AO 120 (Rev. 08/10)

TO:	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the District of Delaware on the following

☐ Trademarks or ☑ Patents. (☐ the patent action involves 35 U.S.C. § 292.):

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
	10/28/2019	for the District of Delaware

PLAINTIFF	DEFENDANT
ADVANCED CLUSTER SYSTEMS, INC.	NVIDIA CORPORATION

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1	US 8,082,289 B2	12/20/2011	Advanced Cluster Systems, Inc.
2	US 8,140,612 B2	3/20/2012	Advanced Cluster Systems, Inc.
3	US 8,676,877 B2	3/18/2014	Advanced Cluster Systems, Inc.
4	US 10,333,768 B2	6/25/2019	Advanced Cluster Systems, Inc.
5			

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT
Dismissed - see attached

CLERK	(BY) DEPUTY CLERK	DATE
John A. Cerino	/s/ Mark D. Buckson, Jr.	4/5/2023

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

IN THE UNITED STATES DISTRICT COURT

FOR THE DISTRICT OF DELAWARE

ADVANCED CLUSTER SYSTEMS, INC.,	)	
	)	
Plaintiff,	)	
	)	
v.	)	C.A. No. 19-2032-MN-CJB
	)	
NVIDIA CORPORATION, NVIDIA SINGAPORE PTE. LTD., and NVIDIA INTERNATIONAL, INC.,	)	
	)	
Defendants.	)	

JOINT STIPULATION OF DISMISSAL OF ALL CLAIMS

WHEREAS Plaintiff Advanced Cluster Systems, Inc. ("ACS") and Defendants NVIDIA Corporation, NVIDIA Singapore PTE, Ltd., and NVIDIA International, Inc. (collectively, "NVIDIA") have resolved all claims concerning ACS' allegations concerning United States Patent Nos, 8,082,289; 8,140,612; 8,676,877; and 10,333,768 (collectively, "the Patents-in-Suit") and have signed a written settlement agreement, dated March 24, 2023.

NOW THEREFORE, the parties stipulate and agree, subject to the approval of the Court, that:

1. All claims asserted by ACS in the above-captioned action are hereby dismissed with prejudice.
2. Each party shall bear its own attorneys' fees, expenses, and costs.

Respectfully submitted,

/s/ Emily S. DiBenedetto
Karen E. Keller (No. 4489)
Nathan R. Hoeschen (No. 6232)
Emily S. DiBenedetto (No. 6779)
SHAW KELLER LLP
1105 North Market Street, 12th Floor
Wilmington, DE 19801
(302) 298-0700
kkeller@shawkeller.com
nhoeschen@shawkeller.com
edibenedetto@shawkeller.com
Attorneys for Plaintiff

/s/ Brian A. Biggs
Brian A. Biggs (No. 5591)
Stephanie E. O'Byrne (No. 4446)
DLA PIPER LLP (US)
1201 North Market Street, Suite 2100
Wilmington, DE 19801-1147
(302) 468-5700
brian.biggs@us.dlapiper.com
stephanie.obyrne@us.dlapiper.com
Attorneys for Defendants

Dated: April 5, 2023

IT IS SO ORDERED this 5th day of April 2023

Maryellen Noreika
The Honorable Maryellen Noreika
United States District Judge

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY.DOCKET NO./TITLE	REQUEST ID
14/181,112	02/14/2014	Zvi Tannenbaum	ZTANN.002P1C3	189446

Acknowledgement of Loss of Entitlement to Entity Status Discount

The entity status change request below filed through Private PAIR on 05/19/2023 has been accepted.

CERTIFICATIONS:

Change of Entity Status:

X Applicant changing to regular undiscounted fee status.

NOTE: Checking this box will be taken to be notification of loss of entitlement to small or micro entity status, as applicable.

This portion must be completed by the signatory or signatories making the entity status change in accordance with 37 CFR 1.4(d)(4).

Signature:	/Jordan M. Cox/
Name:	Jordan M. Cox
Registration Number:	70875

Trials@uspto.gov
571-272-7822

Paper 9
Date: May 5, 2021

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

NVIDIA CORP.,
Petitioner,

v.

ADVANCED CLUSTER SYSTEMS, INC.,
Patent Owner.

IPR2021-00020
Patent 10,333,768 B2

Before KARL D. EASTHOM, ARTHUR M. PESLAK, and SEAN P. O'HANLON, *Administrative Patent Judges.*

PESLAK, *Administrative Patent Judge.*

DECISION
Denying Institution of *Inter Partes* Review
35 U.S.C. § 314

I. INTRODUCTION

Petitioner, NVIDIA Corporation, filed a Petition (Paper1, "Pet.") requesting an *inter partes* review of claims 26, 27, 29, 35–37, and 39 (the "challenged claims") of U.S. Patent No. 10,333,768 B2 (Ex. 1101, "the '768 patent"). Patent Owner, Advanced Cluster Systems, Inc., timely filed a Preliminary Response (Paper 5, "Prelim. Resp."). With our permission, Petitioner filed a Reply to the Patent Owner Preliminary Response (Paper 7, "Pet. Reply"). Patent Owner filed a Sur-reply (Paper 8, "Sur-reply").

We have authority, acting on the designation of the Director, to determine whether to institute an *inter partes* review under 35 U.S.C. § 314(a). *See also* 37 C.F.R § 42.4(a) (2020) ("The Board institutes the trial on behalf of the Director."). Under 35 U.S.C. § 314(a), an *inter partes* review may not be instituted unless the information presented in the Petition shows "there is a reasonable likelihood that the petitioner would prevail with respect to at least 1 of the claims challenged in the petition." Taking into account the Petition, the arguments presented in the Preliminary Response, the additional briefing, as well as all supporting evidence, we deny institution of *inter partes* review because Petitioner fails to establish a reasonable likelihood that at least one of the challenged claims is unpatentable. In particular, we determine that Petitioner fails to show that the cited references disclose the "third node" as recited in independent claims 26, 29, and 35[1] ("the third node limitation").

[1] The relevant third node limitation is identical in claims 26 and 29. The third node limitation in claim 35 is substantially similar.

A. Related Matters

The parties state that the '768 patent is asserted in *Advanced Cluster Systems, Inc. v. NVIDIA Corporation*, Case No. 19-cv-2032-CFC (D. Del.). Pet. 3; Paper 3, 1. The '768 patent is also the subject of IPR2021-00019 filed by Petitioner on October 6, 2020. Paper 3, 1. Patent Owner also identifies IPR2020-01608, IPR2021-00075, and IPR2021-00108 as involving patents related to the '768 patent. *Id.*

B. Real Parties in Interest

Petitioner identifies itself as the only real party-in-interest. Pet. 3. Patent Owner identifies itself as the only real party-in-interest. Paper 3, 1.

C. The '768 Patent

The '768 patent is titled Cluster Computing. Ex. 1101, code (54). The '768 patent issued on June 25, 2019 from an application filed on February 14, 2014. *Id.* at codes (45), (22).

The '768 patent "relates to the field of cluster computing generally and to systems and methods for adding cluster computing functionality to a computer program, in particular." *Id.* at 1:14–17. Embodiments of the '768 patent include enabling a software package to benefit from a plurality of nodes in a cluster. *Id.* at 2:6–10. For example, "[o]ne embodiment allows a user to create applications, using a high-level language such as Mathematica, that are able to run on a computer cluster having supercomputer-like performance." *Id.* at 2:10–13. "One embodiment provides access to such high-performance computing through a Mathematica Front End, a command line interface, one or more high-level commands, or a programming language such as C or FORTRAN." *Id.* at 2:13–17.

Figure 1 of the '768 patent, reproduced below, "is a block diagram of one embodiment of a computer cluster 100 wherein computer systems 110,

120, 130 communicate with one another via a communications network 102." *Id.* at 4:63–66.

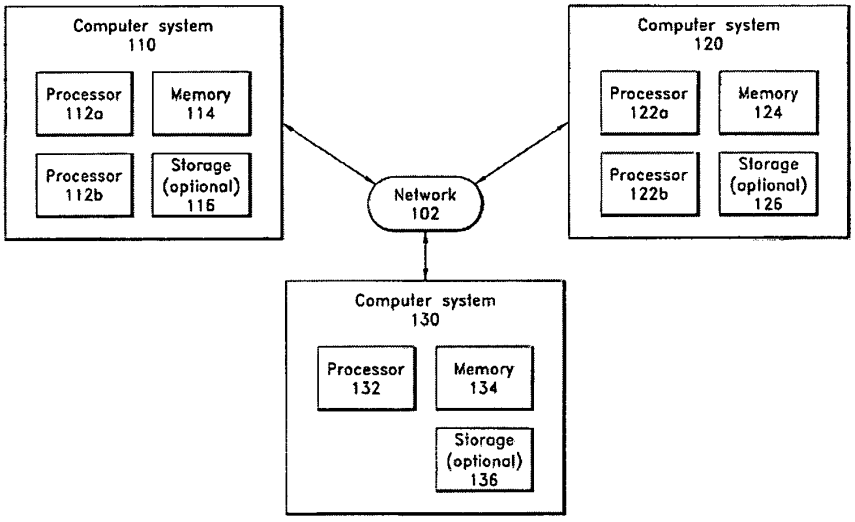


FIG. 1

Figure 1 of the '768 patent illustrates "a computer cluster 100 wherein computer systems 110, 120, 130 communicate with one another via a communications network 102." *Id.* at 4:63–66.

In the embodiment shown in Figure 1, computer cluster 100 comprises first, second, and third computers systems 110, 120, and 130. *Id.* at 6:53–55. Each of the computer systems 110, 120, 130 includes one or more processors (112a, 112b, 122a, 122b, 132), a memory device (114, 124, 134), and an optional storage device (116, 126, 136). *Id.* at 6:56–63. Each of the computer systems 110, 120, 130 includes a network interface (not shown) for connecting to communications network 102, which can include one or more of a LAN, a WAN, an intranet, a wireless network, and/or the Internet. *Id.* at 6:63–67.

Figure 2 of the '768 patent, reproduced below, "is a block diagram showing relationships among software modules running on one embodiment of a computer cluster 100." *Id.* at 5:16–18.

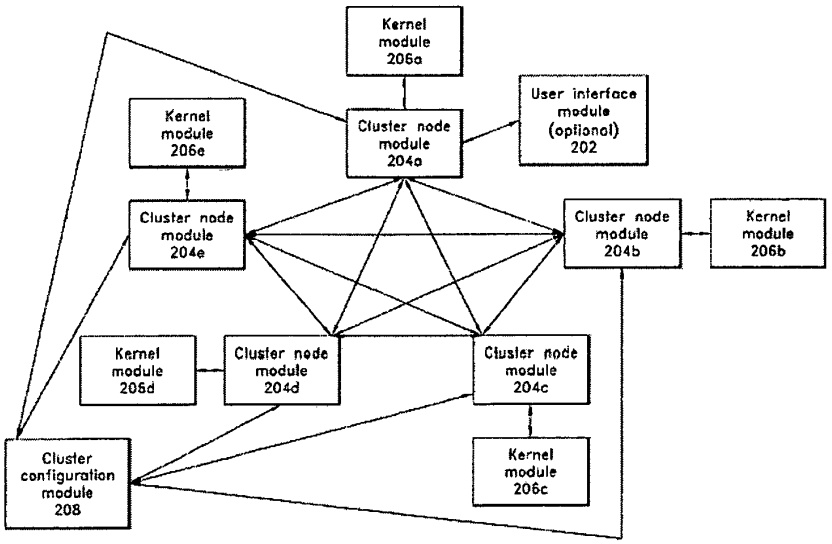


FIG. 2

Figure 2 of the '768 patent illustrates "relationships among software modules running on one embodiment of a computer cluster 100." *Id.* at 5:16–18.

In the embodiment shown in Figure 2, a cluster system 100 includes a user interface module 202 that is able to access a plurality of kernel modules 206a – 206e by communicating with a first cluster node module 204a. *Id.* at 11:15–18. Kernel modules 206a – 206e can reside in the memory of one or more computer systems on which they run. *Id.* at 11:23–25. Kernel modules 206a – 206e include single-threaded program code and are each associated with one of the processors 112a, 112b, 122a, 122b, 132. *Id.* at 11:30–32. In the embodiment shown in Figure 2, each of the kernel modules 206a – 206e is in communication with a single cluster node module 204a – 204e, respectively. *Id.* at 5:33–35. Cluster node module 204a is in communication with each of the other cluster node modules 204b – 204e. *Id.* at 11:21–23. Cluster configuration module 208 is stored on one or more of the computer systems 110, 120, 130 or on a remote computer system, and

can establish communication with the cluster node modules 204a – 204e. *Id.* at 11:32–36. In one embodiment, communication between the cluster configuration module 208 and the cluster node modules 204a – 204e initializes the cluster node modules 204a – 204e to provide cluster computing support for the computer cluster 100. *Id.* at 11:36–40. In one embodiment, the cluster node modules 204a – 204e provide a way for many kernel modules 206a – 206e such as, for example, Mathematica kernels, running on a computer cluster 100 to communicate with one another. *Id.* at 11:42–45.

D. Challenged Claims

Claims 26, 29, and 35 are the independent claims challenged in the Petition. Claim 26 is reproduced below:

> 26. A computer cluster comprising:
>
> a plurality of nodes, wherein one or more of the nodes are configured to receive:
>
> a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes; and
>
> an instruction from a user interface or a script;
>
> and
>
> a mechanism for the nodes to communicate results of mathematical expression evaluation with each other using asynchronous calls;
>
> wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions;
>
> wherein the plurality of nodes comprises:
>
> a first node comprising a first hardware processor configured to access a first memory comprising program code for a user

interface and program code for a first single - node kernel , the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and

a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of mathematical expression evaluation to a third node;

wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node;

wherein the first node is configured to return the result of the second mathematical expression evaluation to the user interface or the script;

wherein one or more of the nodes are configured to:

accept user instructions;

after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other; and

after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels.

Ex. 1101, 32:28–33:12.

E. Prior Art and Asserted Grounds

Petitioner asserts that claims 26, 27, 29, 35–37, and 39 would have been unpatentable on the following grounds:

Claim(s) Challenged	35 U.S.C. §	Reference(s)/Basis
26, 27, 29, 35, 37	§ 103(a)[2]	Schreiner1,[3] Schreiner2,[4] Schreiner3,[5] Maple Guide,[6] Distributed Maple Code,[7] SPARC IV Article,[8] AMD Article[9]
26, 27, 37	§ 103(a)	Schreiner1, Schreiner2, Schreiner3, Maple Guide, Distributed Maple Code, SPARC IV Article, AMD Article, MPI Standard[10]
36, 39	§ 103(a)	Schreiner1, Schreiner2, Schreiner3, Maple Guide, Distributed Maple Code, SPARC IV Article, AMD Article, Maple Reference,[11]

[2] The Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011) ("AIA"), included revisions to 35 U.S.C. § 103 that became effective after the effective filing date of the challenged patent. Therefore, we apply the pre-AIA version of 35 U.S.C. § 103.

[3] Schreiner et al., "Distributed Maple: parallel computer algebra in networked environments," Journal of Symbolic Computation, published in 2003 (Ex. 1108) ("Schreiner1").

[4] Schreiner, "Distributed Maple – User and Reference Manual (V 1.1.12)," July 6, 2001 (Ex. 1109) ("Schreiner2").

[5] Károly Bósa, Wolfgang Schreiner, "Task Logging, Rescheduling and Peer Checking in Distributed Maple," March 18, 2002 (Ex. 1010) ("Schreiner3").

[6] Waterloo Maple Inc., "Maple 5 Learning Guide," 1996, 1998 (Ex. 1111) ("Maple Guide").

[7] Exhibits 1112 – 1118.

[8] "SUN Debuts UltraSPARC IV," EE Times, October 15, 2003 (Ex. 1119) ("SPARC IV Article").

[9] "AMD to Unveil Dual-Core PC Chips," Los Angeles Times, May 31, 2005 (Ex. 1120) ("AMD Article").

[10] MPI: A Message-Passing Interface Standard, 1995 (Ex. 1121) ("MPI Standard").

[11] Michael Kofler, *Maple: An Introduction and Reference*, (1997) (Ex. 1131) ("Maple Reference")

Claim(s) Challenged	35 U.S.C. §	Reference(s)/Basis
		Howard[12]
36	§ 103(a)	Schreiner1, Schreiner2, Schreiner3, Maple Guide, Distributed Maple Code, SPARC IV Article, AMD Article, Maple Reference, Howard, FFTW Manual[13]
39	§ 103(a)	Schreiner1, Schreiner2, Schreiner3, Maple Guide, Distributed Maple Code, SPARC IV Article, AMD Article, PC Magazine1,[14] PC Magazine 2[15]
39	§ 103(a)	Schreiner1, Schreiner2, Schreiner3, Maple Guide, Distributed Maple Code, SPARC IV Article, AMD Article, Nayak[16]

II. ANALYSIS

A. Overview

A petition must show how the construed claims are unpatentable under the statutory grounds it identifies. 37 C.F.R. § 42.104(b)(4). Petitioner bears the burden of demonstrating a reasonable likelihood that it

---

[12] U.S. Publication No. 2003/0195938 A1published October 16, 2003 (Ex. 1128) ("Howard").
[13] Frigo and Johnson, *FFTW User's Manual*, March 16, 2003 (Ex. 1132) ("FFTW Manual").
[14] *Maple V*, PC Magazine, August 1992, pp. 419-425 (Ex. 1134) ("PC Magazine1")
[15] *Maple V Looks Better With Improved Graphics*, PC Magazine, July 1993, p. 50 (Ex. 1135) ("PC Magazine2").
[16] Anshuman Nayak, et al., *A Library based compiler to execute MATLAB Programs on a Heterogeneous Platform*, (Ex. 1136) ("Nayak").

would prevail with respect to at least one challenged claim for a petition to be granted. 35 U.S.C. § 314(a).

A claim is unpatentable under § 103(a) if the differences between the claimed subject matter and the prior art are such that the subject matter, as a whole, would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 406 (2007). The question of obviousness is resolved on the basis of underlying factual determinations, including (1) the scope and content of the prior art; (2) any differences between the claimed subject matter and the prior art; (3) the level of skill in the art; and (4) when in evidence, objective indicia of non-obviousness (i.e., secondary considerations). *Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966).

We analyze the asserted grounds with these principles in mind.

B. Level of Ordinary Skill in the Art

Petitioner contends that a skilled artisan "would have had a Bachelor's degree in computer science, electrical engineering, or an equivalent field, and two years of academic or industry experience working with distributed computing." Pet. 16 (citing Ex. 1105 ¶¶ 38–40).

Patent Owner does not explicitly address the level of ordinary skill in the art. *See generally* Prelim. Resp.

For the purposes of this Decision, we agree with Petitioner's proposed level of ordinary skill in the art because it comports with the level of skill reflected in the '768 patent and the prior art of record.

C. Claim Construction

We apply the same claim construction standard used by Article III federal courts and the ITC, both of which follow *Phillips v. AWH Corp.*, 415

F.3d 1303 (Fed. Cir. 2005) (en banc), and its progeny. 37 C.F.R. § 42.100(b). Accordingly, we construe each challenged claim of the '768 patent to generally be "the ordinary and customary meaning of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent." *Id.*

Petitioner contends that no claim terms require express construction "but reserves the right to present express constructions in response to any argument by the Patent Owner." Pet. 14.

Patent Owner provides contentions concerning various claim limitations. Prelim. Resp. 6–15.

As to the claim 35 limitations of "a mechanism to communicate results of evaluation with other computer cluster nodes using a peer-to-peer architecture" and "wherein the computer cluster node is configured to . . . communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other," Patent Owner contends that "there are two claim construction issues relevant to the Board's decision whether to institute an IPR." *Id.* at 6. First, Patent Owner contends that "[t]he 'wherein' clause refers back to the same 'mechanism for the nodes to communicate . . . with each other using a peer-to-peer architecture' introduced earlier in the claim." *Id.* at 7. Second, Patent Owner contends that "[a] POSITA would understand that, in the context of cluster computing, the adjective 'peer-to-peer' ordinarily means that each node can communicate tasks and data with other nodes without the tasks and data being required to go through a central server or master node." *Id.* at 8.

As to the limitation in claim 27 of "cluster node module," Patent Owner contends that "there was no ordinary meaning of 'cluster node module' in the relevant field at the relevant time." *Id.* at 12. According to

Patent Owner, "a POSITA would understand that the inventors coined the phrase 'cluster node module' to encapsulate essential features of the modules that interconnect the nodes in a preferred embodiment of the invention." *Id.* at 13. Thus, Patent Owner contends that "cluster node module" should be construed in view of the Specification, as "a module that cooperates with other cluster node modules to establish intercommunication among nodes in a computer cluster and to exchange messages such that each node can communicate tasks and data with other nodes without the tasks and data being required to go through a central server or master node." *Id.*

After reviewing the evidence and argument presented by the parties, we determine that it is not necessary to our decision to construe any claim terms. *See Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017) (noting that "we need only construe terms 'that are in controversy, and only to the extent necessary to resolve the controversy'") (citing *Vivid Techs., Inc. v. Am. Sci. & Eng'g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999)).

D. Ground 1: Alleged Obviousness over Schreiner1, Schreiner2, Schreiner3, Distributed Maple Code, Maple Guide, SPARC IV Article, and AMD Article

Petitioner contends that claims 26, 27, 29, 35, and 37 are unpatentable over Schreiner1, Schreiner2, Schreiner3, Distributed Maple Code, Maple Guide, SPARC IV Article, and AMD Article. Pet. 17–65. In support thereof, Petitioner identifies the disclosures in Schreiner1, Schreiner2, Schreiner3, Distributed Maple Code, Maple Guide, SPARC IV Article, and AMD Article alleged to describe the subject matter in the challenged claims. *Id.* Additionally, Petitioner offers declaration testimony from Dr. Henry

Tufo Ph.D ("Tufo Declaration")[17] (Ex. 1105) and Dr. Wolfgang Schreiner Ph.D. ("Schreiner Declaration")[18] (Ex. 1106) in support of the Petition.

1. Schreiner1 – Ex. 1108

Schreiner1 is entitled "Distributed Maple: parallel computer algebra in networked environments" and bears a copyright date of 2003. Ex. 1108. Schreiner1 is a journal article authored by Dr. Schreiner, Christian Mittermaier, and Karoly Bosa. *Id.* at 305. Schreiner1 "gives a comprehensive overview on the design and the use of 'Distributed Maple,' an environment for parallel computer algebra on multiprocessors and heterogeneous computer clusters." *Id.* Schreiner1 explains that Distributed Maple was developed on the basis of the computer algebra system Maple (*id.*) and that Distributed Maple is built on top of the Maple kernel and does not require any kernel extensions. *Id.* at306. According to Schreiner1, Distributed Maple is "so portable that applications can be executed in many different environments" and "so general that it can be applied to schedule tasks of other computer algebra systems (e.g., Mathematica)." *Id.* Schreiner1 describes Distributed Maple as providing "a programming model which is based on functional/logic/dataflow parallelism" that "allows the creation of a large number of implicitly scheduled tasks with automatic resolution of data dependencies and of globally shared data structures with implicit synchronization." *Id.* The authors describe using Distributed Maple "to develop the first parallel versions for a number of non-trivial applications

[17] Dr. Tufo is "presently employed as Professor of Computer Science at the University of Colorado at Boulder ('UCB')." Ex. 1105 ¶ 5.
[18] Dr. Schreiner is "a Professor in the Research Institute for Symbolic Computation ("RISC') at Johannes Kepler University in Austria." Ex. 1106 ¶ 1. Dr. Schreiner "designed and programmed Distributed Maple from 1998-2001." *Id.* ¶ 15.

from algebraic geometry (parallel curve and surface plotting and parallel neighborhood analysis)." *Id.* at 307. Schreiner1 discloses that "[t]he user interacts with Distributed Maple via a conventional Maple frontend (text or graphical)." *Id.* at 309. Schreiner1 explains that "[t]he core of Distributed Maple is a scheduler program which is completely independent and even *unaware* of Maple" and "can in fact embed and schedule tasks from any kind of computation kernels that implement a specific communication protocol." *Id.* at 310. Schreiner1 discloses that a Distributed Maple session comprises two components: a scheduler and a Maple interface. *Id.* Figure 1 of Schreiner1, reproduced below, depicts a software architecture for a Distributed Maple session. *Id.* at 311.

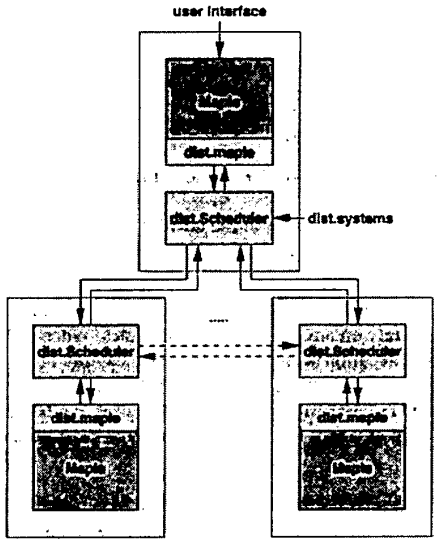


Fig. 1. Software architecture.

Figure 1 of Schreiner1 illustrates a software architecture for a Distributed Maple session. Ex. 1108,311.

As shown in Figure 1, a Distributed Maple session "comprises a set of *nodes* each of which holds a pair of processes: a *kernel* and a *scheduler*." *Id.* at 319. "Initially, a single task runs on the root kernel; this task may

subsequently create new tasks which are distributed via the schedulers to other kernels and may in turn create new tasks." *Id.* With reference to Figure 1, Schreiner1 explains that "every scheduler instance accepts tasks from the attached computation kernel and schedules these tasks among all machines connected to the session." *Id.* at 311. Schreiner1 further explains that "[t]he Maple kernel is a single-threaded process which communicates by a simple communication protocol with the schedule on the same node" and "[a]ll capabilities for parallel and distributed program execution are embedded in this scheduler." *Id.* at 314.

2. Schreiner2 – Ex. 1109

Schreiner2 is entitled "Distributed Maple – User and Reference Manual (V 1.1.12)" and bears a publication date of July 6, 2001. Ex. 1109. Schreiner2 is a publication authored by Dr. Schreiner. *Id.* at 1. Like Schreiner1, Schreiner2 describes the Distributed Maple system. More particularly, Schreiner2 "describes the use of a system for writing distributed Maple applications and sketches its implementation." *Id.* at 4.

3. Schreiner3 – Ex. 1110

Schreiner3 is entitled "Task Logging, Rescheduling and Peer Checking in Distributed Maple" and bears a publication date of March 18, 2002. Ex. 1110. Schreiner3 is a publication authored by Dr. Schreiner and Karoly Bosa. *Id.* at 1. Schreiner3 describes extending the Distributed Maple system by adding "fault tolerance mechanisms such that the time spent in a long running computation is not any [sic] wasted by the eventual occurrence of session failure." *Id.* Schreiner3 describes a first fault tolerance mechanism as "the logging of task return values and of shared object values such that after a failure the newly started session can (transparently to the application program) reuse already computed result." *Id.* A second fault

tolerance mechanism is described as "the migration of tasks such that a session may tolerate the failure of individual nodes without overall failures." *Id.* A third fault tolerance mechanism is described as "the redirection of the messages such that a session may tolerate also the failure of the connections between nodes without overall failure." *Id.* Figure 1 of Schreiner3 illustrates an Execution Model of the Distributed Maple system.

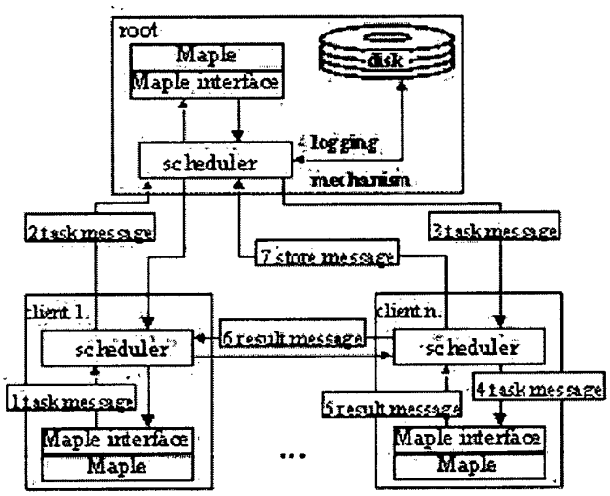


Figure 1: Execution Model

Figure 1 of Schreiner3 illustrates an Execution Model of the Distributed Maple system. Ex. 1110, 5.

Figure 1 depicts the passing of messages between nodes in a Distributed Maple system, and a logging mechanism in the root node. Schreiner3 explains that "[t]he logging mechanism in Distributed Maple is a fault tolerance mechanism for saving the results of intermediate tasks and the values of shared objects during the computation" and allows the system "to restore the results of computed tasks in a later session, if the current session crashes" *Id.* at 3.

4. Distributed Maple Code – Exs. 1112–1118

Distributed Maple Code is a collection of computer code embodying a distribution of Distributed Maple for installation on a computer, the files including: dist.maple5 file (Ex. 1112), source code for parallel versions of Maple functions in the distsoft directory (Ex. 1113), source code for parallel versions of CASA functions in the distsoft directory (Ex. 1114), an "Install" file for Distributed Maple (Ex. 1115), a "ReadMe" file for Distributed Maple (Ex. 1116), an "Install" file for the source code in the distsoft directory (Ex. 1117), and a "ReadMe" file for the source code in the distsoft directory (Ex. 1118).

5. Maple Guide – Ex. 1111

Maple Guide is entitled "Maple V Learning Guide, Release 5" published by Waterloo Maple, Inc., and bears a copyright date of 1998. Ex. 1111, 5.[19] Maple Guide is a manual authored by K. M. Heal, M. L. Hansen, and K. M. Rickard. *Id.* at 4. Maple Guide explains that "Maple V is a *Symbolic Computation System* or *Computer Algebra System*" and that "[b]oth phrases refer to Maple V's ability to manipulate information in a symbolic or algebraic manner." *Id.* at 11.

6. SPARC IV Article – Ex. 1119

The SPARC IV Article is an article published in the EE Times news journal and bears a publication date of October 15, 2003. Ex. 1119. The SPARC IV Article describes Sun Microsystems' "new UltraSPARC IV processor." *Id.* It explains that the UltraSPARC IV processor utilizes "Chip Multithreading (CMT), a design concept that allows the processor to execute

[19] Ex. 1111 contains two sets of page numbers. We refer to the page numbers added by Petitioner.

tens of threads simultaneously, thus enabling increases in application throughput." *Id.*

7. AMD Article – Ex. 1120

The AMD Article is entitled "AMD to Unveil Dual-Core PC Chips." Ex. 1120. The AMD Article was published in the Los Angeles Times newspaper and bears a publication date of May 31, 2005 and a copyright date of 2005. *Id.* The AMD Article discloses that "Advanced Micro Devices Inc. is set to launch its first PC microprocessors with two computing engines on a single chip." *Id.* The article explains that AMD and Intel Corp. had "discovered that performance could be boosted by creating two computing cores and running them at a slower speed on a single chip." *Id.*

8. Analysis of Claim 26

Petitioner provides a number reasons why a skilled artisan would have been motivated to combine the various references relied on in this challenge. Pet. 17–19. Petitioner also provides a limitation-by-limitation analysis of claim 26. *Id.* at 24–53. Petitioner supports its contentions with the testimony of Dr. Tufo. Ex. 1105 ¶¶ 43–120.

Patent Owner contends the Petition fails to show that the prior art discloses the limitation in claim 26 "wherein the third node comprises a third hardware processor . . . configured to receive the result of mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node." Prelim. Resp. 18–21. Patent Owner supports these contentions with the testimony of Dr. Singh. (Ex. 2001 ¶¶ 55–60).

In order to provide the basis for our determination, we set forth Petitioner's contentions concerning how the prior art allegedly satisfies the

limitations of claim 26 preceding the third node limitation. Patent Owner does not dispute these contentions in the Preliminary Response. *See generally* Prelim. Resp.

a) "A computer cluster comprising:"

Petitioner contends that, to the extent the preamble is limiting, Schreiner1 discloses "an environment for executing **parallel computer algebra problems on multiprocessors and heterogeneous clusters**." Pet. 24-25 (citing Ex. 1106 ¶ 61; Ex. 1108, Abstract, Fig. 1, 324–344; Ex. 1109, Abstract; Ex. 1110, 2).

b) "a plurality of nodes, wherein one or more of the nodes are configured to receive: a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes; and an instruction from a user interface or a script; and,"

Petitioner provides the following annotated version of Figure 1 of Schreiner1:

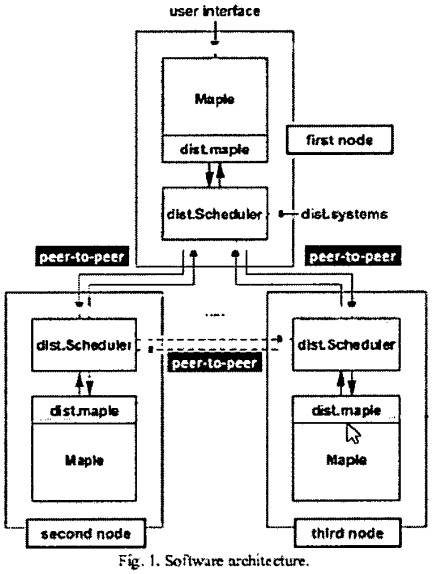


Fig. 1. Software architecture.

The preceding image is Figure 1 of Schreiner illustrating Schreiner's Software Architecture with color annotations added by Petitioner. Pet. 25.

Petitioner contends that "Schreiner1 discloses numerous instances of a cluster comprising a plurality of nodes, such as in Figure 1." Pet. 25 (citing Ex. 1108, 311, Fig. 1). Petitioner further contends that "[e]ach node contains a Maple kernel and the dist.maple and dist.Scheduler components." *Id.* (citing Ex. 1105 ¶ 62). Petitioner further contends that "[a]t least the root node is configured to perform the other portions of this claims limitation, as shown in subections i-ii . . ., though any node can be the root node." *Id.* (citing Ex. 1105 ¶ 63).

i. "a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes"

Petitioner contends that "Schreiner1 teaches that the root node initializes the computer cluster in response to a command received from the user." Pet. 25 (citing Ex. 1008, 314; Ex. 1009, 14). Petitioner further contends that "Schreiner1 provides a specific example of the initialization command dist[initialize] being received by a root node from a user, which causes a cluster of computers nicknamed 'virgo' and 'andromeda' to be configured for Distributed Maple operation." *Id.* (citing Ex. 1105 ¶ 65; Ex. 1008, 310; Ex. 1009, 5, 13–14).

ii. "an instruction from a user interface or a script"

Petitioner contends that Schreiner1's Figure 1 discloses "the root node receives instructions from a 'user interface' and sends them to the first cluster node module. 'The user interacts with Distributed Maple via a conventional Maple frontend . . . she operates within the familiar Maple

environment for writing and executing parallel programs.'" Pet. 26–27 (citing Ex. 1105 ¶ 66; Ex. 1108, 305–306, 309–310).

c) "a mechanism[20] for the nodes to communicate results of mathematical expression evaluation with each other using asynchronous calls;"

Petitioner contends that "[t]he claimed mechanism is disclosed in the Distributed Maple Publications as a combination of two software components: dist.Scheduler, which 'coordinates node interaction,' and dist.maple, which 'implements the interface between kernel and scheduler.'" Pet. 28 (citing Ex. 1008, 310). Petitioner further contends that "[t]his scheduler-based message passing mechanism is installed on each node and is used to communicate results of mathematical expression evaluation between nodes using asynchronous calls." *Id.* (citing Ex. 1105 ¶ 68). Petitioner further contends that the schedulers shown in Figure 1 of Schreiner1 "allow the cluster nodes to communicate results of mathematical expression evaluation with each other. Initially, mathematical expressions are distributed for parallel evaluation by the cluster as task messages." *Id.* at 29 (citing Ex. 1108, 311). Petitioner further contends that "[a]s the cluster evaluates the expression, nodes establish peer-to-peer socket connections to communicate results messages" that "are placed in result tables accessible by all other nodes via peer-to-peer messages." *Id.* (citing Ex. 1105 ¶ 69; Ex. 1108, 316; Ex. 1109, 27). Petitioner further contends that "[t]he scheduler messaging mechanism is asynchronous because, like the preferred embodiment [of the '768 patent], it allows message passing while the

[20] Neither party analyzes whether this limitation should be construed as a means plus function limitation under 35 U.S.C. § 112(f) despite the fact that "mechanism" is a well-known nonce word. *See Williamson v. Citrix Online, LLC*, 792 F. 3d 1339, 1350 (Fed. Cir. 2015). Given our disposition of this case, it is not necessary to reach this question.

kernels are simultaneously proceeding with computations." *Id.* at 32 (citing Ex. 1105 ¶ 75; Ex. 1108, 316); *see also id.* at 31–32 ("the specification of the '768 patent states that, '[a]synchronous calls make it possible for the kernel to do work while communications are proceeding simultaneously'" (citing Ex. 1101, 13:42–43; Ex. 1105 ¶ 74)).

d) "wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions"

Petitioner contends that "Schreiner1 discusses several implementations of Distributed Maple clusters, where each node comprises a hardware processor." Pet. 34. Petitioner further contends that "Schreiner1 discloses a **128 processor** SGI Origin 3800 distributed shared memory multiprocessor cluster, a **24-processor heterogeneous** computer cluster, an **18-processor** Sun HPC 6500 system, and a Linux-based Beowulf cluster with **16 compute** nodes." *Id.* (citing Ex. 1105 ¶¶ 80–81; Ex. 1108, 324–325, 327; Ex. 1110, 2). Petitioner further contends that "[e]ach hardware processor has access to a non-transitory computer-readable medium comprising the Maple code" and "Maple is installed in storage such as a disk drive and loaded in memory when the program runs on one or more processors." *Id.* at 34 (citing Ex. 1105 ¶ 82; Ex. 1106 ¶ 40; Ex. 1111, 5, 30, 88, 92, 187, 262–263). Petitioner further contends that "Maple software is program code for a single-node kernel . . . that ran on a single node – the computer on which it was loaded" and in that regard is similar to the Mathematica kernel referenced in the Specification of the '768 patent. *Id.* at 35 (citing Ex. 1101, 1:42–43, 2:21–24, 4:14–23; Ex. 1111, 88). Petitioner further contends that Schreiner1 uses "Distributed Maple with a

Mathematica kernel, exactly like in the embodiments of the '768 patent." *Id.* at 36 (citing Ex. 1105 ¶ 85; Ex. 1110, 1).

e) "wherein the plurality of nodes comprises: a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and;"

Petitioner posits that "the first node is the Distributed Maple root node, which performs the claimed functions as set forth in sub-sections i–ii." Pet. 37 (citing Ex. 1105 ¶ 87).

i. "configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel,"

Petitioner contends that "Schreiner1 teaches various memory configurations" and "[i]n each instance, each node, including the root node, comprises a hardware processor configured to access memory." Pet. 37 (citing Ex. 1105 ¶ 88; Ex. 1108, 306, 324–325, 327, 332). Petitioner further contends that "Maple comprises program code for a user interface" and "also comprises program code for a first single-node kernel." *Id.* at 38 (citing Ex. 1105 ¶¶ 90–91; Ex. 1108, 309–310).

ii. "the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution;" and

Petitioner contends that the "Maple kernel interprets user instructions" because it "is the base of Maple's system. It contains fundamental and primitive commands: the Maple language **interpreter**." Pet. 38 (citing Ex. 1105 ¶ 93; Ex. 1111, 88). Petitioner further contends that "Schreiner1 describes a user entering the Distributed Maple 'dist[all]' instruction into the root Maple kernel user interface: 'dist[all](*command)* lets the Maple

statement *command* be executed on every Maple kernel connected to the distributed session.'" *Id.* (citing Ex. 1108, 311; Ex. 1009, 30). Petitioner further contends that a skilled artisan would have understood the "root kernel, through its user interface distributes calls – in this case, the Distributed Maple dist[all] command and the Maple *command* to be evaluated – to the other cluster nodes." *Id.* at 39 (citing Ex. 1105 ¶ 94).

f) "a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of mathematical expression evaluation to a third node"

Petitioner posits that "[o]ne of the plurality of nodes discussed above . . . is the 'second node' that performs this claim limitation as set forth in subsections i–ii that follow." Pet. 41.

i. "a plurality of processing cores"

Petitioner contends that "Schreiner1 discloses the use of multi-processor systems such as the '18-processor Sun HPC 6500 system." Pet. 42 (citing Ex. 1108, 324–325, 327).

ii. "the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of mathematical expression evaluation to a third node"

Petitioner contends that Schreiner1's Figure 5 discloses "the second node receives a call from the first node, indicated by task <t,d>, it then executes the mathematical expression indicated by the task call; and the second node then returns a result to the first node in result <t,r>." Pet. 44–45 (citing ex. 1105 ¶ 103; Ex. 1108, 319–320, Fig. 5).

g) *"wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node;"*

Patent Owner contends that the Petition fails to establish this limitation because its fails "to show where the prior art teaches the precise order of operations, involving three specific nodes, specified by the claim." Prelim. Resp. 19. Patent Owner further contends "the Petition refers to the prior art's general disclosure that multiple nodes may be involved in evaluating mathematical expressions and then speculates that the prior art could work in a manner that would meet the claim limitation." *Id.* (citing Pet. 44–47; Ex. 2001 ¶ 67).

Petitioner provides two examples to establish that the prior art references disclose this limitation. Pet. 46–50. Petitioner's first example is based on Figure 5 of Schreiner1, which is reproduced below:

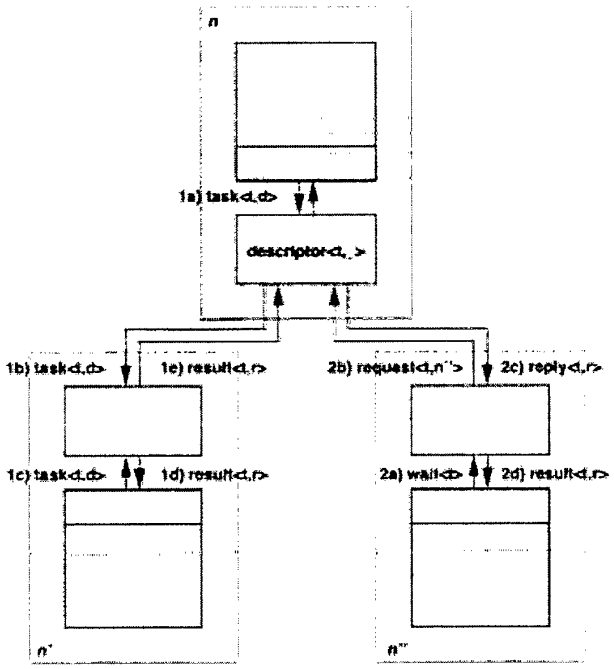


Figure 5 of Schreiner1 is an "execution model" of Distributed Maple. Ex. 1108, 320.

Petitioner contends that Figure 5 teaches that "the third node (in this case, *n''*) requests a result from the first node (*n*), which sends it to the third node after receipt from the second (*n'*)." Pet. 48. Petitioner further contends that the

> third node initiates the request through the wait call, indicating that it was evaluating a second expression that was dependent on the result of a first evaluation performed on the second node. A [skilled artisan] would understand that the third node then *could* execute a second mathematical expression using that result, and in fact, that is *likely* why the third node requested the result in the first place. . . . [A] [skilled artisan] would further understand that the third node *could* send the result of the second mathematical expression to the first (root) node if the first node is the node that originally requested that the second mathematical expression be performed, a common case.

Id. at 48–49 (citing Ex. 1105 ¶ 105) (emphases added).

We note that this block quotation is repeated in the Petition verbatim from Dr. Tufo's Declaration. Further, Dr. Tufo does not provide a citation to any portion of Schreiner1 to support his testimony concerning the execution model shown in Figure 1. Because this testimony is conclusory and not supported by objective evidence, it is entitled to little or no weight. *See Velander v. Garner*, 348 F.3d 1359, 1371 (Fed. Cir. 2003) ("[W]hat the [PTAB] consistently did was accord little weight to broad conclusory statements that it determined were unsupported by corroborating references. It is within the discretion of the trier of fact to give each item of evidence such weight as it feels appropriate." (citation omitted)); *see also In re Am. Acad. of Sci. Tech Ctr.*, 367 F.3d 1359, 1368 (Fed. Cir. 2004) ("[T]he [PTAB] is entitled to weigh the declarations and conclude that the lack of factual corroboration warrants discounting the opinions expressed in the declarations . . . ." (citations omitted)). Because Petitioner has not provided sufficient objective evidence in support of this example, we determine that Petitioner has not shown that this example discloses the third node limitation.

Petitioner's second example is based on Figure 1 of Schreiner3 which is reproduced below:

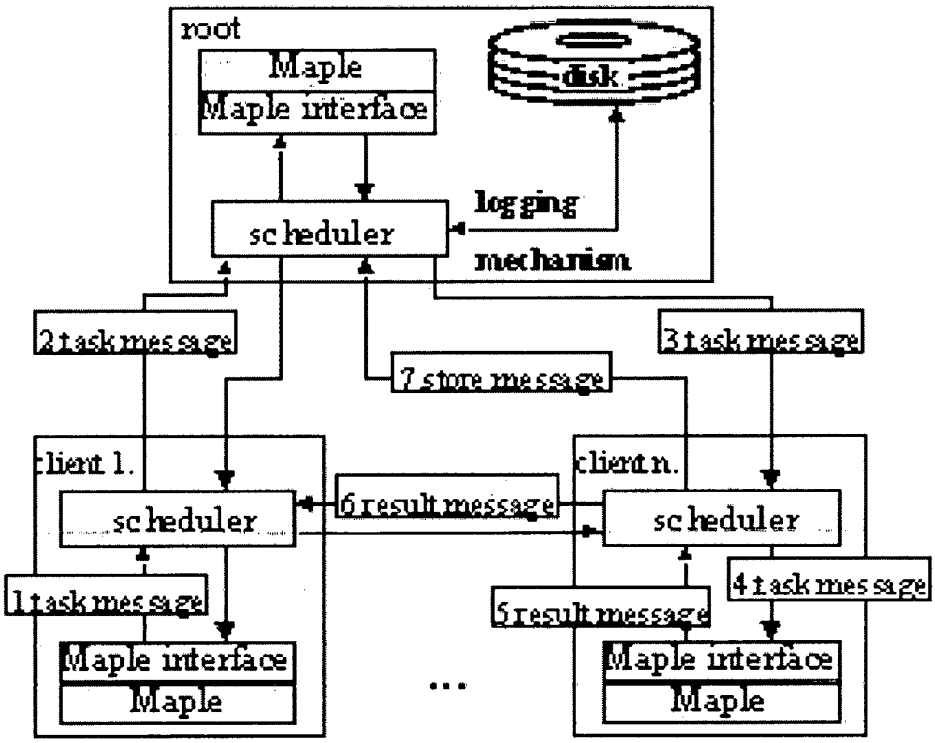


Figure 1: Execution Model

Figure 1 of Schreiner3 is an Execution Model of Distributed Maple. Ex. 1110, 5.

Petitioner contends that in Figure 1 of Schreiner3, "node 'client 1' receives a result from node 'client n.'" Pet. 49. Petitioner notes that in the case of Figure 1, "the initial tasks originating with the root Maple kernel are not shown." *Id.* Petitioner further contends that because "the root initiates the mathematical evaluation by distributing tasks to other nodes, node 'client 1' is generating subsidiary tasks and evaluating expressions in order to eventually provide results to the root." *Id.* Petitioner further contends that "once ['client 1'] receives the result from node 'client n,' it may use it to perform at least a second mathematical expression evaluation in order to

process and return results for tasks sent to it by the root." *Id.* at 49–50 (citing Ex. 1105 ¶ 111).

We note that this portion of the Petition is taken verbatim from Dr. Tufo's Declaration. Further, Dr. Tufo does not provide a citation to any portion of any of Schreiner3 to support his testimony concerning the model shown in Figure 1 of Schreiner3. The testimony is, thus, entitled to little or no weight because it is conclusory and not based on objective evidence. In addition, neither the Petition nor Dr. Tufo reconcile the argument that client 1 returns a result to the root node when the box labelled "result message" only indicates that result messages are sent between client nodes 1–n, not between the root node and client nodes. Because Petitioner has not provided sufficient objective evidence in support of this example, we determine that Petitioner has not shown that this example discloses the third node limitation.

Based on the foregoing, we determine that Petitioner fails to establish a reasonable likelihood that claim 26 is unpatentable.

9. Analysis of Claims 27 and 29

Claim 27 depends from claim 26. Ex. 1101, 33:13–23. We have reviewed Petitioner contentions and additional evidence cited in support of the challenge to claim 27, and determine that it does not cure the deficiencies discussed above in connection with claim 26. *See* Pet. 53–56. Therefore, we determine that Petitioner fails to establish a reasonable likelihood that claim 27 is unpatentable.

Independent claim 29 contains the same third node limitation as claim 26. Ex. 1101, 33:38–34:22. Petitioner relies on the same contentions for claim 29 as for claim 26. Pet. 56. For the same reasons discussed above

with respect to claim 26, we determine that Petitioner fails to establish a reasonable likelihood that claim 29 is unpatentable.

10. Analysis of Claim 35

Claim 35 recites, *inter alia*, "the third node is configured to receive the result of mathematical expression evaluation from the computer cluster node, execute at least a second mathematical expression evaluation using the result of the first mathematical expression evaluation, and communicate a result of the second mathematical expression evaluation to *the first node*." Ex. 1101, 35:40–36:4 (emphasis added). Petitioner correctly notes that "[t]he term 'the first node' lacks antecedent basis." Pet. 63. Petitioner argues "[t]here are two possible ways to construe the 'first node'" and contends that Schreiner1 discloses the limitation under either construction. *Id.* at 63–64. Patent Owner does not address the antecedent basis issue, Petitioner's proposed constructions, or Petitioner's contentions. *See generally* Prelim. Resp.

Petitioner contends that "[i]f the 'first node' is construed as 'the second node' (i.e., the root node in Schreiner1), then the 'third node' of the claim (e.g., the node in the bottom right of Figure 1 of Schreiner1) is configured to perform this limitation" for the same reasons as for claim 26. Pet. 63. For the same reasons discussed above for claim 26, we are not persuaded that Schreiner1 discloses this limitation.

Petitioner alternately contends that "[i]f the 'first node' is construed as 'the computer cluster node,' then it would have been obvious to perform a sequence of tasks that result in the third node of the claim receiving a result from the computer cluster node, performing another mathematical expression evaluation, and providing the result of that evaluation to the computer cluster node." Pet. 63–64. Petitioner further contends that "in the

normal course of Distributed Maple cluster operation, any node, including the third node may receive results from any node, and may send results to any node." *Id.* at 64 (citing Ex. 1105 ¶ 143; Ex. 1018, 312). As with Petitioner's contentions for claim 26, Petitioner and its expert point to no objective evidence but speculate as to what "may" occur.

Based on the foregoing, we determine that Petitioner fails to establish that claim 35 is unpatentable.

11. Analysis of Claim 37

Claim 37 depends from claim 35. Ex. 1101, 36:25–28. We have reviewed Petitioner contentions and additional evidence cited in support of the challenge to claim 37, and determine that it does not cure the deficiencies discussed above in connection with claim 35. *See* Pet. 65. Therefore, we determine that Petitioner fails to establish a reasonable likelihood that claim 37 is unpatentable.

E. Petitioner's Remaining Obviousness Challenges

Petitioner challenges claims 26, 27, 36, 37, and 39 in these grounds. Pet. 13. We have reviewed Petitioner's contentions and additional evidence cited in support of the challenges to claims 26, 27, 36, 37, and 39, and determine that they do not cure the deficiencies discussed above in connection with independent claims 26 and 35. *See* Pet. 65–78. Notably, regarding independent claims 26 and 35, Petitioner appears to rely on the same deficient reasoning as discussed above. *Id.* at 66–67. Therefore, we determine that Petitioner fails to establish a reasonable likelihood that claims 26, 27, 36, 37, and 39 are unpatentable.

III. CONCLUSION

Because Petitioner has not established a reasonable likelihood that any of the challenged claims are unpatentable, we do not institute *inter partes* review.

IV. ORDER

In consideration of the foregoing, it is hereby

ORDERED that institution of *inter partes* is denied.

FOR PETITIONER:

Brent Yamashita
Brennt.yamashita@dlapiper.com

Jonathan Hicks
Jonathan.hicks@dlapiper.com

FOR PATENT OWNER:

Jon Gurka
2jwg@knobbe.com

Ted M. Cannon
2tmc@knobbe.com

Cheryl Burgess
2ctb@knobbe.com

cu

AO 120 (Rev. 08/10)

TO:	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the District of Delaware on the following

☐ Trademarks or ☑ Patents. (☐ the patent action involves 35 U.S.C. § 292.):

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
	10/28/2019	for the District of Delaware

PLAINTIFF	DEFENDANT
ADVANCED CLUSTER SYSTEMS, INC.	NVIDIA CORPORATION

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1	US 8,082,289 B2	12/20/2011	Advanced Cluster Systems, Inc.
2	US 8,140,612 B2	3/20/2012	Advanced Cluster Systems, Inc.
3	US 8,676,877 B2	3/18/2014	Advanced Cluster Systems, Inc.
4	US 10,333,768 B2	6/25/2019	Advanced Cluster Systems, Inc.
5			

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/181,112	06/25/2019	10333768	ZTANN.002P1C3	6874

20995 7590 06/05/2019

KNOBBE MARTENS OLSON & BEAR LLP
2040 MAIN STREET
FOURTEENTH FLOOR
IRVINE, CA 92614

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 524 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Data Management (ODM) at (571)-272-4200.

APPLICANT(s) (Please see PAIR WEB site http://pair.uspto.gov for additional applicants):

Advanced Cluster Systems, Inc., Aliso Viejo, CA;
Zvi Tannenbaum, Newport Beach, CA;
Dean E. Dauger, Huntington Beach, CA;

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The USA offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to encourage and facilitate business investment. To learn more about why the USA is the best country in the world to develop technology, manufacture products, and grow your business, visit SelectUSA.gov.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

20995 7590 11/29/2018
KNOBBE MARTENS OLSON & BEAR LLP
2040 MAIN STREET
FOURTEENTH FLOOR
IRVINE, CA 92614

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/181,112	02/14/2014	Zvi Tannenbaum	ZTANN.002P1C3	6874

TITLE OF INVENTION: CLUSTER COMPUTING

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$500	$0.00	$0.00	$500	02/28/2019

EXAMINER	ART UNIT	CLASS-SUBCLASS
ASRES, HERMON	2449	709-223000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 Knobbe Martens
2 Olson & Bear LLP
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE	(B) RESIDENCE: (CITY and STATE OR COUNTRY)
Advanced Cluster Systems, Inc.	Aliso Viejo, CA

Please check the appropriate assignee category or categories (will not be printed on the patent) : ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. Fees submitted: ☒ Issue Fee ☐ Publication Fee (if required) ☐ Advance Order - # of Copies _____

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

☒ Electronic Payment via EFS-Web ☐ Enclosed check ☐ Non-electronic payment by credit card (Attach form PTO-2038)

☒ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. 11-1410

5. **Change in Entity Status** (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29

☐ Applicant asserting small entity status. See 37 CFR 1.27

☐ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature /Lance D. Smemoe/ Date Feburary 27, 2019

Typed or printed name Lance D. Smemoe Registration No. 66,152

PTOL-85 Part B (08-18) Approved for use through 01/31/2020 OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Electronic Patent Application Fee Transmittal

Application Number:	14181112
Filing Date:	14-Feb-2014
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Filer:	Lance Dean Smemoe/Amy Rodriguez
Attorney Docket Number:	ZTANN.002P1C3

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
UTILITY APPL ISSUE FEE	2501	1	500	500

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**500**

Electronic Acknowledgement Receipt	
EFS ID:	35269940
Application Number:	14181112
International Application Number:	
Confirmation Number:	6874
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Customer Number:	20995
Filer:	Lance Dean Smemoe/Melissa Ramirez
Filer Authorized By:	Lance Dean Smemoe
Attorney Docket Number:	ZTANN.002P1C3
Receipt Date:	27-FEB-2019
Filing Date:	14-FEB-2014
Time Stamp:	15:28:49
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$500
RAM confirmation Number	022819INTEFSW15291400
Deposit Account	111410
Authorized User	Melissa Ramirez
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.16 (National application filing, search, and examination fees) 37 CFR 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Issue Fee Payment (PTO-85B)	IssueFeeTransmittal_ZTANN002P1C3.pdf	208581 a3c07a37eece31a451c1eb2c933fbe84b57db0af	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	30150 01244fc0a8830184348204c4d7b52e7e7dfc9da1	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	238731		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/181,112	02/14/2014	Zvi Tannenbaum	ZTANN.002P1C3	6874

20995 7590 02/08/2019

KNOBBE MARTENS OLSON & BEAR LLP
2040 MAIN STREET
FOURTEENTH FLOOR
IRVINE, CA 92614

EXAMINER	
ASRES, HERMON	

ART UNIT	PAPER NUMBER
2449	

NOTIFICATION DATE	DELIVERY MODE
02/08/2019	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

efiling@knobbe.com
jayna.cartee@knobbe.com

Notice of Allowability	Application No. 14/181,112	Applicant(s) Tannenbaum et al.	
	Examiner HERMON ASRES	Art Unit 2449	AIA Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to 01/25/2019.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☑ The allowed claim(s) is/are 2-25 and 27-41 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information , please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some *c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file areply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☑ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date. _____.
5. ☐ Examiner's Amendment/Comment
6. ☐ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/HERMON ASRES/
Examiner, Art Unit 2449

PTO/SB/08 Equivalent

INFORMATION DISCLOSURE STATEMENT BY APPLICANT	Application No.	14/181112
	Filing Date	February 14, 2014
	First Named Inventor	Zvi Tannenbaum
	Art Unit	2449
(Multiple sheets used when necessary)	Examiner	Asres, Hermon
SHEET 1 OF 1	Attorney Docket No.	ZTANN.002P1C3

U.S. PATENT DOCUMENTS					
Examiner Initials	Cite No.	Document Number *Number - Kind Code (if known)* Example: 1,234,567 B1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear
	1	7,139,882	11-21-2006	Suzuoki et al.	

FOREIGN PATENT DOCUMENTS						
Examiner Initials	Cite No.	Foreign Patent Document *Country Code-Number-Kind Code* Example: JP 1234567 A1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear	T[1]

NON PATENT LITERATURE DOCUMENTS			
Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[1]

29826515

Examiner Signature	/HERMON ASRES/	Date Considered	02/05/2019

*__Examiner:__ Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.A./

T[1] - Place a check mark in this area when an English language Translation is attached.

PTO/SB/08 Equivalent

INFORMATION DISCLOSURE STATEMENT BY APPLICANT	Application No.	14/181112
	Filing Date	February 14, 2014
	First Named Inventor	Zvi Tannenbaum
	Art Unit	2449
(Multiple sheets used when necessary)	Examiner	Asres, Hermon
SHEET 1 OF 1	Attorney Docket No.	ZTANN.002P1C3

U.S. PATENT DOCUMENTS					
Examiner Initials	Cite No.	Document Number *Number - Kind Code (if known)* Example: 1,234,567 B1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear
	1	7,139,882	11-21-2006	Suzuoki et al.	

FOREIGN PATENT DOCUMENTS						
Examiner Initials	Cite No.	Foreign Patent Document *Country Code-Number-Kind Code* Example: JP 1234567 A1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear	**T[1]**

NON PATENT LITERATURE DOCUMENTS			
Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	**T[1]**

29826515

Examiner Signature	Date Considered
***Examiner:** Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.	

T[1] - Place a check mark in this area when an English language Translation is attached.

Electronic Patent Application Fee Transmittal

Application Number:	14181112
Filing Date:	14-Feb-2014
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Filer:	Lance Dean Smemoe/Amy Rodriguez
Attorney Docket Number:	ZTANN.002P1C3

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	120	120
		Total in USD ($)		**120**

Electronic Acknowledgement Receipt

EFS ID:	34968658
Application Number:	14181112
International Application Number:	
Confirmation Number:	6874
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Customer Number:	20995
Filer:	Lance Dean Smemoe/Amy Rodriguez
Filer Authorized By:	Lance Dean Smemoe
Attorney Docket Number:	ZTANN.002P1C3
Receipt Date:	25-JAN-2019
Filing Date:	14-FEB-2014
Time Stamp:	18:28:52
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	IDSTransmittal-ZTANN002P1C3.pdf	20588 d300b9e1cd65971a75fb12864e7205f617260053	no	2

Warnings:

Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	IDSListForm-ZTANN002P1C3.pdf	19308 e6a6d5916532fe4c7c8289ab319abbffa666d62d	no	1
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
3	Fee Worksheet (SB06)	fee-info.pdf	30319 9359ac91ff8000d76c2e6e4dc36a178d983db5f3	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	70215		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

INFORMATION DISCLOSURE STATEMENT

First Inventor	:	Zvi Tannenbaum
App. No.	:	14/181112
Filed	:	February 14, 2014
For	:	CLUSTER COMPUTING
Examiner	:	Asres, Hermon
Art Unit	:	2449
Conf. No.	:	6874

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

References and Listing

Pursuant to 37 CFR 1.56, an Information Disclosure Statement listing references is provided herewith. Copies of any listed foreign and non-patent literature references are being submitted.

