the Cybermedia Center at Osaka University. The Globus grid toolkit has been utilized as a grid technology in the system. An efficient parallel processing method for wavelet cross-correlation analysis was also considered.

Our evaluation shows that our system is highly efficient in reducing analysis time for wavelet cross-correlation analysis. In our simulated grid environment, we also succeeded in integrating computational resources to seamlessly transfer MEG data.

Through building the MEG data analysis system, we arrived at the conviction that grid technology is effective and promising not only for MEG data analysis, but also for other scientific problems. We are planning to continue the research pertaining to grid technology by building an actual application system. We believe that grid technology is essential in the further development of various fields of science and medicine and we hope that this research will become a forerunner for these fields.

Acknowledgments This work was supported in part by Research for the Future Program of the Japan Society for the Promotion of Science (JSPS) under the Project titled "Integrated Network Architecture for Advanced Multimedia Application Systems" (JSPS-RFTF97R16301). And part of the experiment has been conducted over Japan Gigabit Network. And this research includes software developed by and/or derived from the Globus Project (http://www.globus.org/).

References

1) Sato, S.: Magnetoencephalography, *Advances in Neurology*, Vol.54, Raven Press, New York (1990).
2) Li, H. and Nozaki, T.: Wavelet Cross-Correlation Analysis (Analysis of Vortical Structurs in a Turbulent Plane Jet), *JSME Int. J., Ser. B*, Vol.40, No.1, pp.58–66 (1997).
3) Foster, I. and Kesselman, C.: The Globus Project: A Status Report, *Proc.IPPS/SPDP'98 Heterogeneous Computing Workshop*, pp.4–18 (1998).
4) Foster, I. and Karonis, N.: A Grid-Enabled MPI: Message Passing in Heterogeneous Distributed Computing Systems, *Proc. SC'98* (1998).
5) Gropp, W., Lusk, E., Doss, N. and Skjellum, A.: A High-Performance, Portable Implementation of the MPI Message Passing Interface Standard, *Parallel Computing*, Vol.22, No.6, pp.789–828 (1996).
6) Butler, R., Engert, D., Foster, I., Kesselman, C., Tuecke, S., Volmer, J. and Welch, V.: A National-Scale Authentication Infrastructure, *IEEE Computer*, Vol.33, No.12, pp.60–66 (2000).
7) Foster, I., Kesselman, C., Tsudik, G. and Tuecke, S.: A Security Architecture for Computational Grids, *Proc. 5th ACM Conference on Computer and Communications Security Con-*

ference, pp.83–92 (1998).

(Received May 8, 2001)
(Accepted October 16, 2001)

Susumu Date received the B.S. and M.S. degrees in Information and Computer Sciences from Osaka University, Osaka, Japan, in 1997 and 2000, respectively. He is a Ph.D. candidate at Osaka University.

Yuko Mizuno-Matsumoto received the M.D. degree from Shiga University of Medical Science, Japan, in 1991, and the Ph.D. degree from Osaka University Medical School, Japan, in 1996. Since 2000, she has been an Associate Professor of Osaka Jonan Women's University.

Youki Kadobayashi received the B.E., M.S., and Ph.D. degrees in Computer Science from Osaka University, Osaka, Japan, in 1992, 1994, and 1997, respectively. Since 2000, he has been an Associate Professor of Nara Institute of Science and Technology.

Shinji Shimojo received the M.E. and D.E. degrees in Osaka University, Osaka, Japan, in 1983 and 1986, respectively. Since 1998, he has been a Professor with the Cybermedia Center.