No Disclaimers

To the extent that anything in the Information Disclosure Statement or the listed references could be construed as a disclaimer of any subject matter supported by the present application, Applicant hereby rescinds and retracts such disclaimer.

Timing of Disclosure

This Information Disclosure Statement is being filed after the mailing date of a final action or after the mailing date of a Notice of Allowance.

CERTIFICATION UNDER 37 CFR 1.97(e)(2)

I hereby certify that no item of information contained in this Statement was cited in a communication from a foreign Patent Office in a counterpart foreign application, and, to my knowledge after making reasonable inquiry, was known to any individual designated in 37 CFR 1.56(c) more than three (3) months prior to the filing of this Information Disclosure Statement.

This Statement is accompanied by the fee set forth in 37 CFR 1.17(p). The Commissioner is authorized to charge any additional fees which may be required or to credit any overpayment to Account No. 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: January 25, 2019

By: /Lance D. Smemoe/
Lance D. Smemoe
Registration No. 66,152
Registered Practitioner
Customer No. 20995
(949) 760-0404

29826491

Electronic Patent Application Fee Transmittal

Application Number:	14181112
Filing Date:	14-Feb-2014
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Filer:	Lance Dean Smemoe/Amy Rodriguez
Attorney Docket Number:	ZTANN.002P1C3

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	120	120
		Total in USD ($)		**120**

Electronic Acknowledgement Receipt

EFS ID:	34968757
Application Number:	14181112
International Application Number:	
Confirmation Number:	6874
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Customer Number:	20995
Filer:	Lance Dean Smemoe/Amy Rodriguez
Filer Authorized By:	Lance Dean Smemoe
Attorney Docket Number:	ZTANN.002P1C3
Receipt Date:	25-JAN-2019
Filing Date:	14-FEB-2014
Time Stamp:	18:37:26
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Fee Worksheet (SB06)	fee-info.pdf	30320 37387c731a6cd01ab20828f57de455d5b4de4718	no	2

Warnings:

Information:	
Total Files Size (in bytes):	30320

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Electronic Patent Application Fee Transmittal

Application Number:	14181112
Filing Date:	14-Feb-2014
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Filer:	Lance Dean Smemoe/Amy Rodriguez
Attorney Docket Number:	ZTANN.002P1C3

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	120	120
		Total in USD ($)		**120**

Electronic Acknowledgement Receipt	
EFS ID:	34968823
Application Number:	14181112
International Application Number:	
Confirmation Number:	6874
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Customer Number:	20995
Filer:	Lance Dean Smemoe/Amy Rodriguez
Filer Authorized By:	Lance Dean Smemoe
Attorney Docket Number:	ZTANN.002P1C3
Receipt Date:	25-JAN-2019
Filing Date:	14-FEB-2014
Time Stamp:	18:43:03
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$120
RAM confirmation Number	012819INTEFSW18474300
Deposit Account	111410
Authorized User	Amy Rodriguez
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.16 (National application filing, search, and examination fees) 37 CFR 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Fee Worksheet (SB06)	fee-info.pdf	30319 0b5490d464aa3718232657ecfead518ae2798ee1	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	30319		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

20995 7590 11/29/2018

KNOBBE MARTENS OLSON & BEAR LLP
2040 MAIN STREET
FOURTEENTH FLOOR
IRVINE, CA 92614

EXAMINER
ASRES, HERMON

ART UNIT	PAPER NUMBER
2449	

DATE MAILED: 11/29/2018

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/181,112	02/14/2014	Zvi Tannenbaum	ZTANN.002P1C3	6874

TITLE OF INVENTION: CLUSTER COMPUTING

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$500	$0.00	$0.00	$500	02/28/2019

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 1/2 the amount of undiscounted fees, and micro entity fees are 1/2 the amount of small entity fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Maintenance fees are due in utility patents issuing on applications filed on or after Dec. 12, 1980. It is patentee's responsibility to ensure timely payment of maintenance fees when due. More information is available at www.uspto.gov/PatentMaintenanceFees.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

20995 7590 11/29/2018

KNOBBE MARTENS OLSON & BEAR LLP
2040 MAIN STREET
FOURTEENTH FLOOR
IRVINE, CA 92614

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/181,112	02/14/2014	Zvi Tannenbaum	ZTANN.002P1C3	6874

TITLE OF INVENTION: CLUSTER COMPUTING

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$500	$0.00	$0.00	$500	02/28/2019

EXAMINER	ART UNIT	CLASS-SUBCLASS
ASRES, HERMON	2449	709-223000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

❑ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

❑ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 _____
2 _____
3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ❑ Individual ❑ Corporation or other private group entity ❑ Government

4a. Fees submitted: ❑ Issue Fee ❑ Publication Fee (if required) ❑ Advance Order - # of Copies _____

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

❑ Electronic Payment via EFS-Web ❑ Enclosed check ❑ Non-electronic payment by credit card (Attach form PTO-2038)

❑ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. _____

5. **Change in Entity Status** (from status indicated above)

❑ Applicant certifying micro entity status. See 37 CFR 1.29

❑ Applicant asserting small entity status. See 37 CFR 1.27

❑ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature _____ Date _____

Typed or printed name _____ Registration No. _____

PTOL-85 Part B (08-18) Approved for use through 01/31/2020 OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/181,112	02/14/2014	Zvi Tannenbaum	ZTANN.002P1C3	6874

20995 7590 11/29/2018

KNOBBE MARTENS OLSON & BEAR LLP
2040 MAIN STREET
FOURTEENTH FLOOR
IRVINE, CA 92614

EXAMINER	
ASRES, HERMON	

ART UNIT	PAPER NUMBER
2449	

DATE MAILED: 11/29/2018

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

PTOL-85 (Rev. 02/11)

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Notice of Allowability	Application No. 14/181,112	Applicant(s) Tannenbaum et al.	
	Examiner HERMON ASRES	Art Unit 2449	AIA Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to 10/05/2018.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☑ The allowed claim(s) is/are 2-25 and 27-41 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information , please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some *c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file areply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☑ Notice of References Cited (PTO-892)
2. ☑ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☑ Interview Summary (PTO-413), Paper No./Mail Date. 11/21/2018.
5. ☑ Examiner's Amendment/Comment
6. ☑ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/HERMON ASRES/
Examiner, Art Unit 2449

DETAILED ACTION

Claims 2-25, and 27-41 are allowed.

Response to Arguments

Double Patenting rejection

Examiner withdraws the double patenting rejection of claims 2, 28, and 31 in view of the Terminal Disclaimer filed and approved on 11/21/2018

EXAMINER'S AMENDMENT

An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone correspondence with Lance Smemoe (Reg. No. 66,152) on 11/21/2018.

The application has been amended as follows:

1. (Canceled)

2. (**Currently amended**) A computer cluster comprising:

a plurality of nodes, wherein each of the plurality of nodes comprises a hardware processor, wherein one or more of the nodes are configured to receive a command to start a cluster initialization process for the computer cluster, and wherein each of the nodes is

configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing the hardware processor to evaluate mathematical expressions; and

a mechanism for the nodes to communicate results of mathematical expression evaluation with each other using a peer-to-peer architecture;

wherein the plurality of nodes comprises:

a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and

a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of the first mathematical expression evaluation to a third node;

wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of the first mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node;

wherein the first node is configured to return the result of the second mathematical expression evaluation to the user interface;

wherein one or more of the nodes are configured to:

accept user instructions;

after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other; and

after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels.

3. (Previously presented) The computer cluster of claim 2, wherein the single-node kernel comprises a Mathematica kernel.

4. (Previously presented) The computer cluster of claim 2, wherein the mechanism comprises a message passing interface.

5. (Previously presented) The computer cluster of claim 2, wherein each of the nodes comprises one or more cluster node modules.

6. (Previously presented) The computer cluster of claim 2, wherein each single-node kernel is stored in a non-transitory computer-readable medium and configured to accept and execute a request.

7. (Previously presented) The computer cluster of claim 6, wherein each of the nodes comprises one or more cluster node modules, wherein each of the cluster node modules comprises instructions stored in a non-transitory computer-readable medium, and wherein the instructions, when executed by the hardware processor, cause the cluster node module to communicate with the single-node kernel and with one or more other cluster node modules.

8. (Previously presented) The computer cluster of claim 7, wherein the plurality of cluster node modules act as a cluster.

9. (Previously presented) The computer cluster of claim 7, wherein the plurality of cluster node modules communicate with one another to act as a cluster.

10. (Previously presented) The computer cluster of claim 9, wherein the computer cluster includes the user interface.

11. (Previously presented) The computer cluster of claim 10, wherein each cluster node module accepts instructions from the user interface and interprets one or more of the instructions.

12. (Previously presented) The computer cluster of claim 2, wherein the cluster initialization process comprises establishing communication among two or more of the nodes.

13. (Previously presented) The computer cluster of claim 2, wherein the cluster initialization process comprises configuring access by one or more of the nodes to a computer-readable medium comprising program code for the single-node kernel.

14. (Previously presented) The computer cluster of claim 13, wherein the cluster initialization process comprises establishing message-passing support among the nodes in the cluster.

15. (Previously presented) The computer cluster of claim 13, wherein the cluster initialization process comprises launching cluster node modules by the cluster.

16. (Previously presented) The computer cluster of claim 15, wherein the cluster initialization process comprises assigning a processor identification number to each of the cluster node modules.

17. (Previously presented) The computer cluster of claim 2, wherein the cluster initialization process comprises:

launching cluster node modules by the cluster; and

after launching the cluster node modules, configuring access by one or more of the nodes to a non-transitory computer-readable medium comprising program code for the single-node kernel.

18. (Previously presented) The computer cluster of claim 2, wherein the cluster initialization process comprises:

launching cluster node modules by the cluster;

after launching the cluster node modules, establishing communication among two or more of the nodes; and

after establishing communication among two or more of the nodes, assigning a processor identification number to each of the cluster node modules.

19. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to communicate at least some of the user instructions.

20. (Previously presented) The computer cluster of claim 19, wherein one or more of the nodes are configured to communicate at least some of the user instructions to one or more single-node kernels.

21. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to accept user instructions via one or more of the nodes.

22. (Previously presented) The computer cluster of claim 21, wherein one or more of the nodes are configured to communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other.

23. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to transmit at least some of the user instructions that originate from a user interface.

24. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to parallelize at least some of the user instructions before communicating at least some of the user instructions to one or more single-node kernels.

25. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to accept user instructions before communicating at least some of the user instructions to one or more single-node kernels.

26. (Canceled)

27. (Previously presented) The computer cluster of claim 2, wherein the plurality of nodes are configured to communicate with one another to interpret and translate commands for execution by a plurality of single-node kernels.

28. (**Currently amended**) A computer cluster comprising:

a plurality of nodes, wherein one or more of the nodes are configured to receive:

a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes; and

an instruction from a user interface or a script;

and

a mechanism for the nodes to communicate results of mathematical expression evaluation with each other using asynchronous calls;

wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions;

wherein the plurality of nodes comprises:

a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and

a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of mathematical expression evaluation to a third node;

wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node;

wherein the first node is configured to return the result of the second mathematical expression evaluation to the user interface or the script;

wherein one or more of the nodes are configured to:

accept user instructions;

after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other; and

after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels.

29. (Previously presented) The computer cluster of claim 28, wherein the asynchronous calls comprise a first command to create a first packet containing:

an expression to be sent as payload; and

a target node where the expression should be sent;

wherein the first command is configured to be called from within a single-node kernel;

wherein the single-node kernel is configured to send the first packet to a local cluster node module; and

wherein the local cluster node module is configured to forward the expression to the target node.

30. (Previously presented) The computer cluster of claim 29, wherein the asynchronous calls comprise a second command to create a second packet containing:

a location where the expression is expected to be received; and

a sending node from which the expression is expected to be received;

wherein the second command is configured to be called from within a single-node kernel;

wherein the single-node kernel is configured to send the second packet to a local cluster node module; and

wherein the local cluster node module is configured to store the second packet contents in a message receiving queue.

31. (**Currently amended**) A computer cluster comprising:

a plurality of nodes, wherein one or more of the nodes are configured to receive:

a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes; and

an instruction from a user interface or a script;

and

a mechanism for the nodes to communicate results of mathematical expression evaluation with each other;

wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions;

wherein the plurality of nodes comprises:

a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and

a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of mathematical expression evaluation to a third node;

wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node;

and

wherein the first node is configured to return the result of the second mathematical expression evaluation to the user interface or the script;

wherein one or more of the nodes are configured to:

accept user instructions;

after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other; and

after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels.

32. (Previously presented) The computer cluster of claim 5, wherein each of the plurality of nodes implements asynchronous calls that enable the single-node kernel to perform computation tasks while the cluster node modules are simultaneously communicating with one another.

33. (Previously presented) The computer cluster of claim 5, wherein intercommunication among the plurality of single-node kernels during thread execution is enabled by the plurality of cluster node modules, and wherein the computer cluster is configured to permit exchange of information between nodes during the course of a parallel computation.

34. (Previously presented) The computer cluster of claim 5, wherein each of the single-node kernels are configured to call a send command involving creating a packet containing:

an expression to be sent as payload; and

where the expression should be sent;

wherein, upon reception of the send command by a local cluster node module associated with the single-node kernel, the local cluster node module is configured to decode the packet and forward a payload to the cluster node module specified in the packet;

wherein each of the single-node kernels is configured to call a receipt command involving creating a packet specifying which message to test for completion and to then wait for a reply expression to evaluate;

wherein, upon reception of the receipt command by the local cluster node module associated with the single-node kernel, the local cluster node module is configured to decode the packet and use a message specifier to search for any matching expressions listed as completed in a received message queue;

wherein, upon determining that such a completed expression is found, the local cluster node module is configured to send the completed expression to a local single-node kernel associated with the cluster node module in response to the receipt command, and

wherein each of the single-node kernels is configured to receive the completed expression and to update variables used by the single-node kernel during evaluation of expressions.

35. (Previously presented) The computer cluster of claim 2, wherein the plurality of nodes are configured to permit exchange of information between nodes during the course of parallel computation.

36. (Previously presented) The computer cluster of claim 2, wherein each of the plurality of nodes comprises instructions executable by the hardware processor and configured to implement asynchronous behavior, wherein the instructions comprise:

a first instruction to asynchronously send a payload to another node;

a second instruction to asynchronously receive a payload from another node; and

a third instruction to search for a payload matching a message specifier.

37. (**Currently amended**) A computer cluster node for evaluating expressions in parallel with other computer cluster nodes, the computer cluster node comprising:

a hardware processor configured to access one or more non-transitory memory devices comprising program code for a single-node kernel that, when executed, causes the hardware processor to interpret user instructions, to evaluate mathematical expressions, and to produce results of mathematical expression evaluation, wherein the hardware processor comprises multiple processor cores;

a user connection interface configured to receive a command to start a cluster initialization process for a computer cluster;

a mechanism to communicate results of evaluation with other computer cluster nodes using a peer-to-peer architecture; and

~~program code that, when executed, is capable of causing the hardware processor to receive calls from a first node, execute at least a first mathematical expression evaluation,~~

~~and communicate a result of mathematical expression evaluation to another node for use in at least a second mathematical expression evaluation;~~

program code that, when executed, is capable of causing the hardware processor to:

receive calls from a second node comprising a second hardware processor configured to access a second memory comprising program code for a user interface and program code for a second single-node kernel, the second single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution;

execute, using the hardware processor, at least a first mathematical expression evaluation; and

communicate a result of the first mathematical expression evaluation to a third node comprising a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the computer cluster node, execute at least a second mathematical expression evaluation using the result of the first mathematical expression evaluation, and communicate a result of the second mathematical expression evaluation to the first node;

wherein the user connection interface is configured to return at least one result of mathematical expression evaluation to a user interface or a script; and

wherein the computer cluster node is configured to:

accept user instructions;

after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other; and

after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to the single-node kernel.

38. (Previously presented) The computer cluster node of claim 37, wherein the one or more non-transitory memory devices comprise program code for performing a parallel fast Fourier transform, wherein the program code causes the hardware processor to evaluate a command to

perform a Fourier transform on an array comprising a first data portion that is stored on the computer cluster node and a second data portion that is not stored on the computer cluster node.

39. (Previously presented) The computer cluster node of claim 37, wherein the computer cluster node is configured to permit exchange of information with other computer cluster nodes during the course of parallel computation.

40. (Previously presented) The computer cluster node of claim 37, wherein the computer cluster node comprises instructions executable by the hardware processor and configured to implement asynchronous behavior, wherein the instructions comprise:

a first instruction to asynchronously send a payload to another node;

a second instruction to asynchronously receive a payload from another node; and

a third instruction to search for a payload matching a message specifier.

41. (Previously presented) The computer cluster node of claim 37, wherein the hardware processor comprises a special purpose microprocessor.

Reason for Allowance

The following is an examiner's statement of reasons for allowance.

Independent Claims 2 as amended distinguishes itself over the prior art due to the amended limitation in combination with the rest of the limitations. It is to be noted that it is the combination of all limitations that renders the claims allowable. Claims 3-25, and 27-41 are allowed based on the same reason(s).

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HERMON ASRES whose telephone number is (571)272-4257. The examiner can normally be reached on Monday to Friday 9AM to 5PM.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Vivek Srivastava can be reached on (571)272-7304. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/HERMON ASRES/
Examiner, Art Unit 2449

Examiner-Initiated Interview Summary	**Application No.** 14/181,112	**Applicant(s)** Tannenbaum et al.	
	Examiner HERMON ASRES	**Art Unit** 2449	**AIA Status** No

All participants (applicant, applicant's representative, PTO personnel):

(1) HERMON ASRES.

(3) Lance Smemoe (REG NO 66,152).

(2) _____.

(4) _____.

Date of Interview: 21 November 2018.

Type: ☑ Telephonic ☐ Video Conference
☐ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☑ No.
If Yes, brief description: _____.

Issues Discussed ☐ 101 ☐ 112 ☐ 102 ☐ 103 ☑ Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: 2-41.

Identification of prior art discussed: _____.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

See Continuation Sheet.

Applicant recordation instructions: It is not necessary for applicant to provide a separate record of the substance of interview.

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☐ Attachment

/HERMON ASRES/ Examiner, Art Unit 2449	

Continuation of Substance of Interview including description of the general nature of what was agreed to if an agreement was reached, or any other comments: Examiner contacted applicant and proposed incorporating dependent claim 26 into all the Independent claims and also re-write independent claim 37 to have all the limitations of that other independent claims and cancel dependent claim 26 in order to put the application in condition for allowance. Applicant's representative Lance Smemoe (REG NO 66,152) has authorized the cancellation of dependent claim 26 and incorporating it's limitation to all the independent claims and also authorized make independent claim 37 have all the limitations from the other independent claims. Applicant has filed Terminal Disclaimer to overcome the double patenting rejection.

Notice of References Cited	Application/Control No. 14/181,112	Applicant(s)/Patent Under Reexamination Tannenbaum et al.	
	Examiner HERMON ASRES	Art Unit 2449	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-20050021751-A1	01-2005	Block, Timothy Roy	G06F9/54	709/225
*	B	US-20070073705-A1	03-2007	Gray; Theodore W.	G06F9/451	1/1
*	C	US-8601101-B1	12-2013	Singh; Servesh Pratap	H04L69/40	370/254
*	D	US-6968359-B1	11-2005	Miller; Robert	G06F9/5061	709/205
*	E	US-20070094532-A1	04-2007	SENGUPTA; Sabyasachi	G06F11/3632	714/5.1
*	F	US-20030195938-A1	10-2003	Howard, Kevin David	G06F8/45	709/208
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14/181,112	Tannenbaum et al.
	Examiner	Art Unit
	HERMON ASRES	2449

CPC - Searched*		
Symbol	Date	Examiner

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner
709	223	07/07/2016	HA

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
Searched East (US Patents, USPGPubs, EPO, JPO, IBM, Derwent)	07/07/2016	HA
See East Search attached	07/07/2016	HA
Inventor search performed	07/07/2016	HA
Consulted Justin Rider (QAS) regarding Alice 101 rejection	07/08/2016	HA
Searched updated	03/01/2017	HA
Search updated	02/18/2018	HA
Search updated	11/21/2018	HA
Interference search performed	11/21/2018	HA

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner
709	223	11/21/2018	HA

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14/181,112	Tannenbaum et al.
	Examiner	Art Unit
	HERMON ASRES	2449

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14/181,112	Tannenbaum et al.
	Examiner	Art Unit
	HERMON ASRES	2449

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	07/07/2016	03/01/2017	02/18/2018	11/22/2018					
	1	-	-	-	-					
	2	✓	✓	✓	=					
	3	✓	✓	✓	=					
	4	✓	✓	✓	=					
	5	✓	✓	✓	=					
	6	✓	✓	✓	=					
	7	✓	✓	✓	=					
	8	✓	✓	✓	=					
	9	✓	✓	✓	=					
	10	✓	✓	✓	=					
	11	✓	✓	✓	=					
	12		✓	✓	=					
	13		✓	✓	=					
	14		✓	✓	=					
	15		✓	✓	=					
	16		✓	✓	=					
	17		✓	✓	=					
	18		✓	✓	=					
	19		✓	✓	=					
	20		✓	✓	=					
	21		✓	✓	=					
	22		✓	✓	=					
	23		✓	✓	=					
	24		✓	✓	=					
	25		✓	✓	=					
	26		✓	✓	-					
	27		✓	✓	=					
	28		✓	✓	=					
	29		✓	✓	=					
	30		✓	✓	=					
	31		✓	✓	=					
	32			✓	=					
	33			✓	=					
	34			✓	=					
	35			✓	=					
	36			✓	=					
	37			✓	=					
	38			✓	=					
	39			✓	=					
	40			✓	=					
	41			✓	=					

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14/181,112	Tannenbaum et al.
	Examiner	**Art Unit**
	HERMON ASRES	2449

CPC						
Symbol					**Type**	**Version**
H04L	/	41	/	04	F	2013-01-01
G06F	/	9	/	5072	I	2013-01-01
G06F	/	9	/	54	I	2013-01-01
G06F	/	15	/	76	I	2013-01-01

CPC Combination Sets							
Symbol					**Type**	**Set**	**Ranking**	**Version**
	/		/					

None (Assistant Examiner)	(Date)	**Total Claims Allowed:** 39	
/HERMON ASRES/ Examiner, Art Unit 2449 (Primary Examiner)	22 November 2018 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure 1

U.S. Patent and Trademark Office — Part of Paper No.: 20181122

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14/181,112	Tannenbaum et al.
	Examiner	**Art Unit**
	HERMON ASRES	2449

INTERNATIONAL CLASSIFICATION				
CLAIMED				
G06F9/50	/	9	/	50
G06F9/54	/	9	/	54
G06F15/76	/	15	/	76
H04L12/24	/	12	/	24

NON-CLAIMED				
	/		/	

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**
709	223

CROSS REFERENCES(S)						
CLASS	**SUBCLASS (ONE SUBCLASS PER BLOCK)**					

None (Assistant Examiner) (Date)	**Total Claims Allowed:** 39	
/HERMON ASRES/ Examiner, Art Unit 2449 (Primary Examiner) (Date) 22 November 2018	O.G. Print Claim(s) 1	O.G. Print Figure 1

U.S. Patent and Trademark Office Part of Paper No.: 20181122

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14/181,112	Tannenbaum et al.
	Examiner	**Art Unit**
	HERMON ASRES	2449

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIMS

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original
1	2	10	11	19	20	28	30	37	39						
2	3	11	12	20	21	29	31	38	40						
3	4	12	13	21	22	30	32	39	41						
4	5	13	14	22	23	31	33								
5	6	14	15	23	24	32	34								
6	7	15	16	24	25	33	35								
7	8	16	17	25	27	34	36								
8	9	147	18	26	28	35	37								
9	10	18	19	27	29	36	38								

None (Assistant Examiner)	(Date)	**Total Claims Allowed:** 39	
/HERMON ASRES/ Examiner, Art Unit 2449 (Primary Examiner)	22 November 2018 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure 1

U.S. Patent and Trademark Office Part of Paper No.: 20181122

Bibliographic Data

Application No: 14/181,112

Foreign Priority claimed: ○ Yes ◉ No

35 USC 119 (a-d) conditions met: ☐ Yes ☑ No ☐ Met After Allowance

Verified and Acknowledged: Examiner's Signature | Initials

Title: CLUSTER COMPUTING

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
02/14/2014	709	2449	ZTANN.002P1C3
RULE			

APPLICANTS

Advanced Cluster Systems, Inc., Aliso Viejo, CA,

INVENTORS

Zvi Tannenbaum Newport Beach, CA, UNITED STATES

Dean E. Dauger Huntington Beach, CA, UNITED STATES

CONTINUING DATA

This application is a CON of 13423063 03/16/2012 PAT 8676877

13423063 is a CON of 12040519 02/29/2008 PAT 8140612

12040519 is a CIP of 11744461 05/04/2007 PAT 8082289

11744461 has PRO of 60850908 10/11/2006

11744461 has PRO of 60813738 06/13/2006

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

03/04/2014

**** SMALL ENTITY ****

STATE OR COUNTRY

UNITED STATES

ADDRESS

KNOBBE MARTENS OLSON & BEAR LLP
2040 MAIN STREET
FOURTEENTH FLOOR
IRVINE, CA 92614
UNITED STATES

FILING FEE RECEIVED

$1,900

PTO/SB/08 Equivalent

INFORMATION DISCLOSURE STATEMENT BY APPLICANT	Application No.	14/181112
	Filing Date	February 14, 2014
	First Named Inventor	Zvi Tannenbaum
	Art Unit	2449
(Multiple sheets used when necessary)	Examiner	Asres, Hermon
SHEET 1 OF 1	Attorney Docket No.	ZTANN.002P1C3

U.S. PATENT DOCUMENTS

Examiner Initials	Cite No.	Document Number *Number - Kind Code (if known)* Example: 1,234,567 B1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear
	1	2003/195938	10-16-2003	Howard	
	2	2004/015968	01-22-2004	Neiman	
	3	2004/098359	05-20-2004	Bayliss	
	4	2004/098447	05-20-2004	Verbeke	

FOREIGN PATENT DOCUMENTS

Examiner Initials	Cite No.	Foreign Patent Document *Country Code-Number-Kind Code* Example: JP 1234567 A1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear	T[1]
	5	JP2004-247405 (w/ machine translation)	09-02-2004	Semiconductor Energy Lab		x
	6	JP H11-126196 (w/ machine translation)	05-11-1999	Hitachi Ltd.		x

NON PATENT LITERATURE DOCUMENTS

Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[1]
	7	Roman E. Maeder "Mathematica Parallel Computing Toolkit – Unleash the Power of Parallel Computing", January 1, 2005, pages i-x, 1-90. URL:http//media.wolfam.com/documents/ParallelComputingToolkitDoumentation.pdf; retrieved on August 21, 2015.	
	8	European Search Report received in Application No. 17207443.7, dated September 20, 2018, in 9 pages.	

28546208

Examiner Signature	/HERMON ASRES/	Date Considered	11/22/2018

*__Examiner:__ Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.A./

T[1] - Place a check mark in this area when an English language Translation is attached.

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L15	2	("2003195938").PN.	US-PGPUB; USPAT; USOCR; FPRS; JPO; DERWENT; IBM_TDB	OR	OFF	2018/11/22 12:31
L16	6	("2004015968").PN.	US-PGPUB; USPAT; USOCR; FPRS; JPO; DERWENT; IBM_TDB	OR	OFF	2018/11/22 12:32
L17	5	("2004098359").PN.	US-PGPUB; USPAT; USOCR; FPRS; JPO; DERWENT; IBM_TDB	OR	OFF	2018/11/22 12:32
L18	6	("2004098447").PN.	US-PGPUB; USPAT; USOCR; FPRS; JPO; DERWENT; IBM_TDB	OR	OFF	2018/11/22 12:32
S1	6	((Zvi) near2 (Tannenbaum)).INV.	USPAT	OR	OFF	2015/07/24 20:25
S2	7	((Dean) near2 (Dauger)).INV.	USPAT	OR	OFF	2015/07/24 20:25
S3	177	(computer with cluster) same nodes same kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:26
S4	24	S3 and interpret	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:26
S5	28950	709/223.ccls.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:26
S6	5291	H04L41/0213.cpc.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:27
S7	1692	H04L29/08072.cpc.	US-PGPUB;	OR	OFF	2015/07/24

			USPAT; EPO; JPO; DERWENT; IBM_TDB			20:27
S8	27776	H04L29/06.cpc.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:27
S9	6793	H04L41/22.cpc.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S10	8156	H04L41/12.cpc.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S11	0	S3 and S6	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S12	0	S3 and S7	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S13	15	S3 and S8	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S14	0	S3 and S9	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S15	2	S3 and S10	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S16	99	("20020049859" | "20030005266" | "20030051062" | "20030135621" | "20030195931" | "20030195938" | "20040110209" | "20040157203" | "20040252710" | "20050021751" | "20050038852" | "20050060237" | "20050076105" | "20050108394" | "20050180095" | "20060053216" | "20060106931" | "20070044099" | "20070073705" | "20080250347" | "20080281997" | "5423046" | "5881315" | "5930768" | "6108699" | "6202080" | "6546403" | "6578068" | "6751698" | "6782537" | "6968335" | "7093004" | "7136924" | "7174381" | "7249357" | "7334232" | "7437460" | "7472193" | "7554949" | "7937455" | "8082289" |	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/25 12:54

		"8140612" | "8402083" | "8676877").PN.				
S17	177	(computer with cluster) same nodes same kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/25 13:19
S18	10	(computer with cluster) same nodes same kernel and mathematica	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/25 13:20
S19	177	(computer with cluster) same nodes same kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/25 13:23
S20	10	("8402080").PN.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/11/29 23:38
S21	2	("8849889").PN.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/11/29 23:38
S22	92	US-1234567-$.DID. OR US-5423046-$.DID. OR US-5881315-$.DID. OR US-5930768-$.DID. OR US-6108699-$.DID. OR US-6202080-$.DID. OR US-6546403-$.DID. OR US-6578068-$.DID. OR US-6751698-$.DID. OR US-6782537-$.DID. OR US-6968335-$.DID. OR US-7093004-$.DID. OR US-7136924-$.DID. OR US-7174381-$.DID. OR US-7249357-$.DID. OR US-7334232-$.DID. OR US-7437460-$.DID. OR US-7472193-$.DID. OR US-7554949-$.DID. OR US-7937455-$.DID. OR US-8082289-$.DID. OR US-8140612-$.DID. OR US-8402083-$.DID. OR US-8676877-$.DID. OR US-20020049859-$.DID. OR US-20030005266-$.DID. OR US-20030051062-$.DID. OR US-20030135621-$.DID. OR US-20030195931-$.DID. OR US-20030195938-$.DID. OR US-20040110209-$.DID. OR US-20040157203-$.DID. OR US-20040252710-$.DID. OR US-20050021751-$.DID. OR US-20050038852-$.DID. OR US-20050060237-$.DID. OR US-20050076105-$.DID. OR US-20050108394-$.DID. OR US-20050180095-$.DID. OR US-20060053216-$.DID. OR US-20060106931-$.DID. OR US-20070044099-$.DID. OR US-20070073705-$.DID. OR US-20080250347-$.DID. OR US-	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/11/29 23:39

		20080281997-$.DID. OR JP-1234567-$.DID. OR JP-0887473-$.DID. OR JP-2002117010-$.DID. OR US-6131999-$.DID.				
S23	4	(computer with cluster) same (nodes) same (single-node with kernel)	USPAT	OR	OFF	2016/06/17 19:19
S24	10	(computer with cluster) same (nodes) same (single-node with kernel)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/06/17 19:20
S25	11	(computer with cluster) same (nodes) and (single-node with kernel)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/06/17 19:20
S26	11	(computer with cluster) and (nodes) and (single-node with kernel)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/06/17 19:20
S27	224	(computer with cluster) and (nodes) and (single with kernel)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/06/17 19:20
S28	145	(computer with cluster) same (nodes) and (single with kernel)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/06/17 19:21
S29	6	(computer with cluster) same (nodes) and (single with kernel) and 709/223.ccls.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/06/17 19:22
S30	10	(computer with cluster) same (single with node with kernel)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/07 16:08
S31	79	(computer with cluster) and (single with node with kernel)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/07 18:41
S32	2	("8676877").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 12:43
S33	2	("8676877").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 15:47
S34	1505	(computer with cluster) and kernel and nodes and request	US-PGPUB; USPAT; EPO; JPO; DERWENT;	OR	OFF	2016/07/08 15:53

			IBM_TDB			
S35	271	S34 and mathematical	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 15:53
S36	178750	S34 and mathematical smae kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 15:53
S37	10	S34 and mathematical same kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 15:53
S38	2	("20050021751").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 15:54
S39	29	(single with node with kernel) same cluster	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 16:10
S40	22	(mathematical with kernel) same cluster	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 16:11
S41	22	(mathematical with kernel) same cluster	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 17:18
S42	2	("20050021751").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 17:24
S43	180800	mathematica kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 18:39
S44	11	mathematica near5 kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 18:39
S45	13	mathematica with kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 18:40
S46	10	("8402080").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT;	OR	OFF	2016/07/08 18:43

			IBM_TDB			
S47	2	("8849889").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 18:44
S48	101	("20020049859" | "20030005266" | "20030051062" | "20030135621" | "20030195931" | "20030195938" | "20040110209" | "20040157203" | "20040252710" | "20050021751" | "20050038852" | "20050060237" | "20050076105" | "20050108394" | "20050180095" | "20060053216" | "20060106931" | "20070044099" | "20070073705" | "20080250347" | "20080281997" | "5423046" | "5881315" | "5930768" | "6108699" | "6202080" | "6546403" | "6578068" | "6751698" | "6782537" | "6968335" | "7093004" | "7136924" | "7174381" | "7249357" | "7334232" | "7437460" | "7472193" | "7554949" | "7937455" | "8082289" | "8140612" | "8402083" | "8676877").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 18:44
S49	17	mathematica with cluster	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 18:50
S50	43	mathematica same cluster	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 18:51
S51	537678	stephen wolfram mathematica kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 19:42
S52	50	(stephen near2 wolfram).in.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 19:42
S53	1	S52 and mathematica and kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 19:43
S54	215	mathematica and kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 19:43
S55	74	S54 and cluster	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/08 19:43
S56	13	mathematica with kernel	US-PGPUB;	OR	OFF	2016/07/08

			USPAT; EPO; JPO; DERWENT; IBM_TDB			19:44
S57	0	("gridMathematica").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:16
S58	0	("gridwithMathematica").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:16
S59	19	grid with Mathematica	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:16
S60	20	gridMathematica	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:16
S61	14	gridMathematica and wolfram	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:21
S62	20	gridMathematica	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:22
S63	2578	Mathematica	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:22
S64	215	Mathematica and kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:22
S65	15749	wolfram	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:29
S66	8450	wolfram.in.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:29
S67	15	S66 and mathematica	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:30
S68	42	S66 and kernel	US-PGPUB;	OR	OFF	2016/07/09

			USPAT; EPO; JPO; DERWENT; IBM_TDB			10:30
S69	2	("20070073705").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:44
S70	2	("20050021751").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/07/09 10:46
S71	2	("20050021751").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2016/11/07 13:23
S72	2	("20050021751").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/03/01 19:24
S73	0	("(computerwithnodeswithclister)same(peer-to-peer)and@ad<20060101").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/03/01 19:47
S74	0	(computer with nodes with clister) same (peer-to-peer) and @ad<"20060101"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/03/01 19:47
S75	3	(computer with nodes with cluster) same (peer-to-peer) and @ad<"20060101"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/03/01 19:48
S76	2	("20050021751").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/03/02 17:12
S77	12	("8601101").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/03/02 17:12
S78	44	(computer with cluster) and kernel and (mathematical with expression)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/18 22:39
S79	0	(computer with clusters) same kernel same mathematical	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/18 22:47
S80	30	(computer with clusters) and kernel same	US-PGPUB;	OR	OFF	2018/02/18

		mathematical	USPAT; EPO; JPO; DERWENT; IBM_TDB			22:47
S81	9	(computer with clusters) same kernel same mathematica	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/18 22:48
S82	95	(computer with clusters) same mathematical	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/18 22:50
S83	2	(computer with clusters) same mathematical same expressions	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/18 22:50
S84	0	cluster near5 mathematical near5 evaluate	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:19
S85	2	cluster with mathematical with evaluate	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:20
S86	2	("20050038852").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:22
S87	20	("20050021751" | "20070073705" | "20070094532" | "6968359" | "8601101").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:35
S88	70	(US-0005068-$ | US-0015460-$ | US-0026601-$ | US-0059473-$ | US-0097300-$ | US-0098447-$ | US-0124363-$ | US-0154789-$ | US-0188088-$ | US-0254984-$ | US-1234567-$ | US-1368948-$ | US-20030005068-$ | US-20040098447-$ | US-20040254984-$ | US-20050015460-$ | US-20050026601-$ | US-20050097300-$ | US-20050154789-$ | US-20050188088-$ | US-20060059473-$ | US-20070124363-$ | US-25376674-$ | US-5195580-$ | US-6006259-$ | US-6067609-$ | US-6604134-$ | US-7031944-$ | US-7177874-$ | US-7315877-$ | US-7502915-$ | US-7519652-$ | US-7533141-$ | US-7533389-$ | US-7568131-$ | US-7598390-$ | US-7644130-$ | US-7757236-$ | US-7788314-$ | US-7835022-$ | US-8117288-$ | US-9405564-$).DID.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:36
S89	4802	(computer with cluster with nodes)	US-PGPUB; USPAT; EPO; JPO;	OR	OFF	2018/02/19 11:49

			DERWENT; IBM_TDB			
S90	30	(computer with cluster with nodes) same (mathematica$)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:50
S91	12	(computer with cluster with nodes) same (mathematica$) and kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:50
S92	230	(computer with cluster) same (mathematica$)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:51
S93	28	(computer with cluster) same (mathematica$) and kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:52
S94	230	(computer with cluster) same (mathematica$)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:54
S95	17	(computer with cluster) same (mathematica$) same result$3	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:55
S96	8	(computer with cluster) same (mathematica$ with expression)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 11:56
S97	15	(high with performance) same cluster same (mathematics mathematica)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 12:07
S98	114	(high with performance) same cluster and (mathematics mathematica)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 12:08
S99	2	("7418470").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 12:12
S100	2	("20030195938").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 12:13
S101	2	("20050021751").PN.	US-PGPUB; USPAT; EPO; JPO;	OR	OFF	2018/02/19 12:42

			DERWENT; IBM_TDB			
S102	12	("8601101").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 12:42
S103	2	("20030195938").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 12:42
S104	2	("20030195938").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/02/19 13:24

EAST Search History (Interference)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	5233	H04L41/0213.cpc.	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:28
L2	1567	H04L29/08072.cpc.	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:28
L3	26795	H04L29/06.cpc.	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:28
L4	8674	H04L41/22.cpc.	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:28
L5	11500	H04L41/12.cpc.	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:28
L6	0	((cluster with initialization) and (mathematical with expression) and (nodes with mathematical) and (single with node with kernel)).clm.	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:30
L7	0	((cluster with initialization) and (mathematical with expression) and (nodes with mathematical)).clm.	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:30
L8	1	((cluster with initialization) and (nodes with mathematical)).clm.	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:30
L9	177	((cluster with initialization)).clm.	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:31
L10	0	1 and 9	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:31
L11	0	2 and 9	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:31
L12	5	3 and 9	US-PGPUB;	OR	OFF	2018/11/22 12:31

			USPAT			
L13	0	4 and 9	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:31
L14	3	5 and 9	US-PGPUB; USPAT	OR	OFF	2018/11/22 12:31

11/ 22/ 2018 1:03:48 PM
C:\ Users\ hasres\ Documents\ EAST\ Workspaces\ 14181112.wsp

Doc Code: DIST.E.FILE **Document Description: Electronic Terminal Disclaimer - Filed**	PTO/SB/26 U.S. Patent and Trademark Office Department of Commerce

Electronic Petition Request	**TERMINAL DISCLAIMER TO OBVIATE A DOUBLE PATENTING REJECTION OVER A "PRIOR" PATENT**
Application Number	14181112
Filing Date	14-Feb-2014
First Named Inventor	Zvi Tannenbaum
Attorney Docket Number	ZTANN.002P1C3
Title of Invention	CLUSTER COMPUTING

☒ Filing of terminal disclaimer does not obviate requirement for response under 37 CFR 1.111 to outstanding Office Action

☒ This electronic Terminal Disclaimer is not being used for a Joint Research Agreement.

Owner	Percent Interest
Advanced Cluster Systems, Inc.	100%

The owner(s) with percent interest listed above in the instant application hereby disclaims, except as provided below, the terminal part of the statutory term of any patent granted on the instant application which would extend beyond the expiration date of the full statutory term of prior patent number(s)

8676877

as the term of said prior patent is presently shortened by any terminal disclaimer. The owner hereby agrees that any patent so granted on the instant application shall be enforceable only for and during such period that it and the prior patent are commonly owned. This agreement runs with any patent granted on the instant application and is binding upon the grantee, its successors or assigns.

In making the above disclaimer, the owner does not disclaim the terminal part of the term of any patent granted on the instant application that would extend to the expiration date of the full statutory term of the prior patent, "as the term of said prior patent is presently shortened by any terminal disclaimer," in the event that said prior patent later:
- expires for failure to pay a maintenance fee;
- is held unenforceable;
- is found invalid by a court of competent jurisdiction;
- is statutorily disclaimed in whole or terminally disclaimed under 37 CFR 1.321;
- has all claims canceled by a reexamination certificate;
- is reissued; or
- is in any manner terminated prior to the expiration of its full statutory term as presently shortened by any terminal disclaimer.

◉ Terminal disclaimer fee under 37 CFR 1.20(d) is included with Electronic Terminal Disclaimer request.

○ I certify, in accordance with 37 CFR 1.4(d)(4), that the terminal disclaimer fee under 37 CFR 1.20(d) required for this terminal disclaimer has already been paid in the above-identified application.	
Applicant claims the following fee status: ◉ Small Entity ○ Micro Entity ○ Regular Undiscounted I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.	
THIS PORTION MUST BE COMPLETED BY THE SIGNATORY OR SIGNATORIES I certify, in accordance with 37 CFR 1.4(d)(4) that I am: ◉ An attorney or agent registered to practice before the Patent and Trademark Office who is of record in this application Registration Number 70256 ○ A sole inventor ○ A joint inventor; I certify that I am authorized to sign this submission on behalf of all of the inventors as evidenced by the power of attorney in the application ○ A joint inventor; all of whom are signing this request	
Signature	/Damien J. Howard/
Name	Damien J. Howard

*Statement under 37 CFR 3.73(b) is required if terminal disclaimer is signed by the assignee (owner).
Form PTO/SB/96 may be used for making this certification. See MPEP § 324.

Electronic Patent Application Fee Transmittal

Application Number:	14181112
Filing Date:	14-Feb-2014
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Filer:	Damien J. Howard/Leonardo Quintana
Attorney Docket Number:	ZTANN.002P1C3

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
STATUTORY OR TERMINAL DISCLAIMER	2814	1	160	160
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**160**

Doc Code: DISQ.E.FILE
Document Description: Electronic Terminal Disclaimer – Approved

Application No.: 14181112

Filing Date: 14-Feb-2014

Applicant/Patent under Reexamination: Tannenbaum

Electronic Terminal Disclaimer filed on November 21, 2018

☒ APPROVED

This patent is subject to a terminal disclaimer

☐ DISAPPROVED

Approved/Disapproved by: Electronic Terminal Disclaimer automatically approved by EFS-Web

U.S. Patent and Trademark Office

Electronic Acknowledgement Receipt	
EFS ID:	34373895
Application Number:	14181112
International Application Number:	
Confirmation Number:	6874
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Customer Number:	20995
Filer:	Damien J. Howard/Leonardo Quintana
Filer Authorized By:	Damien J. Howard
Attorney Docket Number:	ZTANN.002P1C3
Receipt Date:	21-NOV-2018
Filing Date:	14-FEB-2014
Time Stamp:	14:01:20
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$160
RAM confirmation Number	112318INTEFSW14011800
Deposit Account	111410
Authorized User	Leonardo Quintana
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.16 (National application filing, search, and examination fees) 37 CFR 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Terminal Disclaimer-Filed (Electronic)	eTerminal-Disclaimer.pdf	33382 7d0dd8aa460a29b20acf220bd97b2d20519a9803	no	2
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	30331 a1b8d3930a4f2b702c2d1c7490db50e9837d280b	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	63713		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT	Application No.	14/181112
	Filing Date	February 14, 2014
	First Named Inventor	Zvi Tannenbaum
	Art Unit	2449
(Multiple sheets used when necessary)	Examiner	Asres, Hermon
SHEET 1 OF 1	Attorney Docket No.	ZTANN.002P1C3

U.S. PATENT DOCUMENTS					
Examiner Initials	Cite No.	Document Number *Number - Kind Code (if known)* Example: 1,234,567 B1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear
	1	2003/195938	10-16-2003	Howard	
	2	2004/015968	01-22-2004	Neiman	
	3	2004/098359	05-20-2004	Bayliss	
	4	2004/098447	05-20-2004	Verbeke	

FOREIGN PATENT DOCUMENTS						
Examiner Initials	Cite No.	Foreign Patent Document *Country Code-Number-Kind Code* Example: JP 1234567 A1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear	**T[1]**
	5	JP2004-247405 (w/ machine translation)	09-02-2004	Semiconductor Energy Lab		x
	6	JP H11-126196 (w/ machine translation)	05-11-1999	Hitachi Ltd.		x

NON PATENT LITERATURE DOCUMENTS			
Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	**T[1]**
	7	Roman E. Maeder "Mathematica Parallel Computing Toolkit – Unleash the Power of Parallel Computing", January 1, 2005, pages i-x, 1-90. URL:http//media.wolfam.com/documents/ParallelComputingToolkitDoumentation.pdf; retrieved on August 21, 2015.	
	8	European Search Report received in Application No. 17207443.7, dated September 20, 2018, in 9 pages.	

28546208

Examiner Signature	Date Considered

***Examiner:** Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

T[1] - Place a check mark in this area when an English language Translation is attached.

Espacenet

Bibliographic data: JP2004247405 (A) — 2004-09-02

SEMICONDUCTOR DEVICE

Inventor(s): KATO KIYOSHI; TAKAYAMA TORU; MARUYAMA JUNYA; GOTOU YUUGO; ONO YUMIKO ± (KATO KIYOSHI, ; TAKAYAMA TORU, ; MARUYAMA JUNYA, ; GOTOU YUUGO, ; ONO YUMIKO)

Applicant(s): SEMICONDUCTOR ENERGY LAB ± (SEMICONDUCTOR ENERGY LAB CO LTD)

Classification: **- international:** ***H01L21/02; H01L21/77; H01L21/822; H01L21/84; H01L27/00; H01L27/06; H01L27/12; H01L27/146; H01L27/15; H01L27/32; H01L29/786; H01L31/12;*** (IPC1-7): H01L27/00; H01L27/12; H01L27/146; H01L27/15; H01L29/786; H01L31/12

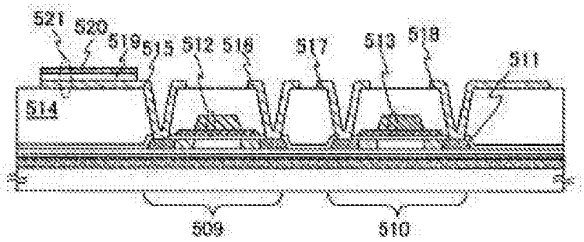
- cooperative: **H01L21/8221**; **H01L27/0688**; **H01L27/1214**; **H01L27/1266**; **H01L27/3227**; **H01L29/78603**; H01L27/3209; H01L29/78621

Application number: JP20030033833 20030212

Priority number(s): JP20030033833 20030212

Also published as: CN1523668 (A) CN1523668 (B) JP4574118 (B2) TW200527693 (A) TWI327377 (B) more

Abstract of JP2004247405 (A)



PROBLEM TO BE SOLVED: To provide a semiconductor device which can deal with the increase of an amount of information by using a low-cost glass board and which has an integrated circuit capable of being operated at a high speed with high performance. ;SOLUTION: The semiconductor device includes semiconductor elements laminated by transferring the semiconductor elements formed on different substrates. In the semiconductor device, a film formed of a resin and a partially provided metal oxide exist between the laminated semiconductor elements. A light emitting element electrically connected to the one of the laminated semiconductor elements converts a first electric signal into an optical signal. A photodetecting element electrically connected to the different one of the

laminated semiconductor elements converts the optical signal into a second electric signal. ;COPYRIGHT: (C)2004,JPO&NCIPI

Patent Translate
Powered by EPO and Google

Notice

This translation is machine-generated. It cannot be guaranteed that it is intelligible, accurate, complete, reliable or fit for specific purposes. Critical decisions, such as commercially relevant or financial decisions, should not be based on machine-translation output.

DESCRIPTION JP2004247405

PROBLEM TO BE SOLVED: To provide a semiconductor device having an integrated circuit capable of coping with an increase in information amount, using an inexpensive glass substrate, and capable of high-speed operation at high speed. SOLUTION: This semiconductor device has a semiconductor element laminated by transferring semiconductor elements formed on different substrates, and between the stacked semiconductor elements, a film formed of resin and a part In the light emitting element electrically connected to one of the stacked semiconductor elements in which a metal oxide provided in advance is present, the first electric signal is converted into an optical signal, and the difference in the stacked semiconductor elements And converting the optical signal into a second electric signal in a light receiving element electrically connected to the other. (FIG.

Semiconductor device

BACKGROUND OF THE INVENTION 1. Field of the Invention The present invention relates to an integrated circuit (thin film integrated circuit) using a semiconductor film having a crystal structure formed on a glass substrate, and more particularly to an integrated circuit (thin film integrated circuit) , A thin film integrated circuit formed by optical interconnection, and a semiconductor device using the thin film integrated circuit. As the minimum dimension of miniaturization approaches the submicron range, the integration of LSIs has been stopping. If the degree of high integration is insufficient, since the wiring resistance can not be suppressed, signal delay, waveform distortion, and signal transmission amount are reduced, and the information processing capability of the integrated circuit is limited. Also, if the parasitic capacitance between the wirings increases along with the increase in wiring length, there is also a problem that power consumption can not be suppressed because charging and discharging energy to the wiring increases. Therefore, in order to further advance the high integration, a

technique concerning a three-dimensional integrated circuit attracts attention. A three-dimensional integrated circuit is a technique of hierarchically stacking semiconductor elements with an interlayer insulating film interposed therebetween, realizing dramatically higher integration of integrated circuits, higher speed and multifunctionalization by shortening wiring length We are trying to do. In order to form a three-dimensional integrated circuit, it is necessary to electrically connect the semiconductor elements formed in each layer. The electrical connection between the semiconductor elements between the respective layers can be performed by a wire bonding method, a flip chip method, or the like. Further, via holes can be formed in the interlayer insulating film separating the respective layers and electrical connection can be made. It is also possible to use a so-called optical interconnection in which an electric signal is converted into an optical signal by using a light emitting element and connection is performed by optical transmission. [0005] Japanese Patent Application Laid-Open Publication No. 2002-208587 discloses a technique for transmitting data between a CPU, a memory controller, and a memory using an optical signal. As described in Japanese Patent Application Laid-Open No. 2000-58882, data transmission is performed using optical signals, so that skew caused by wiring resistance, Unnecessary radio wave radiation can be reduced. As described above, three-dimensional integration of an integrated circuit is an important technique for realizing further higher integration, but further increase in the amount of information processed by a semiconductor device There is a problem that it becomes difficult to connect each layer if the number of layers to be stacked is increased in order to deal with this problem.

For example, in the case of the wire bonding method, it is easy to connect neighboring layers with thin metal wires, but when two layers to be connected are separated, they exist between the layers so as not to hinder the formation of fine metal wires used for connection It is necessary to limit the shape of the layer. Therefore, as multilayering progresses and the number of layers existing between the two layers to be connected increases, it becomes difficult to connect using thin metal wires. Also, because metal thin wires are usually bonded in the vicinity of the edges of each layer, the area where metal thin wires are connected is limited, and the number of terminals of each layer is limited. In the case of the flip chip method, since the area used for electrical connection is not limited, it is possible to provide a relatively large number of terminals of each layer as compared with the wire bonding method. However, as with the wire bonding method, it is suitable for connecting neighboring layers, but if the two layers to be connected are separated, the shape of the layer existing therebetween is changed so as not to hinder the connection by the solder ball It is necessary to limit it, and it is not suitable for multilayering. Also, in the case of forming via holes in the interlayer insulating film and connecting them, if the two layers to be connected are separated, it is necessary to form a via hole penetrating all the layers existing therebetween. As the thickness of the layer penetrating the via hole increases, the diameter of the via hole needs to be increased, which is not preferable because it hinders high integration of the wiring. In the case of using optical transmission, unlike the other cases described above, even if the two layers to be connected are separated by multilayering, by using an optical system such as

an optical fiber, the shape of another layer can be changed It is possible to perform interlayer connection without restriction. In general, however, optical fibers are expensive and, from the viewpoint of the cost of a semiconductor device using an integrated circuit, it is not a preferable member for use. In addition, if a thin film integrated circuit composed of a plurality of layers is continuously formed on one substrate, optical transmission can be easily performed without using an optical system such as an optical fiber. However, when crystallizing a semiconductor film used for a semiconductor element in an upper layer, there is a restriction that it is performed at a temperature lower than the heat resistance temperature of the semiconductor element in the lower layer, and there are many process restrictions, and it is difficult to form a thin film integrated circuit with higher characteristics difficult. In view of the above problems, it is an object of the present invention to provide a semiconductor device having an integrated circuit which can cope with an increase in the amount of information by using an inexpensive glass substrate and which can perform high-speed operation at high speed. In order to solve the above problems, the present invention provides a thin film integrated circuit using a semiconductor thin film formed on a substrate having heat resistance, and the thin film integrated circuit Is transferred onto a separately prepared substrate.

By repeating this transfer, a plurality of thin film integrated circuits are stacked to achieve three-dimensionalization. The substrate for forming the thin film integrated circuit before transfer may have heat resistance enough to withstand the heat treatment in the manufacturing process of the semiconductor element. The thin film integrated circuits (hereinafter referred to as thin film chips) of the respective layers are connected by so-called optical interconnection using optical signals. That is, a light emitting element is provided on the output side of the upper stage circuit of the stacked thin film chips, and a light receiving element is formed on the input side of the subsequent stage circuit so as to face the light emitting element. Then, the electric signal outputted from the upper stage circuit is converted into an optical signal in the light emitting element, the light receiving element which received the optical signal converts it into an electric signal, and inputs it into the circuit in the subsequent stage. Transfer of a thin film chip is carried out by providing a metal oxide film between the substrate and the thin film integrated circuit, weakening the metal oxide film by crystallization, peel off the thin film integrated circuit, and transfer the thin film chip, , The amorphous silicon film is removed by irradiation with laser light or etching so that the substrate and the thin film integrated circuit are peeled off and transferred, a method in which a thin film integrated circuit is formed A method in which the thin film integrated circuit is separated from the substrate by transfer by mechanically removing the substrate by etching or etching with a solution or gas, and transferring the transferred substrate. For example, transfer using a metal oxide film can be performed according to the following method. First, a first substrate having heat resistance enough to withstand the heat treatment in the manufacturing process of the semiconductor device is prepared. Then, a metal film is formed on the first substrate, and the surface of the metal film is oxidized to form a very thin metal oxide film of several nm. Next, an insulating film and a semiconductor film are formed in this order on

the metal oxide film. The insulating film may be a single layer or a stack of a plurality of films. For example, silicon nitride, silicon nitride oxide, silicon oxide, or the like can be used. Then, using the semiconductor film, a semiconductor element used for a thin film integrated circuit is manufactured. After the semiconductor element is formed, a second substrate is bonded so as to cover the semiconductor element, and a state where the semiconductor element is sandwiched between the first substrate and the second substrate is formed. In the case of forming a display device in parallel with the thin film integrated circuit, the second substrate is bonded before the display element is completed. For example, when a liquid crystal cell is used as a display element, a pixel electrode of a liquid crystal cell electrically connected to a thin film transistor (TFT) which is one of semiconductor elements and an alignment film covering the pixel electrode are manufactured, Before pasting the counter substrate on which the counter electrode is formed.

A third substrate is bonded on the side opposite to the side on which the semiconductor element of the first substrate is formed in order to reinforce the rigidity of the first substrate. The higher the rigidity of the first substrate than the second substrate, the less susceptible the damage to the semiconductor element when peeling off the first substrate, so that it can be peeled off smoothly. However, when stiffness of the first substrate is sufficient when peeling off the first substrate from the semiconductor element later, the third substrate does not necessarily need to be bonded. Next, the metal oxide film is crystallized by heat treatment or the like to increase the brittleness of the metal oxide film, so that the first substrate is easily separated from the semiconductor element. Then, the first substrate together with the third substrate is peeled off from the semiconductor element. Note that the heat treatment for crystallizing the metal oxide film may be performed before bonding the third substrate or before bonding the second substrate. Alternatively, the heat treatment performed in the step of forming the semiconductor element may also serve as a step of crystallizing the metal oxide film. Then, the first substrate and the third substrate are peeled off from the semiconductor element. Due to this peeling, a portion to be separated between the metal film and the metal oxide film, a portion to be separated between the insulating film and the metal oxide film, and a portion where the metal oxide film itself is separated are generated. In either case, the semiconductor element is peeled off from the first substrate so as to stick to the second substrate side. Next, by peeling off the first substrate, the semiconductor element stuck to the second substrate side is bonded to a substrate (hereinafter referred to as an element substrate) for forming a semiconductor device Attach them with agents. Then, the second substrate is peeled off, and the semiconductor element is transferred to the element substrate. A thin film chip is formed by a semiconductor element fixed to the element substrate. The light receiving element and the light emitting element for connecting the thin film chip and another chip may be formed on the element substrate in advance or may be formed after the thin film chips are pasted together. Alternatively, a ready-made light emitting element or light receiving element may be attached and electrically connected to the thin film chip. Similarly, a thin film integrated circuit constituting the thin film chip of the next layer is formed on a separately prepared substrate and transferred again so as to

be superimposed on the already transferred thin film chip. By repeating the transfer in this manner, the thin film chip can be laminated on the element substrate. At this time, an interlayer insulating film may be formed of resin or the like between the respective thin film chips.

Further, an adhesive used at the time of transfer may be used instead of the interlayer insulating film. In the case of transferring a display device together with a thin film chip, a display element is manufactured after transfer. This also applies to the case of manufacturing a light-emitting element after transfer. For example, in the case of a liquid crystal display device, specifically, a plastic substrate (hereinafter referred to as a counter substrate) on which a counter electrode, a second polarizing plate and the like are formed is prepared separately, and the liquid crystal To complete the liquid crystal cell. In addition, not only the counter electrode and the second polarizing plate, but also a color filter, an alignment film, a black matrix and the like may be prepared on the opposing substrate. In the present invention, the thickness of one thin film chip can be set to a total film thickness of 5 μm, more desirably 2 μm or less. As described above, according to the present invention, since the thin film chips of the respective layers can be manufactured separately, there is no restriction imposed on the process of the semiconductor element of the upper layer in consideration of the heat resistance of the underlying semiconductor element, Excellent semiconductor devices can be formed. Unlike the case where an integrated circuit (hereinafter referred to as a thin film integrated circuit) formed of a thin film semiconductor film on a glass substrate is superimposed to form a three-dimensional structure, the thin film chips of each layer are formed of a glass substrate or the like Without being interposed therebetween, it is not necessary to consider light diffusion by the glass substrate. Further, since the distance between the light emitting element and the light receiving element can be shortened, even if the light directivity of the light emitting element is inferior, it can be covered to some extent. A plastic substrate having flexibility can be used for the element substrate. As the plastic substrate, ARTON: JSR made of a norbornene resin having a polar group can be used. In addition, it is also possible to use polyethylene terephthalate (PET), polyethersulfone (PES), polyethylene naphthalate (PEN), polycarbonate (PC), nylon, polyetheretherketone (PEEK), polysulfone (PSF), polyetherimide (PEI), poly Plastic substrates such as arylate (PAR), polybutylene terephthalate (PBT), polyimide and the like can be used. By using a plastic substrate having flexibility, it is possible to manufacture a thin film integrated circuit which is superior in mechanical strength to vibration and impact as compared with a glass substrate and which is reduced in weight, and can be used as a thin film integrated circuit It is easy to suppress the thickness of.

Further, since the element substrate has flexibility, the degree of freedom of the shape of the thin film integrated circuit is increased. Therefore, it is also possible to form a thin film integrated circuit in a shape having a curved surface, for example, such as being attached to a columnar bottle or the like. By using the thin film integrated circuit of the present invention in a

semiconductor device, it is possible to secure a wider space for use in an integrated circuit, and to realize a high function without impeding reduction in weight or size of a semiconductor device be able to. Particularly in the case of a portable semiconductor device, convenience is dramatically improved by weight saving and downsizing, so that it is very useful to use the thin film integrated circuit of the present invention. By using optical interconnection, data can be transferred between the thin film chips while reducing the influence of skew caused by the wiring resistance and unnecessary radio wave radiation. Further, in the process of converting an electric signal to an optical signal and converting it from an optical signal to an electric signal again, it is possible to freely control the amplitude of the finally obtained electric signal. Further, since signals of one light emitting element can be received by a plurality of light receiving elements, signals can be transmitted from one thin film chip to a plurality of other thin film chips at the same time, and very high speed light A bus can be formed. Since each thin film chip transmits light unlike a single crystal silicon wafer, it is relatively easy to transmit signals between three or more thin film chips without using a special optical system such as an optical fiber For example. Note that the TFT formed on the glass substrate is inferior in operating speed to the single crystal transistor. However, by using an optical signal for signal transmission between thin film chips, it is possible to take a large bus width of a signal transmitted between the thin film chips and to make efficient parallel operation in a circuit on a plurality of thin film chips It is possible to cover the low operation speed of the TFT formed on the glass substrate. Even if the degree of integration of the circuit formed on the glass substrate is lower than that of the single crystal silicon wafer, by using the optical signal for signal transmission between the substrates, a plurality of thin film chips are stacked So that it is possible to prevent the thin film integrated circuit from being bulky in the horizontal direction. Further, it is possible to prevent the wiring from being extended over a long distance and suppress an increase in power consumption caused by the wiring capacitance. In addition, the process can be easily changed for each thin film chip.

Therefore, unlike the case where all the circuits are simultaneously formed on one substrate, it is possible to select a process according to characteristics required for the thin film chips of each layer. In addition, cost can be reduced by using an inexpensive glass substrate. Since a single integrated circuit is formed by combining separately prepared thin film chips, the yield can be increased as compared with the case where an integrated circuit is formed on one substrate at the same time. Further, by using an optical signal for transmission of signals between the thin film chips, the number of terminals such as FPC for electrically connecting the circuits can be suppressed and the reliability in the mechanical strength can be enhanced . Furthermore, even if the information amount of the signal to be processed increases, it is possible to suppress a decrease in yield due to the occurrence of a contact failure in the terminal portion. Since the light emitting element and the light receiving element which transmit and receive the optical signal do not necessarily need to be arranged at the end of the substrate unlike the terminal of the FPC, restrictions on the layout are reduced and the amount of information to be processed is further

increased It is easy to handle increase. As described above, the present invention makes it possible to provide a semiconductor device having an integrated circuit capable of high performance and high speed operation. DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS A specific method of manufacturing a semiconductor element used in a thin film integrated circuit of the present invention and a method of bonding (transferring) the semiconductor element to an element substrate will be described below. In this embodiment, two TFTs are described as an example of a semiconductor element, but the semiconductor element included in the thin film integrated circuit is not limited thereto, and any circuit element can be used. For example, in addition to the TFT, a storage element (memory), a diode, a photoelectric conversion element, a resistance element, a coil, a capacitive element, an inductor, and the like are representatively exemplified. First, as shown in FIG. 1A, a metal film 501 is formed over a first substrate 500 by a sputtering method. Here, tungsten is used for the metal film 501, and the film thickness is set to 10 nm to 200 nm, preferably 50 nm to 75 nm. Note that in this embodiment mode, the metal film 501 is formed directly on the first substrate 500. After covering the first substrate 500 with an insulating film such as silicon oxide, silicon nitride, silicon nitride oxide, or the like, a metal film 501 may be formed. After the formation of the metal film 501, the oxide film 502 is deposited so as not to be exposed to the atmosphere. Here, a silicon oxide film is formed as the oxide film 502 so as to have a film thickness of 150 nm to 300 nm.

In the case of using the sputtering method, film formation is also performed on the end surface of the first substrate 500. Therefore, in order to prevent the oxide film 502 from remaining on the side of the first substrate 500 at the time of peeling in a later process, the metal film 501 and the oxide film 502 formed on the end face are oxidized by O 2 ashing or the like , Or it is preferable to cut the end portion of the first substrate 500 by dicing or the like. When the oxide film 502 is formed, pre-sputtering is carried out, in which plasma is generated by shutting the space between the target and the substrate with a shutter as a preliminary step of sputtering. The pre-sputtering is performed with a flow rate of 10 sccm for Ar and 30 sccm for O 2, keeping the temperature of the first substrate 500 at 270 ° C. and the deposition power at 3 kW parallel state. By pre-sputtering, a metal oxide film 503 having an extremely thin thickness of about several nanometers (here, 3 nm) is formed between the metal film 501 and the oxide film 502. The metal oxide film 503 is formed by oxidizing the surface of the metal film 501. Therefore, in the present embodiment, the metal oxide film 503 is formed of tungsten oxide. In the present embodiment, the metal oxide film 503 is formed by pre-sputtering, but the present invention is not limited thereto. For example, an inert gas such as Ar may be added to oxygen or oxygen to intentionally oxidize the surface of the metal film 501 by plasma to form the metal oxide film 503. Next, after forming the oxide film 502, the base film 504 is formed by the PCVD method. Here, as the base film 504, a silicon oxynitride film is formed so as to have a film thickness of about 100 nm. After forming the base film 504, the semiconductor film 505 is formed without being exposed to the atmosphere. The film thickness of the semiconductor film 505 is set to 25

to 100 nm (preferably 30 to 60 nm). Note that the semiconductor film 505 may be an amorphous semiconductor or a polycrystalline semiconductor. In addition, not only silicon but also silicon germanium can be used as a semiconductor. When silicon germanium is used, the concentration of germanium is preferably about 0.01 to 4.5 atomic%. Next, the semiconductor film 505 is crystallized by a known technique. Known crystallization methods include a thermal crystallization method using an electrothermal furnace, a laser crystallization method using laser light, and a lamp annealing crystallization method using infrared light. Alternatively, a crystallization method using a catalytic element can be used according to the technique disclosed in JP-A No. 7-130652.

In addition, the semiconductor film 505 which is a polycrystalline semiconductor film may be formed in advance by a sputtering method, a plasma CVD method, a thermal CVD method, or the like. In the present embodiment, the semiconductor film 505 is crystallized by laser crystallization. By using a solid-state laser capable of continuous oscillation and irradiating laser light of second harmonic to fourth harmonic of the fundamental wave, crystals of large grain size can be obtained. For example, typically, it is preferable to use the second harmonic (532 nm) or the third harmonic (355 nm) of an Nd: YVO 4 laser (fundamental wave 1064 nm). Specifically, laser light emitted from a continuous oscillation YVO 4 laser is converted into harmonic waves by a nonlinear optical element, and laser light with an output of 10 W is obtained. There is also a method of emitting harmonics by using a nonlinear optical element. Then, preferably, the semiconductor film 505 is irradiated with a rectangular or elliptical laser beam formed on the irradiation surface by an optical system. The energy density at this time is about 0.01 to 100 MW / cm <2> (preferably 0.1 to 10 MW / cm <2>). Then, the scanning speed is set to about 10 to 2000 cm / s, and irradiation is performed. It is to be noted that laser crystallization may be performed by irradiating continuous wave fundamental wave laser light and continuous wave harmonic laser light, or continuous wave fundamental wave laser light and pulse oscillation high harmonic And a laser beam of a wave may be irradiated. Note that laser light may be irradiated in an inert gas atmosphere such as a rare gas or nitrogen. This makes it possible to suppress roughening of the semiconductor surface due to laser light irradiation and to suppress variations in threshold caused by variations in interface state density. By the irradiation of the semi-laser beam, the crystallinity of the conductor film 505 is further enhanced. Next, as shown in FIG. 1 (B), the semiconductor film 505 having higher crystallinity is patterned to form island-shaped semiconductor films 507 and 508, and the island-shaped semiconductor films 507 and 508 are used To form various semiconductor elements typified by TFTs. Note that in this embodiment, the base film 504 and the island-shaped semiconductor films 507 and 508 are in contact with each other, but depending on the semiconductor element, an electrode or the like is provided between the base film 504 and the island-shaped semiconductor films 507 and 508 Other insulating films or the like may be formed. For example, in the case of a bottom gate type TFT which is one of semiconductor elements, a gate electrode and a gate insulating film are formed between the base film 504 and the island-like semiconductor films 507 and 508.

In the present embodiment, top-gate TFTs 509 and 510 are formed using island-shaped semiconductor films 507 and 508 (FIG. 1C). Specifically, the gate insulating film 511 is formed so as to cover the island-shaped semiconductor films 507 and 508. Then, a conductive film is formed over the gate insulating film 511 and patterned to form gate electrodes 512 and 513. Then, by using the gate electrodes 512 and 513 or a patterned film of a resist as a mask, an impurity imparting n-type conductivity is added to the island-shaped semiconductor films 507 and 508 to form a source region, a drain region, LDD region and the like are formed. Note that although the TFTs 509 and 510 are n-type here, in the case of a p-type TFT, an impurity imparting p-type conductivity is added. The TFTs 509 and 510 can be formed by the above series of steps. Note that the manufacturing method of the TFT is not limited to the above-described steps. Next, a first interlayer insulating film 514 is formed so as to cover the TFTs 509 and 510. Contact holes are formed in the gate insulating film 511 and the first interlayer insulating film 514 and then terminals 515 to 518 connected to the TFTs 509 and 510 through the contact holes are formed so as to be in contact with the first interlayer insulating film 514 Form. A part of the terminal 515 is used as a cathode electrode of a light receiving element (photoelectric conversion element). As the terminals 515 to 518, aluminum formed by a sputtering method is used, but the material is not limited thereto. Other metals such as titanium, tantalum, tungsten, and copper can be used. Alternatively, a laminated film of titanium, aluminum, or titanium may be used. Next, an amorphous silicon film containing hydrogen is formed over the entire surface of the substrate and then patterned to form a photoelectric conversion layer 519. Then, a transparent conductive film is formed on the entire surface of the substrate. In this embodiment, ITO having a thickness of 200 nm is formed as a transparent conductive film by a sputtering method. The transparent conductive film is patterned to form an anode electrode 520 (FIG. 1 (C)). A portion where the terminal 515 used as the cathode electrode, the photoelectric conversion layer 519, and the anode electrode 520 overlap functions as the light receiving element 521. The configuration of the light receiving element is not limited to the configuration shown in FIG. 1 (C). The designer can appropriately determine the configuration of the light receiving element according to the incident direction of the light received by the light receiving element.

Then, a second interlayer insulating film 523 is formed over the first interlayer insulating film 514 so as to cover the light receiving element 521 and the terminals 515 to 518. Note that the second interlayer insulating film 523 is provided to protect the light receiving element 521 in the later bonding step, and it is not always necessary to form a film. The protective layer 524 to be formed later may play its role. Next, a protective layer 524 is formed to cover the second interlayer insulating film 523. The protective layer 524 has a function of protecting the semiconductor element when the second substrate is bonded or peeled later, and a material that can be removed after peeling off the second substrate is used. For example, a protective layer

524 can be formed by applying an epoxy-based, acrylate-based, or silicone-based resin soluble in water or alcohols to the entire surface and baking. In this embodiment, a water-soluble resin (VL-WSHL 10, manufactured by Toagosei Co., Ltd.) is applied by spin coating so as to have a film thickness of 30 μm, exposure is performed for 2 minutes for temporary curing, UV light Is exposed for 2.5 minutes from the back surface and 10 minutes from the front for a total of 12.5 minutes to be fully cured to form a protective layer 524 (FIG. 1 (D)). In the case where a plurality of organic resins are laminated, there is a possibility that organic resins are partly dissolved at the time of coating or firing, or adhesiveness becomes too high depending on the solvent used. Therefore, when an organic resin soluble in the same solvent is used for the second interlayer insulating film 523 and the protective layer 524, in order to smoothly remove the protective layer 524 in a later step, the second interlayer insulating film An inorganic insulating film (SiNx film, SiN x O y film, AlN x film, or AlN x O y film) is preferably formed between the protective film 523 and the protective layer 524 to be formed later. Next, in order to facilitate subsequent peeling, the metal oxide film 503 is crystallized. Due to the crystallization, the metal oxide film 503 easily cracks at the grain boundary, and brittleness can be enhanced. In the present embodiment, heat treatment is performed for about 0.5 to 5 hours at 420 ° C. to 550 ° C. to perform crystallization. Next, the adhesion between the metal oxide film 503 and the oxide film 502 or the adhesion between the metal oxide film 503 and the metal film 501 is partly lowered, and the part which becomes the trigger for the start of peeling is defined as And carries out processing to be performed. Specifically, a local pressure is externally applied along the periphery of the region to be peeled to damage the inside of the layer of the metal oxide film 503 or a part of the vicinity of the interface.

In this embodiment, a hard needle such as a diamond pen or the like is vertically pressed near the end portion of the metal oxide film 503 and moved along the metal oxide film 503 in a state where a load is applied as it is. Preferably, a scriber device is used, the pushing amount is set to 0.1 mm to 2 mm, and pressure is applied to move it. In this manner, by forming a portion with reduced adhesiveness, which will trigger peeling before peeling off, it is possible to reduce defects in the subsequent peeling step, leading to an improvement in yield . Next, using the double-sided tape 525, the second substrate 526 is attached to the protective layer 524, and the third substrate 528 is attached to the first substrate 500 using the double-sided tape 527 (FIG. 1 (E)). Note that an adhesive may be used instead of a double-sided tape. For example, by using an adhesive that is peeled off by ultraviolet rays, it is possible to reduce the burden on the semiconductor element at the time of peeling off the second substrate. By attaching the third substrate 528, it is possible to prevent the first substrate 500 from being damaged in the subsequent peeling step. As the second substrate 526 and the third substrate 528, it is preferable to use a substrate which is almost as rigid as the first substrate 500, for example, a quartz substrate or a semiconductor substrate. Next, the metal film 501 and the oxide film 502 are physically peeled off. In the previous step, the peeling starts from a region in which the adhesion of the metal oxide film 503 to the metal film 501 or the oxide film 502 partially

deteriorates. By the peeling-off, a portion separating between the metal film 501 and the metal oxide film 503, a portion separating between the oxide film 502 and the metal oxide film 503, and the metal oxide film 503 themselves are separated to each other Part. Then, semiconductor elements (here, TFTs 509 and 510) are separated from the second substrate 526 side while the first substrate 500 and the metal film 501 are bonded to the third substrate 528 side, respectively. Peeling can be done with a relatively small force (eg, human hand, wind pressure of gas blown from the nozzle, ultrasonic wave, etc.). FIG. 2 (A) shows the state after peeling. Next, the substrate (element substrate) 530 on which the thin film integrated circuit is finally laminated and the oxide layer 502 to which the metal oxide film 503 is partially adhered are adhered by the adhesive 529 (see FIG. 2 (B)). During this bonding, the adhesive force between the oxide layer 502 and the element substrate 530 by the adhesive 529 is higher than the adhesion force between the second substrate 526 and the protective layer 524 by the double-sided tape 525 , It is important to select the material of the adhesive 529.

Note that if the metal oxide film 503 remains on the surface of the oxide film 502, the adhesion with the element substrate 530 may be deteriorated in some cases. Therefore, after completely removing it by etching or the like, the element substrate 530 So as to enhance the adhesion. Examples of the adhesive 529 include various curable adhesives such as a reaction curing type adhesive, a thermosetting type adhesive, a light curing type adhesive such as an ultraviolet curing type adhesive, and an anaerobic type adhesive. More preferably, the adhesive 529 is also preferably provided with a high thermal conductivity by including a powder of silver, nickel, aluminum, aluminum nitride or a filler. In the present embodiment, an interposer having a solder ball 531 is used as the element substrate 530. The solder balls 531 are exposed on the back surface of the surface of the element substrate 530 on which the thin film integrated circuits are stacked, and are connected to the wiring 532 provided on the surface on which the thin film integrated circuits are stacked via contact holes provided in the element substrate 530 , And is electrically connected. In the present embodiment, the wiring 532 is formed by plating solder, gold, or tin, for example, on copper. Next, as shown in FIG. 2C, the double-sided tape 525 and the second substrate 526 are peeled from the protective layer 524 in order or at the same time. By using an ultraviolet curable adhesive as the adhesive 529 and a tape or an adhesive that is peeled off by ultraviolet rays on the double-sided tape 525, peeling of the double-sided tape 525 and curing of the adhesive 529 can be simultaneously performed . Then, as shown in FIG. 2 (D), the protective layer 524 is removed. Here, since the water-soluble resin is used for the protective layer 524, it is dissolved in water and removed. If the protective layer 524 remains as a cause of failure, it is preferable to subject the surface after the removal to a cleaning treatment or O 2 plasma treatment to remove a part of the remaining protective layer 524. Although tungsten is used as the metal film 501 in this embodiment, the metal film is not limited to this material in the present invention. As long as a metal oxide film 503 is formed on the surface thereof and a metal can be peeled off by crystallizing the metal oxide film 503. For example, TiN, WN, Mo, or the like can be used. When these alloys are used as the metal film, the optimum temperature for the heat

treatment during crystallization varies depending on the composition ratio.

Therefore, by adjusting the composition ratio, the heat treatment can be performed at a temperature that does not interfere with the manufacturing process of the semiconductor element, and the options of the process of the semiconductor element are hardly limited. As shown in FIG. 2 (D), when the transfer of the first-layer thin film integrated circuit 533 to the element substrate 530 is completed, as shown in FIG. 3 (A), first layer thin film integration And transfers the second layer thin film integrated circuit 534 onto the circuit 533. The lamination of the second-layer thin film integrated circuit 534 can be performed by using the same method as the one for laminating the first-layer thin film integrated circuit 533 to the element substrate 530. That is, a second thin film integrated circuit 534 is formed on a separately prepared first substrate, and a second substrate is attached so as to cover the second thin film integrated circuit 534. Then, after reinforcing the first substrate with the third substrate, the metal oxide film provided between the first substrate and the second-layer thin film integrated circuit 534 is crystallized to weaken it, and the first substrate Peeled off together with the third substrate. Then, the second layer thin film integrated circuit 534 stuck to the second substrate is adhered onto the first layer thin film integrated circuit 533 by using the adhesive 535, then the second substrate is removed, and transfer Complete. The distance between the first thin film integrated circuit 533 and the second thin film integrated circuit 534 can be controlled by adjusting the thickness of the adhesive 535. Although the thickness of the adhesive depends on the pressure at the time of bonding, in the present embodiment, the thickness is about 20 μm. In the present embodiment, a case where a light-emitting element is formed after transfer of the second-layer thin film integrated circuit 534 will be described. At the time of transfer to the thin film integrated circuit 533 of the first layer, the pixel electrode 536 of the light emitting element is formed in the second layer thin film integrated circuit 534. The pixel electrode 536 is formed of a transparent conductive film, and in this embodiment, ITO is used. In this embodiment mode, a transparent conductive film is used for the pixel electrode of the light emitting element, but the present invention is not limited thereto. It is desirable to appropriately optimize the configuration of the light emitting element depending on whether the direction of the light from the light emitting element is directed to the element substrate side or the opposite direction. Next, as shown in FIG. 3 (B), a third interlayer insulating film 539 having an opening portion for partially exposing the pixel electrode 536 and the terminal 545 electrically connected to the thin film integrated circuit Is formed. In the present embodiment, the case of transferring the second-layer thin film integrated circuit at the time when the pixel electrode is formed will be described. However, in addition to the pixel electrode, transfer may be performed at the time when the third interlayer insulating film is formed Is also possible.

Next, an electroluminescent layer 537 and a cathode 538 are formed so as to overlap with the pixel electrode 536 in the opening portion of the third interlayer insulating film 539. A portion

where the pixel electrode 536, the electroluminescent layer 537, and the cathode 538 overlap with each other corresponds to the light-emitting element 540. As the transparent conductive film used as the pixel electrode 536, a transparent conductive film in which indium oxide is mixed with 2 to 20% of zinc oxide (ZnO) may be used in addition to ITO. The pixel electrode 536 may be polished by wiping and cleaning with a CMP method or a polyvinyl alcohol-based porous body so that its surface is planarized. After the polishing using the CMP method, ultraviolet irradiation, oxygen plasma treatment, or the like may be performed on the surface of the pixel electrode 536. The electroluminescent layer 537 has a structure in which a plurality of layers including a light-emitting layer alone or a light-emitting layer are stacked. As the cathode 538, any other known material can be used as long as it is a conductive film having a small work function. For example, Ca, Al, CaF, MgAg, AlLi and the like are desirable. Note that the third interlayer insulating film 539 is preferably heated in a vacuum atmosphere in order to remove adsorbed moisture, oxygen, and the like, before the electroluminescent layer 537 is formed. Specifically, heat treatment is performed in a vacuum atmosphere at 100 ° C. to 200 ° C. for 0.5 to 1 hour. Preferably 3 × 10 <-7 Torr or less, and most preferably 3 × 10 <-8 Torr or less if possible. In the case where the electroluminescent layer 537 is formed after heat treatment is performed on the third interlayer insulating film 539 in a vacuum atmosphere, reliability can be further improved by keeping the electroluminescent layer 537 in a vacuum atmosphere just before film formation . Further, it is desirable that the opening portion of the third interlayer insulating film 539 where the pixel electrode 536 is exposed has a rounded end. Since the end portion of the opening portion is rounded, it is possible to prevent the electroluminescent layer 537 from becoming extremely thin at the end portion and to prevent a hole from being formed, and a light emitting element due to a short circuit between the pixel electrode and the cathode Can be suppressed as much as possible. Further, by alleviating the stress of the electroluminescent layer 537 at the end portion, it is possible to reduce defects called shrinkage, in which the light emitting region decreases, and it is possible to enhance reliability. Specifically, it is desirable that the curvature radius of the curve drawn by the cross section of the organic resin film in the opening portion is about 0.2 to 2 μm.

The light emitting element 540 is covered with a protective film 541 in order to prevent substances that cause deterioration of the light emitting element 540 such as moisture and oxygen from entering the light emitting element 540. Typically, for example, a DLC film, a carbon nitride film, a silicon nitride film formed by an RF sputtering method, or the like is preferably used as the protective film 541. It is also possible to laminate a film which is hard to permeate the substance such as moisture or oxygen mentioned above and a film which is more easily permeable to substances such as moisture and oxygen as compared with the film and can be used as a protective film. Then, after forming the protective film 541, in order to further ensure the sealing of the light emitting element, the light emitting element 540 may be covered with a resin to which a desiccant is added. Note that a thin film integrated circuit to be transferred later may be bonded with the resin to which the desiccant is added. In this embodiment, an example

using an organic light emitting device (OLED) using electroluminescence has been described as a light emitting element, but the present invention is not limited thereto. In the present invention, as a light emitting element, in addition to an OLED, an LED (light emitting diode), a semiconductor laser, or the like can be used. The light emitting element is desirably a light emitting element having directivity. In the present embodiment, light from the light emitting element 540 provided in the second layer thin film integrated circuit 534 is incident on the light receiving element 521 provided in the first thin film integrated circuit 533. Therefore, the electric signal outputted from the thin film integrated circuit 534 of the second layer is converted into an optical signal by the light emitting element 540, converted into an electric signal by the light receiving element 521 again, and converted into an electric signal to the thin film integrated circuit 533 of the first layer Is inputted. In the present embodiment, the terminal 545 electrically connected to the second-layer thin film integrated circuit 534 and the wiring 532 provided on the element substrate 530 are connected to each other by the wire bonding method . Specifically, as shown in FIG. 3 (B), the wire 546 is used for connection. Through the wiring 532, power supply voltage and various signals can be supplied to the thin film integrated circuit. Note that the electrical connection to the thin film integrated circuit may be performed by a flip chip method rather than a wire bonding method. Also, the supply of signals from the outside may also be performed by optical transmission. In this embodiment, the wire 546 is connected to the thin film integrated circuit 534 of the second layer, but the present invention is not limited thereto.

For example, even if a solder ball electrically connected to a semiconductor element is provided on the element substrate side of the first-layer thin film integrated circuit 533 and the solder ball is connected to the wiring on the element substrate by the flip chip method good. As described above, by the optical transmission between the thin film integrated circuit using the light emitting element and the light receiving element, it is possible to perform connection such as signal exchange between layers, supply such as supply voltage, etc., and a plurality of stacked thin films One thin film integrated circuit is constructed from the chip (thin film integrated circuit). In the present embodiment, an example in which two layers of thin film integrated circuits are stacked and used as one thin film integrated circuit has been described, but the number of stacked thin film integrated circuits is not limited thereto. Multiple thin film integrated circuits of three or more layers may be stacked. It should be noted that the element substrate is not limited to an interposer using glass epoxy or the like, but may be a substrate using another material. Further, the transfer method of the thin film integrated circuit of the present invention is not limited to the above-described method using the metal oxide film. For example, an amorphous silicon film may be provided between the first substrate and the thin film integrated circuit, and the first substrate and the thin film integrated circuit are formed on the amorphous silicon film by irradiation with laser light or etching It may be peeled off. Next, with reference to FIG. 4 (A), the positional relationship between the light emitting element and the light receiving element between the respective layers when stacking a plurality of thin film integrated circuits will be described. FIG. 4A schematically shows a state in which a thin film integrated circuit is

stacked on the element substrate 100. The thin film chip (thin film integrated circuit) 101 of each layer includes one or a plurality of circuits 103. Further, in the thin film integrated circuit 101 of each layer, one or a plurality of regions in which the light emitting element and the light receiving element for sending and receiving optical signals are laid out are provided. Note that in FIG. 4 (A), a region in which a light emitting element or a light receiving element for performing optical transmission is provided will be referred to as an optical transmission section 102. And an interface 104 for performing processing such as modulation or demodulation on the electrical signal in order to exchange electrical signals between the optical transmission unit 102 and the other circuits 103. In the optical transmission section 102, a light receiving element for receiving an optical signal and a light emitting element for transmitting an optical signal are formed. FIG. 4 (B) shows an enlarged view of the optical transmission section 102 of each layer. 105 corresponds to a light emitting element, and 106 corresponds to a light receiving element. The light emitting element 105 and the light receiving element 106 provided in the thin film integrated circuit of each layer respectively correspond to the light receiving element 106 and the light emitting element 105 provided in the thin film integrated circuit 101 of the other layer.

In FIG. 4A and FIG. 4B, the light emitting element 105 and the light receiving element 106 are laid out in a bundle in the region to be the light transmission section 102, but the present invention is not limited to this configuration. The light emitting element 105 and the light receiving element 106 may be mixed in the circuit 103. A circuit element functioning as the interface 104 may be mixed in the circuit 103 in accordance with the layout of the light emitting element 105 and the light receiving element 106. In FIG. 4 (A), an example is shown in which all the connections between layers are performed by optical transmission, but the present invention is not limited thereto. A variety of signals and power supply voltages may be transmitted and received as a part of an electric signal, and may have a function for transmission and reception, for example, a terminal or the like as an electric signal. FIG. 4 (C) shows an example of a more specific configuration of the optical transmission unit 102 and the interface 104. At least one light receiving element 106 corresponds to one light emitting element 105. In FIG. 4 (B), only the light emitting element 105 and the light receiving element 106 are shown in the light transmission section 102, but actually, as shown in FIG. 4 (C), in order to control the light emission of the light emitting element 105 by an electric signal A driving circuit 110 for the light emitting element, an amplifier circuit 111 for amplifying the electric signal obtained from the light receiving element, a waveform shaping circuit for shaping the waveform of the obtained electric signal, and the like are provided. Note that the interface 104 may have these functions. The electrical signal (output signal) output from the circuit is modulated into a signal conforming to the specifications of the light emitting element driving circuit 110 by the interface 104, and then input to the light emitting element driving circuit 110. Then, the light emitting element driving circuit 110 controls the light emission of the light emitting element 105 according to the input output signal. The light receiving element 106 receives the optical signal emitted from the light emitting element 105 and converts the optical signal into an electric signal. The electric

signal obtained in the light receiving element 106 is amplified in the amplifying circuit 111, its waveform is shaped in the waveform shaping circuit 112, demodulated in accordance with the specifications of each circuit in the interface 114, and input to the circuit in the subsequent stage as an input signal . It should be noted that the amplifier circuit 19 and the waveform shaping circuit 20 are not necessarily provided, and besides these circuits, a circuit for applying some processing to the waveform of the electric signal may be provided. Note that in FIG. 4C, an example in which the light-emitting element and the light-receiving element correspond one-to-one is shown, but the present invention is not limited to this configuration.

Two or more light emitting elements may correspond to one light receiving element, or one light emitting element may correspond to two or more light receiving elements. In FIG. 5A, two light emitting elements 130 and 131 formed in thin film integrated circuits of different layers respectively correspond to one light receiving element 132 formed in a thin film integrated circuit of another layer As shown in FIG. The light emitting element 130 is arranged so that light emitted is transmitted through the layer 133 where the light emitting element 131 is formed and is incident on the light receiving element 132. With the above configuration, while the light signal is transmitted from the light emitting element 130 to the light receiving element 132, another operation can be performed in the thin film integrated circuit of the layer 133 where the light emitting element 131 is formed, and conversely, the light emitting element 131 Another operation can be performed in the thin film integrated circuit of the layer 134 where the light emitting element 130 is formed while the optical signal is being sent from the light receiving element 132 to the light receiving element 132. Conversely, in the case where two light receiving elements formed in thin film integrated circuits of different layers respectively correspond to one light emitting element formed in a thin film integrated circuit of another layer, It is possible to transmit an optical signal simultaneously to the thin film integrated circuit. FIG. 5B shows a case in which one of the electric signals obtained by a plurality of light receiving elements is selected and a selection circuit for transmitting to the circuit of the subsequent stage is provided. In FIG. 5 (B), optical signals emitted from two light emitting elements 135, 136 formed in thin film integrated circuits of different layers are further divided into two light receiving elements formed on a thin film integrated circuit of another layer 137, and 138 into electrical signals. Then, one of the obtained two electric signals is selected by the selection circuit 139 and transmitted to the circuit of the subsequent stage. With the above configuration, as in the case of FIG. 5A, the oscillation frequency per light emitting element can be lowered, and the burden on the light emitting element driving section that controls the driving of the light emitting element can be reduced. In order to perform data transmission by optical signals in parallel between the thin film integrated circuits, it is necessary to make the paths of the optical signals independent from each other. However, depending on the degree of diffusion of light, so-called crosstalk, in which an optical signal enters a corresponding light receiving element, may occur. The degree of light diffusion depends on the directivity of the light emitted from the light emitting element and the refractive index of the medium in the path of the optical signal. Therefore, taking light diffusion

into consideration so that crosstalk is suppressed as much as possible, the light receiving element and the light emitting element are laid out according to the directivity of light of the light emitting element to be used, and the thickness of the thin film integrated circuit, the thickness between the thin film integrated circuits It is desirable to appropriately set the distance and the medium (adhesive or the like) between the thin film integrated circuits.

In order to prevent crosstalk, an optical waveguide such as an optical fiber having a cylindrical shape or a cross section close to it or a thin film waveguide transmitting light along a planar dielectric thin film may be provided in the path of the optical signal. The substrate used in the present invention is not limited to a glass substrate. Of course, it is also possible to use a substrate other than the glass substrate, such as a plastic substrate, as long as it has transparency and can withstand processing temperature in the formation of semiconductor elements and other processes. In this embodiment, an example in which signals are transmitted by optical transmission between different layers has been described. However, signals may be transmitted using optical transmission in a thin film integrated circuit of a certain layer . Note that a light emitting element in which the direction of emitted light is directed toward the element substrate and a light emitting element directed to the side opposite to the element substrate are provided in a thin film integrated circuit formed in the same layer May also be assumed. FIG. 6 shows an example of a cross-sectional view of a thin film integrated circuit in which two kinds of light emitting elements having different light emitting directions are formed in the same layer. The two light-emitting elements 201 and 202 shown in FIG. 6 have the same structure of elements, and include pixel electrodes 203 and 204 formed of a transparent conductive film, electroluminescent layers 205 and 206, a cathode 207, 208. A portion where the pixel electrode 203, the electroluminescent layer 205, and the cathode 207 overlap corresponds to the light-emitting element 201, and a portion where the pixel electrode 204, the electroluminescent layer 206, and the cathode 208 overlap corresponds to the light-emitting element 202. Specifically, the electroluminescent layers 205 and 206 of the light-emitting elements 201 and 202 shown in FIG. 6 are made of copper phthalocyanine (CuPc) having a film thickness of 20 nm as a hole injection layer, 4 (Hereinafter referred to as α-NPD), 4'-bis [N- (1-naphthyl) -N-phenyl- , Alq 3 having a film thickness of 37.5 nm as an electron transporting layer, and CaF 2 having a film thickness of 1 nm as an electron injecting layer are laminated. The cathodes 207 and 208 are formed of a film made of Al having a thickness enough to transmit light. In FIG. 6, Al is evaporated by resistance heating to form a film, and the film is formed to have a film thickness of 20 nm. Note that polythiophene (PEDOT) may be used instead of CuPc as the hole injection layer.

The laminated structure of the light emitting element and its film thickness are not limited to the values shown in this embodiment. In order to obtain light from the cathode side, there is a method of using ITO in which the work function is reduced by adding Li, in addition to the

method of reducing the film thickness. The light-emitting element used in FIG. 6 may be of any configuration as long as light is emitted from both the anode side and the cathode side. In the light-emitting element 201 shown in FIG. 6, a reflective film (or a shielding film) 210 is provided so as to be in contact with the cathode 207. With this reflection film 210, light emitted from the light emitting element 201 is directed toward the element substrate side as indicated by outlined arrows. Note that the reflective film 210 can be formed by depositing a metal film. When a shielding film is used instead of a reflective film, a resin to which a black pigment is added can be used. In any case, a means capable of controlling the direction of light such as a reflective film or a shielding film is provided so that light emitted from the light emitting element 201 is directed only to the element substrate side. In the light emitting element 202 shown in FIG. 6, a metal film 211 for reflecting light emitted from the light emitting element 202 is provided on the element substrate side with respect to the light emitting element 202. With the metal film 211, the light emitted from the light emitting element 202 is directed to the side opposite to the element substrate as indicated by outlined arrows. Note that the metal film 211 can be formed from the same metal film as the gate electrode and the wiring of the TFT formed in the thin film integrated circuit. Further, instead of reflecting light, a shielding film for simply shielding may be provided instead of the metal film 211. The shielding film may be one obtained by adding a black pigment to an insulating film used as an interlayer insulating film. In either case, a means capable of controlling the direction of light such as a reflection film or a shielding film is provided so that light emitted from the light emitting element 201 is directed only to the side opposite to the element substrate. As described above, even in the same layer, it is possible to form light emitting elements with different light directions. Note that the method of manufacturing the light-emitting element and the configuration of the light-emitting element having different light directions in the same layer are not limited to the example shown in FIG. 6. However, when the shielding film or the reflective film is provided at a position closer to the light emitting element, the directivity of light can be further enhanced. In this embodiment, an example in which a thin film integrated circuit is stacked has been described, but a flat panel display using a semiconductor element may be stacked on a thin film integrated circuit. When transferring a flat panel display to a thin film integrated circuit, it is desirable to transfer before the display element is completed and to complete the display element after transfer.

For example, in the case of a light-emitting device using an OLED, after a pixel electrode is formed, it is transferred, and an electroluminescent layer, a cathode, and the like are formed after transfer and sealed. In the case of a liquid crystal display, an alignment film is formed on a pixel electrode, subjected to rubbing treatment, transferred, transferred, sealed with a counter substrate having a counter electrode after transfer, and liquid crystal is injected. As described above, according to the present invention, since the thin film chips of each layer can be manufactured separately, there is no restriction on the process of the semiconductor element of the upper layer in consideration of the heat resistance of the underlying semiconductor element, Excellent semiconductor devices can be formed. Further, since the thin film chips of each layer

can be laminated so as to be in close contact without sandwiching the glass substrate or the like therebetween, it is not necessary to consider the diffusion of light by the glass substrate. Since the distance between the light emitting element and the light receiving element can be further reduced, even if the light directivity of the light emitting element is inferior, it can be covered to some extent. By using the thin film integrated circuit of the present invention for a semiconductor device, it is possible to secure a wider space for use in an integrated circuit, realizing a high function without impeding reduction in weight or size of a semiconductor device be able to. Particularly in the case of a portable semiconductor device, convenience is dramatically improved by weight saving and downsizing, so that it is very useful to use the thin film integrated circuit of the present invention. By using optical interconnection, data can be transferred between the thin film chips while reducing the influence of skew caused by wiring resistance and unnecessary radio wave radiation. Further, in the process of converting an electric signal to an optical signal and converting it from an optical signal to an electric signal again, it is possible to freely control the amplitude of the finally obtained electric signal. EXAMPLES Examples of the present invention will be described below. Embodiment 1 In this embodiment, an example in which a microprocessor is formed using a CPU (Central Processing Unit) constructed of a plurality of thin film chips will be described. A TFT formed on a glass substrate has a slower operation speed than a single crystal transistor. Therefore, in the case of forming a CPU on a glass substrate, if the processing contents become complicated, it is difficult to execute the processing at a sufficient speed in a single CPU. Therefore, the series of processes of the CPU is divided into several processes for each purpose, and the CPU formed by one thin film chip is allocated to each process.

By connecting each thin film chip by optical interconnection, it is possible to perform a series of processing as in the case of using a single CPU. The CPUs formed on the respective thin film chips need only perform the assigned process, and the processing speed is improved as compared with the case where all processes are performed by a single layer CPU. FIG. 7 is a perspective view of the microprocessor of this embodiment. In the microprocessor shown in FIG. 7, a plurality of thin film chips 402 each including a CPU 401, a main memory 403, a clock controller 404, a cache controller 405, a serial interface 406, an I / O port 407, and the like are formed on a substrate 400 . Of course, the microprocessor shown in FIG. 7 is a simplified example, and an actual microprocessor has various configurations depending on its use. Each thin film chip 402 has an optical transmission unit 408, an interface 409, a CPU 401, and a cache memory 410. The light transmission unit 408 may have both a light emitting element having a function of outputting an electric signal as an optical signal and a light receiving element having a function of converting an optical signal into an electric signal, May be provided. Further, it may be laid out so as to coexist among other circuit elements without laying out all the light emitting elements or light receiving elements in the light transmitting section 408. Then, signals or power supply voltages are transmitted and received between the thin film chips 402 or between the thin film chip 402 and various circuits formed on the substrate 400 by using optical

transmission. It should be noted that transmission and reception of all signals and power supply voltages need not be performed by optical transmission, and some of them may be directly transmitted and received by electric signals. In that case, there is a terminal for transmitting and receiving the electric signal as it is without converting it into an optical signal. The cache memory 410 is a small-capacity high-speed memory interposed between the CPU 401 and the main memory 403. In order to operate the CPU faster, it is necessary to have a high speed memory corresponding to it. However, when using a high-speed large-capacity memory having an access time that matches the operation speed of the CPU, the cost generally increases. By accessing the cache memory, the CPU can operate at high speed regardless of the speed of the main memory. An example of the operation of each CPU 401 will be described below.

For example, at the initial stage of execution, for example, the program is downloaded from the main memory 403 or another external memory to the cache memory 410 (SRAM) of each thin film chip 402. Among the CPU 401, a master CPU may do this. Next, among the CPU 401, each slave CPU executes programs stored in the cache memory 410 of the same thin film chip 402 in order. The cache memory 410 of the same thin film chip 402 not only stores the program but also functions as a work area, and temporarily stores calculation results etc. of the CPU 401. When each CPU 401 requires output of another CPU 401 and exchanges of signals with other than the cache memory 410, such as the main memory 403, this is performed via the light transmission unit 408. The overall operation speed is improved according to the number of CPUs 401. In particular, when the signal between the CPU 401 and the signal exchanged outside the board 400 is small, the effect of parallelization is high. Examples of the program include, for example, an optimization problem (for example, automatic wiring, a circulation problem of a salesperson) that searches for a minimum value in a phase space having a very large number of local minima, an evaluation of a variation Simulation, etc), Monte Carlo method, simulated annealing, etc. are applied. Basically, these programs are structured to independently execute the same subprogram many times, and by making each subprogram be assigned to a different CPU 401, practically, each thin film It is possible to execute a program completed by the CPU 401 and the cache memory 410 in the chip 402 and to perform ideal parallel calculation. If the processing speed between the CPUs 401 is varied, inconvenience may occur when viewing the entire process, so that the CPU serving as the master should balance the processing speed among the slave CPUs It is also good. (Embodiment 2) The integrated circuit of the present invention can be applied to various memories. By stacking the memories, it is possible to increase the memory capacity while suppressing the mounting area. By using optical transmission for transmitting and receiving data between the thin film chips, even if the memory is stacked, the number of pins used for mounting can be suppressed. As a stacked memory, a RAM (volatile memory) typified by SRAM (Static Random Access Memory), DRAM (Dynamic Random Access Memory) and the like, a mask ROM, an Erasable Programmable Read Only Memory (EPROM), an EEPROM An electrically erasable programmable ROM (ROM), a flash memory, a ferroelectric memory, or the like can be used.

Also, a combination of these memories may be used. In stacking the memories, an R / W (Read / Write) circuit, a precharge circuit, a refresh control circuit (refresh circuit) for selecting whether to read or write, depending on the type or specification of the memory, Or the like may be provided in the same thin film chip as the memory. Further, the power supply voltage to be supplied may be appropriately set to an optimum value for each thin film chip. In the present embodiment, a configuration of a memory formed by stacking is taken as an example of a DRAM. FIG. 8 (A) shows the configuration of the first thin film chip. In the thin film chip shown in FIG. 8A, a DRAM memory cell array 601, a selector circuit 602, a column decoder 603, a row decoder 604, an R / W circuit 605, and a light transmission unit 606 are formed. The memory cell array 601 is provided with a plurality of memory cells. Then, the memory cell is selected by the selector circuit 602 whose operation is controlled according to the signal from the column decoder 603 and the row decoder 604, and data is read and written. Signal transmission / reception between the connection terminal and the thin film chip of the first layer can be performed directly by an electric signal or via the light transmission section 606. FIG. 8B is an enlarged view of the optical transmission unit 606. The light transmission unit 606 is provided with a light emitting element 610 and a light receiving element 611. The various signals input from the connection terminal or the electric signal output from the thin film chip of the first layer are converted by the light emitting element 610 into an optical signal And is transmitted to the other kth layer (k is a natural number of 2 or more) thin film chips. An optical signal transmitted from the other kth layer thin film chip is converted into an electric signal by the light receiving element 611 and sent to the thin film chip or the connection terminal of the first layer. CE (chip enable) (1) to CE (n) among the signals input to the optical transmission unit 606 correspond to signals for selecting one or more of the first to nth thin film chips . In addition, the electric signals, RE (Read enable), WE (Write enable), Din input via the connection terminals are converted to optical signals via the light emitting element 610 and are sent to the thin film chips selected by the signal FE . Conversely, the optical signal sent from the thin film chip selected by the signal FE is converted into an electric signal in the first layer light receiving element 611, and is output as a signal Dout via the connection terminal.

In this embodiment, RE and WE are signals for selecting reading or writing of data by the R / W circuit 605, Din corresponds to a signal including data written in the memory as information, Dout corresponds to a signal including data read from the memory as information. FIG. 9 shows the configuration of the kth layer thin film chip. In this embodiment, an example in which a DRAM is formed in the kth layer thin film chip as in the first layer will be described. The kth layer thin film chip shown in FIG. 9 includes a k th memory cell array 621, a selector circuit 622, a column decoder 623, a row decoder 624, an R / W circuit 625, an optical transmission unit 626 are formed, respectively. Although not shown, only one of the light emitting element and the light receiving element, and in some cases, only one is formed in the light transmission section 626,

and the signal between the light transmission section 626 and the thin film chip of the other layer Is carried out. Although the example in which the DRAM is formed in the first layer and the kth layer thin film chip has been described in the present embodiment, the present invention is not limited thereto, and it is also possible to form another thin film memory It can be formed in a chip, and the thin film chip can be laminated. In addition, the other circuits related to the driving of the memory are not limited to the circuits shown in this embodiment. In this embodiment, the electric signal input through the connection terminal is input to the thin film chip of the first layer, but the present invention is not limited thereto. An electric signal may be input to the thin film chip of the second layer or later through the connection terminal. (Embodiment 3) In this embodiment, a configuration of a one-chip microcomputer constructed with a plurality of thin film chips will be described. FIG. 10 shows a block diagram of the microprocessor of this embodiment. The microprocessor of this embodiment is composed of two layers of thin film chips. SIO (Serial Input / Output) 631, MMU (Memory Management Unit) 632, a counter 633, a DAC (Digital to Analog converter) 634, a bus controller (bus interface) 635, an optical transmission unit 636 are provided. The second layer thin film chip 640 is provided with a CPU 641, a RAM 642, a ROM 643, and an optical transmission unit 644.

A light emitting element and a light receiving element are formed in the optical transmission units 636 and 644, respectively, and various signals or power supply voltages are transmitted and received by optical transmission. By constructing a thin film chip in which microcomputers are stacked as in this embodiment, it is possible to dramatically reduce the mounting area of the microcomputer. Embodiment 4 In this embodiment, a method of electrical connection between an interposer and a thin film chip will be described. FIG. 11A is a perspective view showing the cross-sectional structure when the laminated thin film chip is connected to the interposer by the wire bonding method. 301 is an interposer, and 302 is a laminated thin film chip. The thin film chip 302 is bonded onto the interposer 301 with an adhesive 304. The interposer 301 shown in FIG. 11A is a ball grid array type in which a solder ball 305 is provided. The solder ball 305 is provided on the side opposite to the side where the thin film chip 302 of the interposer 301 is bonded. The wiring 306 provided in the interposer 301 is electrically connected to the solder ball 305 via a contact hole provided in the interposer 301. In this embodiment, an example is shown in which the wiring 306 for electrically connecting the thin film chip 302 and the solder ball 305 is provided on the surface to which the thin film chip 302 of the interposer 305 is bonded However, the interposer used in the present invention is not limited thereto. For example, the wiring may be provided in a multilayered manner inside the interposer. In FIG. 11 (A), the thin film chip 302 and the wiring 306 are electrically connected by a wire 307. FIG. 11 (B) shows a cross-sectional view of the package shown in FIG. 11 (A). The thin film chip 302 is provided with a semiconductor element 309, and a pad 308 is provided on the side of the thin film chip 302 opposite to the side on which the interposer 301 is provided. In this embodiment, the pad 308 is electrically connected to the uppermost layer of the laminated thin film chip 302. The pad 308 is connected to a wiring 306 provided in the interposer 301 by a wire 307. Reference

numeral 310 corresponds to a part of the printed wiring board and reference numeral 311 corresponds to wiring or electrode provided on the printed wiring board 310.

The wiring 306 is connected to the wiring or the electrode 311 provided on the printed wiring board 310 via the solder ball 305. Note that various methods such as thermocompression bonding, thermocompression bonding with vibration by ultrasonic wave, and the like can be used for connection between the solder ball 305 and the wiring or the electrode 311. In addition, the underfill may fill the gaps between the solder balls after crimping so as to increase the mechanical strength of the connecting portions and the efficiency of diffusion of heat generated in the package. Although the underfill is not necessarily required to be used, it is possible to prevent connection failure from occurring due to a stress generated from a mismatch in thermal expansion coefficient between the interposer and the thin film chip. In the case of applying ultrasonic wave and crimping, poor connection can be suppressed as compared with simply thermocompression bonding. This is particularly effective when the number of solder balls to be connected is more than about 300. Next, FIG. 11 (C) shows a cross-sectional view when the thin film chip is connected to the interposer by using the flip chip method. In FIG. 11 (C), a solder ball 327 is provided on the thin film chip on the lowermost layer of the thin film chip 322. The solder ball 327 is provided on the interposer 321 side of the thin film chip of the lowermost layer and is electrically connected to the thin film chip of the lowermost layer. The solder ball 327 is connected to the wiring 326 provided in the interposer 321. 11C, an underfill 324 that also functions as an adhesive is provided so as to fill the gap between the solder balls 327. Further, the solder ball 325 of the interposer 321 is provided on the side opposite to the side where the thin film chip 322 of the interposer 321 is bonded. The wiring 326 provided in the interposer 321 is electrically connected to the solder ball 325 via a contact hole provided in the interposer 325. In the case of the flip chip method, even if the number of pads to be connected increases, a wider pitch between the pads can be secured as compared with the wire bonding method, so that it is possible to obtain a thin film chip having a large number of terminals It is suitable for connection. Although the thin film chip shown in FIGS. 11A to 11C is transferred to a ball grid array type interposer, the present invention is not limited to this. An interposer in a lead frame type in which terminals are arranged in the periphery may be used.

FIG. 11 (D) is a perspective view of a sectional structure in the case of using a lead frame interposer. In FIG. 11 (D), the thin film chip 351 is connected to the connection terminal 352 on the interposer 350 by the wire bonding method. The connection terminal 352 is disposed on the surface to which the thin film chip 351 of the interposer 350 is bonded. Although the thin film chip is exposed in this embodiment, the thin film chip may be sealed with a mold resin. In this embodiment, an example in which the thin film chips are stacked on the interposer will be described. However, the present invention is not limited to this structure, and the laminated substrate may be a glass substrate, a quartz substrate, a silicon substrate, a plastic substrate, A

known substrate used for forming a semiconductor integrated circuit can be used. In this case, a connection terminal may be provided on the substrate for supply of an external signal or power supply voltage, or may be supplied by optical transmission. (Example 5) Since the process of transferring a thin film chip used in the present invention is performed using various manufacturing apparatuses, an inline type production facility in which transfer of substrates between apparatuses is automated is performed By using it, production efficiency can be improved. FIG. 12 shows a manufacturing method shown in FIGS. 1 to 3 of the embodiment, in which from the step of bonding the second substrate to the first substrate on which the semiconductor element is formed, the third substrate is pasted It shows the configuration of an inline type production facility that automates the conveyance between manufacturing apparatuses until the matching step. In FIG. 12, reference numerals 701 and 714 correspond to means (substrate transport means) for transporting the substrates, and the respective devices are moved in the direction indicated by outlined arrows between the respective devices. Reference numeral 707 corresponds to a substrate transfer station for temporarily stocking a substrate in order to transfer a substrate between the substrate transfer means 701 and the substrate transfer means 714. In the present embodiment, the transportation of the substrate between the respective devices is covered by the two substrate transfer means 701 and 714, but the present invention is not limited thereto. By increasing the number of substrate transfer means, the efficiency of substrate transfer can be increased. Also, unlike the substrate transfer station 707, it is not always necessary to provide equipment for temporarily stocking substrates for the delivery of substrates. For example, the substrate transfer means for transferring the substrate to the apparatus may be made different from the substrate transfer means to be carried out so as to transfer the substrate.

A spinner 702 is a device for coating a water-soluble adhesive by a spin coat method so as to cover the semiconductor element of the first substrate. An exposure apparatus 703 can cure the adhesive applied by the spinner 702 by exposure to ultraviolet light. Reference numeral 704 denotes a scriber, which is an apparatus for cutting an end portion of the substrate so that the semiconductor element is less likely to be peeled off with the metal oxide film as a boundary at the time of peeling in a later process. A double-sided tape sticking apparatus 705 can stick a double-sided tape supplied from the tape supply cassette 706 to the surface of the first substrate coated with the adhesive. A substrate adhering apparatus 708 can attach the second substrate supplied from the substrate supply cassette 709 to the first substrate using the double-sided tape affixed in the previous step. A double-sided tape attaching device 710 can attach the double-sided tape supplied from the tape feeding cassette 711 to a surface opposite to the surface of the first substrate coated with the adhesive. A substrate attaching device 712 can attach the third substrate supplied from the substrate supply cassette 713 to the first substrate using the double-sided tape affixed in the previous step. In the present embodiment, the production equipment in which only the processes described above are in-line has been described. However, the present invention is not limited to this, and the devices used in other processes may be in-line.

(Embodiment 6) The integrated circuit of the present invention can be used for various semiconductor devices. Further, a display device can be bonded to the integrated circuit of the present invention and used for a semiconductor device. Particularly, in the case of a semiconductor device for portable use, convenience is dramatically improved by weight saving and downsizing, so it is very useful to use the integrated circuit of the present invention. FIG. 13A shows a sheet type mobile phone, which includes a main body 2101, a display portion 2103, a voice input portion 2104, an audio output portion 2105, a switch 2106, an external connection port 2107, and the like. A separately prepared earphone 2108 can be connected via the external connection port 2107. A touch panel type display device having a sensor is used for the display portion 2103, and a series of operations can be performed by touching the touch panel type operation keys 2109 displayed on the display portion 2103. The thin film integrated circuit of the present invention can be used as various signal processing circuits provided in the main body 2101. When a display device is bonded to a thin film integrated circuit, the display device can be used for the display portion 2103 is there.

FIG. 13B shows an electronic book, which includes a main body 2201, a display portion 2202, operation keys 2203, and the like. Also, the modem may be incorporated in the main body 2201. Further, the thin film integrated circuit of the present invention can be used as various signal processing circuits. FIG. 13C shows a wristwatch, which includes a main body 2301, a display portion 2302, fasteners 2303, and the like. The thin film integrated circuit of the present invention can be used as various signal processing circuits provided in the main body 2301. When a display device is bonded to a thin film integrated circuit, the display device can be used for the display portion 2302 is there. FIG. 13D shows a sheet type personal computer, which includes a main body 2401, a display portion 2402, a touch panel keyboard 2403, a pointing mouse 2404, an external connection port 2405, a power plug 2406, and the like. A touch panel type display device having a sensor is used for the touch panel type keyboard 2403 and the pointing mouse 2404, and a series of operations can be performed by touching the touch panel type keyboard 2403 and the pointing mouse 2404. The thin film integrated circuit of the present invention can be used as various signal processing circuits, and when a display device is bonded to a thin film integrated circuit, the display device can be used for the display portion 2402, the touch panel type keyboard 2403, and the like . FIG. 13D shows an electronic card, which includes a main body 2601, a display portion 2602, a connection terminal 2603, and the like. The thin film integrated circuit of the present invention can be used as various signal processing circuits provided in the main body 2601. When a display device is bonded to a thin film integrated circuit, the display device can be used for the display portion 2602 is there. As described above, the application range of the present invention is extremely wide and it can be used for semiconductor devices in various fields. Further, in the semiconductor device of this embodiment, any of the integrated circuits shown in Embodiments 1 to 5 may be used. Embodiment 7 In this embodiment, an embodiment in which a display device using an OLED as a display element is full-colored by bonding will be described. FIG. 14A is a cross-sectional view of

a pixel included in the display device of this example. Reference numeral 801 corresponds to an OLED included in the thin film chip of the first layer. In this embodiment, the light emission color is red (R). In addition, 802 corresponds to the OLED included in the second thin film chip, and in this embodiment, the light emission color is blue (B).

Reference numeral 803 denotes an OLED included in the thin film chip of the third layer. In this embodiment, the light emission color is green (G). OLEDs 801 to 803 are provided in each pixel, and the OLEDs 801 to 803 provided in one pixel are arranged so that the light-emitting regions are overlapped with each other so that the light emission thereof appears to overlap when viewed from the viewer . The luminescent color of the OLED included in the thin film chip of each layer is not limited to the form shown in this embodiment. The cathode of the OLED 803 of the uppermost thin film chip is made of a material that reflects light so that the light emission of all the OLEDs 801 to 803 is directed toward the thin film chip side of the lower layer, the OLEDs 801 and 802 are formed of light Are emitted from both sides. With the above structure, the display device using the OLED as a display element can be made full color. Next, a cross-sectional view of a pixel included in the display device according to this embodiment, which is different from that in FIG. 14A, is shown. In FIG. 14B, reference numeral 811 corresponds to the OLED included in the thin film chip of the first layer. In this embodiment, the light emission color is red (R). In addition, 812 corresponds to the OLED included in the second thin film chip, and in this embodiment, the light emission color is blue (B). 813 corresponds to the OLED included in the thin film chip of the third layer, and in this embodiment, the emission color is green (G). OLEDs 811 to 813 are provided in each pixel, and the OLEDs 811 to 813 provided in one pixel are arranged so that the light emitting regions do not overlap with each other, or only partially overlap. The luminescent color of the OLED included in the thin film chip of each layer is not limited to the form shown in this embodiment. The cathodes of the OLEDs 811 to 813 of the thin film chips of each layer are formed of a material that reflects light so that the light emission of all the OLEDs 811 to 813 is directed toward the thin film chip side of the lower layer side. With the above structure, the display device using the OLED as a display element can be made full color. In the display device having the structure shown in this embodiment, since each thin film chip is separately and independently fabricated, it is easy to optimize the manufacturing method for each thin film chip. Therefore, according to the characteristics of the electroluminescent material to be used, it is possible to optimize the configuration of the elements of each color OLED and to appropriately change the manufacturing method accordingly. As described above, according to the present invention, it is possible to separately prepare the thin film chips of each layer, so that there is a restriction on the process of the semiconductor element of the upper layer in consideration of the heat resistance of the underlying semiconductor element So that a semiconductor element having more excellent characteristics can be formed.

Further, since the thin film chips of each layer can be laminated so as to be in close contact

without sandwiching the glass substrate or the like therebetween, it is not necessary to consider the diffusion of light by the glass substrate. Since the distance between the light emitting element and the light receiving element can be further reduced, even if the light directivity of the light emitting element is inferior, it can be covered to some extent. By using the thin film integrated circuit of the present invention in a semiconductor device, it is possible to secure a wider space for use in an integrated circuit, realizing a high function without impeding reduction in weight or size of a semiconductor device be able to. Particularly in the case of a portable semiconductor device, convenience is dramatically improved by weight saving and downsizing, so that it is very useful to use the thin film integrated circuit of the present invention. By using optical interconnection, data can be transferred between the thin film chips while reducing the influence of skew caused by wiring resistance and unnecessary radio wave radiation. Further, in the process of converting an electric signal to an optical signal and converting it from an optical signal to an electric signal again, it is possible to freely control the amplitude of the finally obtained electric signal. BRIEF DESCRIPTION OF THE DRAWINGS FIG. 1 is a view showing a method of manufacturing a thin film integrated circuit of the present invention. FIG. 2 is a view showing a method of manufacturing a thin film integrated circuit of the present invention. FIG. 3 is a view showing a method for manufacturing a thin film integrated circuit of the present invention. FIG. 4 is a diagram showing a configuration of a thin film integrated circuit of the present invention. FIG. 5 is a diagram showing a configuration of a thin film integrated circuit of the present invention. FIG. 6 is a diagram showing a configuration of a light emitting element used in the thin film integrated circuit of the present invention. FIG. 7 is a diagram showing a configuration of a microprocessor which is an example of a thin film integrated circuit of the present invention. FIG. 8 is a diagram showing a configuration of a DRAM which is an example of a thin film integrated circuit of the present invention. FIG. 9 is a diagram showing a configuration of a DRAM which is an example of a thin film integrated circuit of the present invention. FIG. 10 is a diagram showing a configuration of a one-chip microcomputer which is an example of a thin film integrated circuit of the present invention. FIG. 11 is a perspective view and a cross-sectional view showing a state in which the thin film integrated circuit of the present invention is connected to an interposer. FIG. 12 is a view showing an example of a production facility used in the manufacturing process of the thin film integrated circuit of the present invention. FIG. 13 is a diagram showing a configuration of a semiconductor device manufactured using the thin film integrated circuit of the present invention. FIG. 14 is a cross-sectional view of a pixel of a display device which is one of the semiconductor devices of the present invention.

Patent Translate
Powered by EPO and Google

Notice

This translation is machine-generated. It cannot be guaranteed that it is intelligible, accurate, complete, reliable or fit for specific purposes. Critical decisions, such as commercially relevant or financial decisions, should not be based on machine-translation output.

CLAIMS JP2004247405

1.
A semiconductor device having stacked semiconductor elements, wherein a film formed of resin exists between the stacked semiconductor elements, and signal transmission and reception between the stacked semiconductor elements is performed by supplying electricity Emitting element or a light-receiving element which are connected to each other.

2.
A semiconductor device having stacked semiconductor elements, wherein a film formed of a resin and a partially provided metal oxide exist between the stacked semiconductor elements, and the stacked semiconductor Wherein the transmission and reception of signals between the elements is performed by using a light emitting element or a light receiving element electrically connected to each semiconductor element.

3.
A semiconductor device having stacked semiconductor elements, wherein a film formed of resin exists between the stacked semiconductor elements, and a light emitting element electrically connected to one of the stacked semiconductor elements , Converting the first electrical signal into an optical signal and converting the optical signal into a second electrical signal in a light receiving element electrically connected to a different one of the stacked semiconductor elements
.

4.

A semiconductor device having stacked semiconductor elements, wherein a film formed of a resin and a partially provided metal oxide exist between the stacked semiconductor elements, and the stacked semiconductor A light-emitting element electrically connected to one of the elements, the light-receiving element electrically converting a first electric signal into an optical signal and electrically connected to a different one of the stacked semiconductor elements, wherein the light And converting the signal into a second electric signal.

5.

A semiconductor device having a semiconductor element stacked by transferring semiconductor elements formed on different substrates, wherein a film formed of resin exists between the stacked semiconductor elements, and the stacked In a light emitting element electrically connected to one of semiconductor elements, a light receiving element electrically converting a first electric signal into an optical signal and electrically connected to a different one of the stacked semiconductor elements, And converting the optical signal into a second electric signal.

6.

A semiconductor device having semiconductor elements stacked by transferring semiconductor elements formed on different substrates, wherein a film formed of resin and a film formed partly are provided between the stacked semiconductor elements A light emitting element electrically connected to one of the stacked semiconductor elements, wherein the first electric signal is converted into an optical signal, and a different one of the stacked semiconductor elements And converting the optical signal into a second electric signal in a light receiving element electrically connected to the light receiving element.

7.

A semiconductor device formed by peeling off a plurality of semiconductor elements respectively formed on a plurality of substrates and laminating them on an element substrate, the semiconductor device comprising: a plurality of semiconductor elements formed of resin between the stacked semiconductor elements A light emitting element having a film and being electrically connected to one of the plurality of semiconductor elements, the first electric signal is converted into an optical signal and electrically connected to a different one of the plurality of semiconductor elements And converting said optical signal into a second electric signal in said light receiving element.

8.

A semiconductor device formed by separating a plurality of semiconductor elements respectively formed on a plurality of substrates and laminating them on an element substrate, wherein the

semiconductor device is formed of a resin between the stacked semiconductor elements Wherein the first electric signal is converted into an optical signal in a light emitting element in which a film formed on a substrate and a partially provided metal oxide are present and which is electrically connected to one of the plurality of semiconductor elements, And a light receiving element electrically connected to a different one of the plurality of semiconductor elements converts the optical signal into a second electric signal.

9.

A semiconductor device in which a plurality of thin film integrated circuits are stacked, wherein the laminated thin film integrated circuits are bonded with a resin, and a light emitting element electrically connected to one of the laminated thin film integrated circuits , Converting the first electrical signal into an optical signal and converting the optical signal to a second electrical signal in a light receiving element electrically connected to a different one of the stacked thin film integrated circuits .

10.

A semiconductor device in which a plurality of thin film integrated circuits are stacked, wherein the laminated thin film integrated circuits are bonded with a resin, and are partially provided on one surface of each of the stacked thin film integrated circuits A light emitting element electrically connected to one of the laminated thin film integrated circuits in which a metal oxide exists, converting a first electric signal into an optical signal, and converting the different thin film integrated circuits And converting the optical signal into a second electric signal in a light receiving element electrically connected to the semiconductor element.

(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2004-247405
(P2004-247405A)
(43)公開日　平成16年9月2日(2004.9.2)

(51)Int.Cl.[7]	F I		テーマコード（参考）
H01L 27/00	H01L 27/00	301C	4M118
H01L 27/12	H01L 27/12	B	5F089
H01L 27/146	H01L 27/15	T	5F110
H01L 27/15	H01L 31/12	D	
H01L 29/786	H01L 29/78	613Z	

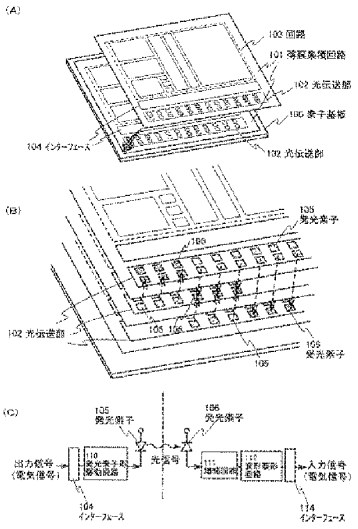
審査請求 未請求　請求項の数 10　OL　（全 30 頁）　最終頁に続く

(21)出願番号　特願2003-33833(P2003-33833)
(22)出願日　平成15年2月12日(2003.2.12)

(71)出願人　000153878
株式会社半導体エネルギー研究所
神奈川県厚木市長谷３９８番地
(72)発明者　加藤　清
神奈川県厚木市長谷３９８番地　株式会社半導体エネルギー研究所内
(72)発明者　高山　徹
神奈川県厚木市長谷３９８番地　株式会社半導体エネルギー研究所内
(72)発明者　丸山　純矢
神奈川県厚木市長谷３９８番地　株式会社半導体エネルギー研究所内
(72)発明者　後藤　裕吾
神奈川県厚木市長谷３９８番地　株式会社半導体エネルギー研究所内
最終頁に続く

(54)【発明の名称】半導体装置

(57)【要約】
【課題】安価なガラス基板を用い、情報量の増加に対応でき、なおかつ高性能で高速動作が可能な集積回路を有する半導体装置の提供を課題とする。
【解決手段】異なる基板上に形成された半導体素子を転写することで積層された半導体素子を有する半導体装置であって、積層された半導体素子の間には、樹脂で形成された膜と、部分的に設けられた金属酸化物が存在し、積層された半導体素子の１つに電気的に接続された発光素子において、第１の電気信号を光信号に変換し、積層された半導体素子の異なる１つに電気的に接続された受光素子において、光信号を第２の電気信号に変換することを特徴とする半導体装置。
【選択図】　図４



【特許請求の範囲】
【請求項1】
積層された半導体素子を有する半導体装置であって、
前記積層された半導体素子の間に樹脂で形成された膜が存在し、
前記積層された半導体素子どうしの信号の送受は、各半導体素子に電気的に接続された発光素子または受光素子を用いて行なわれることを特徴とする半導体装置。
【請求項2】
積層された半導体素子を有する半導体装置であって、
前記積層された半導体素子の間には、樹脂で形成された膜と、部分的に設けられた金属酸化物が存在し、
前記積層された半導体素子どうしの信号の送受は、各半導体素子に電気的に接続された発光素子または受光素子を用いて行なわれることを特徴とする半導体装置。
【請求項3】
積層された半導体素子を有する半導体装置であって、
前記積層された半導体素子の間に樹脂で形成された膜が存在し、
前記積層された半導体素子の１つに電気的に接続された発光素子において、第１の電気信号を光信号に変換し、前記積層された半導体素子の異なる１つに電気的に接続された受光素子において、前記光信号を第２の電気信号に変換することを特徴とする半導体装置。
【請求項4】
積層された半導体素子を有する半導体装置であって、
前記積層された半導体素子の間には、樹脂で形成された膜と、部分的に設けられた金属酸化物が存在し、
前記積層された半導体素子の１つに電気的に接続された発光素子において、第１の電気信号を光信号に変換し、前記積層された半導体素子の異なる１つに電気的に接続された受光素子において、前記光信号を第２の電気信号に変換することを特徴とする半導体装置。
【請求項5】
異なる基板上に形成された半導体素子を転写することで積層された半導体素子を有する半導体装置であって、
前記積層された半導体素子の間に樹脂で形成された膜が存在し、
前記積層された半導体素子の１つに電気的に接続された発光素子において、第１の電気信号を光信号に変換し、前記積層された半導体素子の異なる１つに電気的に接続された受光素子において、前記光信号を第２の電気信号に変換することを特徴とする半導体装置。
【請求項6】
異なる基板上に形成された半導体素子を転写することで積層された半導体素子を有する半導体装置であって、
前記積層された半導体素子の間には、樹脂で形成された膜と、部分的に設けられた金属酸化物が存在し、
前記積層された半導体素子の１つに電気的に接続された発光素子において、第１の電気信号を光信号に変換し、前記積層された半導体素子の異なる１つに電気的に接続された受光素子において、前記光信号を第２の電気信号に変換することを特徴とする半導体装置。
【請求項7】
複数の基板上にそれぞれ形成された複数の半導体素子を剥離し、素子基板上に積層することで形成される半導体装置であって、
前記積層された複数の半導体素子の間に樹脂で形成された膜が存在し、
前記複数の半導体素子の１つに電気的に接続された発光素子において、第１の電気信号を光信号に変換し、前記複数の半導体素子の異なる１つに電気的に接続された受光素子において、前記光信号を第２の電気信号に変換することを特徴とする半導体装置。
【請求項8】
複数の基板上にそれぞれ形成された複数の半導体素子を剥離し、素子基板上に積層することで形成される半導体装置であって、

前記積層された複数の半導体素子の間には、樹脂で形成された膜と、部分的に設けられた金属酸化物が存在し、
前記複数の半導体素子の１つに電気的に接続された発光素子において、第１の電気信号を光信号に変換し、前記複数の半導体素子の異なる１つに電気的に接続された受光素子において、前記光信号を第２の電気信号に変換することを特徴とする半導体装置。
【請求項9】
複数の薄膜集積回路が積層された半導体装置であって、
前記積層された薄膜集積回路どうしは樹脂で接着されており、
前記積層された薄膜集積回路の１つに電気的に接続された発光素子において、第１の電気信号を光信号に変換し、前記積層された薄膜集積回路の異なる１つに電気的に接続された受光素子において、前記光信号を第２の電気信号に変換することを特徴とする半導体装置。
【請求項10】
複数の薄膜集積回路が積層された半導体装置であって、
前記積層された薄膜集積回路どうしは樹脂で接着されており、
前記積層された各薄膜集積回路の一方の面には部分的に設けられた金属酸化物が存在し、
前記積層された薄膜集積回路の１つに電気的に接続された発光素子において、第１の電気信号を光信号に変換し、前記積層された薄膜集積回路の異なる１つに電気的に接続された受光素子において、前記光信号を第２の電気信号に変換することを特徴とする半導体装置。

【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、ガラス基板上に形成された結晶構造を有する半導体膜を用いた集積回路（薄膜集積回路）に係り、該薄膜集積回路内における信号の伝達を、光インターコネクションにより行なう薄膜集積回路及び該薄膜集積回路を用いた半導体装置に関する。
【０００２】
【従来の技術】
微細化の最小寸法がサブミクロン領域に近づくにつれて、ＬＳＩの高集積化に歯止めがかかりつつある。高集積化が不充分だと、配線抵抗が抑えられないため、信号の遅延や波形の乱れ、信号の伝送量の低下を引き起こし、集積回路の情報処理能力が制限される。また、配線の長距離化に伴い配線間の寄生容量が増大すると、配線への充放電エネルギーが増加するため、消費電力が抑えられないという問題もある。
【０００３】
そこで更なる高集積化を進めるために、三次元集積回路に関する技術が注目されている。三次元集積回路は、層間絶縁膜を間に挟んで半導体素子を階層的に積み上げていく技術であり、集積回路の飛躍的な高集積化、配線長の短縮による高速化、多機能化を実現しようとするのものである。
【０００４】
三次元集積回路を形成するには、各層に形成された半導体素子どうしを電気的に接続する必要がある。各層間の半導体素子の電気的な接続は、ワイヤボンディング法や、フリップチップ法などを用いて行なうことができる。また、各層を隔てている層間絶縁膜にビアホールを形成し、電気的な接続を行なうこともできる。また、発光素子を用いて電気信号を光信号に変換して、光伝送により接続を行なう、所謂光インターコネクションを用いることも可能である。
【０００５】
下記、特許文献１に、光信号を用いてＣＰＵ、メモリーコントローラと、メモリ間のデータの伝送を行なう技術について開示されている。
【０００６】
【特許文献１】

特開２０００－５８８８２号公報
【０００７】
特許文献１に記載されているように、光信号を用いてデータの伝送を行なうことで、配線抵抗に起因するスキューや、不要な電波輻射の低減を図ることができる。
【０００８】
【発明が解決しようとする課題】
上述したように、集積回路の三次元化は、さらなる高集積化を実現するにあたって重要な技術であるが、半導体装置の処理する情報量のさらなる増加に対応するために積層数を増やすと、各層間の接続が困難になるという問題点があった。
【０００９】
例えばワイヤボンディング法の場合、近隣の層を金属細線で接続するのは容易だが、接続する２つの層が離れると、接続に用いる金属細線の形成を阻害しないように、間に存在する層の形状を制限する必要がある。そのため、多層化が進み、接続する２つの層の間に存在する層の数が増えると、金属細線を用いての接続は難しくなる。また金属細線は、通常各層のエッジ近傍においてボンディングするため、金属細線を接続するエリアが限られ、各層の端子数に制限が生じる。
【００１０】
フリップチップ法の場合、電気的な接続に用いるエリアに限定がないため、ワイヤボンディング法に比べて各層の端子数を比較的多く設けることができる。しかしワイヤボンディング法と同様に、近隣の層を接続するのに適しているが、接続する２つの層が離れていると、ソルダーボールによる接続を阻害しないように、間に存在する層の形状を制限する必要があり、多層化に適さない。
【００１１】
また層間絶縁膜にビアホールを形成して接続する場合も、接続する２つの層が離れていると、間に存在する全ての層を貫通するビアホールを形成する必要がある。ビアホールは、貫通する層が厚くなればなるほど、その口径を大きくする必要があり、却って配線の高集積化の妨げとなるため好ましくない。
【００１２】
光伝送を用いる場合は、上述した他の場合と異なり、多層化によって接続するべき２つの層が離れていたとしても、光ファイバー等の光学系を用いることで、他の層の形状を制限せずとも層間の接続を行なうことができる。しかし、一般的に光ファイバーは高価であり、集積回路を用いた半導体装置のコストという観点からすると、用いるのにはあまり好ましい部材とは言えない。
【００１３】
また１つの基板上に、複数の層からなる薄膜集積回路を連続して形成すれば、光ファイバー等の光学系を用いずに光伝送を容易に行なうことができる。しかし、上層の半導体素子に用いる半導体膜を結晶化する際、下層の半導体素子の耐熱温度以下で行なうという制限があり、プロセス上の制約が多く、より高い特性の薄膜集積回路を形成するのが難しい。
【００１４】
本発明は上述した問題に鑑み、安価なガラス基板を用い、情報量の増加に対応でき、なおかつ高性能で高速動作が可能な集積回路を有する半導体装置の提供を課題とする。
【００１５】
【課題を解決するための手段】
本発明では上記課題を解決するために、耐熱性を有する基板に成膜した薄膜の半導体膜を用いて薄膜集積回路を作製し、該薄膜集積回路を別途用意した基板上に転写する。この転写を繰り返すことで、複数の薄膜集積回路を積層し、三次元化を図る。転写前の薄膜集積回路を形成するための基板は、半導体素子の作製工程における熱処理に耐え得る程度の耐熱性を有していれば良い。
【００１６】
そして各層の薄膜集積回路（以下、薄膜チップと呼ぶ）どうしの接続は、光信号を用いた

所謂光インターコネクションで行なう。つまり、積層された薄膜チップの、上段の回路の出力側に発光素子を設け、該発光素子と対向するように、後段の回路の入力側に受光素子を形成する。そして上段の回路から出力された電気信号を発光素子において光信号に変換し、該光信号を受けた受光素子が電気信号に変換し、後段の回路に入力する。

【００１７】

薄膜チップの転写は、基板と薄膜集積回路の間に金属酸化膜を設け、該金属酸化膜を結晶化により脆弱化して薄膜集積回路を剥離し、転写する方法、基板と薄膜集積回路の間に水素を含む非晶質珪素膜を設け、レーザ光の照射またはエッチングにより該非晶質珪素膜を除去することで基板と薄膜集積回路とを剥離し、転写する方法、薄膜集積回路が形成された基板を機械的に削除または溶液やガスによるエッチングで除去することで薄膜集積回路を基板から切り離し、転写する方法等、様々な方法を用いることができる。

【００１８】

例えば、金属酸化膜を用いた転写は、以下の手法に従って行なうことができる。

【００１９】

まず半導体素子の作製工程における熱処理に、耐え得るような耐熱性を有する第１の基板を用意する。そして該第１の基板上に金属膜を成膜し、該金属膜の表面を酸化することで数ｎｍの極薄い金属酸化膜を成膜する。次に該金属酸化膜上に絶縁膜、半導体膜を順に積層するように成膜する。絶縁膜は単層であっても、複数の膜を積層したものであっても良い。例えば窒化珪素、窒化酸化珪素、酸化珪素などを用いることができる。そして該半導体膜を用いて、薄膜集積回路に用いられる半導体素子を作製する。

【００２０】

半導体素子を形成したら、該半導体素子を覆うように第２の基板を貼り合わせ、第１の基板と第２の基板の間に半導体素子が挟まれた状態を作る。なお、薄膜集積回路と並行して表示装置を形成する場合は、表示素子を完成させる前に、第２の基板を貼り合わせる。例えば表示素子として液晶セルを用いる場合は、半導体素子の一つである薄膜トランジスタ（ＴＦＴ）に電気的に接続された液晶セルの画素電極や、該画素電極を覆っている配向膜を作製した後、対向電極が形成された対向基板を貼り合わせる前に行なう。

【００２１】

そして第１の基板の半導体素子が形成されている側とは反対の側に、第１の基板の剛性を補強するために第３の基板を貼り合わせる。第２の基板よりも第１の基板の剛性が高いほうが、第１の基板を引き剥がす際に、半導体素子に損傷が与えられにくくスムーズに剥がすことができる。ただし第３の基板は、後に第１の基板を半導体素子から引き剥がす際に、第１の基板の剛性が十分であれば、必ずしも貼り合わせる必要はない。

【００２２】

次に、加熱処理等を施すことで金属酸化膜を結晶化し、金属酸化膜の脆性を高め、第１の基板を半導体素子から剥離しやすくする。そして第１の基板を第３の基板と共に、半導体素子から引き剥がす。なお、金属酸化膜を結晶化するための加熱処理は、第３の基板を貼り合わせる前であってもよいし、第２の基板を貼り合わせる前であってもよい。或いは、半導体素子を形成する工程において行なわれる熱処理が、この金属酸化膜の結晶化の工程を兼ねていても良い。

【００２３】

そして第１の基板を第３の基板と共に、半導体素子から引き剥がす。この引き剥がしによって、金属膜と金属酸化膜の間で分離する部分と、絶縁膜と金属酸化膜の間で分離する部分と、金属酸化膜自体が双方に分離する部分とが生じる。いずれにしろ、半導体素子は第２の基板側に貼り付くように、第１の基板から引き剥がされる。

【００２４】

次に、第１の基板が剥離されることで、第２の基板側に貼り付いた状態の半導体素子を、半導体装置を形成するための基板（以下、素子基板と呼ぶ）に接着剤等で貼り合わせる。そして、第２の基板を剥離し、半導体素子が素子基板に転写された状態にする。該素子基板に固定された半導体素子によって、１つの薄膜チップが形成される。該薄膜チップと他

のチップとの接続を行なうための受光素子及び発光素子は、予め素子基板上に形成しておいてもよいし、薄膜チップを貼り合わせた後に形成しても良い。また既製の発光素子または受光素子を貼り付けて、薄膜チップに電気的に接続するようにしても良い。

【００２５】

そして同様に、別途用意した基板上に、次層の薄膜チップを構成する薄膜集積回路を形成し、既に転写されている薄膜チップに重ねるように、再び転写する。このように転写を繰り返すことで、薄膜チップを素子基板上に積層することができる。このとき、各薄膜チップ間に樹脂等で層間絶縁膜を形成しても良い。また転写の際に用いる接着剤を層間絶縁膜の代わりに用いても良い。

【００２６】

薄膜チップと共に表示装置を転写する場合は、転写した後に表示素子を作製する。転写後に発光素子を作製する場合も同様である。例えば液晶表示装置の場合、具体的には、対向電極、第２の偏光板等が形成されたプラスチック基板（以下、対向基板と呼ぶ）を別途作製しておき、該対向基板を貼り合わせて液晶を注入し、液晶セルを完成させる。なお対向基板には、対向電極、第２の偏光板のみならず、カラーフィルタ、配向膜、ブラックマトリクス等を作製しておいても良い。

【００２７】

本発明では、１層の薄膜チップの厚さをトータルの膜厚５μｍ、より望ましくは２μｍ以下とすることができる。

【００２８】

このように本発明では各層の薄膜チップを別個に作製することができるので、下層の半導体素子の耐熱性を考慮して上層の半導体素子のプロセスに制約がかかることがなく、より特性の優れた半導体素子を形成することができる。

【００２９】

また、ガラス基板上の薄膜の半導体膜で形成された集積回路（以下、薄膜集積回路と呼ぶ）を重ね合わせて三次元化する場合とは異なり、各層の薄膜チップが、ガラス基板などを間に挟まずに、密接するように積層することができるので、ガラス基板による光の拡散を考慮する必要がない。また発光素子と受光素子の間の距離を短くすることができるので、発光素子の光の指向性が劣っていたとしても、ある程度それをカバーすることができる。

【００３０】

なお、素子基板に可撓性を有するプラスチック基板を用いることができる。プラスチック基板としては、極性基のついたノルボルネン樹脂からなるＡＲＴＯＮ：ＪＳＲ製を用いることができる。また、ポリエチレンテレフタレート（ＰＥＴ）、ポリエーテルスルホン（ＰＥＳ）、ポリエチレンナフタレート（ＰＥＮ）、ポリカーボネート（ＰＣ）、ナイロン、ポリエーテルエーテルケトン（ＰＥＥＫ）、ポリスルホン（ＰＳＦ）、ポリエーテルイミド（ＰＥＩ）、ポリアリレート（ＰＡＲ）、ポリブチレンテレフタレート（ＰＢＴ）、ポリイミドなどのプラスチック基板を用いることができる。

【００３１】

可撓性を有するプラスチック基板を用いることで、ガラス基板と比較して振動、衝撃に対する機械的強度に優れ、なおかつ軽量化された薄膜集積回路を作製することができ、また薄膜集積回路の厚さを抑えやすい。また素子基板が可撓性を有しているため、薄膜集積回路の形状の自由度が高まる。よって例えば、円柱状のビンなどに貼り付けられるような、曲面を有する形状に薄膜集積回路を形成することも可能である。

【００３２】

そして本発明の薄膜集積回路を半導体装置に用いることで、集積回路に用いるスペースをより広く確保することができ、半導体装置の軽量化または小型化を妨げることなく高機能化を実現することができる。特に携帯用の半導体装置の場合、軽量化、小型化することで使い勝手が飛躍的に向上するため、本発明の薄膜集積回路を用いることは非常に有用である。

【００３３】

そして、光インターコネクションを用いることで、配線抵抗に起因するスキューや不要な電波輻射の影響を低減しながら、薄膜チップ間でデータの伝送を行うことができる。また、電気信号から光信号へ変換し、光信号から再び電気信号へ変換する過程において、最終的に得られる電気信号の振幅を自由に制御することができる。

【００３４】

また、一つの発光素子の信号を複数の受光素子で受光することができるので、一つの薄膜チップから同時に複数の他の薄膜チップへ信号を伝送することができ、非常に高速な光バスを形成することができる。

【００３５】

そして単結晶のシリコンウェハとは異なり、各薄膜チップは光を透過するので、光ファイバーなどの特殊な光学系を用いずとも、３層以上の薄膜チップ間の信号の伝送を比較的容易に行なうことが可能である。なお、ガラス基板上に形成されたＴＦＴは単結晶トランジスタに比べて動作速度が劣っている。しかし、薄膜チップ間の信号の伝送に光信号を用いることで、薄膜チップ間を伝送する信号のバス幅を大きく取ることができ、複数の薄膜チップ上の回路で効率の良い並列動作をさせることが可能になるので、ガラス基板上に形成されたＴＦＴの動作速度の低さをカバーすることが可能である。

【００３６】

また、単結晶のシリコンウェハの場合に比べてガラス基板上に形成された回路の集積度が低くても、基板間の信号の伝送に光信号を用いることで複数の薄膜チップを積層することができ、薄膜集積回路が水平方向に嵩張るのを防ぐことができる。さらに配線の長距離化を防ぐことができ、配線容量に起因する消費電力の増加を抑えられる。

【００３７】

また、各薄膜チップごとにプロセスを容易に変更することができる。よって、１枚の基板上に全ての回路を同時に形成する場合とは異なり、各層の薄膜チップに求められる特性に合わせたプロセスを選択することができる。また、安価なガラス基板を用いることでコストを抑えることができる。

【００３８】

そして、別途作製された薄膜チップを組み合わせて１つの集積回路を構成するので、１つの基板上に同時に集積回路を形成する場合に比べて、歩留りを高めることができる。また、薄膜チップ間の信号の伝送に光信号を用いることで、回路間を電気的に接続するためのＦＰＣ等の端子の数を抑えることができ、機械的強度における信頼性を高めることができる。さらに、処理する信号の情報量が増加しても、端子の部分における接点不良の発生による歩留りの低下を抑えることができる。

【００３９】

そして、光信号の送受を行なう発光素子及び受光素子は、ＦＰＣの端子と異なり必ずしも基板の端部に配置する必要がないので、レイアウト上の制約が小さくなり、処理する情報量のさらなる増加に対応しやすい。

【００４０】

このように本発明は、高性能で高速動作が可能な集積回路を有する半導体装置の提供を可能とする。

【００４１】

【発明の実施の形態】

次に、本発明の薄膜集積回路に用いられる半導体素子の具体的な作製方法、及び該半導体素子の素子基板への貼り合わせ（転写）方法について説明する。なお本実施の形態では、半導体素子として２つのＴＦＴを例に挙げて説明するが、薄膜集積回路に含まれる半導体素子はこれに限定されず、あらゆる回路素子を用いることができる。例えば、ＴＦＴの他に、記憶素子（メモリ）、ダイオード、光電変換素子、抵抗素子、コイル、容量素子、インダクタなどが代表的に挙げられる。

【００４２】

まず図１（Ａ）に示すように、スパッタ法を用いて第１の基板５００上に金属膜５０１を

成膜する。ここでは金属膜５０１にタングステンを用い、膜厚を１０ｎｍ～２００ｎｍ、好ましくは５０ｎｍ～７５ｎｍとする。なお本実施の形態では第１の基板５００上に直接金属膜５０１を成膜するが、例えば酸化珪素、窒化珪素、窒化酸化珪素等の絶縁膜で第１の基板５００を覆ってから、金属膜５０１を成膜するようにしても良い。

【００４３】

そして金属膜５０１の成膜後、大気に曝すことなく酸化物膜５０２を積層するように成膜する。ここでは酸化物膜５０２として酸化珪素膜を膜厚１５０ｎｍ～３００ｎｍとなるように成膜する。なお、スパッタ法を用いる場合、第１の基板５００の端面にも成膜が施される。そのため、後の工程における剥離の際に、酸化物膜５０２が第１の基板５００側に残ってしまうのを防ぐために、端面に成膜された金属膜５０１と酸化物膜５０２とをO_2アッシングなどで選択的に除去したり、第１の基板５００の端部をダイシング等でカットすることが好ましい。

【００４４】

また酸化物膜５０２の成膜の際に、スパッタの前段階としてターゲットと基板との間をシャッターで遮断してプラズマを発生させる、プレスパッタを行う。プレスパッタはＡｒを１０ｓｃｃｍ、O_2をそれぞれ３０ｓｃｃｍの流量とし、第１の基板５００の温度を２７０℃、成膜パワーを３ｋＷの平行状態に保って行なう。プレスパッタにより、金属膜５０１と酸化物膜５０２の間に極薄い数ｎｍ（ここでは３ｎｍ）程度の金属酸化膜５０３が形成される。金属酸化膜５０３は、金属膜５０１の表面が酸化することで形成される。よって本実施の形態では、金属酸化膜５０３は酸化タングステンで形成される。

【００４５】

なお本実施の形態では、プレスパッタにより金属酸化膜５０３を形成しているが、本発明はこれに限定されない。例えば酸素、または酸素にＡｒ等の不活性ガスを添加し、プラズマにより意図的に金属膜５０１の表面を酸化し、金属酸化膜５０３を形成するようにしても良い。

【００４６】

次に酸化物膜５０２を成膜した後、ＰＣＶＤ法を用いて下地膜５０４を成膜する。ここでは下地膜５０４として、酸化窒化珪素膜を膜厚１００ｎｍ程度となるように成膜する。そして下地膜５０４を成膜した後、大気に曝さずに半導体膜５０５を形成する。半導体膜５０５の膜厚は２５～１００ｎｍ（好ましくは３０～６０ｎｍ）とする。なお半導体膜５０５は、非晶質半導体であっても良いし、多結晶半導体であっても良い。また半導体は珪素だけではなくシリコンゲルマニウムも用いることができる。シリコンゲルマニウムを用いる場合、ゲルマニウムの濃度は０．０１～４．５ａｔｏｍｉｃ％程度であることが好ましい。

【００４７】

次に、半導体膜５０５を公知の技術により結晶化する。公知の結晶化方法としては、電熱炉を使用した熱結晶化方法、レーザー光を用いたレーザ結晶化法、赤外光を用いたランプアニール結晶化法がある。或いは特開平７－１３０６５２号公報で開示された技術に従って、触媒元素を用いる結晶化法を用いることもできる。

【００４８】

また、なお予め多結晶半導体膜である半導体膜５０５を、スパッタ法、プラズマＣＶＤ法、熱ＣＶＤ法などで形成するようにしても良い。

【００４９】

本実施の形態ではレーザ結晶化により、半導体膜５０５を結晶化する。連続発振が可能な固体レーザを用い、基本波の第２高調波～第４高調波のレーザ光を照射することで、大粒径の結晶を得ることができる。例えば、代表的には、Ｎｄ：YVO_4レーザ（基本波１０６４ｎｍ）の第２高調波（５３２ｎｍ）や第３高調波（３５５ｎｍ）を用いるのが望ましい。具体的には、連続発振のYVO_4レーザから射出されたレーザ光を非線形光学素子により高調波に変換し、出力１０Ｗのレーザ光を得る。また非線形光学素子を用いて、高調波を射出する方法もある。そして、好ましくは光学系により照射面にて矩形状または楕円

形状のレーザ光に成形して、半導体膜５０５に照射する。このときのエネルギー密度は０．０１～１００ＭＷ／ｃｍ^{2}程度（好ましくは０．１～１０ＭＷ／ｃｍ^2）が必要である。そして、走査速度を１０～２０００ｃｍ／ｓ程度とし、照射する。

【００５０】

なおレーザ結晶化は、連続発振の基本波のレーザ光と連続発振の高調波のレーザ光とを照射するようにしても良いし、連続発振の基本波のレーザ光とパルス発振の高調波のレーザ光とを照射するようにしても良い。

【００５１】

なお、希ガスや窒素などの不活性ガス雰囲気中でレーザ光を照射するようにしても良い。これにより、レーザ光照射による半導体表面の荒れを抑えることができ、界面準位密度のばらつきによって生じる閾値のばらつきを抑えることができる。

【００５２】

上述した半レーザ光の照射により、導体膜５０５の結晶性がより高められる。次に、図１（Ｂ）に示すように、結晶性がより高められた半導体膜５０５をパターニングし、島状の半導体膜５０７、５０８を形成し、該島状の半導体膜５０７、５０８を用いてＴＦＴに代表される各種の半導体素子を形成する。なお本実施の形態では、下地膜５０４と島状の半導体膜５０７、５０８とが接しているが、半導体素子によっては、下地膜５０４と島状の半導体膜５０７、５０８との間に、電極や他の絶縁膜等が形成されていても良い。例えば半導体素子の１つであるボトムゲート型のＴＦＴの場合、下地膜５０４と島状の半導体膜５０７、５０８との間に、ゲート電極とゲート絶縁膜が形成される。

【００５３】

本実施の形態では、島状の半導体膜５０７、５０８を用いてトップゲート型のＴＦＴ５０９、５１０を形成する（図１（Ｃ））。具体的には、島状の半導体膜５０７、５０８を覆うようにゲート絶縁膜５１１を成膜する。そして、ゲート絶縁膜５１１上に導電膜を成膜し、パターニングすることで、ゲート電極５１２、５１３を形成する。そして、ゲート電極５１２、５１３や、あるいはレジストを成膜しパターニングしたものをマスクとして用い、島状の半導体膜５０７、５０８にｎ型を付与する不純物を添加し、ソース領域、ドレイン領域、さらにはＬＤＤ領域等を形成する。なおここではＴＦＴ５０９、５１０をｎ型とするが、ｐ型のＴＦＴの場合は、ｐ型の導電性を付与する不純物を添加する。

【００５４】

上記一連の工程によってＴＦＴ５０９、５１０を形成することができる。なおＴＦＴの作製方法は、上述した工程に限定されない。

【００５５】

次にＴＦＴ５０９、５１０を覆って第１の層間絶縁膜５１４を成膜する。そして、ゲート絶縁膜５１１及び第１の層間絶縁膜５１４にコンタクトホールを形成した後、コンタクトホールを介してＴＦＴ５０９、５１０と接続する端子５１５～５１８を、第１の層間絶縁膜５１４に接するように形成する。

【００５６】

端子５１５の一部は、受光素子（光電変換素子）のカソード電極として用いる。端子５１５～５１８として、スパッタ法で成膜したアルミニウムを用いたが、材料はこれに限定されない。その他の金属、例えばチタン、タンタル、タングステン、銅を用いることができる。また、チタン、アルミニウム、チタンでなる積層膜を用いてもよい。

【００５７】

次に、水素を含有する非晶質珪素膜を基板全面に成膜した後にパターニングすることで、光電変換層５１９を形成する。そして、基板全面に透明導電膜を形成する。本実施例では透明導電膜として厚さ２００ｎｍのＩＴＯをスパッタ法で成膜する。透明導電膜をパターニングし、アノード電極５２０を形成する（図１（Ｃ））。カソード電極として用いる端子５１５と、光電変換層５１９と、アノード電極５２０とが重なった部分が、受光素子５２１として機能する。

【００５８】

なお、受光素子の構成は、図１（Ｃ）に示した構成に限定されない。受光素子が受ける光の入射方向に合わせて、設計者が適宜受光素子の構成を決めることができる。

【００５９】

そして、受光素子５２１、端子５１５～５１８を覆うように、第１の層間絶縁膜５１４上に第２の層間絶縁膜５２３を成膜する。なお、第２の層間絶縁膜５２３は、後の貼り合わせの工程において受光素子５２１を保護するために設けており、必ずしも成膜する必要はない。後に形成する保護層５２４が、その役割を担っていても良い。

【００６０】

次に、第２の層間絶縁膜５２３を覆って、保護層５２４を形成する。保護層５２４は、後に第２の基板を貼り合わせたり剥離したりする際に、半導体素子を保護する機能を有しており、第２の基板の剥離後に除去することが可能な材料を用いる。例えば、水またはアルコール類に可溶なエポキシ系、アクリレート系、シリコーン系の樹脂を全面に塗布し、焼成することで保護層５２４を形成することができる。

【００６１】

本実施の形態ではスピンコートで水溶性樹脂（東亜合成製：ＶＬ－ＷＳＨＬ１０）を膜厚３０μｍとなるように塗布し、仮硬化させるために２分間の露光を行ったあと、ＵＶ光を裏面から２．５分、表面から１０分、合計１２．５分の露光を行って本硬化させて、保護層５２４を形成する（図１（Ｄ））。

【００６２】

なお、複数の有機樹脂を積層する場合、有機樹脂どうしでは使用している溶媒によって塗布または焼成時に一部溶解したり、密着性が高くなりすぎる恐れがある。従って、第２の層間絶縁膜５２３と保護層５２４を共に同じ溶媒に可溶な有機樹脂を用いる場合、後の工程において保護層５２４の除去がスムーズに行なわれるように、第２の層間絶縁膜５２３と後に成膜される保護層５２４との間に、無機絶縁膜（SiN_X 膜、SiN_XO_Y 膜、AlN_X 膜、またはAlN_XO_Y 膜）を形成しておくことが好ましい。

【００６３】

次に、後の剥離を行ない易くするために、金属酸化膜５０３を結晶化させる。結晶化により、金属酸化膜５０３が粒界において割れやすくなり、脆性を高めることができる。本実施の形態では、４２０℃～５５０℃、０．５～５時間程度加熱処理を行ない、結晶化を行なった。

【００６４】

次に、金属酸化膜５０３と酸化物膜５０２の間の密着性、または金属酸化膜５０３と金属膜５０１の間の密着性を部分的に低下させ、剥離開始のきっかけとなる部分を形成する処理を行う。具体的には、剥離しようとする領域の周縁に沿って外部から局所的に圧力を加えて金属酸化膜５０３の層内または界面近傍の一部に損傷を与える。本実施の形態では、ダイヤモンドペンなどの硬い針を金属酸化膜５０３の端部近傍に垂直に押しつけ、そのまま荷重をかけた状態で金属酸化膜５０３に沿って動かす。好ましくは、スクライバー装置を用い、押し込み量を０．１ｍｍ～２ｍｍとし、圧力をかけて動かせばよい。このように、剥離を行う前に、剥離が開始されるきっかけとなるような、密着性の低下した部分を形成することで、後の剥離工程における不良を低減させることができ、歩留まり向上につながる。

【００６５】

次いで、両面テープ５２５を用い、保護層５２４に第２の基板５２６を貼り付け、さらに両面テープ５２７を用い、第１の基板５００に第３の基板５２８を貼り付ける（図１（Ｅ））。なお両面テープではなく接着剤を用いてもよい。例えば紫外線によって剥離する接着剤を用いることで、第２の基板剥離の際に半導体素子にかかる負担を軽減させることができる。

【００６６】

第３の基板５２８を貼り付けることで、後の剥離工程で第１の基板５００が破損するのを防ぐことができる。第２の基板５２６および第３の基板５２８としては、第１の基板５０

０と同程度が、より剛性の高い基板、例えば石英基板、半導体基板を用いることが好ましい。

【００６７】

次いで、金属膜５０１と酸化物膜５０２とを物理的に引き剥がす。引き剥がしは、先の工程において、金属酸化膜５０３の金属膜５０１または酸化物膜５０２に対する密着性が部分的に低下した領域から開始する。

【００６８】

引き剥がしによって、金属膜５０１と金属酸化膜５０３の間で分離する部分と、酸化物膜５０２と金属酸化膜５０３の間で分離する部分と、金属酸化膜５０３自体が双方に分離する部分とが生じる。そして第２の基板５２６側に半導体素子（ここではＴＦＴ５０９、５１０）が、第３の基板５２８側に第１の基板５００及び金属膜５０１が、それぞれ貼り付いたまま分離する。引き剥がしは比較的小さな力（例えば、人間の手、ノズルから吹付けられるガスの風圧、超音波等）で行なうことができる。剥離後の状態を図２（Ａ）に示す。

【００６９】

次に接着剤５２９で、薄膜集積回路を最終的に積層する基板（素子基板）５３０と、部分的に金属酸化膜５０３が付着している酸化物層５０２とを接着する（図２（Ｂ））。この接着の際に、両面テープ５２５による第２の基板５２６と保護層５２４との間の密着力よりも、接着剤５２９による酸化物層５０２と素子基板５３０との間の密着力の方が高くなるように、接着剤５２９の材料を選択することが重要である。

【００７０】

なお、金属酸化膜５０３が酸化物膜５０２の表面に残存していると、素子基板５３０との密着性が悪くなる場合があるので、完全にエッチング等で除去してから素子基板５３０に接着させ、密着性を高めるようにしても良い。

【００７１】

接着剤５２９としては、反応硬化型接着剤、熱硬化型接着剤、紫外線硬化型接着剤等の光硬化型接着剤、嫌気型接着剤などの各種硬化型接着剤が挙げられる。さらに好ましくは、銀、ニッケル、アルミニウム、窒化アルミニウムからなる粉末、またはフィラーを含ませて接着剤５２９も高い熱伝導性を備えていることが好ましい。

【００７２】

また本実施の形態では、素子基板５３０として、ソルダーボール５３１を備えたインターポーザを用いる。ソルダーボール５３１は、素子基板５３０の薄膜集積回路を積層する面の裏面に露出しており、素子基板５３０に設けられたコンタクトホールを介して、薄膜集積回路を積層する面に設けられた配線５３２と、電気的に接続している。本実施の形態では、配線５３２は、例えば銅にはんだ、金またはスズをメッキすることで形成されている。

【００７３】

次に図２（Ｃ）に示すように、保護層５２４から両面テープ５２５と第２の基板５２６を順に、または同時に剥がす。なお、接着剤５２９として紫外線硬化型接着剤を用い、両面テープ５２５に紫外線によって剥離するテープまたは接着剤を用いることで、一度の紫外線照射により、両面テープ５２５の剥離と接着剤５２９の硬化を同時に行なうことができる。

【００７４】

そして図２（Ｄ）に示すように保護層５２４を除去する。ここでは保護層５２４に水溶性の樹脂が使われているので、水に溶かして除去する。保護層５２４が残留していると不良の原因となる場合は、除去後の表面に洗浄処理やO_2プラズマ処理を施し、残留している保護層５２４の一部を除去することが好ましい。

【００７５】

なお本実施の形態では、金属膜５０１としてタングステンを用いているが、本発明において金属膜はこの材料に限定されない。その表面に金属酸化膜５０３が形成され、該金属酸

化膜５０３を結晶化することで基板を引き剥がすことができるような金属を含む材料であれば良い。例えば、ＴｉＮ、ＷＮ、Ｍｏ等を用いることができる。またこれらの合金を金属膜として用いる場合、その組成比によって結晶化の際の加熱処理の最適な温度が異なる。よって組成比を調整することで、半導体素子の作製工程にとって妨げとならない温度で加熱処理を行なうことができ、半導体素子のプロセスの選択肢が制限されにくい。

【００７６】

図２（Ｄ）に示すように１層目の薄膜集積回路５３３の、素子基板５３０への転写が終了したら、次に図３（Ａ）に示すように、１層目の薄膜集積回路５３３上への、２層目の薄膜集積回路５３４の転写を行なう。２層目の薄膜集積回路５３４の貼り合わせは、１層目の薄膜集積回路５３３を素子基板５３０に貼り合わせるのと同様の手法を用いて、行なうことができる。即ち、別途用意した第１の基板上に２層目の薄膜集積回路５３４を形成し、該２層目の薄膜集積回路５３４を覆うように第２の基板を貼り合わせる。そして第１の基板を第３の基板で補強した後、第１の基板と２層目の薄膜集積回路５３４との間に設けられた金属酸化膜を結晶化して脆弱化し、第１の基板を第３の基板と共に剥離して取り除く。そして第２の基板に貼り付いた状態の２層目の薄膜集積回路５３４を、１層目の薄膜集積回路５３３上に接着剤５３５を用いて接着した後、第２の基板を取り除き、転写を完了させる。

【００７７】

１層目の薄膜集積回路５３３と２層目の薄膜集積回路５３４の間の距離は、接着剤５３５の厚さを調整することで制御することができる。接着剤の厚さは、貼り合わせの際の圧力に依存するが、本実施例では２０μｍ程度とする。

【００７８】

本実施の形態では、２層目の薄膜集積回路５３４の転写後に、発光素子を形成する場合について説明する。２層目の薄膜集積回路５３４には、１層目の薄膜集積回路５３３に転写した時点で、発光素子の画素電極５３６が形成されている。画素電極５３６は透明導電膜で形成されており、本実施の形態ではＩＴＯを用いている。

【００７９】

なお本実施の形態では、発光素子の画素電極に透明導電膜を用いているが、本発明はこれに限定されない。発光素子からの光の方向を、素子基板側に向かせるのか、それとは反対の方向に向かせるのかによって、発光素子の構成を適宜最適化するのが望ましい。

【００８０】

次に、図３（Ｂ）に示すように、画素電極５３６と薄膜集積回路に電気的に接続された端子５４５が一部露出するような開口部を有する第３の層間絶縁膜５３９を成膜する。本実施の形態では、画素電極を形成した時点で２層目の薄膜集積回路を転写する場合について説明するが、画素電極に加えて、第３の層間絶縁膜を形成した時点で転写を行うことも可能である。

【００８１】

次に、第３の層間絶縁膜５３９の開口部において画素電極５３６と重なるように、電界発光層５３７、陰極５３８を積層するように成膜する。画素電極５３６と、電界発光層５３７と、陰極５３８とが重なる部分が、発光素子５４０に相当する。

【００８２】

なお、画素電極５３６として用いる透明導電膜は、ＩＴＯの他、酸化インジウムに２～２０％の酸化亜鉛（ＺｎＯ）を混合した透明導電膜を用いても良い。画素電極５３６は、その表面が平坦化されるように、ＣＭＰ法、ポリビニルアルコール系の多孔質体で拭浄で研磨しても良い。またＣＭＰ法を用いた研磨後に、画素電極５３６の表面に紫外線照射、酸素プラズマ処理などを行ってもよい。

【００８３】

電界発光層５３７は、発光層単独かもしくは発光層を含む複数の層が積層された構成を有している。また陰極５３８は、仕事関数の小さい導電膜であれば公知の他の材料を用いることができる。例えば、Ｃａ、Ａｌ、ＣａＦ、ＭｇＡｇ、ＡｌＬｉ等が望ましい。

【００８４】
なお第３の層間絶縁膜５３９は、電界発光層５３７が成膜される前に、吸着した水分や酸素等を除去するために真空雰囲気下で加熱しておくことが望ましい。具体的には、１００℃～２００℃、０．５～１時間程度、真空雰囲気下で加熱処理を行なう。望ましくは３×１０－７Ｔｏｒｒ以下とし、可能であるならば３×１０－８Ｔｏｒｒ以下とするのが最も望ましい。そして、第３の層間絶縁膜５３９に真空雰囲気下で加熱処理を施した後に電界発光層５３７を成膜する場合、成膜直前まで真空雰囲気下に保つことで、信頼性をより高めることができる。
【００８５】
また、画素電極５３６が露出している、第３の層間絶縁膜５３９の開口部は、その端部に丸みを帯びている方が望ましい。開口部の端部が丸みを帯びていることで、電界発光層５３７が端部において極端に薄くなって穴があいたりするのを防ぐことができ、画素電極と陰極がショートすることによる発光素子の不良を極力抑えることができる。また、端部において電界発光層５３７の応力を緩和させることで、発光領域が減少するシュリンクとよばれる不良を低減させることができ、信頼性を高めることができる。具体的には、開口部における有機樹脂膜の断面が描いている曲線の曲率半径が、０．２～２μｍ程度であることが望ましい。
【００８６】
そして水分や酸素などの発光素子５４０の劣化を促進させる原因となる物質が、発光素子５４０に入るのを防止するために、発光素子５４０を保護膜５４１で覆う。代表的には、例えばＤＬＣ膜、窒化炭素膜、ＲＦスパッタ法で形成された窒化珪素膜等を、保護膜５４１として用いるのが望ましい。また上述した水分や酸素などの物質を透過させにくい膜と、該膜に比べて水分や酸素などの物質を透過させやすい膜とを積層させて、保護膜として用いることも可能である。
【００８７】
そして、保護膜５４１を成膜した後に、さらに発光素子の封止を確かなものにするために、乾燥剤を添加した樹脂で、発光素子５４０を覆うようにしても良い。なお、該乾燥剤を添加した樹脂で、後に転写される薄膜集積回路を貼り合わせるようにしても良い。
【００８８】
なお本実施の形態では、発光素子として、エレクトロルミネッセンスを利用したＯＬＥＤ（ＯＬＥＤ：Ｏｒｇａｎｉｃ　Ｌｉｇｈｔ　Ｅｍｉｔｔｉｎｇ　Ｄｅｖｉｃｅ）を用いた例について説明したが、本発明はこれに限定されない。本発明では、発光素子として、ＯＬＥＤの他に、ＬＥＤ（ｌｉｇｈｔ　ｅｍｉｔｔｉｎｇ　ｄｉｏｄｅ）、半導体レーザ等を用いることが可能である。発光素子は、指向性を有する発光素子であることが望ましい。
【００８９】
本実施の形態では、２層目の薄膜集積回路５３４に設けられた発光素子５４０からの光が、１層目の薄膜集積回路５３３に設けられた受光素子５２１に入射する。よって、２層目の薄膜集積回路５３４から出力される電気信号が、発光素子５４０において光信号に変換され、そして受光素子５２１において再び電気信号に変換されて、１層目の薄膜集積回路５３３に入力される。
【００９０】
また本実施の形態では、２層目の薄膜集積回路５３４と電気的に接続されている端子５４５と、素子基板５３０に設けられた配線５３２とを、ワイヤボンディング法を用いて、接続する。具体的には、図３（Ｂ）に示すように、ワイヤ５４６を用いて接続する。配線５３２を介して、電源電圧や、各種信号を薄膜集積回路に供給することができる。なお、薄膜集積回路への電気的な接続は、ワイヤボンディング法ではなくフリップチップ法を用いていても良い。また外部からの信号の供給をも全て光伝送により行っていても良い。
【００９１】
また本実施の形態では、ワイヤ５４６が２層目の薄膜集積回路５３４に接続されているが

、本発明はこれに限定されない。例えば、１層目の薄膜集積回路５３３の素子基板側に、半導体素子に電気的に接続されたソルダーボールを設け、該ソルダーボールをフリップチップ法で素子基板上の配線と接続するようにしても良い。

【００９２】

このように、発光素子と受光素子を用いた薄膜集積回路間の光伝送により、各層間の信号の授受や電源電圧の供給等の接続を行なうことができ、複数の積層された薄膜チップ（薄膜集積回路）から１つの薄膜集積回路が構築される。なお本実施の形態では２層の薄膜集積回路を積層して、１つの薄膜集積回路として用いる例について説明したが、薄膜集積回路の積層数はこれに限定されない。３層以上の複数の薄膜集積回路を積層していても良い。

【００９３】

なお、素子基板は、ガラスエポキシ等を用いたインターポーザに限定されず、他の材料を用いた基板であっても良い。

【００９４】

また、本発明の薄膜集積回路の転写方法は、上述した金属酸化膜を用いた方法に限定されない。例えば、第１の基板と薄膜集積回路との間に非晶質珪素膜を設けておき、レーザ光の照射またはエッチングにより該非晶質珪素膜に部分において、第１の基板と薄膜集積回路とを剥離するようにしても良い。

【００９５】

次に図４（Ａ）を用いて、複数の薄膜集積回路を積層する際に、各層間の発光素子と受光素子の位置関係について説明する。図４（Ａ）は、素子基板１００上に薄膜集積回路を積層している様子を、模式的に示したものである。

【００９６】

各層の薄膜チップ（薄膜集積回路）１０１には、単数または複数の回路１０３が含まれている。さらに、各層の薄膜集積回路１０１には、光信号の送受を行なう発光素子と受光素子がレイアウトされた領域が、単数または複数設けられている。なお図４（Ａ）には、以下、光伝送を行なうための発光素子または受光素子が設けられた領域を光伝送部１０２として示す。また、光伝送部１０２と、その他の回路１０３との間において電気信号をやり取りするために、該電気信号に変調または復調などの処理を施すインターフェース１０４とを有している。

【００９７】

光伝送部１０２には、光信号を受信するための受光素子と、光信号を送信するための発光素子とが形成されている。図４（Ｂ）に、各層の光伝送部１０２の拡大図を示す。１０５は発光素子に相当し、１０６は受光素子に相当する。各層の薄膜集積回路に設けられている発光素子１０５と、受光素子１０６は、互いに他層の薄膜集積回路１０１に設けられている受光素子１０６と、発光素子１０５とにそれぞれ対応している。なお図４（Ａ）と図４（Ｂ）では、発光素子１０５と受光素子１０６が、光伝送部１０２とする領域にひとまとまりにレイアウトされているが、本発明はこの構成に限定されない。回路１０３の内部に発光素子１０５と受光素子１０６が混在していても良い。そしてインターフェース１０４として機能する回路素子が、発光素子１０５と受光素子１０６のレイアウトに合わせて、回路１０３の内部に混在していても良い。

【００９８】

また図４（Ａ）では、各層間の接続を全て光伝送により行なっている例について示しているが、本発明はこれに限定されない。各種信号や電源電圧を、一部電気信号のまま送受しても良く、電気信号のまま送受用の機能、例えば端子などを有していても良い。

【００９９】

図４（Ｃ）に、光伝送部１０２とインターフェース１０４との、より具体的な構成の一例を示す。１つの発光素子１０５には、少なくとも１つの受光素子１０６が対応している。図４（Ｂ）では、光伝送部１０２に発光素子１０５と受光素子１０６のみ示しているが、実際には図４（Ｃ）に示すように、電気信号により発光素子１０５の発光を制御するため

の発光素子用駆動回路１１０と、受光素子から得られた電気信号を増幅するための増幅回路１１１や、得られた電気信号の波形を整形するための波形整形回路などを設ける。なおこれらの機能をインターフェース１０４が備えていても良い。

【０１００】

回路から出力された電気信号（出力信号）がインターフェース１０４によって、発光素子用駆動回路１１０の仕様に合わせた信号に変調された後、発光素子用駆動回路１１０に入力される。そして発光素子用駆動回路１１０では、入力された出力信号に従って、発光素子１０５の発光を制御する。

【０１０１】

受光素子１０６は、発光素子１０５から発せられた光信号を受けて、該光信号を電気信号に変換する。受光素子１０６において得られた電気信号は、増幅回路１１１において増幅され、波形整形回路１１２においてその波形が整形され、インターフェース１１４において各回路の仕様に合わせて復調され、入力信号として後段の回路に入力される。なお、増幅回路１９と波形整形回路２０は必ずしも設ける必要はなく、またこれらの回路の他に、電気信号の波形に何らかの処理を加える回路を有していても良い。

【０１０２】

なお、図４（Ｃ）では、発光素子と受光素子が一対一で対応している例について示したが、本発明はこの構成に限定されない。２つ以上の発光素子が１つの受光素子に対応していても良いし、１つの発光素子が２つ以上の受光素子に対応していても良い。

【０１０３】

図５（Ａ）に、それぞれ異なる層の薄膜集積回路に形成された２つの発光素子１３０、１３１が、さらに別の層の薄膜集積回路に形成された１つの受光素子１３２に対応している様子を示す。発光素子１３０は、発せられる光が、発光素子１３１の形成されている層１３３を透過して、受光素子１３２に入射するように配置されている。上記構成により、発光素子１３０から光信号を受光素子１３２に送っている間に、発光素子１３１が形成されている層１３３の薄膜集積回路において別の動作を行うことができ、逆に発光素子１３１から光信号を受光素子１３２に送っている間に、発光素子１３０が形成されている層１３４の薄膜集積回路において別の動作を行うことができる。

【０１０４】

逆に、それぞれ異なる層の薄膜集積回路に形成された２つの受光素子が、さらに別の層の薄膜集積回路に形成された１つの発光素子に対応している場合、複数の層の薄膜集積回路へ同時に光信号を送信することができる。

【０１０５】

また図５（Ｂ）に、複数の受光素子で得られた電気信号のいずれかを選択し、後段の回路に伝送する選択回路を設けた場合について示す。図５（Ｂ）では、それぞれ異なる層の薄膜集積回路に形成された２つの発光素子１３５、１３６から発せられた光信号を、さらに別の層の薄膜集積回路上に形成された２つの受光素子１３７、１３８において電気信号に変換する。そして得られる２つの電気信号のいずれか一方を、選択回路１３９において選択し、後段の回路に伝送する。上記構成により、図５（Ａ）の場合と同様に、発光素子１つあたりの発振周波数を低くすることができ、発光素子の駆動を制御する発光素子駆動部の負担を小さくすることができる。

【０１０６】

なお、薄膜集積回路間で光信号によるデータの伝送を並列に行うためには、光信号の経路を夫々独立させる必要がある。しかし、光の拡散の度合いによっては、光信号が対応しない受光素子に入射する所謂クロストークが生じる場合がある。光の拡散の度合いは、発光素子から発せられる光の指向性と、光信号の経路における媒質の屈折率に依存する。よって、クロストークがなるべく抑えられるように光の拡散を考慮し、用いる発光素子の光の指向性に合わせて、受光素子と発光素子をレイアウトし、薄膜集積回路の厚さ、薄膜集積回路間の距離、薄膜集積回路間の媒質（接着剤等）を適宜設定することが望ましい。またクロストークを防ぐために、光信号の経路に、円筒形またはそれに近い断面をもつ光ファ

イバーや、平面状の誘電体薄膜にそって光を伝える薄膜導波路等の光導波路を設けても良い。

【０１０７】

なお本発明で用いる基板はガラス基板に限定されない。透過性を有し、半導体素子の形成やその他のプロセスにおける処理温度に耐え得る基板であれば、プラスチック基板など、ガラス基板以外の基板も用いることは当然可能である。

【０１０８】

なお、本実施の形態では、異なる層間において光伝送により信号を伝送する例について説明したが、ある層の薄膜集積回路内において、光伝送を用いて信号の伝送を行なっていても良い。

【０１０９】

なお、発せられる光の方向が素子基板側に向いてる発光素子と、素子基板とは反対の側に向いてる発光素子とが、同一の層内に形成された薄膜集積回路に設けられている場合も想定される。同一の層内に光の出射方向が異なる２種類の発光素子が形成されている、薄膜集積回路の断面図の一例を、図６に示す。

【０１１０】

図６に示す２つの発光素子２０１、２０２は、素子の構成自体は同じであり、透明導電膜から形成された画素電極２０３、２０４と、電界発光層２０５、２０６と、陰極２０７、２０８とを有している。画素電極２０３、電界発光層２０５、陰極２０７が重なっている部分が、発光素子２０１に相当し、画素電極２０４、電界発光層２０６、陰極２０８が重なっている部分が、発光素子２０２に相当する。

【０１１１】

具体的に図６に示す発光素子２０１、２０２が有する電界発光層２０５、２０６は、正孔注入層として膜厚２０ｎｍの銅フタロシアニン（ＣｕＰｃ）、正孔輸送層として膜厚４０ｎｍの４，４’－ビス［Ｎ－（１－ナフチル）－Ｎ－フェニル－アミノ］－ビフェニル（以下、α－ＮＰＤと示す）、発光層としてキナクリドン（ＤＭＱｄ）が添加された膜厚３７．５ｎｍのAlq_3、電子輸送層として膜厚３７．５ｎｍのAlq_3、電子注入層として膜厚１ｎｍのCaF_2が積層されている。

【０１１２】

そして陰極２０７、２０８は光を透過する程度の膜厚を有するＡｌからなる膜で形成されている。図６では、Ａｌを抵抗加熱により気化することで蒸着させ、２０ｎｍ膜厚となるように成膜する。

【０１１３】

なお、正孔注入層としてＣｕＰｃの代わりにポリチオフェン（ＰＥＤＯＴ）を用ても良い。発光素子の積層構造及びその膜厚は、本実施の形態において示した値に限定されない。なお陰極側から光を得るためには、膜厚を薄くする方法の他に、Ｌｉを添加することで仕事関数が小さくなったＩＴＯを用いる方法もある。図６で用いる発光素子は、陽極側と陰極側の両方から光が発せられる構成であれば良い。

【０１１４】

そして図６に示す発光素子２０１は、陰極２０７に接するように、反射膜（または遮蔽膜）２１０が設けられている。この反射膜２１０によって、白抜きの矢印で示すように、発光素子２０１から発せられる光が素子基板側に向いている。なお反射膜２１０は金属膜を蒸着して形成することができるまた反射膜ではなく遮蔽膜を用いる場合、黒の顔料を添加した樹脂などを用いることができる。いずれにしても、発光素子２０１から発せられる光が素子基板側にのみ向くように、反射膜や遮蔽膜のような光の向きを制御できる手段を設ける。

【０１１５】

また図６に示す発光素子２０２は、発光素子２０２に対して素子基板側に、発光素子２０２から発せられる光を反射する金属膜２１１を設ける。金属膜２１１によって、発光素子２０２から発せられる光が、白抜きの矢印で示すように、素子基板とは反対の側に向いて

いる。なお金属膜２１１は、薄膜集積回路内に形成されたＴＦＴのゲート電極や配線と同じ金属膜から形成することができる。また、光を反射するのではなく、単に遮蔽するための遮蔽膜を、金属膜２１１の替わりに設けても良い。遮蔽膜は層間絶縁膜として用いる絶縁膜に黒の顔料を添加したものであっても良い。いずれにしても、発光素子２０１から発せられる光が素子基板とは反対の側にのみ向くように、反射膜や遮蔽膜のような光の向きを制御できる手段を設ける。

【０１１６】

このように、同じ層内であっても、光の向きの異なる発光素子を形成することが可能である。なお、同じ層内における、光の向きの異なる発光素子の作製方法及びその構成は、図６に示した例に限定されない。ただし、発光素子により近い位置に遮蔽膜または反射膜を設けたほうが、より光の指向性を高めることができる。

【０１１７】

なお、本実施の形態では、薄膜集積回路を積層した例について説明したが、薄膜集積回路に半導体素子を用いたフラットパネルディスプレイを積層する様にしても良い。フラットパネルディスプレイを薄膜集積回路に転写する場合、表示素子を完成させる前に転写し、転写後に表示素子を完成させるようにするのが望ましい。例えばＯＬＥＤを用いた発光装置の場合、画素電極を形成した後で転写し、転写後に電界発光層、陰極等を形成し、封止する。また液晶ディスプレイの場合、画素電極上に配向膜を成膜し、ラビング処理を施した後で、転写し、転写後に対向電極を有する対向基板で封止し、液晶を注入する。

【０１１８】

このように本発明では各層の薄膜チップを別個に作製することができるので、下層の半導体素子の耐熱性を考慮して上層の半導体素子のプロセスに制約がかかることがなく、より特性の優れた半導体素子を形成することができる。また、各層の薄膜チップが、ガラス基板などを間に挟まずに、密接するように積層することができるので、ガラス基板による光の拡散を考慮する必要がない。そして発光素子と受光素子の間の距離をより短くすることができるので、発光素子の光の指向性が劣っていたとしても、ある程度それをカバーすることができる。

【０１１９】

そして本発明の薄膜集積回路を半導体装置に用いることで、集積回路に用いるスペースをより広く確保することができ、半導体装置の軽量化または小型化を妨げることなく高機能化を実現することができる。特に携帯用の半導体装置の場合、軽量化、小型化することで使い勝手が飛躍的に向上するため、本発明の薄膜集積回路を用いることは非常に有用である。

【０１２０】

そして、光インターコネクションを用いることで、配線抵抗に起因するスキューや不要な電波輻射の影響を低減しながら、薄膜チップ間でデータの伝送を行うことができる。また、電気信号から光信号へ変換し、光信号から再び電気信号へ変換する過程において、最終的に得られる電気信号の振幅を自由に制御することができる。

【０１２１】

【実施例】

以下、本発明の実施例について説明する。

【０１２２】

（実施例１）

本実施例では、複数の薄膜チップで構築されるＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）を用いて、マイクロプロセッサを形成する例について説明する。

【０１２３】

ガラス基板上に形成されたＴＦＴは単結晶トランジスタに比べて動作速度が遅い。そのため、ガラス基板上にＣＰＵを形成した場合、処理内容が複雑化すると単一のＣＰＵでは、十分な速度で処理を実行することが困難である。そこでＣＰＵの一連の処理を、その目的別にいくつかの処理に分け、各処理に一つの薄膜チップで形成されたＣＰＵを割り当てる

。そして各薄膜チップを光インターコネクションで接続することで、単一のＣＰＵを用いた場合と同じく一連の処理を行うことができる。それぞれの薄膜チップに形成されたＣＰＵは、割り当てられた処理だけを行えばよく、単層のＣＰＵですべての処理を行う場合にくらべて処理速度が向上する。

【０１２４】

図７に、本実施例のマイクロプロセッサの斜視図を示す。図７に示すマイクロプロセッサは、基板４００上に、ＣＰＵ４０１をそれぞれ含む複数の薄膜チップ４０２、メインメモリ４０３、クロックコントローラ４０４、キャッシュコントローラ４０５、シリアルインターフェース４０６、Ｉ／Ｏポート４０７等が形成されている。勿論、図７に示すマイクロプロセッサは簡略化した一例であり、実際のマイクロプロセッサはその用途によって多種多様な構成を有している。

【０１２５】

各薄膜チップ４０２は、光伝送部４０８と、インターフェース４０９と、ＣＰＵ４０１と、キャッシュメモリ４１０とをそれぞれ有している。光伝送部４０８には、電気信号を光信号として出力する機能を有する発光素子と、光信号を電気信号に変換する機能を有する受光素子の両方を有していても良いし、基板によっては片方だけ有していても良い。また光伝送部４０８に全ての発光素子または受光素子を集めるようにレイアウトせずとも、他の回路素子の間に混在するようにレイアウトされていても良い。そして、薄膜チップ４０２どうしの間で、または薄膜チップ４０２と基板４００上に形成された各種回路との間で、光伝送を用いて信号または電源電圧の送受を行なう。なお、全ての信号、電源電圧の送受を光伝送で行なわなくとも良く、一部直接電気信号で送受するようにしても良い。その場合、電気信号を光信号に変換せずにそのまま送受信するための端子を有する。

【０１２６】

キャッシュメモリ４１０は、ＣＰＵ４０１とメインメモリ４０３の間に介在した、小容量で高速のメモリである。ＣＰＵをより高速に動作させるには、それに見合う程度の高速なメモリを必要とする。しかし、ＣＰＵの動作スピードにあったアクセスタイムをもつ高速の大容量メモリを使用した場合、一般的にコストが高くなってしまう。ＣＰＵはキャッシュメモリをアクセスすることによりメインメモリのスピードによらず、高速で動作することが可能となる。

【０１２７】

以下、各ＣＰＵ４０１の動作の一例について説明する。例えば、まず実行初期において、プログラムをメインメモリ４０３や他の外付メモリなどから、各薄膜チップ４０２のキャッシュメモリ４１０（ＳＲＡＭ）にダウンロードする。ＣＰＵ４０１のうち、マスターとなるＣＰＵがこれを行っても良い。

【０１２８】

次にＣＰＵ４０１のうち、スレーブとなる各ＣＰＵは、同じ薄膜チップ４０２のキャッシュメモリ４１０に格納されたプログラムを順に実行する。同じ薄膜チップ４０２のキャッシュメモリ４１０は、プログラムを格納するだけでなく、ワーク領域としても機能し、ＣＰＵ４０１の計算結果等を一時的に格納する。

【０１２９】

各ＣＰＵ４０１が、他のＣＰＵ４０１の出力結果や、メインメモリ４０３といった、キャッシュメモリ４１０以外との信号のやりとりが必要となる場合には、光伝送部４０８を経由して、これを行う。ＣＰＵ４０１の数に応じて全体の動作速度は向上する。特に、ＣＰＵ４０１間の信号や、基板４００外への信号のやりとりが少ない場合に、並列化の効果が高い。

【０１３０】

プログラム例としては、例えば、非常に多くの極小値をもつ位相空間内において最小値を探すような最適化問題（例えば、自動配線、セールスマンの巡回問題）や、バラツキの評価（回路シミュレーション、等）において、モンテカルロ法やシミュレーテッドアニーリングなどを適用する場合が挙げられる。

【０１３１】
これらのプログラムでは、基本的には、独立に、多数回、同じサブプログラムを実行する構造となっており、各サブプログラムを異なるＣＰＵ４０１に担当させる事で、実質的には、各薄膜チップ４０２内のＣＰＵ４０１とキャッシュメモリ４１０で完結したプログラムを実行することができ、理想的な並列計算を行うことが可能となる。
【０１３２】
なお、ＣＰＵ４０１間の処理速度がまちまちだと処理全体で見たときに不都合が起きる場合があるので、スレーブとなる各ＣＰＵ間の処理速度のバランスを、マスターとなるＣＰＵでとるようにしても良い。
【０１３３】
（実施例２）
本発明の集積回路は、各種のメモリに応用することが可能である。メモリを積層することで、実装面積を抑えつつ、メモリを大容量化することができる。そして、各薄膜チップ間のデータの送受に光伝送を用いることで、メモリを積層しても、実装の際に用いるピン数を抑えることができる。
【０１３４】
積層するメモリとして、ＳＲＡＭ（Ｓｔａｔｉｃ　Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）、ＤＲＡＭ（Ｄｙｎａｍｉｃ　Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）等に代表されるＲＡＭ（揮発性メモリ）や、マスクＲＯＭ、ＥＰＲＯＭ（Ｅｒａｓａｂｌｅ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）、ＥＥＰＲＯＭ（Ｅｌｅｃｔｒｉｃａｌｌｙ　Ｅｒａｓａｂｌｅ　Ｐｒｏｇｒａｍｍａｂｌｅ　ＲＯＭ）、フラッシュ・メモリ（ｆｌａｓｈ　ｍｅｍｏｒｙ）、強誘電体メモリ等に代表されるＲＯＭ（不揮発性メモリ）を用いることができる。また、これらのメモリの組み合わせであっても良い。
【０１３５】
そして、メモリを積層する場合、メモリの種類または仕様に合わせて、読み出しか書き込みかを選択するＲ／Ｗ（Ｒｅａｄ／Ｗｒｉｔｅ）回路、プリチャージ回路、リフレッシュ用のコントロール回路（リフレッシュ回路）等を、メモリと同じ薄膜チップ内に設けるようにしても良い。また供給する電源電圧を、各薄膜チップごとに適宜最適な値に設定するようにしても良い。
【０１３６】
本実施例では、ＤＲＡＭを例に挙げ、積層することで形成されるメモリの構成について説明する。
【０１３７】
図８（Ａ）に、１層目の薄膜チップの構成を示す。図８（Ａ）に示す薄膜チップには、ＤＲＡＭのメモリーセルアレイ６０１と、セレクタ回路６０２、列デコーダ６０３、行デコーダ６０４、Ｒ／Ｗ回路６０５、光伝送部６０６が形成されている。
【０１３８】
メモリーセルアレイ６０１には複数のメモリーセルが備えられている。そして、列デコーダ６０３からの信号に従って動作が制御されているセレクタ回路６０２と、行デコーダ６０４とによって、メモリーセルが選択され、データの読み書きが行なわれる。
【０１３９】
接続端子と１層目の薄膜チップとの間の信号の送受は、直接電気信号で、または光伝送部６０６を介して行なうことができる。図８（Ｂ）に、光伝送部６０６の拡大図を示す。光伝送部６０６には、発光素子６１０と、受光素子６１１が備えられており、接続端子から入力された各種信号または１層目の薄膜チップから出力される電気信号は、発光素子６１０において光信号に変換され、他のｋ層目（ｋは２以上の自然数）の薄膜チップに伝送される。また他のｋ層目の薄膜チップから送信される光信号が、受光素子６１１において電気信号に変換され、１層目の薄膜チップまたは接続端子に送られる。
【０１４０】

光伝送部６０６に入力される信号のうち、ＣＥ（Ｃｈｉｐ　ｅｎａｂｌｅ）（１）～ＣＥ（ｎ）は、１～ｎ層目の薄膜チップの１つまたは複数を選択する信号に相当する。その他、接続端子を介して入力される電気信号、ＲＥ（Ｒｅａｄ　ｅｎａｂｌｅ）、ＷＥ（Ｗｒｉｔｅ　ｅｎａｂｌｅ）、Ｄｉｎは、発光素子６１０を介において光信号に変換され、信号ＦＥによって選択された薄膜チップに送られる。また逆に、号ＦＥによって選択された薄膜チップから送られてきた光信号が、１層目の受光素子６１１において電気信号に変換され、接続端子を介して信号Ｄｏｕｔとして出力される。

【０１４１】

なお本実施例において、ＲＥ、ＷＥはＲ／Ｗ回路６０５によるデータの読み出しまたは書き込みを選択するための信号であり、Ｄｉｎはメモリに書きこまれるデータを情報として含む信号に相当し、Ｄｏｕｔは、メモリから読み出されるデータを情報として含む信号に相当する。

【０１４２】

図９に、ｋ層目の薄膜チップの構成を示す。本実施例では、ｋ層目の薄膜チップに、１層目と同じくＤＲＡＭが形成されている例について示す。図９に示すｋ層目の薄膜チップには、ｋ層目のメモリーセルアレイ６２１と、該メモリーセルアレイ６２１に対応するセレクタ回路６２２、列デコーダ６２３、行デコーダ６２４、Ｒ／Ｗ回路６２５、光伝送部６２６がそれぞれ形成されている。

【０１４３】

図示してはいないが、光伝送部６２６には発光素子と受光素子、場合によってはいずれか一方のみが形成されており、光伝送部６２６において他層の薄膜チップとの間の信号の送受が行なわれる。

【０１４４】

なお本実施例では、１層目とｋ層目の薄膜チップに、共にＤＲＡＭが形成されている例について示したが、本発明はこれに限定されず、他のありとあらゆる半導体メモリーを薄膜チップ内に形成し、該薄膜チップを積層することが可能である。また、メモリの駆動に関わるその他の回路は、本実施例で示した回路に限定されない。

【０１４５】

また本実施例では、接続端子を介して入力される電気信号が、１層目の薄膜チップに入力されているが、本発明はこれに限定されない。２層目以降の薄膜チップに、接続端子を介して電気信号が入力されるようにしても良い。

【０１４６】

（実施例３）

本実施例では、複数の薄膜チップで構築される、ワンチップ・マイコン（ｏｎｅ－ｃｈｉｐ　ｍｉｃｒｏｃｏｍｐｕｔｅｒ）の構成ついて説明する。

【０１４７】

図１０に、本実施例のマイクロプロセッサのブロック図を示す。本実施例のマイクロプロセッサは、２層の薄膜チップで構成されている。１層目の薄膜チップ６３０には、ＳＩＯ（Ｓｅｒｉａｌ　Ｉｎｐｕｔ／Ｏｕｔｐｕｔ）６３１、ＭＭＵ（Ｍｅｍｏｒｙ　Ｍａｎａｇｅｍｅｎｔ　Ｕｎｉｔ）６３２、カウンタ６３３、ＤＡＣ（Ｄｉｇｉｔａｌ　ｔｏ　Ａｎａｌｏｇ　ｃｏｎｖｅｒｔｅｒ）６３４、バスコントローラ（バスインターフェース）６３５、光伝送部６３６が設けられている。２層目の薄膜チップ６４０には、ＣＰＵ６４１、ＲＡＭ６４２、ＲＯＭ６４３、光伝送部６４４が設けられている。

【０１４８】

光伝送部６３６、６４４には、それぞれ発光素子と受光素子とが形成されており、光伝送による各種信号または電源電圧の送受が行なわれる。

【０１４９】

本実施例のように、マイコンを積層した薄膜チップで構築することで、マイコンの実装面積を飛躍的に縮小することができる。

【０１５０】

（実施例４）
本実施例では、インターポーザと薄膜チップとの電気的な接続の仕方について説明する。
【０１５１】
図１１（Ａ）に、積層された薄膜チップを、ワイヤボンディング法でインターポーザに接続した場合の、断面構造を斜視図で示す。３０１はインターポーザ、３０２は積層された薄膜チップに相当する。薄膜チップ３０２はインターポーザ３０１上に、接着剤３０４によって貼り合わされている。
【０１５２】
また図１１（Ａ）に示すインターポーザ３０１は、ソルダーボール３０５が設けられたボールグリッドアレイ型である。ソルダーボール３０５は、インターポーザ３０１の薄膜チップ３０２が貼り合わされている側とは反対の側に設けられている。そしてインターポーザ３０１に設けられた配線３０６は、インターポーザ３０１に設けられたコンタクトホールを介して、ソルダーボール３０５と電気的に接続している。
【０１５３】
なお本実施例では、薄膜チップ３０２とソルダーボール３０５との電気的な接続をするための配線３０６を、インターポーザ３０５の薄膜チップ３０２が貼り合わされている面上に設けた例を示しているが、本発明で用いるインターポーザはこれに限定されない。例えば、インターポーザの内部において配線が多層化されて設けられていても良い。
【０１５４】
そして、図１１（Ａ）では、薄膜チップ３０２と配線３０６とが、ワイヤ３０７によって電気的に接続されている。図１１（Ｂ）に、図１１（Ａ）に示したパッケージの断面図を示す。薄膜チップ３０２には半導体素子３０９が設けられており、また薄膜チップ３０２のインターポーザ３０１が設けられている側とは反対側に、パッド３０８が設けられている。本実施例では、パッド３０８が、積層された薄膜チップ３０２の一番上の層に電気的に接続されている。そしてパッド３０８は、インターポーザ３０１に設けられた配線３０６と、ワイヤ３０７によって接続されている。
【０１５５】
３１０はプリント配線基板の一部に相当し、３１１はプリント配線基板３１０に設けられた配線または電極に相当する。配線３０６はソルダーボール３０５を介して、プリント配線基板３１０に設けられた配線または電極３１１に接続される。なおソルダーボール３０５と、配線または電極３１１との接続は、熱圧着や、超音波による振動を加えた熱圧着等様々な方法を用いることができる。なお、アンダーフィルが圧着後のソルダーボール間の隙間を埋めるようにし、接続部分の機械的強度や、パッケージで発生した熱の拡散などの効率を高めるようにしても良い。アンダーフィルは必ずしも用いる必要はないが、インターポーザと薄膜チップの熱膨張係数のミスマッチから生ずる応力により、接続不良が起こるのを防ぐことができる。超音波を加えて圧着する場合、単に熱圧着する場合に比べて接続不良を抑えることができる。特に、接続するソルダーボールが３００程度よりも多い場合に有効である。
【０１５６】
次に図１１（Ｃ）に、フリップチップ法を用いて薄膜チップがインターポーザに接続されている場合の、断面図を示す。図１１（Ｃ）では、薄膜チップ３２２の一番下層の薄膜チップに、ソルダーボール３２７が設けられている。ソルダーボール３２７は、一番下層の薄膜チップのインターポーザ３２１側に設けられており、一番下層の薄膜チップと電気的に接続されている。
【０１５７】
ソルダーボール３２７は、インターポーザ３２１に設けられた配線３２６と接続されている。そして図１１（Ｃ）では、ソルダーボール３２７間の隙間を埋めるように、接着剤としても機能するアンダーフィル３２４が設けられている。またインターポーザ３２１のソルダーボール３２５は、インターポーザ３２１の薄膜チップ３２２が貼り合わされている側とは反対の側に設けられている。そしてインターポーザ３２１に設けられた配線３２６

は、インターポーザ３２５に設けられたコンタクトホールを介して、ソルダーボール３２５と電気的に接続している。

【０１５８】

フリップチップ法の場合、接続するべきパッドの数が増加しても、ワイヤボンディング法に比べて、比較的パッド間のピッチを広く確保することができるので、端子数の多い薄膜チップの接続に向いている。

【０１５９】

なお図１１（Ａ）～図１１（Ｃ）において示した薄膜チップは、ボールグリッドアレイ型のインターポーザに転写されているが、本発明はこれに限定されない。端子が周辺に配置されているリードフレーム型でのインターポーザを用いていても良い。図１１（Ｄ）に、リードフレーム型のインターポーザを用いた場合の、断面構造の斜視図を示す。

【０１６０】

図１１（Ｄ）では、薄膜チップ３５１がワイヤボンディング法により、インターポーザ３５０上の接続端子３５２と接続されている。接続端子３５２は、インターポーザ３５０の薄膜チップ３５１が貼り合わされている面上に配置されている。

【０１６１】

なお本実施例では薄膜チップが剥き出しの状態を示しているが、モールド樹脂で薄膜チップを封止しても良い。

【０１６２】

また本実施例では、薄膜チップをインターポーザ上において積層した例について示すが、本発明は必ずしもこの構成に限定されず、積層する基板はガラス基板、石英基板、シリコン基板、プラスチック基板など、半導体集積回路の形成に用いられる公知の基板を用いることができる。この場合、外部からの信号または電源電圧の供給用に、基板上に接続端子を設けても良いし、光伝送により供給を行っても良い。

【０１６３】

（実施例５）

本発明で用いられる薄膜チップの転写のプロセスは、様々な製造装置を用いて処理が行なわれるので、装置間における基板の搬送が自動化されたインライン型の生産設備を用いることで、生産効率を高めることができる。

【０１６４】

図１２に、実施の形態の図１～図３に示した作製方法において、半導体素子が形成された第１の基板に第２の基板を貼り合わせる工程から、第３の基板を貼り合わせる工程まで、製造装置間における搬送を自動化した、インライン型の生産設備の構成を示す。

【０１６５】

図１２において、７０１と７１４は基板を搬送するための手段（基板搬送手段）に相当し、それぞれ各装置間を、白抜きの矢印で示した方向に行き来する。７０７は、基板搬送手段７０１と基板搬送手段７１４との間において基板の受け渡しをするために、基板を一次的にストックする基板受け渡しステーションに相当する。

【０１６６】

なお本実施例では、各装置間における基板の搬送を、２つの基板搬送手段７０１、７１４で賄っているが、本発明はこれに限定されない。基板搬送手段の数を増やせば、基板搬送の能率を高めることができる。また基板受け渡しステーション７０７のように、基板の受け渡しのために一次的に基板をストックするような設備を、必ずしも設ける必要はない。例えば、装置へ基板を搬入する基板搬送手段と、搬出する基板搬送手段とを異なるようにすることで、基板の受け渡しを行なっても良い。

【０１６７】

７０２はスピナーであり、第１の基板の半導体素子を覆うように、水溶性の接着剤をスピンコート法により塗布する装置である。７０３は露光装置であり、スピナー７０２において塗布された接着剤を紫外線の露光により硬化させることができる。７０４はスクライバーであり、後の工程における剥離の際に、金属酸化膜を境にして半導体素子が剥離しすく

なるように、基板の端部をカットするための装置である。
【０１６８】
７０５は両面テープ貼付装置であり、テープ供給カセット７０６から供給される両面テープを、第１の基板の接着剤が塗布された面に貼り付けることができる。７０８は基板貼付装置であり、基板供給カセット７０９から供給された第２の基板を、前工程で貼り付けられた両面テープを用い、第１の基板に貼り付けることができる。
【０１６９】
７１０は両面テープ貼付装置であり、テープ供給カセット７１１から供給される両面テープを、第１の基板の接着剤が塗布された面とは反対の面に貼り付けることができる。７１２は基板貼付装置であり、基板供給カセット７１３から供給された第３の基板を、前工程で貼り付けられた両面テープを用い、第１の基板に貼り付けることができる。
【０１７０】
なお本実施例では、上述した工程のみインライン化された生産設備について説明したが、本発明はこれに限定されず、他の工程において用いる装置をインライン化するようにしても良い。
【０１７１】
（実施例６）
本発明の集積回路は様々な半導体装置に用いることが可能である。また本発明の集積回路に表示装置を貼り合わせ、半導体装置に用いることも可能である。特に携帯用の半導体装置の場合、軽量化、小型化することで使い勝手が飛躍的に良くなるため、本発明の集積回路を用いることは非常に有用である。
【０１７２】
図１３（Ａ）はシート型の携帯電話であり、本体２１０１、表示部２１０３、音声入力部２１０４、音声出力部２１０５、スイッチ２１０６、外部接続ポート２１０７等を含む。外部接続ポート２１０７を介して、別途用意したイヤホン２１０８を接続することができる。表示部２１０３には、センサを備えたタッチパネル式の表示装置が用いられており、表示部２１０３に表示されたタッチパネル式操作キー２１０９に触れることで、一連の操作を行なうことができる。本発明の薄膜集積回路は、本体２１０１内に設けられた各種信号処理回路として用いることができ、また薄膜集積回路に表示装置を貼り合わせる場合、該表示装置を表示部２１０３に用いることが可能である。
【０１７３】
図１３（Ｂ）は電子ブックであり、本体２２０１、表示部２２０２、操作キー２２０３等を含む。またモデムが本体２２０１に内蔵されていても良い。また本発明の薄膜集積回路は、各種信号処理回路として用いることができる。
【０１７４】
図１３（Ｃ）は腕時計であり、本体２３０１、表示部２３０２、留具２３０３等を含む。本発明の薄膜集積回路は、本体２３０１内に設けられた各種信号処理回路として用いることができ、また薄膜集積回路に表示装置を貼り合わせる場合、該表示装置を表示部２３０２に用いることが可能である。
【０１７５】
図１３（Ｄ）はシート型のパーソナルコンピュータであり、本体２４０１、表示部２４０２、タッチパネル式キーボード２４０３、ポインティングマウス２４０４、外部接続ポート２４０５、電源プラグ２４０６等を含む。タッチパネル式キーボード２４０３、ポインティングマウス２４０４には、センサを備えたタッチパネル式の、表示装置が用いられており、タッチパネル式キーボード２４０３、ポインティングマウス２４０４に触れることで、一連の操作を行なうことができる。本発明の薄膜集積回路は、各種信号処理回路として用いることができ、また薄膜集積回路に表示装置を貼り合わせる場合、該表示装置を表示部２４０２、タッチパネル式キーボード２４０３などに用いることが可能である。
【０１７６】
図１３（Ｄ）は電子カードであり、本体２６０１、表示部２６０２、接続端子２６０３等

を含む。本発明の薄膜集積回路は、本体２６０１内に設けられた各種信号処理回路として用いることができ、また薄膜集積回路に表示装置を貼り合わせる場合、該表示装置を表示部２６０２に用いることが可能である。

【０１７７】

以上の様に、本発明の適用範囲は極めて広く、あらゆる分野の半導体装置に用いることが可能である。また、本実施例の半導体装置は実施例１～５に示したいずれの構成の集積回路を用いても良い。

【０１７８】

（実施例７）

本実施例では、ＯＬＥＤを表示素子として用いた表示装置を、貼り合わせによりフルカラー化する実施例について説明する。

【０１７９】

図１４（Ａ）に、本実施例の表示装置が有する画素の断面図を示す。８０１は１層目の薄膜チップが有するＯＬＥＤに相当し、本実施例では発光色は赤色（Ｒ）とする。また８０２は２層目の薄膜チップが有するＯＬＥＤに相当し、本実施例では発光色は青色（Ｂ）とする。８０３は３層目の薄膜チップが有するＯＬＥＤに相当し、本実施例では発光色は緑色（Ｇ）とする。

【０１８０】

各画素にＯＬＥＤ８０１～８０３が設けられており、１画素に設けらたＯＬＥＤ８０１～８０３は、その発光が観察者から見て重なって見えるように、発光領域を互いに重ね合わせるように配置する。なお、各層の薄膜チップが有するＯＬＥＤの発光色は、本実施例で示した形態に限定されない。

【０１８１】

そして、全てのＯＬＥＤ８０１～８０３の発光が、下層側の薄膜チップ側に向かうように、最も上層の薄膜チップが有するＯＬＥＤ８０３の陰極が光を反射する材料で形成し、ＯＬＥＤ８０１、８０２は光が両側から出射されるような構成の素子とする。

【０１８２】

上記構成により、ＯＬＥＤを表示素子として用いた表示装置をフルカラー化することができる。

【０１８３】

次に、図１４（Ａ）とは異なる形態の、本実施例の表示装置が有する画素の断面図を示す。

【０１８４】

図１４（Ｂ）において、８１１は１層目の薄膜チップが有するＯＬＥＤに相当し、本実施例では発光色は赤色（Ｒ）とする。また８１２は２層目の薄膜チップが有するＯＬＥＤに相当し、本実施例では発光色は青色（Ｂ）とする。８１３は３層目の薄膜チップが有するＯＬＥＤに相当し、本実施例では発光色は緑色（Ｇ）とする。

【０１８５】

各画素にＯＬＥＤ８１１～８１３が設けられており、１画素に設けらたＯＬＥＤ８１１～８１３は、発光領域が互いに重ならないか、もしくは重なっても一部のみにとどまるように配置されている。なお、各層の薄膜チップが有するＯＬＥＤの発光色は、本実施例で示した形態に限定されない。

【０１８６】

そして、全てのＯＬＥＤ８１１～８１３の発光が、下層側の薄膜チップ側に向かうように、各層の薄膜チップが有するＯＬＥＤ８１１～８１３の陰極が光を反射する材料で形成されている。

【０１８７】

上記構成により、ＯＬＥＤを表示素子として用いた表示装置をフルカラー化することができる。

【０１８８】

本実施例で示した構成の表示装置は、各薄膜チップを別個に独立して作製するので、薄膜チップごとに作製方法を最適化するの容易である。よって、用いる電界発光材料の特性に合わせて、各色のＯＬＥＤの素子の構成を最適化し、それに合わせて作製方法を適宜変更することが可能である。

【０１８９】

【発明の効果】

このように本発明では各層の薄膜チップを別個に作製することができるので、下層の半導体素子の耐熱性を考慮して上層の半導体素子のプロセスに制約がかかることがなく、より特性の優れた半導体素子を形成することができる。また、各層の薄膜チップが、ガラス基板などを間に挟まずに、密接するように積層することができるので、ガラス基板による光の拡散を考慮する必要がない。そして発光素子と受光素子の間の距離をより短くすることができるので、発光素子の光の指向性が劣っていたとしても、ある程度それをカバーすることができる。

【０１９０】

そして本発明の薄膜集積回路を半導体装置に用いることで、集積回路に用いるスペースをより広く確保することができ、半導体装置の軽量化または小型化を妨げることなく高機能化を実現することができる。特に携帯用の半導体装置の場合、軽量化、小型化することで使い勝手が飛躍的に向上するため、本発明の薄膜集積回路を用いることは非常に有用である。

【０１９１】

そして、光インターコネクションを用いることで、配線抵抗に起因するスキューや不要な電波輻射の影響を低減しながら、薄膜チップ間でデータの伝送を行うことができる。また、電気信号から光信号へ変換し、光信号から再び電気信号へ変換する過程において、最終的に得られる電気信号の振幅を自由に制御することができる。

【図面の簡単な説明】

【図１】本発明の薄膜集積回路の作製方法を示す図。

【図２】本発明の薄膜集積回路の作製方法を示す図。

【図３】本発明の薄膜集積回路の作製方法を示す図。

【図４】本発明の薄膜集積回路の構成を示す図。

【図５】本発明の薄膜集積回路の構成を示す図。

【図６】本発明の薄膜集積回路に用いられる発光素子の構成を示す図。

【図７】本発明の薄膜集積回路の一例であるマイクロプロセッサの構成を示す図。

【図８】本発明の薄膜集積回路の一例であるＤＲＡＭの構成を示す図。

【図９】本発明の薄膜集積回路の一例であるＤＲＡＭの構成を示す図。

【図１０】本発明の薄膜集積回路の一例であるワンチップマイコンの構成を示す図。

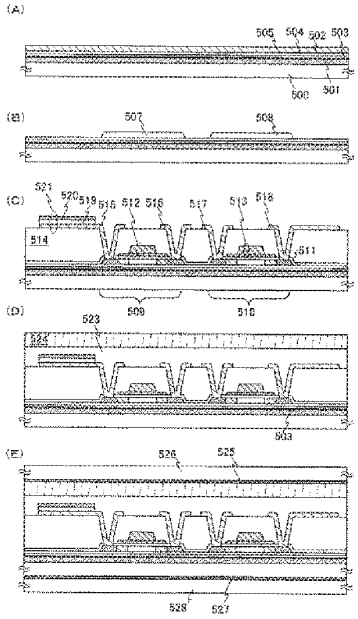
【図１１】本発明の薄膜集積回路をインターポーザ上に接続した様子を示す斜視図及び断面図。

【図１２】本発明の薄膜集積回路の作製工程において用いる生産設備の一例を示す図。

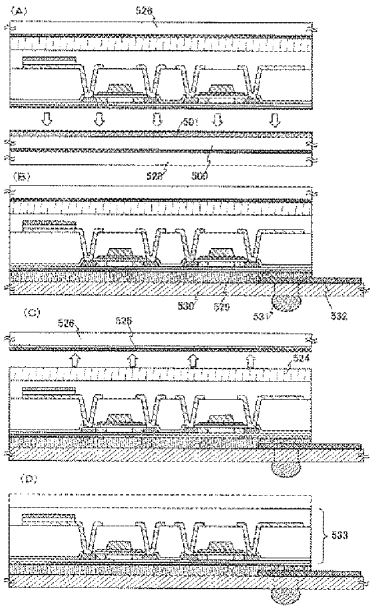
【図１３】本発明の薄膜集積回路を用いて作製される半導体装置の構成を示す図。

【図１４】本発明の半導体装置の１つである表示装置の画素の断面図。

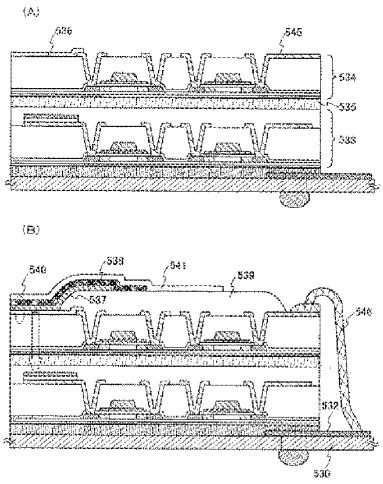
【図1】



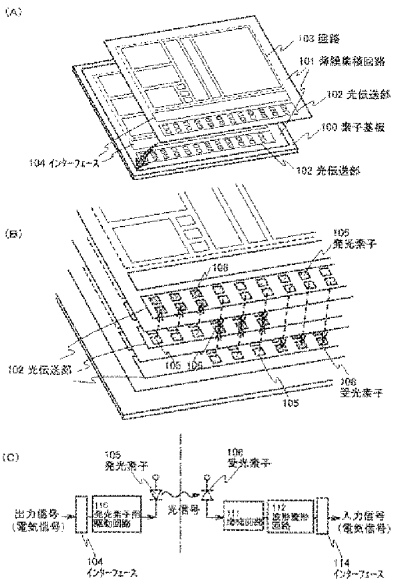
【図2】



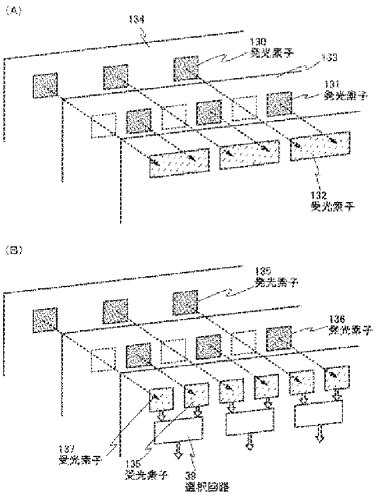
【図3】



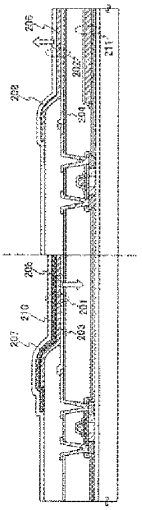
【図4】



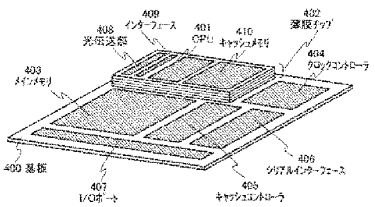
【図5】


(A)
134
130
発光素子
133
131
発光素子
132
受光素子
(B)
135
発光素子
136
発光素子
137
受光素子
135
受光素子
38
選択回路

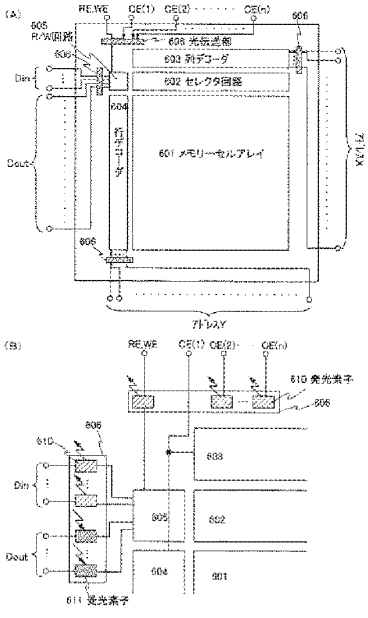
【図6】


205
211
202
204
206
210
207
203
201

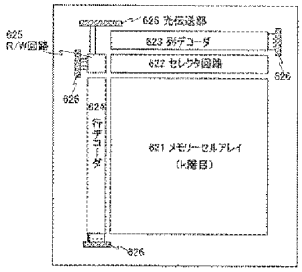
【図7】


409
インターフェース
408
光伝送器
401
CPU
410
キャッシュメモリ
402
基板イップ
404
クロックコントローラ
403
メインメモリ
406
シリアルインターフェース
400 基板
407
I/Oポート
405
キャッシュコントローラ

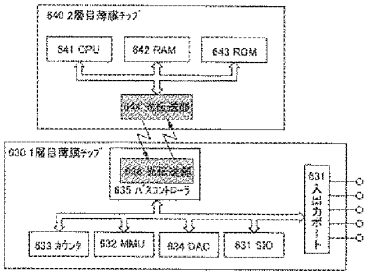
【図8】


(A)
RE,WE
CE(1)
CE(2)
CE(n)
605
R/W回路
606
606 光伝送器
603 列デコーダ
602 セレクタ回路
Din
604
601 メモリーセルアレイ
Dout
606
アドレスY
アドレスX
(B)
RE,WE
CE(1)
CE(2)
CE(n)
610 発光素子
606
610
Din
603
605
602
Dout
604
601
611 受光素子

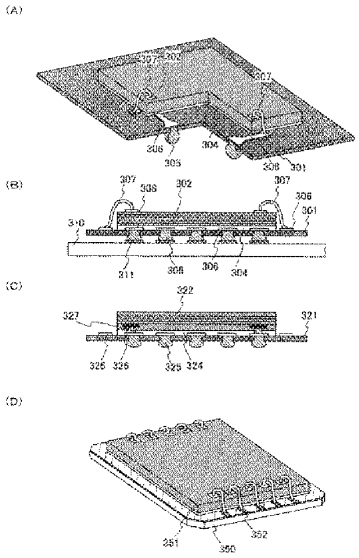
【図9】



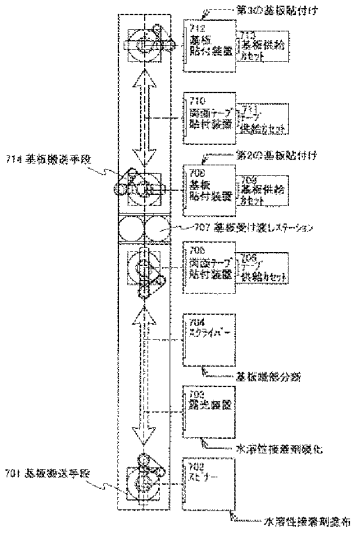
【図10】



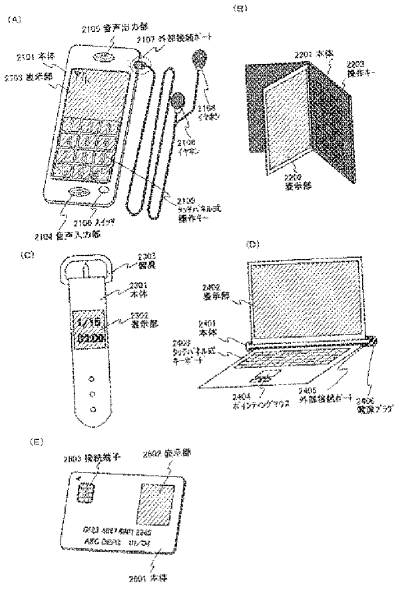
【図11】



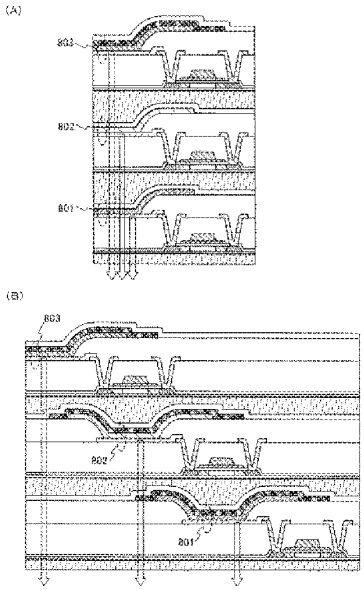
【図12】



【図13】



【図14】



(51)Int.Cl.[7]	F I		テーマコード（参考）
Ｈ０１Ｌ　31/12	Ｈ０１Ｌ　27/14	Ｃ	

(72)発明者　大野　由美子
神奈川県厚木市長谷３９８番地　株式会社半導体エネルギー研究所内

Ｆターム(参考) 4M118 BA02 BA05 CA02 CA32 CB06 CB14 FB13 FB24 FC02 FC06
FC15 HA02 HA11 HA19 HA21 HA24 HA25 HA29 HA30 HA31
HA35
5F089 AA06 AB03 AB11 AC19 CA12 EA01 FA03 FA05 GA08
5F110 AA01 AA04 AA09 BB03 BB06 BB07 BB11 CC02 CC07 DD01
DD12 DD13 DD14 DD15 DD17 DD30 GG01 GG02 GG13 GG25
GG43 GG44 GG45 HL02 HL03 HL04 HL12 HL23 HM15 NN03
NN22 NN24 NN27 NN36 NN71 NN72 PP01 PP02 PP03 PP05
PP06 PP07 PP13 PP34 QQ09 QQ11 QQ16 QQ30

Espacenet

Bibliographic data: JPH11126196 (A) — 1999-05-11

DATA TRANSFER METHOD AND COMPUTER SYSTEM SUITABLE FOR THE SAME

Inventor(s): YASUDA YOSHIKO; FUJII KEIMEI ± (YASUDA YOSHIKO, ; FUJII KEIMEI)

Applicant(s): HITACHI LTD ± (HITACHI LTD)

Classification: **- international:** ***G06F13/00; G06F15/16; G06F15/163; G06F15/17; G06F15/173; G06F15/177; G06F9/46;*** *H04L29/08;* (IPC1-7): G06F13/00; G06F15/163
- cooperative: **G06F9/544**; H04L67/10

Application number: JP19970290597 19971023

Priority number(s): JP19970290597 19971023

Also published as: JP3601955 (B2) US6338095 (B1)

Abstract of JPH11126196 (A)

PROBLEM TO BE SOLVED: To accelerate the inter-memory direct transfer through a standard MPI library. SOLUTION: A transmission side PUT/GET library generates a transfer control information 200 in response to a data transmission request command issued from a user process to the MPI library. While using a transmission data address and a transmission flag address in that information, a transmission side network interface circuit reads additional information 714 respectively corresponding to user data 713 and user data to be used by the MPI library out of a local memory and sends it to a network 105 as one message 705, and a transmission end flag 715 is written while using the transmission flag address. A reception side network interface circuit respectively writes the user data and the additional information in local memory areas 721 and 722 designated by a reception data address and a reception flag address in a message and writes a reception completion flag 723 in an area 723.

Espacenet

Description: JPH11126196 (A) — 1999-05-11

DATA TRANSFER METHOD AND COMPUTER SYSTEM SUITABLE FOR THE SAME

Description not available for JPH11126196 (A)
Description of corresponding document: US6338095 (B1)

A high quality text as facsimile in your desired language may be available amongst the following family members:

US6338095 (B1)

The EPO does not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO; in particular, the EPO does not guarantee that they are complete, up-to-date or fit for specific purposes.

BACKGROUND OF THE INVENTION

This invention relates to a data transfer method between plural element processors connected through an interconnecting network and a computer system suitable therefor.

A conventional parallel computer system often has structure that plural element processors each of which includes a local memory and an instruction processor are connected by an interconnecting network. In general, the parallel computer system of this type is called as a parallel computer system of a distributed memory type. Each element processor exchanges data stored in its local memory with other element processors by way of the interconnecting network and executes processing in parallel with each other.

In the parallel computer system of the distributed memory type, data transfer is generally achieved by using the programming model called message passing. In the message passing model, a sending (SEND) procedure and a receiving (RECEIVE) procedure are described explicitly in the parallel program (which is called as a user process hereinafter) executed in each element processor. The element processor of the sending side transfers a message which contains send data designated by a send procedure (SEND) when the procedure has been executed. The element processor of the receiving side receives the message when a receive procedure has been executed. The instruction processor in each element processor analyzes these communication procedures included in the user process under execution, and advances the processing by transferring data to the interconnecting network or receiving data therefrom. The element processor of the sending side designates the number of the destination

element processor and transfers the message. Some specific data transfer methods are proposed by which each element processor actually processes the data send request and the data receive request from such a user process. From a practicable viewpoint, it is preferable that the extra processing (send overhead) executed from issuing of the send request by the user process of the sending side till the start of transferring of the user data and the extra processing (receive overhead) executed from issuing of the receive request by the user process of the receiving side till handing over of the data to the user process are small.

Recently the following data transfer method is used in many plural parallel computer systems, in order to reduce the send overhead and the receive overhead. The send/receive circuit of the element processor of the sending side directly read user data to be sent from a region for a user process provided in the local memory, and transfers the user data and generates a message which contains the user data. The send/receive circuit in the element processor of the receiving side directly writes the user data within the message into a region provided for a user process of the receiving side within the local memory. There are some specific methods for achieving such data transfer, but, in the present specification, those data transfer methods are generically called as a direct inter-memory data transfer method.

Specific examples of the communication method for executing the method are the PUT communication method for transferring send data to another element processor and GET communication method for receiving data from another element processor. For instance, refer to "Proceedings of Information Processing Society of Japan Parallel Processing Symposium JSPP '95" and PP. 233-240 (May, 1995) or refer to a manual "HI-UX/MPP Remote DMA User's Guide -C-" published by the present assignee for a parallel computer system SR2201 developed thereby. The PUT communication method is called as a direct remote memory access method in this manual.

To be specific, in the PUT communication method, the instruction processor in the element processor of the sending side notifies the send/receive circuit of an address (send data address) of the local memory of the sending side to read the send data, an address (receive data address) of the local memory of the receiving side to write the send data, and some other addresses. The send/receive circuit of the sending side reads the user data from the local memory based on the send data address, generates a message which contains the user data, the receive data address, and other addresses, and transfers the message. The send/receive circuit of the receiving side writes the user data within the message into the local memory in the element processor of the receiving side according to the receive data address in the message. Thus, in the PUT communication method, the send/receive circuit of the sending side directly accesses the local memory of the sending side according to the send data address and the send/receive circuit of the receiving side directly accesses the local memory of the receiving side according to the receive data address. Other communication methods which differ from such a PUT communication method but execute similar processing will be called below as a PUT communication method or a remote memory directly access method. The PUT communication method will be discussed below as a representative method of the direct inter-memory data transfer method, when the latter method is to be discussed.

In direct inter-memory data transfer method, the send/receive circuit which generates or receives a message may directly read the send data from a region for a user process in the local memory of each element processor or writes the received data into a region for a user process in the local memory. So, in the element processor of the sending side, the send data does not need to be copied from a region for a user

process onto a buffer area (controlled by the Operating System), and in the element processor of the receiving side, the received data need not be copied from a buffer area (controlled by the Operating System) onto a region for a user process, either. Therefore, in this communication method, the send overhead and the receive overhead which derive from the copy can be reduced.

The communication library which executes the direct inter-memory data transfer method of such as the PUT communication or the GET communication and so on is developed by each parallel computer system maker or a research organization as a library specific to them. So, it is difficult to port the parallel programs programmed so as to use the library to other machines. On the other hand, a library (message passing library) with a standard interface specification to use the message passing model between user programs has come in use. It is possible to run such a parallel program programmed so as to use the library on different computers equipped with the library without change. The representative of that interface specification is MPI (Message Passing Interface). Universities in the United States of America and parallel computer system makers has organized the Message Passing Interface Forum (MPI Forum), and MPI is specification decided by the forum as a result of the research. It is expected that the library produced based on the specification will become a main current of the parallel program development support libraries in the future. Hereinafter, the library will be called as the MPI library.

In order to execute a data send request or a data receive request issued from a user process to the MPI library at high speed, it is effective to execute those request by using the direct inter-memory data transfer method like the PUT communication method and so on. Thus, parallel computer system makers and so on have developed MPI libraries to execute the send request or the receive request from the user process, by using the direct inter-memory data transfer library such as the PUT communication library. For instance, such a MPI library is used in the parallel computer system SR2201 developed by the assignee of the present application. The MPI library of this kind responds to the data send request or the data receive request from the user program, issues a suitable command to the PUT communication library, and requests execution of the transfer or receive operation requested by the user process. The PUT communication library orders transfer of a message to the network interface circuit which includes a message send/receive circuit. The communication method which processes the data send request or the data receive request issued from the user process to a MPI library by using a direct inter-memory data transfer library such as a PUT communication library may be called below as a communication method of using these two libraries or using the two libraries together or using combination of these two libraries. The MPI specification does not defines the specification concerning the interface between the MPI library and the direct inter-memory data transfer library such as the PUT communication library. Therefore, the MPI library to use the direct inter-memory data transfer library such as the PUT communication library together is different depending upon each computer maker. However, the interface between a user program and a MPI library is the same with any MPI library produced by any computer maker. Therefore, the user program can run on a computer by any maker as far as it is installed with a MPI library, and a data send request from the user program will be processed at a high-speed by the data transfer by direct inter-memory data transfer library such as the PUT communication library.

When the user program directly uses the PUT communication library without using the MPI library together, the user program should designate information necessary for an operation of the data send/receive circuit, such as an address of a memory where the user data to be transferred exists and the length of the user data as arguments of a call

sentence to the PUT communication library. When the user program uses the MPI library for data transfer, the user program should designate the additional information, in addition to the two arguments designated for the PUT communication library, as arguments of a call sentence to the MPI library. The additional information includes information necessary for processing data send/receive protocol defined by the MPI specification. The additional information includes plural pieces of information of plural kinds predetermined by the MPI library. Specifically, the additional information is data of fixed length and contains the identifier of the destination process and the process group identifier and so on. The additional information is used in the destination element processor of the message to identify whether the message is one requested by a receive request issued by a user process under execution. The additional information may be called below as the MPI additional information.

In the conventional parallel computer system which uses a direct inter-memory data transfer library such as a PUT communication library together with a MPI library, the user data and the MPI additional information designated by a send request issued by a user process were transferred by two different messages. That is, in the element processor of the sending side, the MPI library notifies the PUT communication library of information necessary for transfer of the user data and requests the PUT communication library to generate transfer control information which contains plural pieces of information, based on those notified information. The put communication library generates the transfer control information, writes it into the local memory, and requests the network interface circuit to execute data transfer according to the transfer control information. The network interface circuit read the transfer control information from the local memory, read the send data from the local memory according to the send data address in the transfer control information, and generates and transfers a message which contains the data. In addition, the network interface circuit writes the send completion flag into the local memory according to the send data flag address in the transfer control information. The MPI library, the PUT communication library and the network interface circuit execute the above-mentioned processing to the MPI additional information also. In the element processor of the receiving side, the network interface circuit, the PUT communication library, and the MPI library processes a receive request issued by the user process of the receiving side based on the two transferred messages.

When the user data and the MPI additional information are transferred to the same destination processor in a form of two messages like the prior art, different pieces of transfer control information are needed for transfer of each message. As a result, it is necessary to generate the transfer control information twice. Therefore, the delay time (called transfer latency) from requesting by a user program, transfer of data till start of the transfer is large in the conventional data transfer using a MPI library.

In addition, the send processing and the receive processing are executed to each of these two messages in the processor element of the sending side and in the processor element of the receiving side, respectively. So, the access frequency to the local memory in these processor elements increases in proportion to the number of the messages. For instance, the element processor of the sending side needs to execute the operations such as reading of the transfer control information, reading of data to be transferred, and writing of a send completion flag for each message.

SUMMARY OF THE INVENTION

The object of the present invention is to provide a data transfer method which can decrease the above-mentioned problems and transfer data at a higher speed, and a computer system suitable therefor.

A more concrete object of the present invention is to provide a data transfer method which can transfer both user data and additional information which a user process of the sending side designate, and a computer system suitable therefor.

To achieve those objects, in a computer system according to the present invention, a network interface circuit in each element processor is provided with a memory read circuit connected to the memory in the element processor and connected to the memory read circuit which generates a message which should be transferred to the interconnecting network. The memory read circuit reads the first data and the second data both to be transferred from the memory, based on the first and second pieces of address information designated by the processor in the element processor. The message assembly circuit generates a message which comprises a header and send data which includes both the first data and the second data. Therefore, each element processor can transfer plural data at a higher speed than the conventional apparatus.

A data transfer method according to the present invention is achieved by using a network circuit which has the function to generate a message which contains plural data as send data, just like the network circuit mentioned above.

That is, first and second pieces of address information and other information are generated by cooperate between a message passing library and a direct inter-memory data transfer library provided in an element processor of a sending side, in response to a data send request issued by a user process of a sending side running in said element processor of a sending side;

wherein the data send request designates user data and additional information to be transferred, and data length information for the user data;

wherein the first and second pieces of address information designate two storage regions within the memory of the element processor of a sending side where said user data and said additional information are stored;

wherein the additional information comprises a plurality of pieces of information each of which is of one of a plurality of kinds predetermined by the message passing library of a sending side and relates to transferring of the user data to a user process of a receiving side running in the element processor of a receiving side;

wherein the other information comprises a plurality of pieces of information each of which is of one of a plurality of kinds predetermined by the direct inter-memory data transfer library and relates to transferring of the user data and the additional information to the element processor of a receiving side.

The network interface circuit reads the user data and the additional information from the memory of the element processor of a sending side, based on the first and second pieces of address information.

The network interface circuit further assembles a message send data of which includes the user data and the additional information, and a header of which includes the other information and the data length information.

The network interface circuit transfers the message to the element processor of a receiving side by way of the interconnecting network.

In a more specific embodiment of the present invention, the user process not only can use the above mentioned data transfer which uses both the message passing library and the message passing library, but also can issue a data send request directly to the direct inter-memory data transfer library. In the latter case, the direct inter-memory data transfer library and the network interface circuit execute processing basically the same to the above-mentioned processing except that the additional information does not exist.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a block diagram of a parallel computer system according to the present invention.

FIG. 2 shows an example of transfer control information used in the apparatus shown in FIG. 1.

FIG. 3 shows a block diagram of an element processor used in the apparatus shown in FIG. 1.

FIG. 4 shows a block diagram of the storage controller and the network interface circuit used in the element processor shown in FIG. 3.

FIG. 5 shows a block diagram of the message generating circuit used in the network interface circuit shown in FIG. 4.

FIG. 6 shows a block diagram of the message disassembly circuit used in the network interface circuit shown in FIG. 4.

FIG. 7 shows the outline of data transfer processing executed in the apparatus shown in FIG. 1.

FIG. 8 shows the outline of the prior art data transfer processing.

DESCRIPTION OF THE EMBODIMENTS

A computer system according to the present invention will be explained in more detail with reference to one or several embodiments or their modifications shown in the drawings.

FIG. 1 shows schematic structure of a parallel computer system according to the present invention. In the drawing, the reference numerals 101-104 show element processors, which compose the parallel computer system, and 105 shows an interconnecting network. The element processor 101-104 are connected to the interconnecting network 105 and exchange data with other element processors through the interconnecting network 105. There are various kinds of structure (topology) for the interconnecting network such as crossbar interconnection, lattice interconnection, ring interconnection, and multistage interconnection and so on. The present invention is not limited to an interconnecting network of a specific S but can be applied to an interconnecting network of any of these kinds of S.

FIG. 3 shows schematic structure of the element processor 101 , , , or 104. In the figure, 301 shows an instruction processor, 302 shows a cache memory, 303 shows a storage controller, 304 shows a local memory, 305 shows a network interface circuit, 306 shows an I/O interface circuit and 106 shows an I/O device. The I/O device 106 is

not shown in FIG. 1 for simplification. The I/O device 106 does not need to be installed in all of the element processors 101 to 104. The parallel computer system is one of a distributed local memory type where each element processor has a local memory 304 of its own.

Each element processor is composed so as to execute communication with other element processors by message passing. That is, each element processor includes a message passing library which has a standard message passing interface, MPI for instance, (which may be called as a MPI library hereinafter), a library which communicates with the MPI library and executes PUT/GET communication (which may be called as a PUT/GET library hereafter) and the network interface circuit 305 invoked by the PUT/GET library. The PUT/GET library is installed as a library for executing direct inter-memory data transfer which directly transfers data with the local memory in the element processor where the library is installed. The present invention is not limited to this specific message passing library but other message passing libraries can be applied such as PVM (Parallel Virtually Machine) proposed by Oak Ridge National Laboratory, Inc. and PARMACS proposed by PALLAS company. Moreover, the PUT communication library that executes only the PUT communication can be used in place of the PUT/GET communication library. It is needless to say that other library can be used as a library for executing direct inter-memory data transfer, in place of the PUT/GET communication library or the PUT communication library.

In the present embodiment, when a user process in each element processor issues a data send request to the MPI library, the MPI library and the PUT communication library decide by cooperation in response to the data send request, plural pieces of address information and other information. The pieces of address information are used to read the user data and the MPI additional information, and the other information is information to be transferred to the element processor of the receiving side together with the user data and the MPI additional information. The network interface circuit 305 read the user data and the MPI additional information based on these pieces of address information from the local memory and transfers these data and the other information in one message. In addition, when a message which contains user data and MPI additional information is received from another element processor, the network interface circuit 305 writes these two data in the local memory 304 separately. The additional information is used for communication through the MPI library. It comprises plural pieces of information which relate to transferring of data and are of the kinds designated by the MPI library. Each of the plural pieces information is designated by the user process. Specifically, the additional information contains the identifier of the user process of the receiving side and the process group identifier and so on, as illustrated already.

More specifically, when a data send request command is issued from a user process of the user being executed in an element processor to the MPI library, the MPI library requests the PUT/GET communication library to store the MPI additional information. The PUT/GET communication library stores the information designated by the command into a communication region within the local memory predetermined as one directly accessible by the network interface circuit 305. The MPI library requests the PUT/GET communication library to generate transfer control information. The PUT/GET communication library generates transfer control information for transfer of both the user data and the additional information in one message, and writes it into the local memory 304. The transfer control information contains two pieces of address information about two storage locations of the destination local memory at which the user data and the additional information are to be written, and other address information. Thereafter, the PUT/GET library request the network interface circuit 305

to transfer the user data and the additional information, based on the transfer control information. Then, the MPI library requests the PUT/GET communication library to start the transfer based upon the transfer control information. The PUT/GET communication library requests the network interface circuit 305 to execute the requested transfer. If the user data is located at a location within the local memory 304 which the network interface circuit 305 cannot access directly, the PUT/GET communication library copies the user data into a location within the communication area within the local memory.

The network interface circuit 305 responds to this command from the PUT/GET communication library, and assembles, according to the transfer control information, one message which contains the user data and the MPI additional information, and transfers the message to the element processor of the receiving side through the interconnecting network 105. The header of the message contains two pieces of address information and other information. The two pieces of address information designate the two locations in the destination local memory where the user data and the additional information in the message are to be written, respectively. When the network interface circuit 305 in the element processor of the receiving side receives the message, it writes the user data and the additional information into two regions of the local memory, based upon the two pieces of address information included in the header of the message, and delivers them over to the user process and the MPI library of the receiving side, respectively.

FIG. 2 shows an example of the transfer control information. The transfer control information 200 contains plural pieces of information used at the PUT or GET operation, and the kinds of these pieces of information are predetermined by the PUT/GET communication library. To be more specifically, in the present embodiment, the transfer control information 200 includes the number of the destination processor 201, the mode bit202, the send data address 203, the send flag address 204, the transfer data length 205, the receive data address 206, the receive flag address 207, other information 208 necessary for communication processing, and so on. The send data address 203 is the start address of the local memory region where the user data to be transferred at the PUT operation is stored. The transfer data length 205 is the length of data (user data) transferred at PUT or GET operation. The receive data address 206 is the start address of the local memory region into which the user data in the received message is written, within the element processor of the receiving side.

The mode bit 202 is a bit which shows whether use of the send flag address field 204 is expanded according to the embodiment. That is, the mode bit 202 is set to one in the transfer control information generated when the user program uses the PUT/GET communication library via the MPI library. On the other hand, the mode bit 202 is reset to 0 in the transfer control information generated when the user program directly uses the PUT/GET communication library. When the mode bit 202 is not set, the send flag address 204 designates the location of the local memory of the element processor of the sending side where the send completion flag is to be written at completion of the transfer of the user data at the message sending operation in the element processor of the sending side of the message. Similarly, the receive flag address 207 is the address of the location of the local memory of the element processor of the receiving side into which the receive completion flag corresponding to the user data is written at completion of the receipt of the user data transferred at the message receiving operation in the element processor of the receiving side of the message. However, when the mode bit 202 is set to one, the send flag address field 204 is used to designate the start address of the region of the local memory of the element processor of the sending side from which the MPI additional information should be read at the message sending operation in the element processor of the sending side of the

message. Similarly, the receive flag address field 207 is used to designate the start address of the region of the local memory of the element processor of the receiving side at which the MPI additional information included in the received message should be written at the message receiving operation in the element processor of the receiving side of the message.

As a result, the local memory address where the send completion flag should be written is not designated explicitly by the transfer control information 200 at the message sending operation in the element processor of the sending side of the message. In the present embodiment, the network interface circuit 305 of the element processor of the sending side adds the predetermined size of the MPI additional information to the address of in the send flag address field 204 and writes the send completion flag common to the user data and the MPI additional information into the location of the local memory, having the address obtained as a result of the addition. Similarly, the network interface circuit 305 of the element processor of the receiving side adds the predetermined size of the MPI additional information to the address of in the receive flag address field 207 and writes the receive completion flag common to the user data and the MPI additional information included in the received message, into the location of the local memory, having the address obtained as a result of the addition, at the message receiving operation in the element processor of the receiving side of the message. As a result, two kinds of data, that is, the user data and the MPI additional information can be transferred or received in one message by using one transfer control information 200, and yet the send completion flag or the receive completion flag required in execution of the PUT operation can be written into the local memory of the element processor of the sending side and the local memory of the element processor of the receiving side, respectively.

The instruction processor 301 is a unit which executes program processing. The cache memory 302 is a memory which accompanies the instruction processor 301 and has a relatively high speed and a relatively small capacity. The network interface circuit 305 is a unit which is connected to the interconnecting network 105 and executes data transfer processing according to the commands from the instruction processor 301. The data transfer processing is executed independently of the program processing in the instruction processor 301, which is the characteristic of the parallel computer system of the distributed memory type. The storage controller 303 accesses suitable storage medium according to a data access request issued from the instruction processor 301, the network interface circuit 305, or the I/O interface circuit 306. The local memory 304 stores data and so on under control by the storage controller 303. The instruction processor 301 and the network interface circuit 305 operate independently from each other, so the storage controller 303 processes an access request to the local memory 304 required by data transfer from the network interface circuit 305, in parallel to processing of the access request to the local memory 304 from the instruction processor 301. The I/O interface circuit 306 accesses the I/O device 106 according to an access request from the storage controller 303. In an element processor to which the I/O device 106 is not added, the I/O interface circuit 306 may not exist.

As shown in FIG. 4, the storage controller 303 is composed of the instruction processor interface circuit 401, the address analyzing circuit 402, the memory access interface circuit 403 and the data transfer interface circuit 404. The instruction processor interface circuit 401 receives a transaction such as an access request to the local memory 304 or a command which should be sent to the network interface circuit 305, from the instruction processor 301 by line 401S. The command is usually an access request to one of the control registers (not shown in the drawings) in the network

interface circuit 305. The instruction processor interface circuit 401 transfers to the instruction processor 301 through line 402S, an response to the transaction generated by the memory access interface circuit 403 or the network interface circuit 305 or an interrupt transfer request which requests transfer of the interrupt factor detected by the storage controller 303 or the network interface circuit 305. The response to the transaction of the access request to the local memory 304 is data read therefrom.

The address analyzing circuit 402 receives by line 403S, the local memory access request or a command to be sent to the network interface circuit 305, both issued by the instruction processor 301, and analyzes the destination address which the access request or the command designates. The address analyzing circuit 402 transfers the local memory access request to the memory access interface circuit 403 through line 404S and the command which should be sent to the network interface circuit 305 to the data transfer interface circuit 404 through line 406S.

The memory access interface circuit 403 receives the local memory access request from the address analyzing circuit 402 through line 404S and transfers it to the local memory 304 through line 407S. When the local memory access request is a read request to the local memory, the data which the read request designates is transferred from the local memory 304 to line 408S. In addition, the read data is transferred from the memory access interface circuit 403 to the instruction processor interface circuit 401 through line 409S, and is further transferred to the instruction processor 301 through line 402S. The memory access interface circuit 403 also processes a local memory access related to inter-processor data transfer. A request for the local memory access related to the inter-processor data transfer is transferred from the network interface circuit 305 through the data transfer interface circuit 404 and line 410S. The memory access interface circuit 403 issues the local memory access request transferred from the network interface circuit 305 to the local memory 304 like a local memory access request transferred from the instruction processor 301. When the access request is a read access, the memory access interface circuit 403 transfers data read out from the local memory 304 back to the network interface circuit 305 via line 411S and the data transfer interface circuit 404.

The data transfer interface circuit 404 receives a command which should be transferred to the network interface circuit 305 as transferred from the address analyzing circuit 402 through line 406S, and transfers it to the network interface circuit 305 through line 412S. A response to the command or a local memory access request related to the inter-processor data transfer is transferred to the data transfer interface circuit 404 from the network interface circuit 305 through line 413S. The data transfer interface circuit 404 transfers the response to the above-mentioned command to the instruction processor interface circuit 401 through line 414S, and to the instruction processor 301 through line 402S. The data transfer interface circuit 404 transfers the local memory access request to the memory access interface circuit 403 through line 410S. When the access request is a read request, the data read from the local memory 304 is transferred from the memory access interface circuit 403 to the data transfer interface circuit 404 through line 411S and is further transferred from the data transfer interface circuit 404 to the network interface circuit 305 through line 412S. The data transfer interface circuit 404 occasionally receives from the network interface circuit 305, an interrupt transfer request generated therein through line 413S. The interrupt transfer request is transferred to the instruction processor interface circuit 401 through line 414S and is further transferred to the instruction processor 301.

The network interface circuit 305 is composed of the command receive circuit 405, the command processing circuit 406, the message generating circuit 407, the message

send circuit 408, the message receive circuit 409, the message disassembly circuit 410, and the command send circuit 411. The command receive circuit 405 receives a command which requests an access to a control register (not shown in the drawings) within the network interface circuit 305 from the storage controller 303 through line 412S, the data requested by the network interface circuit 305 and read from the local memory 304, and so on. The read data is transferred to the command processing circuit 406 through line 415S. The read data is, for example, data which should be transferred to another element processor, data to be set in the control register (not shown in the drawings) within the network interface circuit 305 as operation control data therefor, or the transfer control information used to generate a message at transfer of data. The transfer control information contains the number of the destination processor, the send data address, the send flag address, the transfer data length, the receive data address, and the receive flag address and so on.

The command processing circuit 406 executes the command which requests access to a control register in the network interface circuit 305 (not shown in the drawing) and is transferred from the command receive circuit 405 through line 415S. When the access is a read access, the command processing circuit 406 executes the read access and transmits data as read from the local memory 304 to the command send circuit 411 through line 417S. When the access is a write access, the command processing circuit 406 executes writing requested by the command. When the access to the control register is a write request to the control register (not shown in the drawing) for invocation of the message transfer, the network interface circuit 305 starts message send processing. In the message send processing, the command processing circuit 406 generates a local memory access request for readout of the transfer control information 200 stored in the local memory 304, to get the local memory address of a location where information required to produce the header of the message and transfer data is stored. The access request is transferred to the command send circuit 411 through line 417S, is further supplied to the local memory 304 through line 413S, the data transfer interface circuit 404, line 410S, the memory access interface circuit 403, and line 407S, and the transfer control information is read from the local memory 304. The read transfer control information is transferred to the message generating circuit 407 via line 408S, the memory access interface circuit 403, line 411S, the data transfer interface circuit 404, line 412S, the command receive circuit 405 and the command processing circuit 406.

The message generating circuit 407 is one of the characteristic circuits of the present embodiment. When the mode bit 202 in the transfer control information 200 is 1, the message generating circuit 407 reads the user data and the MPI additional information from the local memory 304, based upon the send data address 203 and the send flag address field 204 in the transfer control information 200, respectively. The message generating circuit 407 assembles a message, and delivers it to the message send circuit 408. The header of the message contains information within the transfer control information 200 other than the above-mentioned two addresses. The send data of the message comprises combination of the user data and the additional information as read. When the mode bit 202 is 0, the message generation circuit 407 assembles a message which contains the same header as above and the send data which includes the user data as read.

FIG. 5 shows internal structure of the message generating circuit 407. The message generating circuit 407 has the registers 501, 502, and 503 which respectively maintain the send data address 203, the address in the send flag address field 204, the transfer data length 205 in the transfer control information 200 read from the local memory 304. The message generating circuit 407 further has a register 504 which holds the amount

of data which has already been read from the local memory 304 for a message under generation, a register 505 which holds the size of the MPI additional information and a register 506 which maintains the mode bit 202 in the transfer control information 200. The latter two registers are ones of characteristic circuits of the present embodiment. Access to the local memory 304 is executed by repeating accesses to different data each of a predetermined unit of amount. The amount of already read data as held in the register 504 is counted up whenever the data (the user data and the additional information) to be sent is read from the local memory 304 by predetermined amount, and is reset after all the data to be sent has been transferred to the message send circuit 408. Therefore, the amount of the already read data held in the register 504 can be regarded as total amount of data already sent. Similarly, the send data address held in the register 501 is counted up whenever the data to be sent is read out from the local memory 304 by predetermined amount. The send flag address held in the register 502 is counted up whenever the MPI additional information to be sent is read from the local memory 304 by predetermined amount. The size of the MPI additional information is set into the register 505, when the MPI library and the PUT/GET communication library are initialized by the initialization request issued from a user process. The mode bit 202 in the transfer control information 200 is set into the register 506 when the information is read from the local memory 304, and the mode bit in the register 506 is reset, when all the data is transferred to the message send circuit 408. The message generating circuit 407 includes a memory read request issue circuit 511, a memory write request issue circuit 512, and a message assembly circuit 513. The message generating circuit 407 further includes an AND gate 507, adders 508, 509, and a comparator 510, which are ones of characteristic circuits of the present embodiment. The memory read request issue circuit 511 and circuits connected thereto composes a memory read circuit. The memory write request issue circuit 512 and circuits connected thereto composes a memory write circuit.

The And gate 507 outputs the size of the MPI additional information held in the register 505, in case the mode bit in the register 506 is one, and outputs the value 0 of the mode bit, in case the mode bit is 0. The adder 509 adds the output of the AND gate 507 to the value of the send flag address field 204 held in the register 502. The result of the addition it is stored in the control register (not shown in the drawing) within the memory write request issue circuit (512 (FIG. 5)) at start of the send operation in the element processor of the sending side of the message, and is used later on as the send completion flag address. At send operation in the element processor of the sending side of the message, the head address of the MPI additional information stored in the local memory 304 is included in the send flag address field 204 held the register 502, in case the mode bit is set to 1. The resultant address obtained by the addition indicates the address of the storage location next to the storage region where the MPI additional information is stored, and is used as the write address of the send completion flag at the send operation in the element processor of the sending side of the message. The adder 508adds the output of the AND gate 507 to the transfer data length held in the register 503. The result of this addition is a sum of the length of the user data and the size of the MPI additional information, in case one is set into the mode bit, and indicates the total amount of data which is to be read from the local memory 304 and to be included in the message to be transferred at the send operation in the element processor of the sending side of the message.

The comparator 510 compares the amount of the already read data as held in the register 504 with each of the transfer data length held in the register 503 and the total amount of the send data outputted from the adder 508. In case the amount of already read data held in the register 504 does not exceed the transfer data length held in the register 503, the comparator 510 outputs a comparison result signal indicative of that to

the memory read request issue circuit 511. In case the amount of already read data held in the register 504 exceeds the transfer data length held in the register 503 but not the total amount of data to be read from the local memory 304 as given from the adder 508, the comparator 510 outputs a comparison result signal indicative of that to the memory read request issue circuit 511. When the amount of already read data held in the register 504 has reached the total amount of data to be read, the comparator 510 outputs a comparison result signal indicative of that to the local memory write request issue circuit 512.

The memory read request issue circuit 511 generates a local memory read request, from the comparison result signal from the comparator 510, the send data address in the register 501, and the send flag address in the register 502 and transfers it to the command send circuit 411. That is, the memory read request issue circuit 511 generates a local memory read request to read an unread portion of the user data, based on the send data address in the register 501, when the comparison result signal shows that the already read data does not exceed the transfer data length. The memory read request issue circuit 511 generates a local memory read request to read an unread portion of the MPI additional information, based on the send flag address in the register 502, when the total amount of already read data exceeds the transfer data length but not the total amount of the data to be read. The memory read request issue circuit 511 repeats the above-mentioned operation until all of the user data and the MPI additional information are read.

The memory write request issue circuit 512 transfers the local memory write request to the command send circuit 411, based upon the comparison result signal provided from the comparator 510 and the send completion flag address which was outputted from the adder 509 and is stored in the control register (not shown in the drawing), as mentioned before. That is, the memory write request issue circuit 512 generates a write request which requests writing of the send completion flag at a storage location of the send completion flag address given from the control register (not shown), when the comparison result signal shows that the amount of already read data has reached the total amount of data to read.

The message assembly circuit 513 assembles a message having structure decided according to the value of the mode bit and sends a send request of the message to the message send circuit 408. When the mode bit is one, the selector 514 assembles a message which contains a header which contains part of the transfer control information 200 read from the local memory 304 and sent via the command processing circuit 406, and a data portion which contains the user data and the MPI additional information, and sets the assembled message into the register 515. When the mode bit is 0, the selector 514 assembles a message with the header and a data portion including the user data, and sets the message into the register 516. Information other than the send data address 203 and the send flag address 204 are included in the header, among the transfer control information 200 read from the local memory 304.

In FIG. 4, the message send circuit 408 receives the message send request from the message generating circuit 407 and sends the message to the interconnecting network 105 through line 420S. The message is transferred to a destination element processor through the interconnecting network 105 according to the number of the destination processor in the header. The send processing in the message generating circuit 407 is executed to plural message send requests transferred to the network interface circuit 305 sequentially according to the order of arrival of those send requests.

Next, the message receive circuit 409 will be explained. The message receive circuit

409 sequentially receives messages from the interconnecting network 105 by line 421S and transfers them to the message disassembly circuit 410 through line 422S. The message disassembly circuit 410 disassembles the messages into the header portion and the data portion and further separates the user data and the additional information included in the data portion, based upon the transfer data length in the header portion and the predetermined size of the MPI additional information. The transfer length is included in the transfer control information as included in the header portion of the message. The message disassembly circuit 410 transfers write requests to the local memory 304 to write the user data, to write the additional information and to write a receive completion flag to the command send circuit 411 through the request line 423S.

FIG. 6 shows internal structure of the message disassembly circuit 410. In the message disassembly circuit 410, the registers 601, 602, 603, and 606 are provided which respectively hold the receive data address, the receive flag address, the transfer data length, and the mode bit in the message header. The register 606 is one of registers characteristic of the present embodiment. The mode bit in the header is set into the register 606 when the header has been received, and the mode bit is reset when all the data in the message has been received. In addition, the registers 604 and 605 are installed in the present embodiment. The register 604 holds the total amount of data already written into the local memory within the data in received the message, and the register 605 holds the size of the MPI additional information. The register 605 is one of registers characteristic to the present embodiment. The predetermined size of the MPI additional information is set into the register 605 at MPI initialization or at job invocation, and this size information is reset when all the data in the message has been received. The amount of the data already written as held in the register 604 is counted up whenever predetermined amount is written into the local memory 304 within the data (the user data and the additional information) in the received message and is reset after all the received data has been written into the local memory 304. Therefore, the total amount of already written data can be regarded as the total amount of already received data. The receive data address in the register 602 is counted up, whenever the user data in the message is written into the local memory 304 by predetermined amount, and similarly the receive flag address in the register 601 is counted up, whenever the MPI additional information in the message is written into the local memory 304 by predetermined amount.

The AND gate 607, the adders 608, 609 and the comparator 610, which are characteristic to the present embodiment, are further installed in the message disassembly circuit 410, in addition to the memory write request issue circuit 611. The And gate 607 outputs the size of the MPI additional information in the register 605 when the mode bit in the register 606 is one, and outputs 0 when the mode bit is 0. The operations of the adders 608 and 609 is similar to the adders 508 and 509 (FIG. 5). The output of the adder 609 represents a receive completion flag address and is stored in the control register (not shown) within the memory write request issue circuit 611 at the start of the receive operation at the element processor of the receiving side. The comparator 610 compares the total amount of already received data as held in the register 604 with each of the transfer data length and the output result data of the adder 608, and outputs the comparison result signal to the memory write request issue circuit 611. That is, the comparator judges whether the total amount of already received data is smaller than the transfer data length, whether the total amount of already received data is larger than the transfer data length and at the same time is smaller than a sum of the transfer data length and the size of the MPI additional information as provided by the AND gate 607, or whether the total amount of already received data is equal to that sum.

The memory write request issue circuit 611 and circuits connected thereto composes the second memory write circuit. Specifically, the memory write request issue circuit 611 generates a local memory write request, based upon the comparison result signal of the comparator 610, the receive data address in the register 601, the receive completion flag address as provided from the adder 609 and already stored in the control register (not shown) within the memory write request issue circuit 611, and the value of the receive flag address field in the register 602, and transfers it to the command send circuit 411 through line 424S. When the mode bit is one, the adder 609 outputs the address obtained by adding the size of the MPI additional information to the value of the receive flag address field in the register 602 as the local memory address of a storage location where the receive completion flag should be written.

When the output of the comparator 610 shows that the total amount of the received data is smaller than the transfer data length, the memory write request issue circuit 611 generates a local memory write request, based on the receive data address in the register 601, to write the received data at a storage location next to the storage locations where already received data has been written.

When the output of the comparator 610 shows that the total amount of the received data is larger than the transfer data length, but is smaller than the sum of the transfer data length and the size of the MPI additional information as provided by the AND gate 607, newly received data belongs to the MPI additional information. Therefore, the memory write request issue circuit 611 generates a local memory write request, based on the address of the receive flag address in the register 602, to write the received data at a storage location next to the storage locations where already received MPI additional information has been written.

When the output of the comparator 610 shows that the total amount of the received data coincides with the sum of the transfer data length and the size of the MPI additional information as provided by the AND gate 607, it means that all data has been received. Therefore, the memory write request issue circuit 611 generates a local memory write request which requests writing of the receive completion flag at a storage location next to the storage locations of the MPI additional information, based on the receive completion flag address as stored in the control register (not shown) within the memory write request issue circuit 611.

The command send circuit 411 writes the received data or the receive completion flag into the local memory 304 according to these write requests. The command send circuit 411 transmits the following data or signals transferred thereto to the data transfer interface circuit 404 in the storage controller 303 through line 413S:

data read from a control register (not shown in the drawings) within the network interface circuit 305, as transferred through line 417S;

a read request to the local memory 304 for the send data used in the message send processing, as transferred through line 424S;

a write request to the local memory 304 for the send completion flag, as transferred through line 424S according to completion of the message send processing;

a write request to the local memory 304 for the receive data or the receive completion flag, as transferred through line 423S according to the message receive processing; and

an interrupt request generated in the network interface circuit 305.

Moreover, the command send circuit 411 transmits the read request to the local memory 304 for information related to operation control of the network interface circuit 305 and the read request to the local memory 304 for the transfer control information 200 used in the message send processing, to the data transfer interface circuit 404 through line 413S.

Next, the flow of the message transfer in the present embodiment will be explained. First, the PUT processing will be explained to write the user data and the MPI additional information stored in the local memory of an element processor of the sending side, into the local memory of an element processor of a receiving side, by referring to FIGS. 4, 5, and 7. In the communication method of the present embodiment, the user process of the sending side and the user process of the receiving side issue a command which calls the initialization routine within the MPI library, for instance, MPI_init, they use the MPI library. When the MPI library receives the call to the initialization routine is, it calls some communication preparation procedures in the PUT/GET communication library. These communication preparation procedures are predetermined by the PUT/GET communication library used, but only the processing which relates to the explanation given below or the processing newly executed in the present embodiment will be explained.

Both the PUT/GET communication library of the sending side and the PUT/GET communication library of the receiving side execute the following processing in these communication preparation procedures in the present embodiment. The local memory 701 is mapped into the user space, and the communication regions 703 or 704 (FIG. 7) is secured. The send data region and the send completion flag region corresponding to it are secured in each communication region. In FIG. 7, the reference numeral 713 shows an example of the send data region. In FIG. 7, the send data region in the communication region 704 of the receiving side is not shown. The present embodiment differs from the prior art in that a continuous region which has the size enough to store both the MPI additional information 714 and the send completion flag 715 is secured as the send completion flag region. Similarly, the receive data region and the receive completion flag region are secured in each communication region. The reference numeral 721 in FIG. 7 shows an example of the receive data region in the present embodiment. In FIG. 7, the receive data region in the communication region 703 of the sending side is not shown. The present embodiment differs from the prior art in that a continuous region which has the size enough to store both the MPI additional information 722 and the receive completion flag 723 is secured as the receive completion flag region. The length of the MPI additional information 714 or 722 is about 64 bytes, for instance. The region 715 or 723 where the send completion flag or the receive completion flag is written is eight bytes, for instance. The size of the MPI additional information is set into the registers 505 and 605 during these communication preparation procedures.

There are both a case in which the PUT/GET communication library presumes that the user data is resident in the local memory 701 and a case in which the PUT/GET communication library does not do so. In the former case, the send data region 713 is decided so that the send data region 713 coincides with this resident region of the user data. On the other hand, in the latter case, the send data region 713 is decided independently of the region in the local memory allocated to user data which the user process of the sending side uses. The present invention can be applied to any of the two kinds of PUT/GET communication library. In the latter case, however, it is necessary to copy data in the region of the local memory allocated to user data which

the user process uses onto the send data region 713, as will be described later on. In the former case, this copy operation is not necessary, which results in speed up of the data transfer operation.

When the processing of the user process of the sending side advances after the communication preparation procedures are executed as mentioned above, the user process issues a data send request command, for instance, MPI_send to the MPI library of the sending side, after writing data to be transferred into the user data region 713 in the communication region 703 of the sending side. The name of this command is predetermined by the message passing library used and the arguments to be designated by the command comprises plural pieces of information of plural kinds similarly predetermined by the library. The arguments of the command for the MPI library used in the present embodiment comprise a start address of the user data to be transferred, length of the user data, and additional information. The start address is a virtual address of the user data in the virtual memory space allocated to the user process. The additional information contains an identifier of the user process of the receiving side and the identifier of the process group, and so on. The first two arguments are the same as the arguments designated by a data send request command which the user process issue directly to the PUT request/GET communication library so as to request data transfer without using the MPI library. The additional information are arguments which the data transfer request deignates newly to the MPI library.

The MPI library of the sending side responds to the data send request command MPI_send and issues a request the PUT/GET communication library to send data designated by the command. The MPI library of the sending side issues the request in the form of one or plural commands predetermined by the MPI library and the PUT/GET communication library. Only one main command used in the present embodiment among those commands will be explained below.

The MPI library of the sending side issues a command to request acquisition of the transmission right. The put/GET communication library of the sending side responds to this command, communicates with the user process of the receiving side and obtains the transmission right of data to the process. The MPI library of the sending side decides the user process of the receiving side and the number of the element processor which executes it, based upon the process identification number and the process group identification number of the receiving side as included in the additional information designated by the data send request command MPI_send.

The MPI library of the sending side communicates with the user process of the receiving side (substantially, the MPI library of the receiving side) and receives the receive data address and the receive flag address which respectively indicates the head locations of the receive data region and the receive flag region for the user process of the receiving side both controlled by the MPI library of the receiving side. These addresses are real addresses. When the user process of the sending side issues the same data send request command MPI_send later on, the command to request acquisition of the transmission right needs not be executed again.

If the PUT/GET communication library does not presume, as already described, that the user data to be transferred is resident in the local memory 701, the MPI library of the sending side issues a command which requests copying of the user data to be transferred onto the user data region 713. The put/GET communication library of the sending side copies data in a region of the local memory as designated by the command to the user data region 713. The arguments of the command designate the

address and the data length of the data to be copied. There may be a case where it is desirable for the MPI library to execute the copy operation directly, but the copy operation is executed in the present embodiment by the PUT/GET communication library in response to a request by the MPI library, as described above. If the PUT/GET communication library presumes that the user data to be transferred is resident in the local memory 701, the copy operation is unnecessary.

Next, the MPI library of the sending side issues a command which requests writing of the additional information into the local memory. The PUT/GET communication library of the sending side responds to this command and writes the MPI additional information designated by the argument of this command into the head region 714 of the send flag region already decided in correspondence to the send data region 713. There may be a case where it is desirable for the MPI library to execute the write operation directly, but the write operation is executed in the present embodiment by the PUT/GET communication library in response to a request by the MPI library, as described above.

The sending side issues a command which requests generation of the transfer control information. The command designates the destination processor number (which is determined by the MPI library based upon the message destination information included in the additional information designated by the MPI_send command), the address of the user region 713 controlled by the MPI library, the address of the region 714 where the additional information is stored, the length of the user data as designated by the MPI command, the addresses of the receive data region 721 and the receive flag region 722 for the process of the receiving side and so on. The PUT/GET communication library of the sending side responds to this command and writes the transfer control information into an appropriate region of the local memory 701. Information included in the transfer control information 200 are as follows. The number of the destination processor 201 is a number of the processor in which the user process of the receiving side is executed. This number is decided in the processing of acquisition of the transmission right, as already described. The mode bit 202 is a bit which shows whether use of the send flag address field 204 is expanded according to the present embodiment. The put/GET communication library of the sending side sets the mode bit to one, when data transfer is requested by the MPI library of the sending side. The value 1 of the mode bit 202 requests transfer of the user data and the MPI additional information in one message. In case, the user process of the sending side issues the data send request command directly to the PUT/GET communication library and not to the MPI library in the present embodiment, the PUT/GET communication library responses to the command and generates the transfer control information which has the same structure as the transfer control information 200 shown in FIG. 7 and in which the value of the mode bit 202 is 0. Various addresses included in the transfer control information are real addresses.

The send data address 203 is the start address of the send data region 713 which stores the user data or its copy transfer of which has been requested by the user process of the sending side. The transfer data length 205 is the data length of the user data which the data send request command MPI_send designates, and the MPI library which has received the command notifies the PUT/GET communication library of the length.

In the conventional apparatus, the local memory address to write the send completion flag is stored in the send flag address field 204, the start address of the MPI additional information 714 in the local memory 701 is stored in the send flag address field 204 when the mode bit 202 is one in the present embodiment. It means that the address of

the region 715 in the local memory which stores the send completion flag will not be designated explicitly by the send flag address field 204, when the mode bit 202 is one. In the present embodiment, the network interface circuit 305 stores the send completion flag common to the user data and the MPI additional information, into the region 715 of the next address to the last address of the MPI additional information 714 in the local memory 701. Therefore, the PUT/GET communication library of the sending side notifies the user process of the sending side of send completion when the send completion flag is written, as a response to the data send request command from the user process (substantially the MPI library) of the sending side. When the mode bit 202 is 0, the PUT/GET communication library sets into the send flag address field 204, the address where the send completion flag is to be written as decided by the library.

The receive data address 206 is the address of the local memory region 721 (FIG. 7) to store the received data in the element processor of the receiving side. The receive flag address field 207 is the start address of the region which stores the additional information received with the receive data and the receive completion flag indicative of completion of a receive operation. In the present embodiment, the receive completion flag is written into the region 723 which succeeds to the region 722 where the additional information is stored. Therefore, the start address in the region 722 is written into the receive flag address field 207. Both the receive data address 206 and the receive flag address 207 are notified by the destination PUT/GET communication library in the MPI initialization routine. These addresses are obtained by the MPI library of the sending side after communication with the user process of the receiving side and are included in the arguments of the command which is issued by the MPI library and requests generation of the transfer control information. These addresses are used at the receive operation in the processor element of the receiving side.

The other control information 208 contains other information necessary for communication processing, such as one indicating a type of PUT message, GET message or GET request message or one indicating a type of communication like one to one communication or multicast communication and so on.

In the transfer control information 200, it is possible to install the field which designates separately the local memory address where the MPI additional information should be read and the field which designates the local memory address where received the MPI additional information should be written. However, the switching of two addresses which the send flag address 204 and the receive flag address 207 designate by using the mode bit 202 as in the present embodiment has an advantage that the structure and the size of the transfer control information 200 can be the same as the case where the user process uses only the PUT/GET communication library to transfer or to receive data.

When the transfer control information 200 has been generated, the MPI library of the sending side issues to the network interface circuit 305, a command which requests transferring of data according to the generated transfer control information. The PUT/GET communication library responds to this command and issues a write command which request writing of the start addresses of the transfer control information 200 into a message transfer activation register (not shown in the drawings) as provided in the network interface circuit 305. The command processing circuit 406 executes the write command and writes the start address of the transfer control information 200 into the message transfer activation register (not shown in the drawings) within the network interface circuit 305. The network interface circuit 305 starts message send processing by this writing.

As shown in FIG. 4, the command processing circuit 406 transmits a local memory access request which reads the transfer control information 200 stored in the local memory 304 to the command send circuit 411 through line 417S, to make the header of the message. The transfer control information 200 as read from the local memory 304 based on the address written into the message transfer activation register (not shown in the drawings) is transmitted through line 413S, the data transfer interface circuit 404, line 410S, and the memory access interface circuit 403 and is further transferred to the message generating circuit 407 through the memory access interface circuit 403, line 411S, the data transfer interface circuit 404, line 412S, and the command receive circuit 405, and the command processing circuit 406.

In FIG. 5, the message generating circuit 407 sets the send data address 203, the value of the send flag address field 204, and the transfer data length 205 and the mode bit 202 in the transfer control information 200 into the registers 501, 502, 503, 506, respectively. Other information is held in the registers not shown in the drawing. The mode bit in the register 506 is one in the present case. The mode bit in the register 506 is 1, so the AND gate 507 is turned on and outputs the size of the MPI additional information in the register 505. The adder 508 adds the transfer data length in the register 503 to the MPI additional information outputted from the AND gate 507, and outputs total amount of data to be read from the local memory 701. When the comparator 510 judges that the total amount of already read data as held in the register 504 is smaller than the transfer data length held in the register 503, the memory read request issue circuit 511 transmits a local memory read request to the command send circuit 411 to read the user data 713 from the storage location of the send data address held in the register 501. The user data is read from the local memory 701 through the command send circuit 411 and the memory access interface circuit 403 in the storage controller 303. The read data is transferred to the command processing circuit 406 through the memory access interface circuit 403 in the storage controller 303, the data transfer interface circuit 404, and the command receive circuit 405, and is further transferred to the message generating circuit 407 through line 418S.

When the comparator 510 judges that the total amount of already read data as held in the register 504 is larger than the transfer data length held in the register 503 but is not greater than a sum of the transfer data length and the size of the MPI additional information held in the register 505, the memory read request issue circuit 511 transmits a local memory read request to the command send circuit 411, to read the MPI additional information 714 from a storage location in the local memory 701 which has the send flag address held in the register 502. The read request is sent from the command send circuit 411 to the local memory 701 through the data transfer interface circuit 404 and the memory access interface circuit 403 in the storage controller 303. The MPI additional information 714 is read from the local memory 701. The MPI additional information 714 is transferred to the command processing circuit 406 through the memory access interface circuit 403, the data transfer interface circuit 404, and the command receive circuit 405 in the storage controller 303 and is further transferred to the message generating circuit 407 through line 418S.

The selector 514 in the message assembly circuit 513 in the message generating circuit 407 receives the mode bit is as a control signal. As the mod is one, the selector stores the part of the transfer control information 200 read from the local memory 304 other than the send data address and the send flag address into the header portion in the register 515. Similarly, the selector 514 stores the user data 713 and the MPI additional information 714 into the data portion of the register 515.

When the comparator 510 judges that the total amount of already read data as held in

the register 504 has become equal to the sum of the transfer data length and the size of the MPI additional information, the memory write request issue circuit 512 transmits to the command send circuit 411, a local memory write request which requests writing of the send completion flag. The request requests writing of the flag into a storage location of an address equal to the sum of the start address of the MPI additional information as held in the send flag address field in the register 602 and the size of the MPI additional information in the register 505. The sum is already provided by the adder 509 and is already stored in the control register (not shown) within the memory write request issue circuit 611, as was explained before. The send completion flag 715 is written into the local memory 304 from the command send circuit 411 through the memory access interface circuit 403 in the storage controller 303. In the present embodiment, completion of reading of the user data and the MPI additional information which the message should include is regarded as completion of transfer of the message, and the send completion flag 715 is written at completion of the reading. The flag, however, may be written when the message has been transferred to the interconnecting network 105.

Thus, one message 705 which contains the user data and the MPI additional information which relates to it is generated within the register 515. The message 705 includes the number of the destination processor 716, the mode bit 717, the transfer data length 718, the receive data address 719, the receive flag address 720, and the other control information 725 all included in the transfer control information 200, and further includes the user data 726 and the additional information 727 in place of the send data address 203 and the send flag address 204 included in the transfer control information 200. The message generating circuit 407 transfers the message 705 to the message send circuit 408. When the message generating circuit 407 finishes sending the message 705 to the message send circuit 408, the message send circuit 408 resets the register 504 and 506. The message send circuit 408 sends the message 705 to the interconnecting network 105. The interconnecting network 105 transfers the message to a processor designated by the number of the destination processor 716 in the message.

Next, the receive processing will be explained. The message 705 transferred from the interconnecting network 105 is received by the message receive circuit 409 and is transferred to the message disassembly circuit 410. The message disassembly circuit 410 writes the receive data address 719, the receive flag address 720, the transfer data length 718, and the mode bit 717 included in the message header into the register 601, 602, 603, and 606 (FIG. 6), respectively. The register 604 is reset to 0 beforehand and the size of the MPI additional information is set into the register 605 beforehand. The memory write request issue circuit 611 in the message disassembly circuit 410 generates a local memory access request to write the received user data into the local memory region designated by the receive data address 719 in the header, and transfers it to the command send circuit 411 through line 423S. The received user data is written into the region 721 in the communication region 704 of the local memory 304 according to the receive data address 719 in the message header, by way of the memory access interface circuit 403 within the storage controller 303. The mode bit in the register 606 is 1, so the AND gate 607 is turned on and supplies the size of the MPI additional information held in the register 605 to the adder 608. The adder 608 adds the transfer data length 607 in the register 603 to the size of the MPI additional information provided from the AND gate 607, and supplies total amount of data to be written into the local memory. Whenever the message receive circuit 409 receives one of different portions of the user data in the message from the interconnecting network 105, the amount of the received data held in the register 604 is updated.

Whenever a different portion of the user data is received by the message receive circuit 409, the comparator 610 in the message disassembly circuit 410 judges whether the total amount of received data as held in the register 604 is smaller than the transfer data length held in the register 603. When the comparator 610 judges that the former is smaller than the latter, the memory write request issue circuit 611 requests the command send circuit 411 to write the received user data into the local memory 702 of the receiving side according to the receive data address held in the register 601.

When the comparator 610 judges that the total amount of already read data as held in the register 604 is larger than the transfer data length held in the register 503 but is not greater than a sum of the transfer data length and the size of the MPI additional information held in the register 605, the memory read request issue circuit 611 transmits to the command send circuit 411, a local memory write request which requests writing of the received MPI additional information into a storage location of the local memory 701 which has an address included in the receive flag address field as held in the register 602.

As already mentioned, the adder 609 adds the value of the receive flag address field in the register 602 to the size of the MPI additional information given from the AND gate 607, and decides a memory address at which the receive completion flag should be written. The output of the adder 609 was already stored in the control register (not shown) within the memory write request issue circuit611 at start of the receive operation in the element processor of the receiving side. When the comparator 610 detects that the amount of already receives data as held in the register 604 has become equal to a sum of the transfer data length and the size of the MPI additional information, as given from the adder 608, the memory write request issue circuit 611 transmits to the command send circuit 411, a local memory write request which requests writing of the receive completion flag at the receive completion flag address given by the control register (not shown) within the memory write request issue circuit 611. As a result, the receive completion flag 723 is written into the local memory 304 from the command send circuit 411 through the memory access interface circuit 403 in the storage controller 303. Thus, the data receive processing is completed. Moreover, the data transfer processing ends at completion of the data receive processing.

The received user data and additional information are used in the element processor of the receiving side in the following manner. The user process of the receiving side issues a command which requests receiving of data transferred from another element processor, MPI_recv, for instance. This command designates arguments which include an address indicative of user data to be received, the maximum length of the user data and the additional information predetermined by the MPI library. The additional information includes an identifier of the user process of the sending side and other information. The address designated by this command is a virtual address which belongs to the address space allocated to the user process of the receiving side. The MPI library of the receiving side judges, based on the additional information 722, and the receive completion flag 723 written in the local memory 702, whether the user data as designated by the argument of this receive command has been written into the local memory 702. The MPI library detects a valid receive completion flag 723, and compares the additional information 722 corresponding to the flag with the additional information designated by the MPI_recv command. If coincidence is detected with respect to necessary information as determined by the MPI library, the MPI library judges that the user data is already written. If the user data has been written into the local memory 702, the MPI library of the receiving side notifies the user process of the receiving side of completion of receipt of the user data. If the user data has not been written into the local memory 702 yet, the MPI library repeats the judgement until the

judgement succeeds.

If data region which has the virtual address designated by the user process of the receiving side is not resident in the local memory 702, the MPI library of the receiving side copies the received user data 721 onto a local memory region allocated to the address which the user process designates, before executing the notifying operation. If data region which has the virtual address designated by the user process of the receiving side is resident in the local memory 702, the copy operation is unnecessary.

The user process of the receiving side executes an instruction to read the received data, when the user process receives the notification of the receive completion. Therefore, the receive processing of the user data in the receiving side differs from the prior art in that the common receive completion flag 723 is used to detect whether the user data 721 and the additional information 722 has been written into the local memory 702.

The explanation given above is about the case where the user process has issued a send request command of data to the MPI library. In the present embodiment a user process can issue a data send request command to the PUT/GET communication library directly. In this case, the data send request command designates the virtual address indicative of the user data to be transferred, the data length, the destination element processor, the send data address and so on to the PUT/GET communication library. The PUT/GET communication library responds to this command, generates the transfer control information 200 which has the same structure as one described above and writes it into the local memory 701. The value of the mode bit 202 in this information is 0. The send flag address field 204 designates the address of a write region 715 of the send completion flag. It is the same with the receive flag address field 207. The additional information storage regions 714, 722 are unnecessary for the data send request command. The addresses such as the send data address are all real addresses as in the case of the transfer control information generated by the instruction from the MPI library.

When the mode bit 202 in the transfer control information 200 is 0, the operations of the message generating circuit 407 and the message disassembly circuit 410 differ from the case when the mode bit 202 in the transfer control information 200 is one. That is, as the mode bit in the register 506 is 0, the output of the adder 508 is equal to the transfer data length held in the register 503, and the output of the adder 509 is equal to the send completion flag address held in the register 502. The message generating circuit 407 finishes reading of data to be transferred, when the comparator 510 has detected that the total amount of already read data as held in the register 504 has become equal to the transfer data length held in the register 504 and given from the AND gate 507. When entirety of the message which contains the read out data has been sent to the interconnecting network 105, the memory write request issue circuit 512 in the message generating circuit 407 transmits to the command send circuit 411, a write request which requests writing of the send completion flag into the local memory region designated by the send flag address held in the register 502 and given from the adder 509.

In the element processor of the receiving side, 0 is set into the register 606 (FIG. 6), as the mode bit 717 in the header in the message transferred from the interconnecting network 105 is 0. As the mode bit in the register 606 is 0, the output of the adder 608 is equal to the transfer data length held in the register 603 and the output of the adder 609 is equal to the receive flag address held in the register 602. Therefore, the memory write request issue circuit 611 repeats issuing of a write request which requests writing

of the receive data into the local memory until the comparator 610 detects that the total amount of received data as held in the register 604 has become equal to the transfer data length as provided by the adder 608. After all the receive data has been written into the local memory, the memory write request issue circuit 611 issues a write request to write the receive completion flag into the local memory 304 according to the receive flag address held in the register 602 and provided from the adder 609. Thus, the data transfer requested by the MPI library ends.

For the comparison, in the prior art PUT processing, as shown in FIG. 8, the PUT/GET communication library of the sending side generates two pieces of transfer control information A and B (800, 801) to transfer the user data and the additional information therefor respectively, and the user data, the send completion flag A therefor, the additional information, and the send completion flag B therefor are stored in the communication region 802 of the local memory of the sending side. The network interface circuit of the sending side transfers the two messages A and B (804, 805) based on these two pieces of transfer control information. The network interface circuit of the receiving side writes the user data, the receive completion flag A therefor, the additional information, and the receive completion flag B therefor into the communication region 803. The mode bit used in the present embodiment does not exist in the transfer control information A or B.

On the other hand, in the present embodiment, two different data such as the user data and the additional information can be transferred by one message according to the direct inter-memory data transfer method. Therefore, the transfer latency and the local memory access frequency can be reduced and the parallel processing efficiency can be improved, in comparison to the prior art which needs transfer of two messages. In addition, the prior art direct inter-memory data transfer method which does not use the MPI library can be executed.

In addition, the network interface circuit of the sending side uses the send completion flag address to write both the additional information and control information which shows the send completion. So, the amount of information which the PUT/GET communication library should designate for the network interface circuit can be less. Moreover, the network interface circuit in the receiving side uses the receive completion address to write the additional information and control information which shows the receive completion. So, the amount of information included in a PUT message can be less. As a result, the congestion of the network can be prevented, and the transfer time through the network and the receive time in the destination element processor can be shortened. Transfer time of a message can be shortened as a whole. In the prior art method as mentioned above, the network interface circuit 305 needs to access the local memory six times for the send processing of the two different kinds of data (reading of two pieces of transfer control information, reading of two kinds of transfer data, and writing of two flags). The network interface circuit 305 needs to access the local memory four times for the receive processing (writing of two kinds of transfer data, and writing of two flags). In the present embodiment, the total number of accesses to the local memory of the sending side and the total amount of aceses to the local memory of the receiving side are reduced respectively to four times and three times. So, the processing overhead due to the local memory accesses can be reduced.

The PUT processing through the MPI library was explained above but there are other communications such as GET processing in which an element processor of the sending side fetches user data stored in the local memory of the destination element processor. The embodiment can be similarly applied also to the GET processing through the MPI library. That is, the user data and the additional information therefor

can be got (transferred) in one message at the same time by using the MPI library together with the PUT/GET library, as will be explained below.

The user process of the request side issues a GET request command to the MPI library. This command designates arguments which are similar to that designated by the send request command explained above in that the arguments include the address of requested data, the data length and the additional information decided by the MPI library and so on. The MPI library, the PUT/GET communication library and the network interface circuit responds to this command, and transfers the GET request message to the destination element processor as in the case of the PUT operation. The send data address for data to be got and the send completion flag for the data are included in the message instead of the receive data address and the receive completion flag address included in the message at the PUT operation, and neither the user data nor the additional information are included. The point where the message is different from the prior art GET request message is to have the mode bit.

The destination element processor reads the user data from the local memory region designated by the data address included in the GET request message, sends back reads the additional information from the local memory region designated by the send completion flag address, and sends back to the processor element of the sending side, the GET message which contains the user data and the additional information. The message contains the receive data address and the receive completion flag. After sending the entirety of the message the interconnecting network, the destination element processor writes the send completion flag at a location in the local memory next to the storage region of the additional information.

The element processor of the sending side receives the GET message, writes the user data in the message into the local memory according to the receive data address in the message, and writes the additional information in the message into the local memory according to the receive completion flag address in the message. After the element processor of the sending side has received all of the user data and the additional information, it writes the receive completion flag into the local memory. The processing overhead due to the local memory access can be reduced by reducing the read frequency of the transfer control information from the local memory and the writing frequency of the flag into the local memory, even in the GET processing as in the PUT processing. Moreover, in the prior art, the user data and the additional information are transferred separately, but they are transferred together in the present embodiment. So, the latency of the data transfer can be reduced.

MODIFICATIONS AND VARIATIONS

The present invention is not limited to the embodiment shown above, but can be realized as various modifications or variations, including the following.

For instance, the extended size field can be installed in the transfer control information stored in the local memory 304. The size of the extended flag region is set into the extended size field. When the mode bit is set, at the send processing, a message is generated which includes part of the transfer control information, data of the transfer data length as read from the local memory region according to the send data address, and another data of the size set in the extended size field as read from the local memory region according to the send flag address, and the message is transferred to the interconnecting network 105. In the receive processing, the message is disassembled into the header part and the data part. The data portion in the message is written by the amount equal to the transfer data length, into the local memory region designated by the receive data address included in the header. Another data portion in

the message is written by the amount equal to the size included in the extended size field, into the local memory region designated by the receive flag address included in the header. When the mode bit is not set, the value set into the extended size field is neglected.

According to the data transfer method by the present invention, when the direct inter-memory data transfer is executed through the message passing library like MPI, the user data and the additional information predetermined by the message passing library can be transferred in one message. So, the data transfer processing can be executed at high speed.

Moreover, the computer system according to the present invention has a network interface circuit which can generate one message which contains different data such as user data and additional information as send data, so the computer system is suitable to execute the data transfer method.

Espacenet

Claims: JPH11126196 (A) — 1999-05-11

DATA TRANSFER METHOD AND COMPUTER SYSTEM SUITABLE FOR THE SAME

Claims not available for JPH11126196 (A)
Claims of corresponding document: US6338095 (B1)

A high quality text as facsimile in your desired language may be available amongst the following family members:

US6338095 (B1)

- Original claims
- Claims tree

The EPO does not accept any responsibility for the accuracy of data and information originating from other authorities than the EPO; in particular, the EPO does not guarantee that they are complete, up-to-date or fit for specific purposes.

What is claimed is:

1. A data transfer method in a computer system which includes a plurality of element processors connected by an interconnecting network, wherein each element processor includes a processor, a memory, and a network interface circuit for exchanging messages with said interconnecting network, wherein each element processor is provided with a message passing library for communicating with user processes running therein and a direct inter-memory data transfer library for communicating with said message passing library and controlling said network interface circuit, said method comprising the steps of:informing, by said message passing library provided in an element processor of a sending side, first to fourth pieces of address information and data length information, to said direct inter-memory data transfer library provided in said element processor of a sending side, in response to a data send request issued by a user process of a sending side running in said element processor of a sending side; wherein said user process designates user data and additional information to be transferred and said data length information, said data length information relating to said user data; wherein said additional information comprises a plurality of pieces of information required at transfer of said user data to a user process of a receiving side running in an element processor of a receiving side; wherein said first and second pieces of address information designate two storage regions within said memory of said element processor of a sending side where said user data and said additional information are stored; wherein said third and fourth pieces of address information designate two storage regions within said memory of said element processor of a receiving side where said user data and said additional information are to be stored; requesting said direct inter-memory data transfer library, by said message passing

library, to transfer a message which includes said user data, said additional information, said data length information, and said third to fourth pieces of address information; requesting said network interface circuit of said element processor of a sending side, by said direct inter-memory data transfer library, to transfer said user data, said additional information, said data length information, and said third and fourth pieces of address information, to said element processor of a receiving side, in response to said requesting by said message passing library; reading, by said network interface circuit of said element processor of a sending side, said user data and said additional information from said memory of said element processor of a sending side, based on said first and second pieces of address information; assembling a message, by said network interface circuit of said element processor of a sending side, send data included in said message including said user data and said additional information as read, a header of said message including said third and fourth pieces of address information and said data length information; transferring, by said network interface circuit, said message to said element processor of a receiving side by way of said interconnecting network; and writing by said network interface circuit in said element processor of said receiving side, said user and said additional information within said message into two storage regions within said memory of said element processor of a receiving side respectively designated by said third and fourth pieces of address information.

2. A data transfer method according to claim 1, further comprising the steps of:writing by said network interface circuit in said element processor of a sending side after said message has been transmitted, control information indicative of send completion at a storage location in said memory in said element processor of a sending side which is designated by said second piece of address information and has a predetermined relation to one of said two storage regions where said additional information is stored; and writing by said network interface circuit in said element processor of said receiving side after said user data and said additional information within said message has been written into said memory of said element processor of said receiving side, control information indicative of receive completion at a storage location in said memory in said element processor of said receiving side having a predetermined relation to one of said two storage regions which is designated by said fourth piece of address information in said message and where said additional information has been stored.

3. A data transfer method in a computer system which includes a plurality of element processors connected by an interconnecting network, wherein each element processor includes a processor, a memory, and a network interface circuit for exchanging messages with said interconnecting network, wherein each element processor is provided with a message passing library for communicating with user processes running therein and a direct inter-memory data transfer library for communicating with said message passing library and controlling said network interface circuit, said method comprising the steps of:informing, by said message passing library provided in an element processor of a sending side, first to fourth pieces of address information and data length information, to said direct inter-memory data transfer library provided in said element processor of a sending side, in response to a data send request issued by a user process of a sending side running in said element processor of a sending side; wherein said user process designates user data and additional information to be transferred and said data length information, said data length information relating to said user data; wherein said additional information comprises a plurality of pieces of information required at transfer of said user data to a user process of a receiving side running in an element processor of a receiving side; wherein said first and second pieces of address information designate two storage regions within said memory of said element processor of a sending side where said user data and said additional

information are stored; wherein said third and fourth pieces of address information designate two storage regions within said memory of said element processor of a receiving side where said user data and said additional information are to be stored; requesting said direct inter-memory data transfer library, by said message passing library, to transfer a message which includes said user data, said additional information, said data length information, and said third to fourth pieces of address information; requesting said network interface circuit of said element processor of a sending side, by said direct inter-memory data transfer library, to transfer said user data, said additional information, said data length information, and said third and fourth pieces of address information, to said element processor of a receiving side, in response to said requesting by said message passing library; reading, by said network interface circuit of said element processor of a sending side, said user data and said additional information from said memory of said element processor of a sending side, based on said first and second pieces of address information; assembling a message, by said network interface circuit of said element processor of a sending side, send data included in said message including said user data and said additional information as read, a header of said message including said third and fourth pieces of address information and said data length information; transferring, by said network interface circuit, said message to said element processor of a receiving side by way of said interconnecting network; and generating by said direct inter-memory data transfer library provided in said element processor of a sending side, transfer control information which includes said first to fourth pieces of address information and said data length information; wherein said requesting step executed by said direct inter-memory data transfer library in said element processor of a sending side to said network interface circuit in said element processor of a sending side includes a step of designating said transfer control information.

4. A computer system comprising:a plurality of element processors; and an interconnecting network for connecting said plurality of element processors; wherein each element processor includes: a processor, a memory, and a network interface circuit for exchanging messages with said interconnecting network; wherein said network interface circuit includes: a memory read circuit connected to said memory, and a message assembly circuit connected to said memory read circuit for generating a message to be transferred to said interconnecting network; wherein said memory read circuit reads user data and additional information to be transferred from said memory, respectively based on first and second pieces of address information as designated by said processor; wherein said message assembly circuit generates a message which includes a header and send data which includes said user data and said additional information; wherein said header includes third and fourth pieces of address information as designated by said processor which indicates storage regions in said memory in a destination element processor where said user data and additional information within said send data are to be stored, respectively.

5. A computer system comprising:a plurality of element processors; and an interconnecting network for connecting said plurality of element processors; wherein each element processor includes: a processor, a memory, and a network interface circuit for exchanging messages with said interconnecting network; wherein said network interface circuit includes: a memory read circuit connected to said memory, and a message assembly circuit connected to said memory read circuit for generating a message to be transferred to said interconnecting network; wherein said memory read circuit reads user data and additional information to be transferred from said memory, respectively based on first and second pieces of address information as designated by said processor; wherein said message assembly circuit generates a message which includes a header and send data which includes said user data and said additional

information, wherein said processor writes said first to fourth pieces of address information in said memory; wherein said memory read circuit reads said first and second pieces of address information from said memory; wherein said memory read circuit uses said first and second pieces of address information as read for readout of said user data and said additional information, respectively; and wherein said message assembly circuit uses said third and fourth pieces of address information for said header.

6. A computer system according to claim 5, wherein said network interface circuit includes a circuit for writing control information indicative of send completion at a storage location in said memory which has a predetermined relation to a storage region which is designated by said second piece of address information and where said additional information is stored.

7. A computer system according to claim 6,wherein said memory read circuit reads from said memory, transfer control information as having been written into said memory for said message; wherein said transfer control information includes said first to fourth pieces of address information, data length information related to said first data, and other information; wherein said memory read circuit uses said first and second pieces of address information and said data length information as included in said transfer control information, for readout of said first and second data; wherein said memory write circuit uses said second piece of address information to write control information indicative of send completion; wherein said message assembly circuit uses for said header, said third and fourth pieces of address information, said data length information and said other information as included in said transfer control information.

8. A computer system, according to claim 7, wherein said transfer control information further includes a mode bit;wherein said memory read circuit executes readout of said user data and said additional information, when said mode bit is of a first value; wherein said memory read circuit executes readout of said user data and does not reads out said additional information, when said mode bit is of a second value; wherein said message assembly circuit generates a message send data of which includes said user data and said additional information, when said mode bit is of said first value, and generates a message send data of which includes said first data but not said additional information, when said mode bit is of said second value; wherein said memory write circuit writes control information indicative of send completion, based on said second piece of address information, at a storage location in said memory which has a predetermined relation to said storage region where said additional information is stored, when said mode hit is of said first value, and at a storage location designated by said second piece of address information, when said mode bit is of said second value.

9. A computer system comprising:a plurality of element processors; and an interconnecting network for connecting said plurality of element processors; wherein each element processor includes: a processor, a memory, and a network interface circuit for exchanging messages with said interconnecting network; wherein said network interface circuit includes: a memory read circuit connected to said memory, and a message assembly circuit connected to said memory read circuit for generating a message to be transferred to said interconnecting network; wherein said memory read circuit reads user data and additional information to be transferred from said memory, respectively based on first and second pieces of address information as designated by said processor; wherein said message assembly circuit generates a message which includes a header and send data which includes said user data and said additional information, wherein said network interface circuit further includes: a receive circuit for

receiving a message transferred from another element processor by way of said interconnecting network, and another memory write circuit; wherein send data in said received message includes another user data and another additional information, and a header in said received message includes another third piece of address information and another fourth piece of address information; and wherein said another memory write circuit writes said another first data within said received message into said memory, based on said another third piece of address information within said received message, and writes said another additional within said received message into said memory, based on said another fourth piece of address information included in said received message.

10. A computer system according to claim 9, wherein said another memory write circuit writes control information indicative of receive completion of said another user data after writing said another user data and said another user within said received message into said memory, at a storage location within said memory having a predetermined relation to a storage region which is designated by said another fourth piece of address information within said received message and where said received another user has been stored.

11. A computer system according to claim 10, wherein said received message further includes a mode bit;wherein said another memory write circuit executes the following write operations when said mode bit is of a first value: (a) a write operation to write said another user data and said another additional information within said received message into said memory, respectively based upon said another third and another fourth pieces of address information included in said received message, and (b) a write operation to write, after said another user data and said another second data within said received message have been written, control information indicative of receive completion at a storage location in said memory which has a predetermined relation to a storage region which is designated by said another fourth piece of address information and where said another additional information has been stored; wherein said another memory write circuit executes the following write operations when said mode bit is of a second value: (a) a write operation to write said another user data within said received message into said memory, based upon said another third piece of address information included in said received message, and (b) a write operation to write, after said another user data within said received message has been written, control information indicative of receive completion at a storage location in said memory as designated by said another fourth piece of address information.

12. A computer system , comprising:a plurality of element processors; and an interconnecting network for connecting said plurality of element processors; wherein each element processor includes: a processor, a memory, and a network interface circuit for exchanging messages with said interconnecting network; wherein each element processor is provided with a message passing library for communicating with user processes running therein and a direct inter-memory data transfer library for communicating with said message passing library and controlling said network interface circuit; wherein said message passing library provided in each element processor is programmed so as to inform first to fourth pieces of address information and data length information, to said direct inter-memory data transfer library provided in said each element processor, in response to a data send request issued by a user process of a sending side running in said each element processor; wherein said user process designates user data and additional information to be transferred and said data length information, said data length information relating to said user data; wherein said additional information comprises a plurality of pieces of information required at transfer of said user data to a user process of a receiving side running in an element processor

of a receiving side; wherein said first and second pieces of address information designate two storage regions within said memory of said element processor of a sending side where said user data and said additional information are stored; wherein said third and fourth pieces of address information designate two storage regions within said memory of said element processor of a receiving side where said user data and said additional information are to be stored; wherein said message passing library provided in said each element processor is further programmed so as to request said direct inter-memory data transfer library to transfer a message which includes said user data, said additional information, said data length information, and said third to fourth pieces of address information; wherein said direct inter-memory data transfer library provided in said each element processor is programmed so as to request said network interface circuit of said element processor to transfer said user data, said additional information, said data length information, and said third and fourth pieces of address information, to said element processor of a receiving side, in response to said requesting by said message passing library; wherein said network interface circuit includes: a memory read circuit for reading said user data and said additional information from said memory of said each element processor, based on said first and second pieces of address information, a message assembly circuit for assembling a message, send data included in said message including said user data and said additional information, a header included in said message including said other information and said data length information, and a message transmit circuit for transferring said message to said element processor of a receiving side by way of said interconnecting network; wherein said network interface circuit further includes: a message receive circuit for receiving from another element processor, a message which includes user data, additional information, and another third and fourth pieces of address information; wherein said network interface circuit in said each element processor further includes a memory write circuit for writing said user data and said additional information included in said received message into two storage regions respectively designated by said another third and another fourth pieces of address information included in said received message.

13. A computer system according to claim 12, wherein said network interface circuit in said each element processor further includes:another memory write circuit for writing, after said message has been transmitted, control information indicative of send completion at a storage location in said memory in said each element processor which has a predetermined relation to one of said two storage regions which is designated by said second piece of address information and where said additional information is stored; and a further another memory write circuit for writing, after writing by said memory write circuit, of said user data and said additional information included in said received message, control information indicative of receive at a storage location in said memory in said each element processor which has a predetermined relation to one of said two storage regions which is designated by said another fourth piece of address information and where said received additional information is stored.

(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（Ａ）　(11)特許出願公開番号

特開平11－126196

(43)公開日　平成11年(1999)５月11日

(51)Int.Cl.[6]		識別記号		FI		
Ｇ０６Ｆ	15/163			Ｃ０６Ｆ	15/16	３２０Ａ
	13/00	３５３			13/00	３５３Ｃ

審査請求　未請求　請求項の数９　ＯＬ　（全 19 頁）

(21)出願番号　特願平9－290597

(22)出願日　平成９年(1997)10月23日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　保田　淑子
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内

(72)発明者　藤井　啓明
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内

(74)代理人　弁理士　高橋　明夫　（外１名）

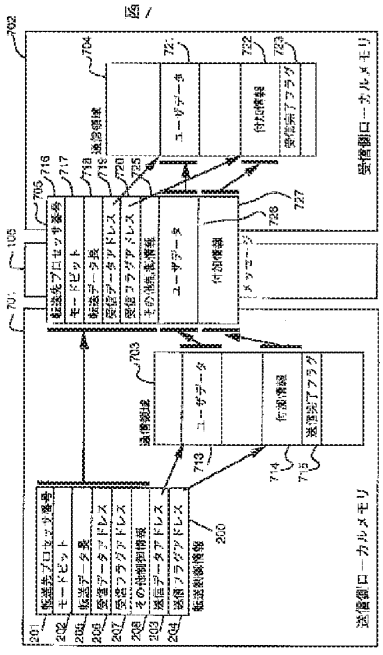
(54)【発明の名称】　データ転送方法およびそれに適した計算機システム

(57)【要約】

【課題】標準ＭＰＩライブラリを介したメモリ間直接転送を高速化する。

【解決手段】送信側ＰＵＴ／ＧＥＴライブラリは、ユーザプロセスがＭＰＩライブラリに対して発行したデータ送信要求コマンドに応答して転送制御情報２００を生成する。送信側ネットワークインタフェース回路は、その中の送信データアドレスと送信フラグアドレスを用いてそれぞれユーザデータ７１３とＭＰＩライブラリにより使用されるユーザデータに対応する付加情報７１４をローカルメモリ３０４から読み出し、一つのメッセージ７０５としてネットワーク１０５に送出し、送信完了フラグ７１５を送信フラグアドレスを用いて書き込む。受信側のネットワークインタフェース回路は、メッセージ内の受信データアドレスと受信フラグアドレスで指定されるローカルメモリ領域７２１、７２２にそれぞれユーザデータと付加情報を書き込み、受信完了フラグ７２３を領域７２３に書き込む。

図１



【特許請求の範囲】
【請求項１】相互結合ネットワークで接続された複数の要素プロセッサを有し、各要素プロセッサ内には、ユーザプロセスと交信するメッセージパッシングライブラリと、そのメッセージパッシングライブラリと交信するメモリ間直接転送ライブラリとが組み込まれ、各要素プロセッサは、プロセッサと、メモリと、上記相互結合ネットワークとの間でメッセージを交換するためのネットワークインタフェース回路とを有する計算機システムにおいて、
送信側の要素プロセッサで走行中のユーザプロセスにより、その要素プロセッサ内のメッセージパッシングライブラリに対して発行されたデータ送信要求が要求する送信データと、そのデータ送信要求が要求する、メッセージパッシングライブラリにより定められた、受信側の要素プロセッサに送信されるべき付加情報との送信を、そのメッセージパッシングライブラリからその要素プロセッサ内のメモリ間直接転送ライブラリに対して要求し、
そのメモリ間直接転送ライブラリにより、上記要求された送信データおよび付加情報と、それらの受信の完了を示す制御情報を書き込むべき領域を指定するための、受信側の要素プロセッサがあらかじめ決定した受信側アドレス情報とを含むメッセージの送信を上記ネットワークインタフェース回路に対して要求し、
上記ネットワークインタフェース回路により、上記メッセージを組立て、受信側の要素プロセッサに宛てて上記相互結合ネットワークに送信し、
上記メッセージの送信後に、そのメモリ間直接転送ライブラリが決定したアドレスを有する、上記メモリ内の記憶位置に送信完了を示す制御情報を上記ネットワークインタフェース回路により書き込むステップからなるデータ転送方法。
【請求項２】受信側の要素プロセッサ内のネットワークインタフェース回路により、上記メッセージ内の送信データと付加情報を、受信側の要素プロセッサのメモリ内の、上記受信側アドレス情報により定まる領域に書き込み、
上記書き込みの終了後に、受信側の要素プロセッサ内のネットワークインタフェース回路により、受信完了を示す制御情報を、上記メモリ内の、上記受信側アドレス情報により定まる領域に書き込むステップをさらに有する請求項１記載のデータ転送方法。
【請求項３】上記データ送信要求は、上記送信データに関する第１のアドレス情報とデータ長情報と上記付加情報を指定し、
上記メッセージの送信を要求するステップは、
送信側の要素プロセッサ内の上記メモリ間直接転送ライブラリにより、上記要求された付加情報を、そのメモリ間直接転送ライブラリが付加情報と送信完了を示す制御情報を書き込むための領域として決定した、上記メモリ内の領域に書き込み、
そのメモリ間直接転送ライブラリにより、上記第１のアドレス情報、上記送信データ長情報、上記付加情報が書き込まれた上記領域を指定する第２のアドレス情報および上記受信側アドレス情報とを指定するメッセージ送信要求を送信側の要素プロセッサ内の上記ネットワークインタフェース回路に対して発行するステップからなり、
上記メッセージを送信するステップは、
上記ネットワークインタフェース回路により、そのメッセージ送信要求で指定された上記第１のアドレス情報と上記データ長情報との組および上記第２のアドレス情報にそれぞれ基づいて、上記送信データおよび上記付加情報を送信側の要素プロセッサのメモリより読み出し、
上記ネットワークインタフェース回路により、そのメッセージ送信要求が指定した上記受信側アドレス情報と上記データ長情報、上記読み出された送信データおよび付加情報を含むメッセージを生成し、上記ネットワークを介して受信側の要素プロセッサに送信するステップを有し、
上記送信完了を示す制御情報を書き込むステップは、上記第２のアドレス情報に基づいて、上記付加情報が記憶された領域と異なる記憶位置に送信完了を示す制御情報を書き込むステップを有する請求項１記載のデータ送信方法。
【請求項４】受信側の要素プロセッサ内のネットワークインタフェース回路により、上記メッセージ内の送信データと付加情報を、受信側の要素プロセッサのメモリ内の、上記受信側アドレス情報により定まる領域に書き込み、
上記書き込みの終了後に、受信側の要素プロセッサ内のネットワークインタフェース回路により、受信完了を示す制御情報を、上記メモリ内の、上記受信側アドレス情報により定まる領域に書き込むステップをさらに有し、
上記受信側のアドレス情報は、上記メッセージ内の送信データを書き込むべき第３のアドレス情報と受信完了を示す制御情報を書き込む第４のアドレス情報からなり、
上記メッセージ内の送信データと付加情報を書き込むステップは、上記第３、第４のアドレス情報に基づいて、上記送信データおよび上記付加情報を、受信側の要素プロセッサのメモリに書き込むステップからなり、
上記受信完了を示す制御情報を書き込むステップは、上記第４のアドレス情報に基づいて、受信側の要素プロセッサのメモリ内の、上記付加情報が書き込まれた領域と異なる領域に書き込むステップからなる請求項３記載のデータ転送方法。
【請求項５】送信側の要素プロセッサで走行中の他のユーザプロセスによりその要素プロセッサ内の上記メモリ間直接転送ライブラリに対して発行された他のデータ送信要求が要求する送信データと、受信側の要素プロセッサがあらかじめ決定した、受信データと受信完了を示す

制御情報を書き込むべき領域を指定するための受信側アドレス情報とを含むメッセージの送信を、そのメモリ間直接転送ライブラリにより、送信側の要素プロセッサの上記ネットワークインタフェース回路に対して要求し、上記メッセージを上記ネットワークインタフェース回路により組立て、受信側の要素プロセッサに宛てて上記相互結合ネットワークに送信し、

上記メッセージの送信後に、そのメモリ間直接転送ライブラリが決定したアドレスを有する、上記メモリ内の記憶位置に送信完了を示す制御情報を上記ネットワークインタフェース回路により書き込み、

受信側の要素プロセッサ内のネットワークインタフェース回路により、上記メッセージ内の送信データを、受信側の要素プロセッサのメモリ内の、上記受信側アドレス情報により定まる領域に書き込み、

上記書き込みの終了後に、受信側の要素プロセッサ内のネットワークインタフェース回路により、受信完了を示す制御情報を、上記メモリ内の、上記受信側アドレス情報により定まる領域に書き込むステップをさらに有する請求項２記載のデータ転送方法。

【請求項６】相互結合ネットワークで接続された複数の要素プロセッサを有し、

各要素プロセッサは、プロセッサと、メモリと、上記相互結合ネットワークとの間でメッセージを交換するためのネットワークインタフェース回路とを有し、上記ネットワークインタフェース回路は、

転送すべきメッセージに関する情報として上記プロセッサにより供給される、第１のアドレス情報とデータ長情報との組および第２のアドレス情報とに基づいて、それぞれ送信されるべき第１のデータおよびそのデータとともに送信されるべき第２のデータを上記メモリから読み出すメモリアクセス回路と、

読みだされた第１，第２のデータと、転送すべきメッセージに関する他の情報として上記プロセッサにより供給される第３，第４のアドレス情報を含む一つのメッセージを生成し、上記相互結合ネットワークに送信する回路とを有し、

上記メモリアクセス回路は、上記メッセージの送信後に、上記第２のアドレス情報に基づいて、上記送信完了を示す制御情報を上記メモリ内の、上記第２のデータが記憶されている記憶位置と異なる記憶位置に書き込み、該第１のアドレス情報は、上記第１のデータを保持する、上記メモリ内の領域のアドレスを指示し、上記データ長情報は、上記第１のデータの長さを指定し、該第２のアドレス情報は、上記第２のデータを保持し、かつ、該第１のデータの送信完了を示す制御情報をさらに保持すべき、上記メモリ内の領域のアドレスを指示し、該第３のアドレス情報は、第１のデータを書き込むべき、受信側の要素プロセッサのメモリ内の領域のアドレスを指示し、該第４のアドレス情報は、上記第１のデータおよび該第１のデータの受信完了を示す制御情報とを格納するための、受信側の要素プロセッサのメモリ内の領域のアドレスを指示する計算機システム。

【請求項７】上記ネットワークインタフェース回路は、他の要素プロセッサから送信されたメッセージを上記相互結合ネットワークから受信する回路をさらに有し、

上記メモリアクセス回路は、受信されたメッセージ内の上記第３のアドレス情報に基づいて、受信されたメッセージ内の上記第１のデータを上記メモリに書き込み、受信されたメッセージ内の上記第４のアドレス情報に基づいて、受信されたメッセージ内の上記第２のデータを上記メモリに書き込み、上記受信されたメッセージ内の上記第１，第２のデータの書き込みの終了後に、上記第４のアドレス情報に基づいて、上記第１のデータの受信完了を示す制御情報を、上記メモリ内の、受信された上記第２のデータが記憶されている記憶位置と異なる記憶位置に書き込む回路を有する請求項６記載の計算機システム。

【請求項８】上記転送制御情報はモードビットをさらに有し、

上記メモリアクセス回路は、上記モードビットが第１の値の時に、上記第２のデータの読み出しを実行し、上記モードビットが第２の値の時に、上記第２のデータの読み出しを実行せず、

上記メッセージ送信回路は、上記モードビットが第１の値の時に、上記第２のデータを含むメッセージを生成し、上記モードビットが第２の値の時に、上記第２のデータを含まないメッセージを生成し、

上記メモリアクセス回路は、上記モードビットが第１の値の時には、上記メッセージの送信後に、上記第２のアドレス情報に基づいて、上記送信完了を示す制御情報を、上記メモリ内の、上記第２のデータが記憶されている記憶位置と異なる記憶位置に書き込み、上記モードビットが第２の値の時には、上記メッセージの送信後に、上記送信完了を示す制御情報を、上記第２のアドレス情報に依存する、上記メモリ内の記憶位置に書き込む回路を有する請求項６記載の計算機システム。

【請求項９】上記メッセージ送信が生成するメッセージは、上記モードビットを含み、

上記ネットワークインタフェース回路は、他の要素プロセッサから送信されたメッセージを上記相互結合ネットワークから受信する回路をさらに有し、

上記メモリアクセス回路は、

受信されたメッセージ内のモードビットが第１の値の時には、受信されたメッセージ内の上記第３のアドレス情報に基づいて、受信されたメッセージ内の上記第１のデータを上記メモリに書き込み、上記受信されたメッセージ内の上記第４のアドレス情報に基づいて、受信されたメッセージ内の上記第２のデータを上記メモリに書き込み、上記受信されたメッセージ内の上記第１，第２のデ

ータの書き込みの終了後に、上記第４のアドレス情報に基づいて、上記第１のデータの受信完了を示す制御情報を、上記メモリ内の、受信された上記第２のデータが記憶されている記憶位置と異なる記憶位置に書き込み、受信されたメッセージ内のモードビットが第２の値の時には、受信されたメッセージ内の上記第１のデータを、受信されたメッセージ内の上記第３のアドレス情報に基づいて、上記メモリ内の記憶位置に書き込み、上記受信されたメッセージ内の上記第１のデータの書き込みの終了後に、上記第１のデータの受信完了を示す制御情報を、上記第４のアドレス情報に依存する、上記メモリ内の記憶位置に書き込む回路を有する請求項８記載の計算機システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、相互結合ネットワークを介して接続された複数の要素プロセッサ間でのデータ転送方法およびそれに適した計算機システムに関する。

【０００２】

【従来の技術】従来、並列計算機は、ローカルメモリと命令プロセッサから構成される複数の要素プロセッサを相互結合ネットワークで結合した構成をとっている。一般に、このような形態の並列計算機は、分散メモリ型並列計算機と呼ばれる。各要素プロセッサは、相互結合ネットワークを介して個々のローカルメモリに格納されているデータの授受を行い、並列に処理を実行する。

【０００３】一般に、分散メモリ型の並列計算機では、メッセージパッシングと呼ぶプログラミングモデルを用いてデータ転送を実現する。メッセージパッシングモデルでは、ユーザが並列プログラム中に明示的に送信（ＳＥＮＤ）手続きおよび受信（ＲＥＣＥＩＶＥ）手続きを記して、要素プロセッサ間で必要になるデータの授受をメッセージのやりとりという形で行う。命令プロセッサはこれらの通信手続きを解析して、相互結合ネットワークにデータを送信したり、相互結合ネットワークからデータを受信しながら処理を進める。送信元の要素プロセッサは、転送先の要素プロセッサ番号を指定してメッセージを転送し、転送先の要素プロセッサでメッセージをバッファリングする。メッセージパシングモデルでは、メッセージ通信に伴ってデータのバッファリングやフロー制御が必要となり、送受信オーバヘッドが大きくなってしまう。

【０００４】この送受信オーバヘッドを削減するために、近年複数の並列計算機において、要素プロセッサのローカルメモリの内容を、メッセージの生成あるいは受信を行う送受信回路が直接アクセスするメモリ間直接転送方法が使用されている。この方法を実行する代表例は、ＰＵＴ／ＧＥＴ通信である。例えば、”情報処理学会並列処理シンポジウムＪＳＰＰ’９５、”ＰＰ．２３３－２４０（１９９５年５月）参照。メモリ間直接転送方法では、各要素プロセッサのローカルメモリからの送信データの読み出しあるいはそのメモリへの受信データの書き込みをメッセージの生成あるいは受信を行う送受信回路が直接実行するため、これらのデータをＯＳ管理の領域にコピーする必要がなく、このコピーに由来するオーバーヘッドを削減できる。

【０００５】しかしながら、このようなＰＵＴ／ＧＥＴ通信を実際に実行する通信ライブラリは、各並列計算機メーカーや研究機関が独自に開発しているため、それを用いて作成された並列プログラムを他機種へ移植することは困難であった。この問題点を解決するために、ＭＰＩ（Ｍｅｓｓａｇｅ　Ｐａｓｓｉｎｇ　Ｉｎｔｅｒｆａｃｅ）に代表される、メッセージパッシングライブラリの標準化が進みつつある。ＭＰＩは、米国各大学および並列計算機メーカーがメッセージパッシングインクフェース標準化団体ＭＰＩ　Ｆｏｒｕｍを組織し、その研究成果をまとめた仕様である。この仕様に基づいて作成されたライブラリ（以下ＭＰＩライブラリと呼ぶことがある）は、今後の並列プログラム開発支援ライブラリの主流になると考えられる。ＭＰＩライブラリを用いて記述された並列プログラムは、異機種間で変更なしに走らせることができる。各計算機メーカーは、自社の並列計算機上で高性能を達成するようにＭＰＩライブラリを開発している。

【０００６】上記ＭＰＩ仕様は、ＰＵＴ／ＧＥＴ通信に関する仕様を含んでいない。しかし、データ転送の高速化のためには、ＰＵＴ／ＧＥＴ通信を併用することが重要である。このために、並列計算機メーカー等は、ＰＵＴ／ＧＥＴ通信ライブラリを使用可能にしたＭＰＩライブラリを開発している。たとえば、本出願人による、「並列計算機ＳＲ２２Ｄ支援ライブラリ」参照。したがって、ＭＰＩライブラリは、各計算機メーカーごとに異なるものであるが、ユーザプログラムから見れば、ＭＰＩライブラリとの間のインタフェースは、いずれの計算機メーカのＭＰＩライブラリに対しても同じである。従って、そのユーザプログラムは、いずれの計算機のメーカの上でも実行できることになる。

【０００７】ＭＰＩライブラリを用いるユーザプログラムは、データを送信する時点で、ＭＰＩライブラリをコールする。従来は、ユーザプログラムがＰＵＴ／ＧＥＴ通信ライブラリを使用するときには、ユーザプログラムは送信すべきユーザデータおよびデータ長をこのコール文の引数でもって指定すればよい。しかし、ＭＰＩライブラリを使用するときには、ユーザプログラムは、このコール文の中でこのＭＰＩライブラリにより定められた付加情報を引数としてさらに指定する必要がある。この付加情報は、送信先のプロセスの識別子、プロセスグループ識別子等を含む固定長のデータであり、メッセージの送信先の要素プロセッサにおいて、受信したメッセー

ジがそこで実行中のユーザプロセスが発行する受信要求が要求したメッセージか否かの識別に使用される。以下、この情報をＭＰＩ付加情報とも呼ぶ。従来のＭＰＩライブラリとＰＵＴ／ＧＥＴライブラリを併用した通信方法では、ユーザデータおよびＭＰＩ付加情報を異なる二つのメッセージにより転送していた。

【０００８】

【発明が解決しようとする課題】従来のように、同じ転送先プロセッサに対しユーザデータおよびＭＰＩ付加情報を２つのメッセージとして転送する場合、各メッセージの転送に異なる転送制御情報が必要になる。その結果、これらの情報の生成も２度行わなければならない。このために、従来のＭＰＩライブラリを使用したデータ転送では、ユーザプログラムがデータの転送を要求してから、実際に転送が開始されるまでの遅延時間（転送レイテンシと呼ばれる）が大きい。

【０００９】さらに、これらの２種のデータに対して別々のメッセージとして送信処理、受信処理を行うと、メッセージ数に比例してローカルメモリに対するアクセス回数（転送制御情報の読み出し、転送データの読み出し、フラグの書き込み）が増加してしまう。

【００１０】本発明の目的は、以上の問題を減少させ、より高速にデータを転送できるデータ転送方法を提供することにある。

【００１１】

【発明を解決するための手段】上記の目的を達成するために、本発明によるデータ転送方法は、送信側の要素プロセッサで走行中のユーザプロセスから発行されたデータ送信要求が要求する、送信データとそれに関連する付加情報とを、送信元のメッセージパッシングライブラリから送信元のメモリ間直接転送ライブラリに通知し、そのメモリ間直接転送ライブラリにより、上記送信データおよび付加情報と、それらの受信の完了を示す制御情報を書き込むべき領域を指定するための、受信側の要素プロセッサがあらかじめ決定した受信側アドレス情報とを含むメッセージの送信をネットワークインタフェース回路に対して要求し、上記ネットワークインタフェース回路により、上記メッセージを組立て、受信側の要素プロセッサに宛てて上記相互結合ネットワークに送信し、上記メッセージの送信後に、そのメモリ間直接転送ライブラリが決定したアドレスを有する、上記メモリ内の記憶位置に送信完了を示す制御情報を上記ネットワークインタフェース回路により書き込む。

【００１２】より具体的には、上記メモリ間直接転送ライブラリは、上記付加情報を上記メモリ内の領域に書き込み、ユーザが指定した送信データの記憶位置を示す第１のアドレスとそのデータの長さと、上記付加情報の記憶位置を示す第２のアドレスと、送信先の要素プロセッサのメモリにおける、送信データの記憶位置を指定する第３のアドレスと、上記付加情報を記憶する位置を示す第４のアドレス等を指定し、これらの情報を含むメッセージの送信を上記ネットワークインタフェース回路に要求する。

【００１３】この回路は、上記第１のアドレスと上記データ長により送信データを読み出し、上記第２のアドレスにより付加情報を読み出し、これらの送信データ、付加情報および上記第３，第４のアドレスを含むメッセージを生成し、送信先にあてて送信する。この送信の完了後に、上記第２のアドレスを使用して、送信完了を示す制御情報をメモリ内の、上記付加情報の書き込み位置と異なる位置、具体的には、付加情報が書き込まれた記憶位置の次の記憶位置に書き込む。

【００１４】さらに、受信側の要素プロセッサにおいても、ネットワークインタフェース回路が、上記メッセージ内の上記送信データと付加情報とをそれぞれ上記第３、第４のメモリアドレスが指定する記憶位置に書き込むとともに、この書き込みの完了後に、受信完了を示す制御情報を上記第４のアドレスの基づいて上記付加情報の書き込み位置と異なる記憶位置に書き込む。

【００１５】本発明のより具体的な態様では、ユーザプロセスが、メッセージパッシングライブラリを介して行う上記の転送とともに、他のユーザプロセスはメッセージパッシングライブラリを介さないでメモリ間直接転送ライブラリにデータの送信要求を発行することができるようになっている。この場合には、メモリ間直接転送ライブラリとネットワークインタフェース回路は、上に述べた処理における、付加情報が存在しない場合の処理と基本的に同じ処理を行う。

【００１６】

【発明の実施の形態】以下、本発明に係る計算機システムを図面に示した実施の形態を参照してさらに詳細に説明する。

【００１７】＜発明の実施の形態＞図１に、本発明における並列計算機の概略構成を示す。図中、１０１～１０４は並列計算機を構成する要素プロセッサ、１０５は相互結合ネットワークである。要素プロセッサ１０１～１０４は、相互結合ネットワーク１０５に接続し、相互結合ネットワーク１０５を介して、要素プロセッサ間でデータの授受を行う。相互結合ネットワーク１０５の構成方法（トポロジ）は、クロスバ結合、格子結合、リング結合、多段結合等多種存在するが、本発明は、これらのいずれにも適用可能であり、特定の相互結合ネットワークトポロジに限定されない。図３に要素プロセッサ１０１の概略構成を示す。図中、３０１は命令プロセッサ、３０２はキャッシュローカルメモリ、３０３はストレージコントローラ、３０４はローカルメモリ、３０５はネットワークインタフェース回路、３０６はＩ／Ｏインタフェース回路である。この並列計算機は、個々の要素プロセッサがそれぞれ固有のローカルメモリ３０４を有する分散ローカルメモリ型の並列計算機である。

【００１８】各要素プロセッサは、他の要素プロセッサとの間でメッセージパッシングによる通信を実行するように構成されている。すなわち、各要素プロセッサは、標準のメッセージパッシングインタフェース、たとえばＭＰＩを有するメッセージパッシングライブラリ（以下、ＭＰＩライブラリと呼ぶ）と、このライブラリと交信して自要素プロセッサ内のローカルメモリとの間で直接データの授受を行うメモリ間直接転送を実行するライブラリとして、ＰＵＴ／ＧＥＴ型通信を実行するためのライブラリ（以下、ＰＵＴ／ＧＥＴ型ライブラリと呼ぶ）と、ＰＵＴ／ＧＥＴ型ライブラリからのコマンドにより起動されるネットワークインタフェース回路３０５を有している。なお、本発明は、この特定のメッセージパッシングライブラリに限定されるのではなく、他のメッセージパッシングライブラリたとえばＰＶＭ、ＰＡＲＭＡＣＳとして知られているライブラリも適用できる。

【００１９】本実施の形態では、各要素プロセッサ内のユーザプロセスがＭＰＩライブラリに対してデータ送信要求を発行したときに、ＭＰＩライブラリ、ＰＵＴ／ＧＥＴ型ライブラリおよびネットワークインタフェース回路３０５は、協同してユーザデータとＭＰＩ付加情報を一つのメッセージにて転送し、さらに他の要素プロセッサからユーザデータとＭＰＩ付加情報を含むメッセージを受信したときに、ネットワークインタフェース回路３０５は、これらのデータを区分してローカルメモリに書き込むところに特徴がある。この付加情報は、ＭＰＩライブラリを介した通信のために使用されるもので、MPIライブラリが指定した形式の、データ送信に関連する複数の情報からなり、それぞれの情報は、ユーザプロセスにより指定される。具体的には、既に例示したように、この付加情報は、受信側のユーザプロセスの識別子、プロセスグループの識別子等を含む。

【００２０】より具体的には、各要素プロセッサで実行中のユーザプロセスからＭＰＩライブラリに対する送信要求が発行されたときに、ＭＰＩライブラリは、その送信要求が指定するユーザデータと付加情報の送信をＰＵＴ／ＧＥＴ型ライブラリに要求する。ＰＵＴ／ＧＥＴ型ライブラリにその送信要求が指定する付加情報をローカルメモリに書き込み、そのユーザデータとＭＰＩ付加情報の両方を一つのメッセージとして転送するための転送制御情報を生成し、ローカルメモリ３０４に書き込み、その後転送制御情報によるユーザデータおよび付加情報の送信をネットワークインタフェース回路３０５に要求する。

【００２１】ネットワークインタフェース回路３０５は、この送信要求に応答して、転送制御情報に従ってユーザデータとＭＰＩ付加情報を含む一つのメッセージを組み立て、受信側の要素プロセッサに相互結合ネットワーク１０５を介して転送する。受信側の要素プロセッサでは、ネットワークインタフェース回路３０５は、このメッセージを受信すると、メッセージに含まれたユーザデータおよびＭＰＩ付加情報をメッセージのヘッダ内の転送制御情報が指定するローカルメモリ内の二つのアドレスに書き込み、それぞれを受信側のユーザプロセスおよび受信側のＭＰＩライブラリに引き渡す。

【００２２】図２に転送制御情報の例を示す。転送制御情報２００には、ＧＥＴあるいはＰＵＴ動作の場合に使用される転送先プロセッサ番号２０１、ＰＵＴ動作の場合に送信されるユーザデータが格納されているローカルメモリ領域の先頭アドレスである送信データアドレス２０３、ＰＵＴ動作の場合に送信完了フラグを書き込むローカルメモリ領域の先頭アドレスである送信フラグアドレス２０４、ＧＥＴあるいはＰＵＴ動作の場合に使用される、転送されるデータの長さである転送データ長２０５、受信側の要素プロセッサにおいて受信データを書き込むローカルメモリ領域の先頭アドレスである受信データアドレス２０６、その要素プロセッサにおいて、その受信データに対する受信完了フラグを書き込むローカルメモリ領域の先頭アドレスである受信フラグアドレス２０７、その他通信処理に必要な情報２０８等を格納する。

【００２３】さらに、本実施の形態では、モードビット２０２がセットされていない場合、送信フラグアドレス２０４および受信フラグアドレス２０７は、それぞれＰＵＴ動作時の送信完了フラグおよびＧＥＴ動作時の受信完了フラグを書き込むローカルメモリアドレスを指定する。しかし、モードビット２０２がセットされている場合、送信フラグアドレスフィールド２０４は、ＰＵＴ動作時にＭＰＩ付加情報を読み出すべきローカルメモリアドレスを指定するのに使用され、受信フラグアドレスフィールド２０７は、ＧＥＴ動作時に受信したＭＰＩ付加情報を書き込むべきローカルメモリアドレスを指定するのに使用される。

【００２４】この結果、ＰＵＴ動作時に送信完了フラグを書き込むべきローカルメモリアドレスが転送制御情報２００により指定されなくなるが、本実施の形態では、あらかじめセットされたＭＰＩ付加情報サイズを送信フラグアドレスに加算し、その結果得られるアドレスにユーザデータおよびＭＰＩ付加情報という二つのデータの送信完了フラグを書き込む。同様に、ＧＥＴ動作時には、ＧＥＴ動作時には、あらかじめセットされたＭＰＩ付加情報サイズを受信フラグアドレスフィールドに加算し、その結果得られるアドレスに２種類のデータの受信完了フラグを書き込む。これにより、１つの転送制御情報２００を用いてユーザデータとＭＰＩ付加情報という２つの種類のデータを１つのメッセージで送信または受信し、従来と同様に送信完了フラグあるいは受信完了フラグもローカルメモリに書き込むことができる。

【００２５】命令プロセッサ３０１は、プログラム処理を行うユニットである。キャッシュローカルメモリ３０

２は、命令プロセッサ３０１に付随する、高速かつ小容量のローカルメモリである。ネットワークインタフェース回路３０５は、相互結合ネットワーク１０５に接続し、命令プロセッサ３０１からの指示に従って、ローカルメモリ分散型の並列計算機の特徴であるデータ転送処理を命令プロセッサ３０１のプログラム処理とは独立して行うユニットである。ストレージコントローラ３０３は、命令プロセッサ３０１、ネットワークインタフェース回路３０５およびＩ／Ｏインタフェース回路３０６から発行されるデータアクセス要求に従って、適当な記憶媒体にアクセスする。ローカルメモリ３０４は、ストレージコントローラ３０３で制御され、データ等を格納する。命令プロセッサ３０１およびネットワークインタフェース回路３０５は独立に動作するため、ストレージコントローラ３０３は、命令プロセッサ３０１からローカルメモリ３０４へのアクセス要求を処理するのと同時に、ネットワークインタフェース回路３０５からのデータ転送に伴うローカルメモリ３０４へのアクセスも処理する。Ｉ／Ｏインタフェース回路３０６は、ストレージコントローラ３０３からのアクセス要求に従って、Ｉ／Ｏ装置にアクセスする。Ｉ／Ｏインタフェース回路３０６は、要素プロセッサの構成によっては、存在しない場合もある。

【００２６】図４に示すように、ストレージコントローラ３０３は、命令プロセッサインタフェース回路４０１、アドレス解析部４０２、メモリアクセスインタフェース回路４０３およびデータ転送インタフェース回路４０４で構成される。命令プロセッサインタフェース回路４０１は、命令プロセッサ３０１からローカルメモリ３０４へのアクセスおよび命令プロセッサ３０１からネットワークインタフェース回路３０５へのコマンド発行というトランザクションを線４０１Ｓから受け取る。通常、このコマンドは、ネットワークインタフェース回路３０５内部の制御レジスタへのアクセス要求である。命令プロセッサインタフェース回路４０１は、このトランザクションへの返答、ストレージコントローラ３０３やネットワークインタフェース回路３０５で検出した割り込み要因を線４０２Ｓを介して命令プロセッサ３０１へ伝える。このトランザクションの応答は、たとえば、ローカルメモリからの読み出しデータである。

【００２７】アドレス解析部４０２は、命令プロセッサ３０１が発行した、ローカルメモリアクセス要求およびネットワークインタフェース回路３０５へのコマンドを線４０３Ｓを介して受け取り、そのアクセス要求あるいはコマンドが指定するアクセス先アドレスを解析する。ローカルメモリアクセス要求は線４０４Ｓを介してメモリアクセスインタフェース回路４０３に伝えられる。また、ネットワークインタフェース回路３０５へのコマンドは、線４０６Ｓを介してデータ転送インタフェース回路４０４に伝達される。

【００２８】メモリアクセスインタフェース回路４０３は、アドレス解析部４０２からのローカルメモリアクセス要求を線４０４Ｓを介して受け、線４０７Ｓを介してローカルメモリ３０４に伝達する。ローカルメモリアクセス要求がローカルメモリからの読み出し要求であった場合、この読み出し要求が指定するデータがローカルメモリ３０４から線４０８Ｓを介して伝達される。読み出しデータは、メモリアクセスインタフェース回路４０３から線４０９Ｓを介して命令プロセッサインタフェース回路４０１に伝達され、線４０２Ｓを介して命令プロセッサ３０１に伝達される。また、ローカルメモリアクセスインターフェース４０３は、データ転送に関わるローカルメモリアクセスも処理する。データ転送処理に関わるローカルメモリアクセスは、データ転送インタフェース回路４０４から線４１０Ｓを介して伝達される。メモリアクセスインタフェース回路４０３は、アドレス解析部４０２からローカルメモリアクセス要求が伝達された時と同様に、ローカルメモリアクセス要求をローカルメモリ３０４に対して発行し、読み出しアクセスに対しては、読み出しデータを線４１１Ｓを介してデータ転送インタフェース回路４０４に返送する。

【００２９】データ転送インタフェース回路４０４は、アドレス解析部４０２から線４０６Ｓを介して伝達されるネットワークインタフェース回路３０５へのコマンドを受け取り、線４１２Ｓを介してネットワークインタフェース回路３０５に伝達する。ネットワークインタフェース回路３０５からは、線４１３Ｓを介してコマンドに対する返答およびデータ転送に関わるローカルメモリアクセス要求が伝達される。データ転送インタフェース回路４０４は、前記コマンドに対する返答を線４１４Ｓを介して命令プロセッサインタフェース回路４０１に伝達し、線４０２Ｓを介して命令プロセッサ３０１に伝達する。ローカルメモリアクセス要求は、線４１０Ｓを介してメモリアクセスインタフェース回路４０３に伝達する。ローカルメモリ読み出しデータは、線４１１Ｓを介してデータ転送インタフェース回路４０４に伝達され、データ転送インタフェース回路４０４から線４１２Ｓを介してネットワークインタフェース回路３０５に伝達される。データ転送インタフェース回路４０４は、ネットワークインタフェース回路３０５内部で発生した割り込み伝達要求を受ける場合もある。この場合、割り込み伝達要求は線４１４Ｓを介して命令プロセッサインタフェース回路４０１に伝達され、さらに命令プロセッサ３０１に伝達される。

【００３０】ネットワークインタフェース回路３０５は、コマンド受信部４０５、コマンド処理部４０６、メッセージ生成部４０７、メッセージ送信部４０８、メッセージ受信部４０９、メッセージ分解部４１０およびコマンド送信部４１１で構成される。コマンド受信部４０５は、ストレージコントローラ３０３から線４１２Ｓを

介して、ネットワークインタフェース回路３０５内部の制御レジスタへのアクセスあるいはネットワークインタフェース回路３０５が要求したローカルメモリ３０４から読み出されたデータ等を受け取る。このデータは、線４１５Ｓを介してコマンド処理部４０６に伝達され、転送データとして使用されたり、ネットワークインタフェース回路３０５の動作制御用データとしてネットワークインクフェース回路内部の制御レジスタに設定されたり、データ送信時にメッセージ生成用データ（転送先プロセッサ番号、送信データアドレス、送信フラグアドレス、転送データ長、受信データアドレス、受信フラグアドレス等）として使用される。

【００３１】コマンド処理部４０６は、線４１５Ｓを介してコマンド受信部４０５から伝達されるネットワークインタフェース回路３０５内部の制御レジスタへのアクセスを行う。制御レジスタ読み出しアクセスを受けた場合、コマンド処理部４０６は、読み出し結果を線４１７Ｓを介してコマンド送信部４１１に伝達する。また、制御レジスタ書き込みアクセスを受けた場合には、コマンド処理部４０６は、その書き込みを実行する。メッセージの送信処理は、上記の制御レジスタへのアクセスがメッセージ送信起動用レジスタへの書き込み要求である場合に開始する。メッセージの送信処理では、データを宛先の要素プロセッサに転送するのに必要な情報であるヘッダを作成したり、転送データが存在するローカルメモリアドレスを知るために、ローカルメモリ３０４に格納されている転送制御情報２００を読み出すローカルメモリアクセス要求が発生される。このアクセス要求は線４１７Ｓを介してコマンド送信部４１１に伝達され、線４１３Ｓ、データ転送インタフェース回路４０４、線４１０Ｓ、メモリアクセスインタフェース回路４０３、線４０７Ｓを介してローカルメモリ３０４から読み出される。ローカルメモリ３０４からの読み出し結果は、線４０８Ｓ、メモリアクセスインタフェース回路４０３、線４１１Ｓ、データ転送インタフェース回路４０４、線４１２Ｓ、コマンド受信部４０５を介してコマンド処理部４０６に伝達され、メッセージ生成部４０７に伝達される。

【００３２】メッセージ生成部４０７は、本実施の形態での特徴的な回路の一つであり、４１８Ｓを介して伝達された転送データと転送制御情報２００を含むヘッダからメッセージを生成し、線４１９Ｓを介してメッセージ送信部４０８に送出する。転送制御情報２００内のモードビット２０２が１である場合、メッセージ生成部４０７は、転送制御情報２００から生成されるヘッダと、転送制御情報２００内の送信データアドレス２０３に従ってローカルメモリ３０４から読み出した送信データと、送信フラグアドレス２０４に従ってローカルメモリ３０４から読み出したＭＰＩ付加情報からメッセージを組み立て、メッセージ送信部４０８に送出する。

【００３３】図５に、メッセージ生成部４０７の内部構成を示す。メッセージ生成部４０７は、ローカルメモリ３０４から読み出した転送制御情報２００内の、送信データアドレス、送信フラグアドレスフィールド内のアドレス、転送データ長をそれぞれ保持するレジスタ５０１、５０２、５０３を有する。メッセージ生成部４０７は、さらに、ローカルメモリ３０４から生成中のメッセージのためにローカルメモリからすでに読みだされたデータの量を保持するレジスタ５０４と、本実施の形態に特徴的な回路として、ＭＰＩ付加情報のサイズを保持するレジスタ５０５と、転送制御情報２００内モードビットを保持するレジスタ５０６を有する。レジスタ５０４に保持された読み出し済みのデータの量はデータがローカルメモリ３０４から読み込まれるたびにカウントアップされ、送信すべきすべてのデータがメッセージ送信部４０８に伝達された後リセットされる。従って、この読み出し済みのデータの量は、送信済みのデータの総量と考えることができる。ＭＰＩ付加情報のサイズは、ユーザアプロセスからの初期化要求によりＭＰＩライブラリとＰＵＴ／ＧＥＴライブラリが初期化される時にレジスタ５０５にあらかじめセットされる。レジスタ５０６には、転送制御情報２００の読み出し時にその情報内のモードビットがセットされ、すべてのデータがメッセージ送信部４０８に伝達された後このレジスタ内のモードビットがリセットされる。メッセージ生成部４０７は、さらに、ローカルメモリ読み出し要求発行部５１１と、ローカルメモリ書込み要求発行部５１２と、メッセージ組み立て部５１３の他に、本実施の形態に特徴的な回路として、アンドゲート５０７と、加算器５０８および５０９と比較回路５１０とを有する。

【００３４】アンドゲート５０７は、モードビットが１である場合には、レジスタ５０５内のＭＰＩ付加情報サイズを出力し、０である場合にはモードビットの値０を出力する。加算器５０９は、アンドゲート５０７の出力とレジスタ５０２に保持されている送信フラグアドレスフィールドの値を加算する。ＰＵＴ動作時にモードビットが１にセットされている場合、レジスタ５０２に保持されている送信フラグアドレスフィールドには、ローカルメモリ３０４内の、ＭＰＩ付加情報７１４の先頭アドレスが含まれているので、この加算の結果アドレスは、そのＭＰＩ付加情報の次のアドレスを指すことになり、ＰＵＴ動作時の送信完了フラグの書込みアドレスとして使用される。加算器５０８は、アンドゲート５０７の出力と、レジスタ５０３に保持された転送データ長を加算する。この加算の結果は、モードビットに１がセットされた場合、ＰＵＴ動作時に送信すべきメッセージに含まれるべき、ローカルメモリ３０４から読み出すべきデータの総量を示す。

【００３５】比較回路５１０は、レジスタ５０４に保持された読み出し済みのデータ量を、レジスタ５０３に保

持された転送データ量と加算器５０８から出力される転送データの総量とを比較する。比較の結果として、レジスタ５０４に保持された読み出し済みデータ量がレジスタ５０３内の転送データ長を越えていない場合には、そのことを示す比較結果信号をローカルメモリ読み出し要求発行部５１１に出力する。レジスタ５０４に保持された読み出し済みデータ量がレジスタ５０３内の転送データ長を越えているが、加算器５０８から与えられる、ローカルメモリ３０４から読み出すべきデータの総量を越えていない場合には、そのことを示す比較結果信号をローカルメモリ読み出し要求発行部５１１に出力する。レジスタ５０４に保持された読み出し済みデータ量が読み出すべきデータの総量に達した場合、そのことを示す比較結果信号をローカルメモリ書込み要求発行部５１２に出力する。

【００３６】ローカルメモリ読み出し要求発行部５１１は、比較回路５１０からの比較結果信号と、レジスタ５０１内の送信データアドレスおよびレジスタ５０２内の送信フラグアドレスとからローカルメモリ読み出し要求を生成し、コマンド送信部４１１に送信する。すなわち、比較結果信号が、読み出し済みのデータが転送データ長を超えていないことを示すときには、送信データアドレスを元に後続の未読み出しのユーザデータを読み出すためのローカルメモリ読み出し要求を生成し、読み出し済みのデータが転送データ長を越えているが、読み出すべきデータの総量を超えていないときには、送信フラグアドレスを元にして未読み出しのＭＰＩ付加情報を読み出すためのローカルメモリ読み出し要求を生成する。

【００３７】ローカルメモリ書込み要求発行部５１２は、比較回路５１０からの比較結果信号と、加算器５０９から出力される送信完了フラグアドレスとを元にローカルメモリ書込み要求をコマンド送信部４１１に送信する。すなわち、比較結果信号が、読み出し済みデータ量が読み出すべきデータの総量に達したことを示す場合、加算器５０８より与えられる送信完了フラグアドレスに送信完了フラグを書き込むことを要求する書き込み要求を生成する。

【００３８】メッセージ組み立て部５１３では、本実施の形態に特徴的なセレクタ５１４がモードビットの値に従ってメッセージを組み立て、メッセージ送信部４０８にそのメッセージの送信要求を送付する。モードビットが１である場合、ヘッダとデータとＭＰＩ付加情報を含むメッセージ５１５を組立て、モードビットが０である場合には、ヘッダとデータのみからなるメッセージ５１６を組立る。

【００３９】図４において、メッセージ送信部４０８は、メッセージ生成部４０７からのメッセージ送信要求を受けて線４２０Ｓを介して相互結合ネットワーク１０５にメッセージを送出する。送出されたメッセージは、相互結合ネットワーク１０５を介してそのヘッダ情報に従って宛先に転送される。メッセージ生成部４０７における送信処理は、ネットワークインタフェース回路３０５へ送信される複数のメッセージ送信要求に対してそれらの送信要求の到着順に順次行われる。

【００４０】次に、メッセージ受信部４０９について説明する。メッセージ受信部４０９は、線４２１Ｓを介して、相互結合ネットワーク１０５からメッセージを順次受け取り、メッセージ分解部４１０に線４２２Ｓを介して転送する。メッセージ分解部４１０はこのメッセージのヘッダ部に含まれる転送制御情報に従ってこのメッセージをデータ部とヘッダ部に分解し、ローカルメモリ３０４へこのデータ部や受信完了フラグの書き込みを要求する書き込み要求を線４２３Ｓを介してコマンド送信部４１１に伝達する。

【００４１】図６にメッセージ分解部４１０の内部構成を示す。メッセージ分解部４１０には、メッセージヘッダ内の受信データアドレス、受信フラグアドレス、転送データ長、モードビットをそれぞれ保持するレジスタ６０１、６０２、６０３、６０６が設けられている。レジスタ６０６は本実施の形態で特徴的なレジスタであり、レジスタ６０６には、ヘッダ受信時にヘッダ内のモードビットがセットされ、メッセージ内の全データを受信したときにそのモードビットがリセットされる。さらに、受信したメッセージ内のデータの内、ローカルメモリに書き込み済みのデータの総量を保持するレジスタ６０４と、ＭＰＩ付加情報のサイズを保持する、本実施の形態に特徴的なレジスタ６０５が設けられている。レジスタ６０５には、ＭＰＩ初期化時あるいはジョブ起動時にあらかじめ定められたＭＰＩ付加情報サイズがセットされ、このサイズ情報はメッセージ内の全データが受信されたときにリセットされる。レジスタ６０４に保持された書き込み済みのデータの量は受信されたデータがローカルメモリ３０４に書き込まれるたびにカウントアップされ、受信すべきすべてのデータがローカルメモリ３０４に書き込まれた後リセットされる。従って、この書き込み済みのデータの総量は、受信済みのデータの総量であるとも考えることができる。

【００４２】メッセージ分解部４１０には、ローカルメモリ書込み要求発行部６１１の他に、本実施の形態で特徴的な、アンドゲート６０７、加算器６０８、６０９と比較回路６１０とがさらに設けられている。アンドゲート６０７は、レジスタ６０６内のモードビットが１である場合にレジスタ６０５内のＭＰＩ付加情報サイズを出力し、モードビットが０である場合には０を出力する。加算器６０８および６０９の動作は、加算器５０８および５０９（図５）と同様である。比較回路６１０は、レジスタ５０４内の受信済みデータの総量を、転送データ長、アンドゲート６０７の出力結果データと比較し、比較結果信号をローカルメモリ書込み要求発行部６１１に出力する。すなわち、この比較回路は、受信済みのデー

タが転送データ長よりも短いか、受信済みのデータの総量が転送データ長より大きいが、転送データ長とアンドゲート607から出力されるMPI付加情報のサイズの和より小さいか、あるいは受信済みのデータの総量が転送データ長とMPI付加情報サイズの和より大きいかを判別する。

【0043】メモリ書込み要求発行部611は、比較回路610の比較結果信号と、レジスタ601内の受信データアドレスと、加算器609の出力とレジスタ602内の受信フラグアドレスフィールドの値とから、ローカルメモリ書込み要求を生成し、コマンド送信部411に線42Sを介して伝達する。加算器609は、モードビットが1である場合に、レジスタ602内の受信フラグアドレスフィールドの値にMPI付加情報サイズを加算したアドレスを受信完了フラグを書き込むべきローカルメモリアドレスとして出力する。

【0044】比較回路610の出力が、受信済みのデータの総量が転送データ長よりも大きいことを示す場合、メモリ書き込み要求発行部611は、受信されたデータをそれまでに受信したデータの書き込み位置に続けて書き込むためのローカルメモリ書き込み要求をレジスタ601内の受信データアドレスに基づいて生成する。

【0045】比較回路610の出力が、受信済みのデータの総量が転送データ長よりも大きく転送データ長とアンドゲート607から出力されるMPI付加情報のサイズの和より小さいことを示す場合、新たに受信されたデータは、MPI付加情報である。従って、この場合には、メモリ書き込み要求発行部611は、受信されたデータをそれまでに受信したMPI付加情報の書き込み位置に続けて書き込むためのローカルメモリ書き込み要求を、レジスタ602内の受信フラグアドレスフィールドに含まれる、MPI付加情報の書き込みアドレスに基づいて生成する。

【0046】比較回路610の出力が、受信済みのデータの総量が転送データ長とアンドゲート607から出力されるMPI付加情報のサイズの和より大きいことを示す場合、すべてのデータが受信されたことになる。従って、この場合には、メモリ書き込み要求発行部611は、受信完了フラグをMPI付加情報の書き込み位置に続けて書き込むためのローカルメモリ書き込み要求を、加算器609の出力に基づいて生成する。

【0047】コマンド送信部411は、これらの書き込み要求に従って、受信されたデータあるいは受信完了フラグをローカルメモリ304に書き込む。コマンド送信部411は、線417Sを介して伝達される、ネットワークインタフェース回路305内部の制御レジスタからの読み出しデータ、線424Sを介して伝達される、メッセージ送信処理において使用される送信データのローカルメモリ304からの読み出し要求、線424Sを介して伝達される、メッセージ送信処理の完了に伴う送信完了フラグのローカルメモリ304への書き込み要求、線423Sを介して伝達される、メッセージ受信処理に伴う受信データあるいは受信完了フラグのローカルメモリ304への書き込み要求およびネットワークインタフェース回路305内で発生した割り込み要求を、線413Sを介してストレージコントローラ303内のデータ転送インタフェース回路404に伝達する。また、コマンド送信部411は、ネットワークインタフェース回路305の動作制御に関わる情報のローカルメモリ304からの読み出し要求、メッセージ送信処理に使用する転送制御情報200のローカルメモリ304からの読み出し要求を線413Sを介してデータ転送インタフェース回路404に伝達する。

【0048】次に本実施の形態におけるメッセージ転送の流れを説明する。最初に自プロセッサのローカルメモリに格納されているユーザデータおよびMPI付加情報を転送先の要素プロセッサのローカルメモリに直接書き込むPUT処理について図4、5、7を用いて説明する。まず、本実施の形態に係わる通信方式では、送信側のユーザプロセスおよび受信側のユーザプロセスは、MPIライブラリを使用する前に、MPIライブラリ内の初期化ルーチンたとえばMPI_initをコールするコマンド発行する。この初期化ルーチンのコールを受けると、MPIライブラリは、PUT／GETライブラリ内のいくつかの通信準備手続きをコールする。これらの通信準備手続きは、使用するPUT／GETライブラリにより予め定められているが、以下では、後の説明に関連する部分および本実施の形態で新規に行われる処理のみを説明する。

【0049】本実施の形態では、送信側のPUT／GETライブラリおよび受信側のPUT／GETライブラリは、いずれもこれらの通信準備手続において以下の処理をすると仮定する。すなわち、ローカルメモリ701をユーザ空間にあらかじめマップし、さらに、通信領域703、704（図7）を確保する。さらに、それぞれの通信領域内に送信データ領域およびそれに対応する送信完了フラグ領域を確保する。図7では、713は送信データ領域の例を示す。図では、受信側の通信領域704内の送信データ領域は図示していない。本実施の形態では、送信完了フラグ領域として、MPI付加情報714および送信完了フラグ715の両方を格納する連続した領域を確保する点で従来と異なる。同様に、各通信領域内に、受信データ領域および受信完了フラグ領域を確保する。本実施の形態では図7では、721は受信データ領域の例を示す。図では、送信側の通信領域703内の受信データ領域は図示していない。本実施の形態では、受信完了フラグ領域として、MPI付加情報722および受信完了フラグ723の両方を格納する連続した領域を確保する点で従来と異なる。なお、MPI付加情報714の長さは例えば64バイト程度である。なお、送信完

了フラグあるいは受信完了フラグを書き込む領域７１５，７２３は例えば８バイトである。

【００５０】ＰＵＴ／ＧＥＴ型ライブラリとして、送信すべきユーザデータを、ローカルメモリ７０１に常駐させることを前提とする場合とそうでない場合とがある。前者の場合には、送信データ領域７１３は、この常駐されたユーザデータの領域と一致するように、送信データ領域７１３が決定される。一方、後者の場合には、送信データ領域７１３は、送信側のユーザプロセスが使用するユーザデータに割り当てられたローカルメモリ内の領域とは独立に決定される。本発明はいずれの構造のＰＵＴ／ＧＥＴ型ライブラリにも適用可能である。しかし、後者の場合には、後に述べるように、ユーザプロセスが使用しているユーザデータに割り当てられたローカルメモリ内の領域のデータを送信データ領域７１３にコピーする処理が必要となる。しかし、前者の場合にはこのコピー動作が必要でなく、それだけデータ転送動作が高速化される。

【００５１】以上のようにして、通信準備手続きが実行された後に、送信側のユーザプロセスの処理が進むと、そのユーザプロセスは、送信すべきデータを送信側の通信領域７０３内のユーザデータ領域７１３に書き込んだ後に、データ送信要求コマンド、たとえばＭＰＩ_ｓｅｎｄを送信側のＭＰＩライブラリに対して発行する。このコマンドの名称は、使用するメッセージパッシングライブラリにより定まり、それが指定する引数も同様にそのライブラリにより定められた複数の種類の情報からなる。ここで仮定するＭＰＩライブラリの場合には、このコマンドの引数は、送信すべきユーザデータの先頭アドレス、ユーザデータ長と付加情報からなり、この先頭アドレスは、ユーザプロセスに割り当てられた仮想メモリ空間内での、そのユーザデータに対する仮想アドレスである。この付加情報は、受信側のユーザプロセスの識別子、プロセスグループの識別子等を含む。最初の二つの引数は、ＭＰＩライブラリを介さないでデータ転送をユーザプロセスがＰＵＴ／ＧＥＴ型ライブラリに直接要求するためのデータ送信要求が指定する引数と同じであり、付加情報がＭＰＩライブラリに対するデータ転送要求が新たに指定する引数である。

【００５２】送信側のＭＰＩライブラリは、このデータ送信要求コマンドＭＰＩ_ｓｅｎｄに応答して、送信側のＰＵＴ／ＧＥＴ型ライブラリに、この送信要求が指定するデータの送信を要求する。送信側のＭＰＩライブラリは、この要求を、ＭＰＩライブラリとＰＵＴ／ＧＥＴ型ライブラリにより予め定められた一つまたは複数のコマンドの形で発行する。以下では、それらのコマンドの内、本実施の形態で使用すると仮定する主なコマンドのみを説明する。

【００５３】まず、送信側のＭＰＩライブラリは、送信権の取得を要求するコマンドを発行する。送信側のＰＵＴ／ＧＥＴ型ライブラリは、このコマンドに応答して、受信側のユーザプロセスと交信してそのプロセスに対するデータの送信権を得る。受信側のユーザプロセスおよびそのプロセスを実行している要素プロセッサの番号は、上記データ送信要求コマンドＭＰＩ_ｓｅｎｄが指定する付加情報中の、受信側のプロセス識別番号とプロセスグループ識別番号とにより決定される。

【００５４】送信側のＭＰＩライブラリは、さらに、受信側のユーザプロセスの受信データ領域および受信フラグ領域のそれぞれの先頭位置を示す受信データアドレスおよび受信フラグアドレスを受信する。但し、この後に再度同じデータ送信要求コマンドＭＰＩ_ｓｅｎｄを送信側のユーザプロセスが発行したときには、このコマンドを実行する必要はない。

【００５５】既に述べたように、ＰＵＴ／ＧＥＴ型ライブラリが、送信すべきユーザデータがローカルメモリ７０１に常駐されることを前提としない場合には、送信側のＭＰＩライブラリは、送信すべきユーザデータを、ユーザデータ領域７１３にコピーすることを要求するコマンドを発行し、送信側のＰＵＴ／ＧＥＴ型ライブラリにより、このコマンドの引数で指定されるユーザデータのアドレスとデータ長で指定されるユーザデータに割り当てられた、ローカルメモリ内の領域のデータを、先に決定された送信データ領域７０３にコピーする。ＰＵＴ／ＧＥＴ型ライブラリが、送信すべきユーザデータをローカルメモリ７０１に常駐させることを前提とする場合には、このコピー動作は不要である。

【００５６】次に、送信側のＭＰＩライブラリは、付加情報のローカルメモリへの書き込みを要求するコマンドを発行する。送信側のＰＵＴ／ＧＥＴ型ライブラリは、このコマンドに応答して、このコマンドの引数で指定されるＭＰＩ付加情報を、先に決定された送信データ領域７０３に対応して決定された送信フラグ領域の先頭の領域７１４に書き込む。

【００５７】送信側のＭＰＩライブラリは、付加情報のローカルメモリへの書き込みを要求するコマンドを発行する。送信側のＰＵＴ／ＧＥＴ型ライブラリは、このコマンドに応答して、このコマンドの引数で指定されるＭＰＩ付加情報を、先に決定された送信データ領域７０３に対応して決定された送信フラグ領域の先頭の領域７１４に書き込む。

【００５８】送信側のＭＰＩライブラリは、転送制御情報を生成することを要求するコマンドを発行する。送信側のＰＵＴ／ＧＥＴ型ライブラリは、このコマンドに応答して、転送制御情報７００を生成して、ローカルメモリ７００内の適当な領域に書き込む。この転送制御情報７００に含まれた情報は以下の通りである。転送先プロセッサ番号２０１は、受信側のユーザプロセスが実行されているプロセッサの番号であり、この番号は、すでに述べたように、ＭＰＩ初期化ルーチンにおいて決定され

ている。モードビット２０２は、送信フラグアドレスフィールド２０４を本実施の形態に従って拡張して使用するか否かを示すビットである。送信側のＰＵＴ／ＧＥＴ型ライブラリは、送信側のＭＰＩライブラリから、データ転送を要求されたときに、ＰＵＴ／ＧＥＴ型ライブラリはこのモードビットを１にセットする。モードビット２０２が１であることは、ユーザデータとＭＰＩ付加情報とを一つのメッセージで送信することを指示する。なお、本実施例の形態では、送信側のユーザプロセスが、ＭＰＩライブラリに対してでなく、ＰＵＴ／ＧＥＴ型ライブラリに対して直接データ送信要求コマンドを発行した場合には、ＰＵＴ／ＧＥＴ型ライブラリは、そのデータ要求に対して、図7に示す転送制御情報７００と同じ構造を有し、モードビット２０２の値が０である転送制御情報を生成する。

【００５９】送信データアドレス２０３は、送信側のユーザプロセスにより転送が要求されたユーザデータまたはそのコピーを保持する送信データ領域７１３の先頭アドレスである。転送データ長２０５は、データ送信要求コマンドＭＰＩ_ｓｅｎｄが指定した、ユーザデータのデータ長であり、そのコマンドを受けたＭＰＩライブラリがＰＵＴ／ＧＥＴ型ライブラリに通知する。

【００６０】送信フラグアドレスフィールド２０４には、従来では送信完了フラグを書き込むためのローカルメモリアドレスが格納されるが、本実施の形態では、モードビット２０２が１である場合には、ローカルメモリ７０１内のＭＰＩ付加情報７１４の先頭アドレスをこの送信フラグアドレスフィールド２０４に格納する。なお、モードビット２０２が１の場合には、送信完了フラグを格納するローカルメモリ内の領域７１５のアドレスは送信フラグアドレスフィールド２０４によっては明には指定されないことになる。本実施の形態では、ネットワークインタフェース回路３０５が、ユーザプロセスとＭＰＩ付加情報とに対する共通の送信完了フラグを、ローカルメモリ７０１内のＭＰＩ付加情報７１４の最終のアドレスの次のアドレスの領域７１５に格納するようになっている。このため、送信型のＰＵＴ／ＧＥＴ型ライブラリは、ユーザプロセスからの先のデータ送信要求コマンドに対する応答として、この送信完了フラグが書き込まれた時点で、送信完了を送信側のユーザプロセスに通知するようになっている。なお、モードビット２０２が０である場合には、ＰＵＴ／ＧＥＴ型ライブラリは、送信フラグアドレスフィールド２０４に、そのライブラリが決定した送信完了フラグを書き込むアドレスをセットする。

【００６１】受信データアドレス２０６は、受信側の要素プロセッサにおいて、受信したデータを格納するためのローカルメモリ領域７２１（図7）のアドレスである。受信フラグアドレスフィールド２０７には、この受信データとともに受信した付加情報と受信完了を示す受信完了フラグを格納する領域の先頭アドレスである。本実施の形態では、付加情報を記憶する領域７２２の後続の領域７２３に受信完了フラグを書き込む。したがって、受信フラグアドレスフィールド２０７には、この領域７２２の先頭アドレスが書き込まれる。受信データアドレス２０６と受信フラグアドレス２０７は、いずれもＭＰＩ初期化ルーチンにて送信先のＰＵＴ／ＧＥＴライブラリにより通知される。これらのアドレスは、ＧＥＴ動作時に送信側のＰＵＴ／ＧＥＴ型ライブラリにより使用される。

【００６２】その他制御情報２０８は、ＰＵＴメッセージあるいはＧＥＴメッセージあるいはＧＥＴ要求メッセージの種別を示したり、１対１通信あるいは１対多通信等の通信形態を示すといった、その他通信処理に必要な情報を含む。

【００６３】なお、転送制御情報２００の中に、ＭＰＩ付加情報を読み出すべきローカルメモリアドレスを指定するフィールドおよび受信したＭＰＩ付加情報を書き込むべきローカルメモリアドレスを指定するフィールドを別に設けることも可能であるが、本実施の形態のようにモードビット２０２を用いて、送信フラグアドレス２０４および受信フラグアドレス２０７が指定する二つのアドレスを切り替えることにより転送制御情報２００の構造と大きさを、ユーザプロセスがＰＵＴ／ＧＥＴ型ライブラリのみを使用して送信あるいは受信する場合と同じとすることができる。

【００６４】こうして、転送制御情報７００が生成されると、送信側のＭＰＩライブラリは、生成された転送制御情報に従って、データの送信を行うことをネットワークインタフェース回路３０５に要求するコマンドを発行する。ＰＵＴ／ＧＥＴライブラリは、このコマンドに応答して、ネットワークインタフェース回路３０５内部の送信起動用レジスタ（図示せず）にその転送制御情報７００のアドレスを書き込むことを要求する書き込みコマンドを発行する。コマンド処理部４０６は、この書き込みコマンドを実行して、ネットワークインタフェース回路３０５内のメッセージ送信起動用レジスタ（図示せず）へ転送制御情報７００の先頭アドレスを書き込む。この書き込みによりネットワークインタフェース回路３０５はメッセージ送信処理を開始する。

【００６５】図4に示すように、コマンド処理部４０６は、メッセージのヘッダを作成するためにローカルメモリ３０４に格納されている転送制御情報７００を読み出すローカルメモリアクセス要求を線４１７Ｓを介してコマンド送信部４１１に伝達する。線４１３Ｓ、データ転送インタフェース回路４０４、線４１０Ｓ、メモリアクセスインタフェース回路４０３を介して、送信起動用レジスタ（図示せず）に書き込まれたアドレスをもとにローカルメモリ３０４から読み出された転送制御情報７００は、メモリアクセスインタフェース回路４０３、線４

１１Ｓ、データ転送インタフェース回路４０４、線４１２Ｓ、コマンド受信部４０５を介してコマンド処理部４０６に伝達され、メッセージ生成部４０７に伝達される。

【００６６】図５において、メッセージ生成部４０７は、転送制御情報７００内の送信データアドレス２０３、送信フラグアドレスフィールド２０４の値、転送データ長２０５、モードビット２０２をそれぞれレジスタ５０１、５０２、５０３、５０６にセットする。その他の情報は図示しないレジスタに保持される。レジスタ５０６内のモードビットは今の場合には１である。レジスタ５０６内のモードビットが１であるため、アンドゲート５０７がＯＮになり、レジスタ５０５内のＭＰＩ付加情報サイズを出力する。加算器５０８は、レジスタ５０３内の転送データ長とアンドゲート５０７から出力されたＭＰＩ付加情報を足しあわせ、ローカルメモリ７０１から読み出すべきデータの総量を出力する。比較回路５１０がレジスタ５０４に保持された、読み出し済みデータの総量がレジスタ５０３内の転送データ長よりも小さいと判断した場合には、ローカルメモリ７０１内の、レジスタ５０１に保持された送信データアドレスの記憶位置からユーザデータ７１３を読み出すために、コマンド送信部４１１にローカルメモリ読み出し要求を伝達する。このユーザデータは、コマンド送信部４１１からストレージコントローラ３０３内のメモリアクセスインタフェース回路４０３を介してローカルメモリ３０４から読み出される。読み出されたデータは、ストレージコントローラ３０３内のメモリアクセスインタフェース回路４０３、データ転送インタフェース回路４０４およびコマンド受信部４０５を介してコマンド処理部４０６に伝達され、線４１８Ｓを介してメッセージ生成部４０７に伝達される。

【００６７】メッセージ生成部４０７では、比較回路５１０が、レジスタ５０４内の読み出し済みのデータの量が、レジスタ５０３内の転送データ長より大きいが、（転送データ長＋レジスタ５０５内のＭＰＩ付加情報サイズ）以下であると判断したときには、メモリ読み出し要求発行部５１１は、ローカルメモリ７０１内の、レジスタ５０２に保持された送信フラグアドレスを有する記憶位置からＭＰＩ付加情報７１４を読み出すためのローカルメモリ読み出し要求をコマンド送信部４１１に伝達する。この読み出し要求は、コマンド送信部４１１からストレージコントローラ３０３内の、データ転送インタフェース回路４０４およびメモリアクセスインタフェース回路４０３を介してローカルメモリ３０４に送られ、ＭＰＩ付加情報７１４がそこから読み出される。読み出されたＭＰＩ付加情報７１４は、ストレージコントローラ３０３内のメモリアクセスインタフェース回路４０３、データ転送インタフェース回路４０４およびコマンド受信部４０５を介してコマンド処理部４０６に伝達され、線４１８Ｓを介してメッセージ生成部４０７に伝達される。

【００６８】メッセージ生成部４０７では、メッセージ組み立て部５１３内のセレクタ５１４は、モードビットが１であることから、ローカルメモリ３０４から読み出された転送制御情報７００内の送信データアドレス、送信フラグアドレス以外の部分をレジスタ５１５内のヘッダ部の格納する。同様に、ユーザデータ７１３およびＭＰＩ付加情報７１４をレジスタ５１５のデータ部に格納する。

【００６９】比較回路６１０が、読み出したユーザデータの総量が、転送データ長とＭＰＩ付加情報サイズの和に等しくなったことを検出したとき、メモリ書き込み要求発行部５１２は、送信完了フラグを書き込むことを要求するローカルメモリ書込み要求をコマンド送信部４１１に伝達する。このコマンドは、加算器５０９により与えられる、レジスタ６０２内の送信フラグアドレスフィールドに保持されたＭＰＩ付加情報の先頭アドレスと、レジスタ５０５内のＭＰＩ付加情報サイズとの和に等しいアドレスにこのフラグを書き込むことを要求する。送信完了フラグ７１５は、コマンド送信部４１１からストレージコントローラ３０３内のメモリアクセスインタフェース回路４０３を介してローカルメモリ３０４に書き込まれる。本実施の形態では、メッセージに含まれるべきユーザデータとＭＰＩ付加情報の読み出しが完了した時点で、メッセージの送信が完了したと見なして、送信完了フラグ７１５を書き込む。しかし、このメッセージが実際に相互結合ネットワーク１０５に送信された時点でこのフラグを書き込むようにしてもよい。

【００７０】こうして、レジスタ５１５内にユーザデータとそれに関連するＭＰＩ付加情報を含む一つのメッセージ７０５が生成される。このメッセージ７０５には、転送制御情報７００に含まれていたのと同じ転送先プロセッサ番号７１６、モードビット７１７、転送データ長７１８、受信データアドレス７１９、受信フラグアドレス７２０、その他の制御情報２０８をそのまま含み、転送制御情報７００に含まれていた送信データアドレス２０３と送信フラグアドレス２０４に代えて、ユーザデータ７２６、付加情報７２７を含むことになる。送信データアドレスメッセージ生成部４０７はそのメッセージ７０５をメッセージ送信部４０８に送信する。メッセージ生成部４０７は、メッセージ７０５の生成に使用される上記３つの情報をメッセージ送信部４０８へ送出し終ると、メッセージ送信部４０８はレジスタ５０４と５０６をリセットする。メッセージ送信部４０８はそのメッセージ７０５を相互結合ネットワーク１０５に送出する。相互結合ネットワーク１０５はそのメッセージ内の転送先プロセッサ番号２０１により指定されるプロセッサにそのメッセージを転送する。

【００７１】次に受信処理について説明する。相互結合

ネットワーク１０５から転送されたメッセージ７０５は、まずメッセージ受信部４０９で受け取られ、メッセージ分解部４１０に転送される。メッセージ分解部４１０は、メッセージヘッダ内の受信データアドレス７１９、受信フラグアドレス７２０、転送データ長７１８、モードビット７１７をそれぞれレジスタ６０１、６０２、６０３、６０６（図６）に書き込む。レジスタ６０４はあらかじめ０にリセットされ、レジスタ６０５にはあらかじめＭＰＩ付加情報サイズがセットされている。メッセージ分解部４１０では、メモリ書き込み要求発行部６１１は、受信されたユーザデータをヘッダ内の受信データアドレス７１９が示すローカルメモリ領域に書き込むためのローカルメモリアクセス要求を生成し、線４２３Ｓを介してコマンド送信部４１１に伝達する。受信されたユーザデータは、メッセージヘッダ内の受信データアドレス７１９に従って、ストレージコントローラ３０３内のメモリアクセスインタフェース回路４０３を介して、ローカルメモリ３０４の通信領域７０４内の領域７２１に書き込まれる。レジスタ６０６内のモードビットは１であるため、アンドゲート６０７がＯＮになり、レジスタ６０５内のＭＰＩ付加情報サイズを加算器６０８に供給する。加算器６０８では、レジスタ６０３内の転送データ長とアンドゲート６０７から与えられるＭＰＩ付加情報サイズを足しあわせ、ローカルメモリに書込むべきデータの総量を得る。レジスタ６０４内の受信データ量は、メッセージ受信部４０９が相互結合ネットワーク１０５からメッセージ内のデータの異なる部分を受信するごとに更新される。

【００７２】メッセージ分解部４１０では、ユーザデータの異なる部分がメッセージ受信部４０９により受信されるごとに、比較回路６１０が、レジスタ６０４内の受信されたデータの総量がレジスタ６０３内の転送データ長よりも小さいか否かを判断し、前者が後者より小さいと判断したときには、メモリ書き込み要求発行部６１１は、レジスタ６０１に保持された受信データアドレスにしたがって受信されたデータを受信側のローカルメモリ７０２に書き込むことを要求するコマンド送信部４１１にローカルメモリ書込み要求を伝達する。

【００７３】その後比較回路６１０が、レジスタ６０４内の、受信されたデータの総量がレジスタ６０３内の転送データ長よりも大きいが、転送データ長とレジスタ６０５内のＭＰＩ付加情報サイズの和以下と判断したならば、メモリ書き込み要求発行部６１１は、レジスタ６０２に保持された受信フラグアドレスフィールドの値のアドレスに、受信されたＭＰＩ付加情報を書き込むことを要求するローカルメモリ書込み要求を伝達する。

【００７４】加算器６０９は、レジスタ６０２内の受信フラグアドレスフィールドの値とアンドゲート６０７から与えられるＭＰＩ付加情報サイズを加算し、受信完了フラグを書き込むべきメモリアドレスを決定する。比較回路６１０が、レジスタ６０４内の受信データ数が、加算器６０８より与えられる、転送データ長＋ＭＰＩ付加情報サイズに等しくなったことを検出すると、メモリ書き込み要求発行部６１１は、加算器６０９により与えられるアドレスに受信完了フラグを書き込むことを要求するローカルメモリ書き込み要求をコマンド送信部４１１に伝達する。その結果、受信完了フラグ７２３は、コマンド送信部４１１から、ストレージコントローラ３０３内のメモリアクセスインタフェース回路４０３を介してローカルメモリ３０４に書き込まれる。こうして、データ受信処理が完了する。また、受信の完了でもってデータ転送処理が終了する。

【００７５】なお、受信されたデータおよび付加情報は、以下のようにして受信側の要素プロセッサ７０２で使用される。受信側のユーザプロセスが、他の要素プロセッサから送信されたデータの受信を要求するコマンド、例えば、ＭＰＩ＿ｒｅｃｖを発行する。このコマンドは、受信すべきユーザデータを指定するアドレスと、そのデータの最大長、およびＭＰＩライブラリにより定められた付加情報とからなる引数を指定する。この付加情報は、送信元ユーザプロセスの識別子その他の情報からなる。このコマンドで指定される上記アドレスは、受信側のユーザプロセスに割り当てられたアドレス空間に属する仮想アドレスである。受信側のＭＰＩライブラリは、この受信コマンドの引数で指定されるユーザデータがローカルメモリ７０４に書き込み済みであるか否かを、ローカルメモリ７０４に書き込まれたユーザデータ７２１、付加情報７２２、受信完了フラグ７２３に基づいて判別する。もし、この要求されたデータがローカルメモリ７０４に書き込み済みであるときには、受信側のＭＰＩライブラリは、受信側のユーザプロセスに受信完了を通知する。この要求されたデータがローカルメモリ７０４に書き込み済みでないときには、ＭＰＩライブラリは、上記判別が成功するまでその判別を繰り返す。

【００７６】なお、受信側のユーザプロセスが指定した仮想アドレスを有するデータ領域が、ローカルメモリ７０４に常駐していない場合には、受信側のＭＰＩライブラリは、上記通知を行う前に、受信されたユーザデータ７２１を、ユーザプロセスが指定するアドレスに割り当てられたローカルメモリ領域にコピーする。もし、受信側のユーザプロセスが指定した仮想アドレスを有するデータ領域が、ローカルメモリ７０４に常駐している場合には、このコピーは不要である。

【００７７】受信側のユーザプロセスは、この受信完了の通知を受けると、受信データを読みだす命令を実行する。したがって、この受信側でのユーザデータの受信判別処理では、ユーザデータ７２１と付加情報７２２が書き込み済みであるか否かを検出するのに、共通の受信完了フラグ７２３を使用するところが従来と異なる。

【００７８】以上は、ユーザプロセスがデータの送信要

求コマンドをＭＰＩライブラリに対して発行した場合である。本実施の形態では、他のユーザプロセスは、ＰＵＴ／ＧＥＴライブラリに対してデータ送信要求を発行することもできるようになっている。この場合には、このデータ送信要求コマンドは、送信すべきユーザデータを示す仮想アドレスと、データ長を指定する。送信側のＭＰＩライブラリは、このコマンドに応答して、先に述べたと同じようにして、転送制御情報７００を生成する。但し、この情報の中のモードビット２０２の値は0である。さらに、送信フラグアドレスフィールド２０４は、送信完了フラグの書き込み領域７１５のアドレスを指定する。受信フラグアドレスフィールド７２３についても同じである。上記データ送信要求コマンドの場合には、付加情報記憶領域７１４，７２２は不要である。

【００７９】転送制御情報７００内のモードビット２０２が0である場合、転送制御情報７００内のモードビット２０２が1である場合と比べると、メッセージ生成部４０７およびメッセージ分解部４１０の動作が異なる。すなわち、レジスタ５０６内のモードビットが0であるため、加算器５０８の出力は、レジスタ５０３内の転送データ長に等しく、加算器５０９の出力は、レジスタ５０２内の送信完了フラグアドレスに等しい。したがって、メッセージ生成部４０７では、比較回路５１０が、レジスタ５０４内の読み出し済みのデータの総量がアンドゲート５０７より与えられる、レジスタ５０３内の転送データ長に等しくなったことを検出したときに、メモリ読み出し要求発行部５１１は、ローカルメモリ７０１からの送信すべきデータの読み出しを終了する。この読み出されたデータを含むメッセージがすべて相互結合ネットワーク１０５へ送出されると、メッセージ生成部４０７内のメモ書き込み要求発行部５１２は、加算器５０９より与えられる、レジスタ５０２内の送信フラグアドレスに示されるローカルメモリ領域に送信完了フラグを書き込むことを要求する書き込み要求をコマンド送信部４１１に伝達する。

【００８０】受信側の要素プロセッサでは、相互結合ネットワーク１０５から転送されたメッセージ内のヘッダ内のモードビット７１７は0であるため、レジスタ６０６（図６）には0がセットされる。レジスタ６０６内のモードビットが0であるため、加算器６０８の出力はレジスタ６０３内の転送データ長に等しく、加算器６０９の出力は、レジスタ６０２にセットされる受信フラグアドレスに等しい。したがって、比較回路６１０が、レジスタ６０４内の受信済みのデータの総量が加算器６０８から出力される、転送データ長に等しくなったことを検出するまで、メモリ書き込み要求発行部６１１は、受信データのローカルメモリへの書き込みを要求する書き込み要求を発行する。すべての受信データがローカルメモリに書き込まれた後、メモリ書き込み要求発行部６１１は加算器６０９が出力する、レジスタ６０２内の受信フラグアドレスに従ってローカルメモリ３０４に受信完了フラグを書き込むための書き込み要求を発行する。こうして、ＭＰＩライブラリが要求したデータ転送が終了する。

【００８１】なお、比較のために、従来のＰＵＴ処理では、ＭＰＩ_ｓｅｎｄに示すように、送信側のＰＵＴ／ＧＥＴライブラリは、ユーザデータとそれに対する付加情報をそれぞれ転送するための転送制御情報Ａ、Ｂ（８００、８０１）を作成し、通信領域８０２内にユーザデータとそれに対する送信完了フラグＡ、付加情報とそれに対する送信完了フラグＢを記憶するように構成され、送信側のネットワークインタフェース回路はこえらの転送制御情報に基づいて、二つのメッセージＡ、Ｂ（８０４、８０５）を送信する。受信側のネットワークインタフェース回路は、通信領域８０３内にユーザデータおよびそれに対する受信完了フラグ、付加情報とそれに対する受信完了フラグを書き込むように構成される。なお、転送制御情報Ａ、Ｂには、本実施の形態で言うモードビットが存在しない。

【００８２】これに対して、本実施の形態では、ユーザデータおよびその付加情報という異なる２つのデータをローカルメモリ間直接転送にしたがって１回のメッセージ転送で行なえる。したがって、２回のメッセージ転送が必要であった従来よりも転送レイテンシおよびローカルメモリアクセス回数を削減して並列処理効率を向上できる。さらに、ＭＰＩライブラリを介さない従来のメモリ間直接転送も実行できる。

【００８３】さらに、送信側のネットワークインタフェース回路は、送信完了フラグアドレスを付加情報と送信完了を示す制御情報の書き込みの両方に使用しているので、ＰＵＴ／ＧＥＴライブラリがネットワークインタフェース回路に対して指定すべき情報量が少なくて済む。また、受信側においても、ネットワークインタフェース回路は、受信完了アドレスを、付加情報と受信完了を示す制御情報の書き込みの両方に使用しているので、ＰＵＴメッセージに含まれる情報量が少なくて済み、それだけネットワークの混雑を防ぐことができ、さらに、メッセージの、送信元の要素プロセッサでの送信時間、ネットワーク上の転送時間および送信先の要素プロセッサにおける受信時間が短くでき、全体としてメッセージの転送時間が短くなる。

【００８４】以上のように、従来の方式では、異なる２種類のデータ送信処理を行う場合、ネットワークインタフェース回路３０５がローカルメモリを６回アクセス（転送制御情報の読み出し×２と転送データの読み出し×２とフラグの書き込み×２）し、受信処理の場合も送信と同様に、ネットワークインタフェース回路３０５がローカルメモリを４回アクセス（転送データの書き込み×２とフラグの書き込み×２）することになり、本実施の形態に比べて処理オーバーヘッドが大きい。本実施の

形態では、転送制御情報にモードビットを持つことで、転送制御情報の読み出し回数、およびフラグのローカルメモリへの書き込み回数を送信側4回、受信側3回に削減してローカルメモリアクセスに関わる処理オーバヘッドを小さくできる。また、従来は2回にわけて転送していたメッセージを1回で転送できるため、データ転送のレイテンシを削減できる。

【００８５】以上ではＭＰＩライブラリを介したＰＵＴ処理について説明したが、他の通信として、送信元の要素プロセッサが宛先の要素プロセッサのローカルメモリに格納されているユーザデータを取ってくるＧＥＴ処理がある。本実施の形態は、ＭＰＩライブラリを介したＧＥＴ処理にも同様に適用できる。すなわち、ＭＰＩライブラリとＰＵＴ／ＧＥＴ処理を併用して、ユーザデータとそれに対する付加情報を同時に一つのメッセージでＧＥＴ（転送）できる。

【００８６】まず、要求元のユーザプロセスは、ＧＥＴ要求コマンドをＭＰＩライブラリに対して発行する。このコマンドは、要求するデータのアドレスとデータ長およびＭＰＩライブラリが定めた付加情報等からなる点で、先に説明した送信要求コマンドと同様の引数を指定する。このコマンドを受けて、ＭＰＩライブラリとＰＵＴ／ＧＥＴライブラリとネットワークインタフェース回路は、ＰＵＴ動作の場合と同様にしてＧＥＴ要求メッセージを宛先要素プロセッサに対して転送する。このメッセージには、ＰＵＴのときのメッセージにおける受信データアドレス、受信完了フラグアドレスの代わりに、ＧＥＴすべきデータに対する送信データアドレスおよび送信完了フラグが含まれ、ユーザデータと付加情報は含まれない。このメッセージが従来のＧＥＴ要求メッセージと異なる点は、モードビットを有することである。

【００８７】宛先プロセッサは、ＧＥＴ要求メッセージに含まれたデータアドレスで指定されるローカルメモリ領域からユーザデータを読み出し、送信完了フラグアドレスで指定されるローカルメモリ領域から付加情報を読み出し、送信元プロセッサにそのユーザデータと付加情報を含むＧＥＴメッセージを送り返す。このメッセージは、受信データアドレスと受信完了フラグも含む。宛先プロセッサではそのメッセージを全て相互結合ネットワークに送出した後、送信完了フラグをローカルメモリ内の、上記付加情報の記憶領域の次のアドレス位置に書き込む。

【００８８】送信元要素プロセッサは、ＧＥＴメッセージを受信し、メッセージ内の受信データアドレスに従ってローカルメモリにそのメッセージ内のデータを書き込み、そのメッセージ内の受信完了フラグアドレスに従って、メッセージ内の付加情報をローカルメモリに書き込む。これらのユーザデータと付加情報をすべて受信した後に受信完了フラグをローカルメモリに書き込む。

【００８９】ＰＵＴ処理と同様に、ＧＥＴ処理でも、転送制御情報の読み出し回数およびフラグのローカルメモリへの書き込み回数を削減してローカルメモリアクセスに関わる処理オーバヘッドを小さくできる。また、従来はユーザデータとそれに対する付加情報とを２回にわけて転送していたが、本実施の形態では、一つのＧＥＴメッセージでこれらの二つのデータを転送できるため、データ転送のレイテンシを削減できる。

【００９０】＜変形例＞本発明は以上の実施の形態に限定されるのではなく、以下に示す変形例を含むいろいろの実施の形態により実施可能である。たとえば、ローカルメモリ３０４に格納する転送制御情報に拡張サイズフィールドを設ける。拡張サイズフィールドには、拡張したいフラグ領域のサイズをセットする。モードビットがセットされた場合、送信処理において、転送制御情報を含むヘッダと、送信データアドレスに従ってローカルメモリ領域からよみだされる転送データ長分のデータと、送信フラグアドレスに従ってローカルメモリ領域から読み出される拡張サイズフィールドに設定されたサイズ分の別のデータからメッセージを生成し、相互結合ネットワーク１０５に送出する。受信処理においては、メッセージをヘッダ部とデータ部に分解し、ヘッダ内の転送制御情報に含まれる受信データアドレスで示されるローカルメモリ領域にデータを転送データ長分書き込み、さらに、ヘッダ内の転送制御情報に含まれる受信フラグアドレスで示されるローカルメモリ領域に対し、別のデータを拡張サイズフィールドに設定されたサイズ分書き込む。モードビットがセットされていない場合、拡張サイズフィールドにセットされた値は無視される。

【００９１】

【発明の効果】本実施の形態によれば、ＭＰＩのようなメッセージパッシングライブラリを介してメモリ間直接転送を行う場合におけるデータ転送処理をより高速に行うことができる。

【図面の簡単な説明】

【図１】本実施の形態が対象とする並列計算機の概略構成を示す図である。

【図２】本実施の形態で使用する転送制御情報の例を示す図である。

【図３】図１の装置に使用する要素プロセッサの概略構成を示す図である。

【図４】図３の要素プロセッサに使用するストレージコントローラおよびネットワークインタフェース回路の構成を示す図である。

【図５】図４のネットワークインタフェース回路内のメッセージ生成部の内部構成を示す図である。

【図６】図４のネットワークインタフェース回路内のメッセージ分解部の内部構成を示す図である。

【図７】本実施の形態におけるデータ転送処理の概要を説明する図である。

【図８】従来例のデータ転送処理の概要を説明する図で

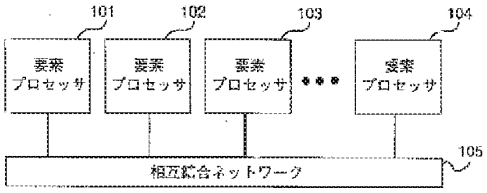
ある。

１０５…相互結合ネットワーク

【符号の説明】

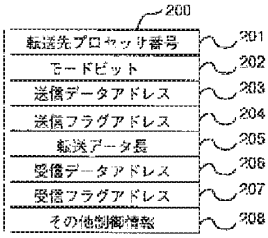
【図１】

図１



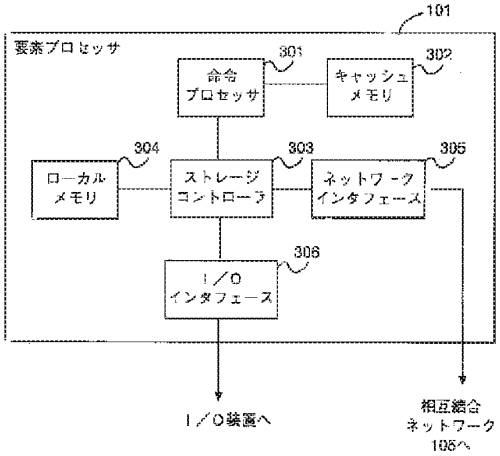
【図２】

図２



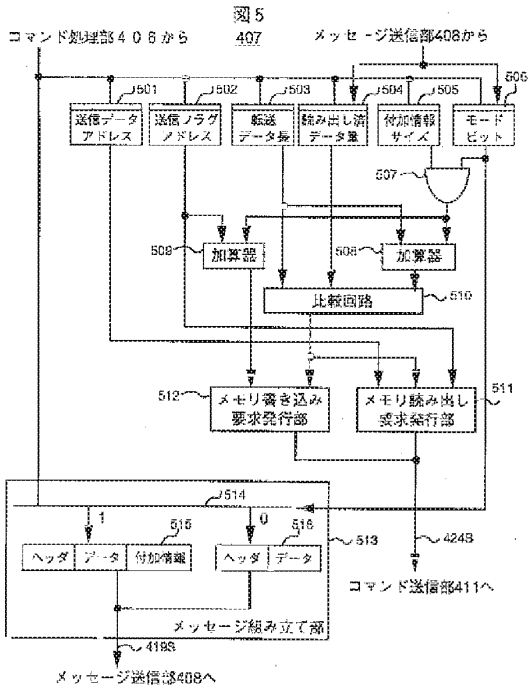
【図３】

図３



【図５】

図５



【図4】

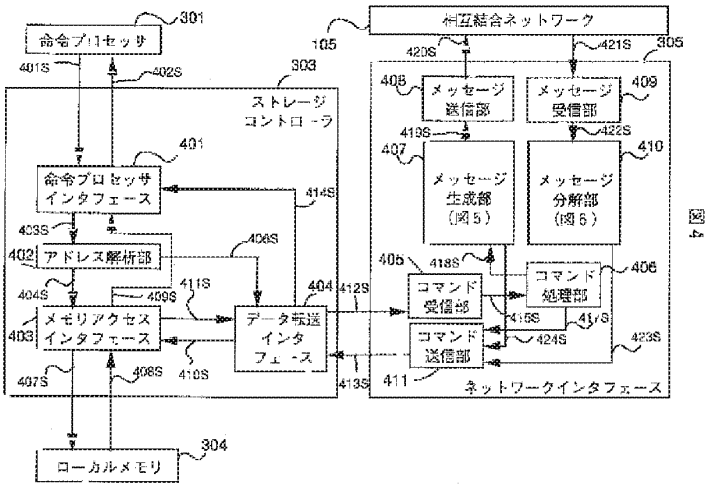
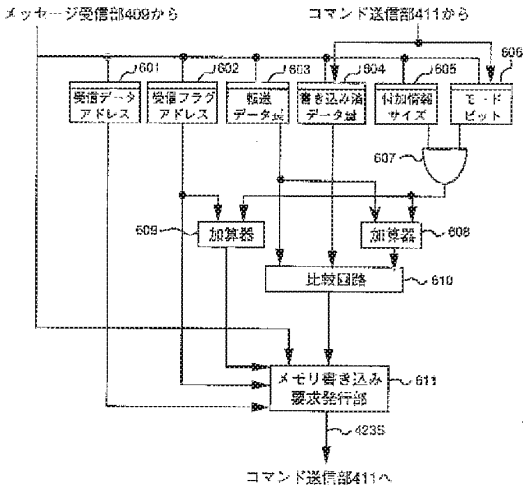

相互結合ネットワーク
105
305
301
命令プロセッサ
401S
402S
303
ストレージ
コントローラ
401
命令プロセッサ
インタフェース
414S
403S
402
アドレス解析部
406S
404S
409S
411S
404
メモリアクセス
インタフェース
403
データ転送
インタ
フェース
412S
407S
408S
410S
413S
420S
421S
408
メッセージ
送信部
409
メッセージ
受信部
419S
422S
407
メッセージ
生成部
（図5）
410
メッセージ
分解部
（図6）
405
418S
コマンド
受信部
406
コマンド
処理部
415S
417S
コマンド
送信部
424S
423S
411
ネットワークインタフェース
304
ローカルメモリ

図4

【図6】

図6


410
メッセージ受信部409から
コマンド送信部411から
601
602
603
604
605
606
受信データ
アドレス
受信フラグ
アドレス
転送
データ長
書き込み済
データ量
付加情報
サイズ
モード
ビット
607
609
加算器
加算器
608
比較回路
610
メモリ書き込み
要求発行部
611
423S
コマンド送信部411へ

【図7】

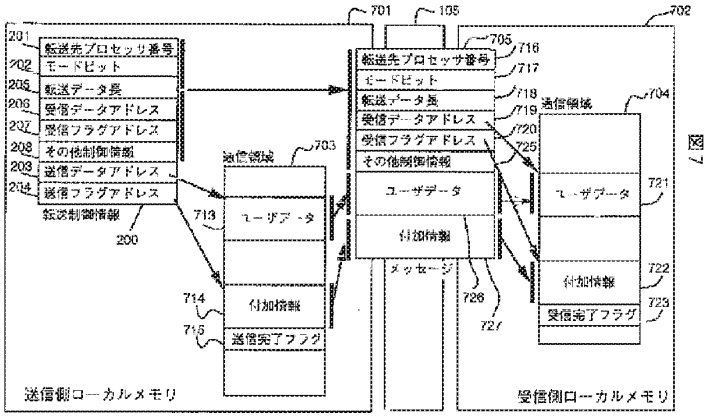

701
105
702
201
転送先プロセッサ番号
202
モードビット
203
転送データ長
206
受信データアドレス
207
受信フラグアドレス
208
その他制御情報
209
送信データアドレス
214
送信フラグアドレス
転送制御情報
200
通信領域
703
ユーザデータ
713
付加情報
714
送信完了フラグ
715
送信側ローカルメモリ
705
716
717
718
719
720
725
メッセージ
726
727
通信領域
704
ユーザデータ
721
付加情報
722
受信完了フラグ
723
受信側ローカルメモリ

図7

【図8】

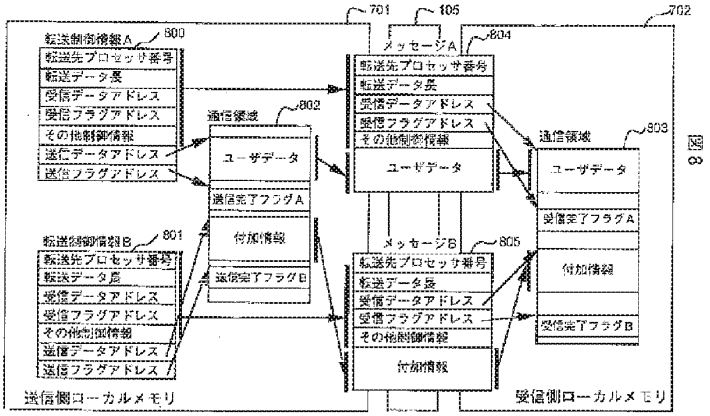

701
105
702
転送制御情報A
800
転送先プロセッサ番号
転送データ長
受信データアドレス
受信フラグアドレス
その他制御情報
送信データアドレス
送信フラグアドレス
通信領域
802
ユーザデータ
送信完了フラグA
付加情報
転送制御情報B
801
転送先プロセッサ番号
転送データ長
受信データアドレス
受信フラグアドレス
その他制御情報
送信データアドレス
送信フラグアドレス
送信完了フラグB
送信側ローカルメモリ
メッセージA
804
転送先プロセッサ番号
転送データ長
受信データアドレス
受信フラグアドレス
その他制御情報
ユーザデータ
メッセージB
805
転送先プロセッサ番号
転送データ長
受信データアドレス
受信フラグアドレス
その他制御情報
付加情報
通信領域
803
ユーザデータ
受信完了フラグA
付加情報
受信完了フラグB
受信側ローカルメモリ

図8

Electronic Patent Application Fee Transmittal

Application Number:	14181112
Filing Date:	14-Feb-2014
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Filer:	Lance Dean Smemoe
Attorney Docket Number:	ZTANN.002P1C3

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension - 3 months with $0 paid	2253	1	700	700
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	120	120
		Total in USD ($)		**820**

Electronic Acknowledgement Receipt	
EFS ID:	33929242
Application Number:	14181112
International Application Number:	
Confirmation Number:	6874
Title of Invention:	CLUSTER COMPUTING
First Named Inventor/Applicant Name:	Zvi Tannenbaum
Customer Number:	20995
Filer:	Lance Dean Smemoe/Wendi Manzanares
Filer Authorized By:	Lance Dean Smemoe
Attorney Docket Number:	ZTANN.002P1C3
Receipt Date:	05-OCT-2018
Filing Date:	14-FEB-2014
Time Stamp:	14:25:48
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$820
RAM confirmation Number	100918INTEFSW14262000
Deposit Account	111410
Authorized User	Wendi Manzanares
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.16 (National application filing, search, and examination fees) 37 CFR 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		OAResponse-ZTANN002P1C3.pdf	78043 7fa5f76c6a47798cbe756f5cc51486949af0ce9e	yes	15
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Amendment/Req. Reconsideration-After Non-Final Reject		1	1	
	Claims		2	11	
	Applicant summary of interview with examiner		12	12	
	Applicant Arguments/Remarks Made in an Amendment		13	15	
Warnings:					
Information:					
2	Transmittal Letter	IDSTransmittal-ZTANN002P1C3.pdf	20412 11abab3cdf25dabe264d326972eb2bf46cc8292e	no	2
Warnings:					
Information:					
3	Information Disclosure Statement (IDS) Form (SB08)	IDSListForm-ZTANN002P1C3.pdf	22006 56930b17a7f1a7c68bf174c0dac37753a744907b	no	1
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
4	Foreign Reference	JP2004247405_w_MT.pdf	3163341 47bcfa33d5580e933fbdde5cfbe5c1f3f018c536	no	62
Warnings:					
Information:					

5	Foreign Reference	JPH11126196_w_MT.pdf	2783060 44a06d09b35baa990bb07fa70a43f7bccad78db9	no	52
Warnings:					
Information:					
6	Non Patent Literature	NPL_ParallelComputingToolkit.pdf	6604030 20a928005a5a8f0f90ff52a66b4ba021821d46fb	no	100
Warnings:					
Information:					
7	Non Patent Literature	NPL_SearchReport.pdf	986210 cfb4dcc3927ad8cd331ab3bdb8866df2e6d239b1	no	9
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	32353 362c3dd354b5760fdbdd73de6ea0a733a88d228c	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	13689455		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor	:	Zvi Tannenbaum
App. No.	:	14/181112
Filed	:	February 14, 2014
For	:	CLUSTER COMPUTING
Examiner	:	Asres, Hermon
Art Unit	:	2449
Conf. No.	:	6874

RESPONSE TO NON-FINAL OFFICE ACTION DATED APRIL 13, 2018

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Commissioner:

In response to the Office Action dated April 13, 2018, please reconsider the rejections in view of the following remarks.

Amendments to the Claims are reflected in the listing of claims which begins on page 2.

Summary of Interview begins on page 12.

Remarks/Arguments begin on page 13.

AMENDMENTS TO THE CLAIMS

1. (Canceled)

2. (**Currently amended**) A computer cluster comprising:

a plurality of nodes, wherein each of the plurality of nodes comprises ~~comprising~~ a hardware processor, wherein one or more of the nodes are configured to receive a command to start a cluster initialization process for the computer cluster, and wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing the hardware processor to evaluate mathematical expressions~~, when executed, causes the hardware processor to interpret user instructions, to evaluate mathematical expressions, and to produce results of mathematical expression evaluation~~; and

a mechanism for the nodes to communicate results of mathematical expression evaluation with each other using a peer-to-peer architecture;

wherein the plurality of nodes comprises:

a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and

a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of the first mathematical expression evaluation to a third node;

wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of the first mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node;

wherein ~~at least one of the nodes~~ the first node is configured to return ~~at least one~~ the result of the second mathematical expression evaluation to ~~a user interface or a script~~ the user interface.

3. (Previously presented) The computer cluster of claim 2, wherein the single-node kernel comprises a Mathematica kernel.

4. (Previously presented) The computer cluster of claim 2, wherein the mechanism comprises a message passing interface.

5. (Previously presented) The computer cluster of claim 2, wherein each of the nodes comprises one or more cluster node modules.

6. (Previously presented) The computer cluster of claim 2, wherein each single-node kernel is stored in a non-transitory computer-readable medium and configured to accept and execute a request.

7. (Previously presented) The computer cluster of claim 6, wherein each of the nodes comprises one or more cluster node modules, wherein each of the cluster node modules comprises instructions stored in a non-transitory computer-readable medium, and wherein the instructions, when executed by the hardware processor, cause the cluster node module to communicate with the single-node kernel and with one or more other cluster node modules.

8. (Previously presented) The computer cluster of claim 7, wherein the plurality of cluster node modules act as a cluster.

9. (Previously presented) The computer cluster of claim 7, wherein the plurality of cluster node modules communicate with one another to act as a cluster.

10. (Previously presented) The computer cluster of claim 9, wherein the computer cluster includes the user interface.

11. (Previously presented) The computer cluster of claim 10, wherein each cluster node module accepts instructions from the user interface and interprets one or more of the instructions.

12. (Previously presented) The computer cluster of claim 2, wherein the cluster initialization process comprises establishing communication among two or more of the nodes.

13. (Previously presented) The computer cluster of claim 2, wherein the cluster initialization process comprises configuring access by one or more of the nodes to a computer-readable medium comprising program code for the single-node kernel.

14. (Previously presented) The computer cluster of claim 13, wherein the cluster initialization process comprises establishing message-passing support among the nodes in the cluster.

15. (Previously presented) The computer cluster of claim 13, wherein the cluster initialization process comprises launching cluster node modules by the cluster.

16. (Previously presented) The computer cluster of claim 15, wherein the cluster initialization process comprises assigning a processor identification number to each of the cluster node modules.

17. (Previously presented) The computer cluster of claim 2, wherein the cluster initialization process comprises:

launching cluster node modules by the cluster; and

after launching the cluster node modules, configuring access by one or more of the nodes to a non-transitory computer-readable medium comprising program code for the single-node kernel.

18. (Previously presented) The computer cluster of claim 2, wherein the cluster initialization process comprises:

launching cluster node modules by the cluster;

after launching the cluster node modules, establishing communication among two or more of the nodes; and

after establishing communication among two or more of the nodes, assigning a processor identification number to each of the cluster node modules.

19. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to communicate at least some of the user instructions.

20. (Previously presented) The computer cluster of claim 19, wherein one or more of the nodes are configured to communicate at least some of the user instructions to one or more single-node kernels.

21. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to accept user instructions via one or more of the nodes.

22. (Previously presented) The computer cluster of claim 21, wherein one or more of the nodes are configured to communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other.

23. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to transmit at least some of the user instructions that originate from a user interface.

24. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to parallelize at least some of the user instructions before communicating at least some of the user instructions to one or more single-node kernels.

25. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to accept user instructions before communicating at least some of the user instructions to one or more single-node kernels.

26. (Previously presented) The computer cluster of claim 2, wherein one or more of the nodes are configured to:

accept user instructions;

after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other; and

after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels.

27. (Previously presented) The computer cluster of claim 2, wherein the plurality of nodes are configured to communicate with one another to interpret and translate commands for execution by a plurality of single-node kernels.

28. (**Currently amended**) A computer cluster comprising:

a plurality of nodes, wherein one or more of the nodes are configured to receive:

a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes; and

an instruction from a user interface or a script;

and

a mechanism for the nodes to communicate results of mathematical expression evaluation with each other using asynchronous calls;

wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions ~~configured to interpret user instructions, to evaluate mathematical expressions, and to produce results of mathematical expression evaluation~~;[[and]]

wherein the plurality of nodes comprises:

a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and

a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of mathematical expression evaluation to a third node;

wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node;

wherein ~~at least one of the nodes~~ the first node is configured to return ~~at least one~~ the result of the second mathematical expression evaluation to the user interface or the script.

29. (Previously presented) The computer cluster of claim 28, wherein the asynchronous calls comprise a first command to create a first packet containing:

an expression to be sent as payload; and

a target node where the expression should be sent;

wherein the first command is configured to be called from within a single-node kernel;

wherein the single-node kernel is configured to send the first packet to a local cluster node module; and

wherein the local cluster node module is configured to forward the expression to the target node.

30. (Previously presented) The computer cluster of claim 29, wherein the asynchronous calls comprise a second command to create a second packet containing:

a location where the expression is expected to be received; and

a sending node from which the expression is expected to be received;

wherein the second command is configured to be called from within a single-node kernel;

wherein the single-node kernel is configured to send the second packet to a local cluster node module; and

wherein the local cluster node module is configured to store the second packet contents in a message receiving queue.

31. (**Currently amended**) A computer cluster comprising:

a plurality of nodes, wherein one or more of the nodes are configured to receive:

a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes; and

an instruction from a user interface or a script;

and

a mechanism for the nodes to communicate results of mathematical expression evaluation with each other;

wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that, when executed, is capable of causing a hardware processor to evaluate mathematical expressions

~~configured to interpret user instructions, to evaluate mathematical expressions, and to produce results of mathematical expression evaluation~~;[[and]]

wherein the plurality of nodes comprises:

a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution; and

a second node comprising a second hardware processor with a plurality of processing cores, wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of mathematical expression evaluation to a third node;

wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node;

and

wherein ~~at least one of the nodes~~ the first node is configured to return ~~at least one~~ the result of the second mathematical expression evaluation to the user interface or the script.

32. (Previously presented) The computer cluster of claim 5, wherein each of the plurality of nodes implements asynchronous calls that enable the single-node kernel to perform computation tasks while the cluster node modules are simultaneously communicating with one another.

33. (Previously presented) The computer cluster of claim 5, wherein intercommunication among the plurality of single-node kernels during thread execution is enabled by the plurality of cluster node modules, and wherein the computer cluster is configured to permit exchange of information between nodes during the course of a parallel computation.

34. (Previously presented) The computer cluster of claim 5, wherein each of the single-node kernels are configured to call a send command involving creating a packet containing:

an expression to be sent as payload; and

where the expression should be sent;

wherein, upon reception of the send command by a local cluster node module associated with the single-node kernel, the local cluster node module is configured to decode the packet and forward a payload to the cluster node module specified in the packet;

wherein each of the single-node kernels is configured to call a receipt command involving creating a packet specifying which message to test for completion and to then wait for a reply expression to evaluate;

wherein, upon reception of the receipt command by the local cluster node module associated with the single-node kernel, the local cluster node module is configured to decode the packet and use a message specifier to search for any matching expressions listed as completed in a received message queue;

wherein, upon determining that such a completed expression is found, the local cluster node module is configured to send the completed expression to a local single-node kernel associated with the cluster node module in response to the receipt command, and

wherein each of the single-node kernels is configured to receive the completed expression and to update variables used by the single-node kernel during evaluation of expressions.

35. (Previously presented) The computer cluster of claim 2, wherein the plurality of nodes are configured to permit exchange of information between nodes during the course of parallel computation.

36. (Previously presented) The computer cluster of claim 2, wherein each of the plurality of nodes comprises instructions executable by the hardware processor and configured to implement asynchronous behavior, wherein the instructions comprise:

a first instruction to asynchronously send a payload to another node;

a second instruction to asynchronously receive a payload from another node; and

a third instruction to search for a payload matching a message specifier.

37. (**Currently amended**) A computer cluster node for evaluating expressions in parallel with other computer cluster nodes, the computer cluster node comprising:

a hardware processor configured to access one or more non-transitory memory devices comprising program code for a single-node kernel that, when executed, causes the hardware processor to interpret user instructions, to evaluate mathematical expressions, and to produce results of mathematical expression evaluation;

a user connection interface configured to receive a command to start a cluster initialization process for a computer cluster;[[and]]

a mechanism to communicate results of evaluation with other computer cluster nodes using a peer-to-peer architecture; and

program code that, when executed, is capable of causing the hardware processor to receive calls from a first node, execute at least a first mathematical expression evaluation, and communicate a result of mathematical expression evaluation to another node for use in at least a second mathematical expression evaluation;

wherein the user connection interface is configured to return at least one result of mathematical expression evaluation to a user interface or a script.

38. (Previously presented) The computer cluster node of claim 37, wherein the one or more non-transitory memory devices comprise program code for performing a parallel fast Fourier transform, wherein the program code causes the hardware processor to evaluate a command to perform a Fourier transform on an array comprising a first data portion that is stored on the computer cluster node and a second data portion that is not stored on the computer cluster node.

39. (Previously presented) The computer cluster node of claim 37, wherein the computer cluster node is configured to permit exchange of information with other computer cluster nodes during the course of parallel computation.

40. (Previously presented) The computer cluster node of claim 37, wherein the computer cluster node comprises instructions executable by the hardware processor and configured to implement asynchronous behavior, wherein the instructions comprise:

a first instruction to asynchronously send a payload to another node;

a second instruction to asynchronously receive a payload from another node; and

a third instruction to search for a payload matching a message specifier.

41. (Previously presented) The computer cluster node of claim 37, wherein the hardware processor comprises a special purpose microprocessor.

Application No.: 14/181112
Filing Date: February 14, 2014

SUMMARY OF INTERVIEW

Attendees, Date and Type of Interview

An in-person interview was conducted on May 15, 2018 and attended by the Examiner, Mr. Hermon Asres and the Applicant's representative, Mr. Lance D. Smemoe (Reg. No. 66,152).

Exhibits and/or Demonstrations

None

Identification of Claims Discussed

Claim 1

Identification of Cited/Disclosed Art Discussed

The art of record, including Block (US 2005/0021751), Singh (US 8,601,101), and Howard (US 2003/0195938), was discussed.

Proposed Amendments

The Applicant's representative presented a proposed amendment substantially as set forth in claim 1 as amended herein.

Principal Arguments and Other Matters

While reserving the right to pursue the subject matter as previously presented in the claims, the Applicant's representative argued that the claims as amended distinguished the references cited in the outstanding office action.

Results of Interview

The Examiner indicated that the proposed amendment when formally filed would overcome the cited references. The Applicant's representative will file a response, and the Examiner will reconsider the allowability of the claims after conducting an updated search.

REMARKS

In the Office Action dated April 13, 2018, Claims 2-41 were rejected, as discussed below. In this Amendment, claims 2, 28, 31, and 37 are amended. No new matter is added. After the amendments, claims 2-41 remain pending for consideration. Claims 2, 28, 31, and 37 are independent claims.

In-person Interview

Applicant thanks the Examiner for the courteous and helpful in-person interview held on May 15, 2018 (summarized above).

Response to Double Patenting Rejection

Claims 2, 28, and 31 are rejected on the ground of non-statutory (obviousness-type) double patenting over claims 1 and 8 of U.S. Patent No. 8,676,877. Applicant respectfully requests that the double patenting rejections be held in abeyance until the other rejections are overcome.

Response to Rejections under 35 U.S.C. § 103

The Office Action rejects Claims 2 and 4-41 under 35 U.S.C. § 103, arguing that they are unpatentable in view of U.S. Publication No. 2005/0021751 to Block in view of U.S. Patent No. 8,601,101 to Singh and further in view of U.S. Publication No. 2003/0195938 to Howard. The Office Action rejects Claim 3 under 35 U.S.C. § 103, arguing that it is unpatentable in view of U.S. Publication No. 2005/0021751 to Block, U.S. Patent No. 8,601,101 to Singh, U.S. Publication No. 2007/0073705 to Gray, and U.S. Publication 2003/0195938 to Howard. Applicant respectfully traverses the rejections.

Applicant has amended the independent claims in an effort to advance prosecution of the application. No new matter is added. The application as filed explains that a first node of a computer cluster can be connected to a user interface or a script and configured to receive user instructions from the user interface or the script. *E.g.,* application as filed at ¶¶ [0012], [0025]. The first node can distribute calls to other cluster nodes. *E.g., id.* at ¶ [0026]. The cluster nodes can evaluate expressions according to a set of rules. *E.g., id.* at ¶ [0079]. The application

describes an intercommunication architecture and functions that permit intermediate results to be passed from one node to another node without involvement of the first node. *E.g., id.* at ¶¶ [0027], [0121]. For example, in a non-limiting embodiment, the application describes a function that allows elements near the edge of a list, which can include intermediate results of evaluation, to be copied to neighboring processors so that neighboring edges of each node partition can interact. *Id.* at ¶¶ [0080], [0137]. Results of evaluations can be communicated back to the first node and returned to the user interface or script. *E.g., id.* at ¶¶ [0026], [0133].

As explained in the interview of May 15, 2018, the cited references do not disclose the combination of features recited in the claims as amended. For example, the references even if combined do not disclose a first node configured to interpret user instructions and distribute calls to other nodes for execution; a second node configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result to a third node; and a third node configured to receive the result, execute at least a second mathematical expression evaluation using the result, and communicate the result of the second mathematical expression evaluation to the first node, in combination with the other features recited in Claim 2.

Like Claim 2, independent Claims 28, 31, and 37 similarly recite subject matter not anticipated or rendered obvious by the combination of cited references. The dependent claims are patentable because each depends directly or indirectly from a patentable independent claim and because of the additional features recited in each claim. For at least these reasons, Applicant requests withdrawal of the rejections under 35 U.S.C. § 103.

No Disclaimers or Disavowals

Applicant respectfully submits that the claims are in condition for allowance. Furthermore, any remarks in support of patentability of one claim should not be imputed to any other claim, even if similar terminology is used. Any remarks referring to only a portion of a claim should not be understood to base patentability on that portion or that the limitation discussed is essential or critical; rather, patentability must rest on each claim taken as a whole. Applicant respectfully traverses each of the Examiner's rejections and each of the Examiner's assertions regarding what the prior art shows or teaches, even if not expressly discussed herein. Although the present communication may include alterations to the application or claims, or

characterizations of claim scope or referenced art, applicant is not conceding in this application that previously pending claims are not patentable over the cited references. Rather, any alterations or characterizations are being made to facilitate expeditious prosecution of this application. Applicant reserves the right to pursue at a later date any previously pending or other broader or narrower claims that capture any subject matter supported by the present disclosure, including subject matter found to be specifically disclaimed herein or by any prior prosecution. Accordingly, reviewers of this or any parent, child or related prosecution history shall not reasonably infer that applicant has made any disclaimers or disavowals of any subject matter supported by the present application.

CONCLUSION

For the foregoing reasons, it is respectfully submitted that the rejections set forth in the outstanding Office Action are inapplicable to the present claims. Accordingly, issuance of a Notice of Allowance is requested.

The undersigned has made a good faith effort to respond to all of the rejections in the case and to place the claims in condition for immediate allowance. Nevertheless, if any undeveloped issues remain or if any issues require clarification, the Examiner is respectfully requested to call the undersigned at 949-721-5280.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: October 5, 2018

By: /Lance D. Smemoe/
Lance D. Smemoe
Registration No. 66,152
Registered Practitioner
Customer No. 20995
(949) 760-0404

28065276

INFORMATION DISCLOSURE STATEMENT

First Inventor	:	Zvi Tannenbaum
App. No.	:	14/181112
Filed	:	February 14, 2014
For	:	CLUSTER COMPUTING
Examiner	:	Asres, Hermon
Art Unit	:	2449
Conf. No.	:	6874

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

References and Listing

Pursuant to 37 CFR 1.56, an Information Disclosure Statement listing references is provided herewith. Copies of any listed foreign and non-patent literature references are being submitted. Any foreign references may also include English abstract(s) and/or machine translation(s), but no representation is made as to their accuracy.

If the Examiner would like additional information regarding these references or if anything is unclear, the Examiner is invited to contact the undersigned for assistance.

No Disclaimers

To the extent that anything in the Information Disclosure Statement or the listed references could be construed as a disclaimer of any subject matter supported by the present application, Applicant hereby rescinds and retracts such disclaimer.

Timing of Disclosure

This Information Disclosure Statement is being filed after receipt of a First Office Action, but before the mailing date of a Final Action and before the mailing date of a Notice of Allowance. This Statement is accompanied by the fees set forth in 37 CFR 1.17(p).

The Commissioner is hereby authorized to charge any additional fees which may be required or to credit any overpayment to Account No. 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: October 5, 2018

By: /Lance D. Smemoe/
Lance D. Smemoe
Registration No. 66,152
Registered Practitioner
Customer No. 20995
(949) 760-0404

28546233

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 14/181,112	Filing Date 02/14/2014	☐ To be Mailed

ENTITY: ☐ LARGE ☑ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $40 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $210 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	10/05/2018	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 40	Minus	** 40	= 0		x $50 =	0
	Independent (37 CFR 1.16(h))	* 4	Minus	*** 4	= 0		x $230 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	LIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	Moliki I May
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/181,112	02/14/2014	Zvi Tannenbaum	ZTANN.002P1C3	6874

20995 7590 05/22/2018

KNOBBE MARTENS OLSON & BEAR LLP
2040 MAIN STREET
FOURTEENTH FLOOR
IRVINE, CA 92614
UNITED STATES OF AMERICA

EXAMINER	
ASRES, HERMON	

ART UNIT	PAPER NUMBER
2449	

NOTIFICATION DATE	DELIVERY MODE
05/22/2018	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

jayna.cartee@knobbe.com
efiling@knobbe.com

Applicant-Initiated Interview Summary	**Application No.** 14/181,112	**Applicant(s)** TANNENBAUM ET AL.	
	Examiner HERMON ASRES	**Art Unit** 2449	

All participants (applicant, applicant's representative, PTO personnel):

(1) *HERMON ASRES*.

(3) *Lance Smemoe (REG NO 66,152)*.

(2) _____.

(4) _____.

Date of Interview: *15 May 2018*.

Type: ☐ Telephonic ☐ Video Conference
☒ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☒ No.
If Yes, brief description: _____.

Issues Discussed ☐101 ☐112 ☐102 ☒103 ☐Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: *1*.

Identification of prior art discussed: _____.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Applicant discussed proposed amendment to independent claim 1 including "wherein the plurality of nodes comprises: a first node comprising at least one hardware processor configured to access a memory comprising program code for a user interface and program code for a single-node kernel configured to interpret user code and distribute computational instructions to at least one of a plurality of other nodes for execution; a second node comprising at least one hardware processor with a plurality of processing cores, wherein the second node is configured to receive computational instructions from the first node, execute the instructions, and communicate a result of the computation to at least one other node; a third node comprising at least one processor with a plurality of processing cores, wherein the third node is configured to receive the result of an executed computation from another node, execute a second computation based in part on the received result , and communicate the result of the second computation to the first node". Examiner indicated that the proposed amendment when formally filed would overcome the cited references and also that a search will be updated. No other agreement was reached.

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☒ Attachment

/HERMON ASRES/ Examiner, Art Unit 2449	/VIVEK SRIVASTAVA/ Supervisory Patent Examiner, Art Unit 2449

Summary of Record of Interview Requirements

Manual of Patent Examining Procedure (MPEP), Section 713.04, Substance of Interview Must be Made of Record
A complete written statement as to the substance of any face-to-face, video conference, or telephone interview with regard to an application must be made of record in the application whether or not an agreement with the examiner was reached at the interview.

Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews
Paragraph (b)

In every instance where reconsideration is requested in view of an interview with an examiner, a complete written statement of the reasons presented at the interview as warranting favorable action must be filed by the applicant. An interview does not remove the necessity for reply to Office action as specified in §§ 1.111, 1.135. (35 U.S.C. 132)

37 CFR §1.2 Business to be transacted in writing.
All business with the Patent or Trademark Office should be transacted in writing. The personal attendance of applicants or their attorneys or agents at the Patent and Trademark Office is unnecessary. The action of the Patent and Trademark Office will be based exclusively on the written record in the Office. No attention will be paid to any alleged oral promise, stipulation, or understanding in relation to which there is disagreement or doubt.

———

The action of the Patent and Trademark Office cannot be based exclusively on the written record in the Office if that record is itself incomplete through the failure to record the substance of interviews.

It is the responsibility of the applicant or the attorney or agent to make the substance of an interview of record in the application file, unless the examiner indicates he or she will do so. It is the examiner's responsibility to see that such a record is made and to correct material inaccuracies which bear directly on the question of patentability.

Examiners must complete an Interview Summary Form for each interview held where a matter of substance has been discussed during the interview by checking the appropriate boxes and filling in the blanks. Discussions regarding only procedural matters, directed solely to restriction requirements for which interview recordation is otherwise provided for in Section 812.01 of the Manual of Patent Examining Procedure, or pointing out typographical errors or unreadable script in Office actions or the like, are excluded from the interview recordation procedures below. Where the substance of an interview is completely recorded in an Examiners Amendment, no separate Interview Summary Record is required.

The Interview Summary Form shall be given an appropriate Paper No., placed in the right hand portion of the file, and listed on the "Contents" section of the file wrapper. In a personal interview, a duplicate of the Form is given to the applicant (or attorney or agent) at the conclusion of the interview. In the case of a telephone or video-conference interview, the copy is mailed to the applicant's correspondence address either with or prior to the next official communication. If additional correspondence from the examiner is not likely before an allowance or if other circumstances dictate, the Form should be mailed promptly after the interview rather than with the next official communication.

The Form provides for recordation of the following information:
- Application Number (Series Code and Serial Number)
- Name of applicant
- Name of examiner
- Date of interview
- Type of interview (telephonic, video-conference, or personal)
- Name of participant(s) (applicant, attorney or agent, examiner, other PTO personnel, etc.)
- An indication whether or not an exhibit was shown or a demonstration conducted
- An identification of the specific prior art discussed
- An indication whether an agreement was reached and if so, a description of the general nature of the agreement (may be by attachment of a copy of amendments or claims agreed as being allowable). Note: Agreement as to allowability is tentative and does not restrict further action by the examiner to the contrary.
- The signature of the examiner who conducted the interview (if Form is not an attachment to a signed Office action)

It is desirable that the examiner orally remind the applicant of his or her obligation to record the substance of the interview of each case. It should be noted, however, that the Interview Summary Form will not normally be considered a complete and proper recordation of the interview unless it includes, or is supplemented by the applicant or the examiner to include, all of the applicable items required below concerning the substance of the interview.

A complete and proper recordation of the substance of any interview should include at least the following applicable items:
1) A brief description of the nature of any exhibit shown or any demonstration conducted,
2) an identification of the claims discussed,
3) an identification of the specific prior art discussed,
4) an identification of the principal proposed amendments of a substantive nature discussed, unless these are already described on the Interview Summary Form completed by the Examiner,
5) a brief identification of the general thrust of the principal arguments presented to the examiner,
   (The identification of arguments need not be lengthy or elaborate. A verbatim or highly detailed description of the arguments is not required. The identification of the arguments is sufficient if the general nature or thrust of the principal arguments made to the examiner can be understood in the context of the application file. Of course, the applicant may desire to emphasize and fully describe those arguments which he or she feels were or might be persuasive to the examiner.)
6) a general indication of any other pertinent matters discussed, and
7) if appropriate, the general results or outcome of the interview unless already described in the Interview Summary Form completed by the examiner.

Examiners are expected to carefully review the applicant's record of the substance of an interview. If the record is not complete and accurate, the examiner will give the applicant an extendable one month time period to correct the record.

Examiner to Check for Accuracy

If the claims are allowable for other reasons of record, the examiner should send a letter setting forth the examiner's version of the statement attributed to him or her. If the record is complete and accurate, the examiner should place the indication, "Interview Record OK" on the paper recording the substance of the interview along with the date and the examiner's initials.

Applicant-Initiated Interview Summary	Application No.	Applicant(s)	
	14/181,112	TANNENBAUM ET AL.	
	Examiner	Art Unit	
	HERMON ASRES	2449	

All participants (applicant, applicant's representative, PTO personnel):

(1) *HERMON ASRES*.

(3) *Lance Smemoe (REG NO 66,152)*.

(2) _____.

(4) _____.

Date of Interview: *15 May 2018*.

Type: ☐ Telephonic ☐ Video Conference
☒ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☒ No.
If Yes, brief description: _____.

Issues Discussed ☐101 ☐112 ☐102 ☒103 ☐Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: *1*.

Identification of prior art discussed: _____.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Applicant discussed proposed amendment to independent claim 1 including "wherein the plurality of nodes comprises: a first node comprising at least one hardware processor configured to access a memory comprising program code for a user interface and program code for a single-node kernel configured to interpret user code and distribute computational instructions to at least one of a plurality of other nodes for execution; a second node comprising at least one hardware processor with a plurality of processing cores, wherein the second node is configured to receive computational instructions from the first node, execute the instructions, and communicate a result of the computation to at least one other node; a third node comprising at least one processor with a plurality of processing cores, wherein the third node is configured to receive the result of an executed computation from another node, execute a second computation based in part on the received result , and communicate the result of the second computation to the first node". Examiner indicated that the proposed amendment when formally filed would overcome the cited references and also that a search will be updated. No other agreement was reached.

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☒ Attachment

/HERMON ASRES/ Examiner, Art Unit 2449	/VIVEK SRIVASTAVA/ Supervisory Patent Examiner, Art Unit 2449

Summary of Record of Interview Requirements

Manual of Patent Examining Procedure (MPEP), Section 713.04, Substance of Interview Must be Made of Record
A complete written statement as to the substance of any face-to-face, video conference, or telephone interview with regard to an application must be made of record in the application whether or not an agreement with the examiner was reached at the interview.

Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews
Paragraph (b)

In every instance where reconsideration is requested in view of an interview with an examiner, a complete written statement of the reasons presented at the interview as warranting favorable action must be filed by the applicant. An interview does not remove the necessity for reply to Office action as specified in §§ 1.111, 1.135. (35 U.S.C. 132)

37 CFR §1.2 Business to be transacted in writing.
All business with the Patent or Trademark Office should be transacted in writing. The personal attendance of applicants or their attorneys or agents at the Patent and Trademark Office is unnecessary. The action of the Patent and Trademark Office will be based exclusively on the written record in the Office. No attention will be paid to any alleged oral promise, stipulation, or understanding in relation to which there is disagreement or doubt.

———

The action of the Patent and Trademark Office cannot be based exclusively on the written record in the Office if that record is itself incomplete through the failure to record the substance of interviews.

It is the responsibility of the applicant or the attorney or agent to make the substance of an interview of record in the application file, unless the examiner indicates he or she will do so. It is the examiner's responsibility to see that such a record is made and to correct material inaccuracies which bear directly on the question of patentability.

Examiners must complete an Interview Summary Form for each interview held where a matter of substance has been discussed during the interview by checking the appropriate boxes and filling in the blanks. Discussions regarding only procedural matters, directed solely to restriction requirements for which interview recordation is otherwise provided for in Section 812.01 of the Manual of Patent Examining Procedure, or pointing out typographical errors or unreadable script in Office actions or the like, are excluded from the interview recordation procedures below. Where the substance of an interview is completely recorded in an Examiners Amendment, no separate Interview Summary Record is required.

The Interview Summary Form shall be given an appropriate Paper No., placed in the right hand portion of the file, and listed on the "Contents" section of the file wrapper. In a personal interview, a duplicate of the Form is given to the applicant (or attorney or agent) at the conclusion of the interview. In the case of a telephone or video-conference interview, the copy is mailed to the applicant's correspondence address either with or prior to the next official communication. If additional correspondence from the examiner is not likely before an allowance or if other circumstances dictate, the Form should be mailed promptly after the interview rather than with the next official communication.

The Form provides for recordation of the following information:
- Application Number (Series Code and Serial Number)
- Name of applicant
- Name of examiner
- Date of interview
- Type of interview (telephonic, video-conference, or personal)
- Name of participant(s) (applicant, attorney or agent, examiner, other PTO personnel, etc.)
- An indication whether or not an exhibit was shown or a demonstration conducted
- An identification of the specific prior art discussed
- An indication whether an agreement was reached and if so, a description of the general nature of the agreement (may be by attachment of a copy of amendments or claims agreed as being allowable). Note: Agreement as to allowability is tentative and does not restrict further action by the examiner to the contrary.
- The signature of the examiner who conducted the interview (if Form is not an attachment to a signed Office action)

It is desirable that the examiner orally remind the applicant of his or her obligation to record the substance of the interview of each case. It should be noted, however, that the Interview Summary Form will not normally be considered a complete and proper recordation of the interview unless it includes, or is supplemented by the applicant or the examiner to include, all of the applicable items required below concerning the substance of the interview.

A complete and proper recordation of the substance of any interview should include at least the following applicable items:
1) A brief description of the nature of any exhibit shown or any demonstration conducted,
2) an identification of the claims discussed,
3) an identification of the specific prior art discussed,
4) an identification of the principal proposed amendments of a substantive nature discussed, unless these are already described on the Interview Summary Form completed by the Examiner,
5) a brief identification of the general thrust of the principal arguments presented to the examiner,
   (The identification of arguments need not be lengthy or elaborate. A verbatim or highly detailed description of the arguments is not required. The identification of the arguments is sufficient if the general nature or thrust of the principal arguments made to the examiner can be understood in the context of the application file. Of course, the applicant may desire to emphasize and fully describe those arguments which he or she feels were or might be persuasive to the examiner.)
6) a general indication of any other pertinent matters discussed, and
7) if appropriate, the general results or outcome of the interview unless already described in the Interview Summary Form completed by the examiner.

Examiners are expected to carefully review the applicant's record of the substance of an interview. If the record is not complete and accurate, the examiner will give the applicant an extendable one month time period to correct the record.

Examiner to Check for Accuracy

If the claims are allowable for other reasons of record, the examiner should send a letter setting forth the examiner's version of the statement attributed to him or her. If the record is complete and accurate, the examiner should place the indication, "Interview Record OK" on the paper recording the substance of the interview along with the date and the examiner's initials.

M E M O R A N D U M

TO: Examiner Hermon Asres
Direct: 571-272-4257
Email: Hermon.Asres@uspto.gov

FROM: Lance Smemoe, Reg. #66,152

RE: Agenda for in-person interview on May 15, 2018 at 2 p.m. Eastern time

Application No.: 14/181112

Attorney Docket No.: ZTANN.002P1C3

INFORMAL DOCUMENT FOR DISCUSSION ONLY – DO NOT ENTER

I. Discuss Applicant's disclosure and claimed inventions

II. Discuss references cited in Office Action

A. Block et al. (US 2005/0021751)
B. Singh (US 8,601,101)
C. Howard et al. (US 2003/0195938)

III. Discuss proposed claim amendments

1. (Canceled)

2. (**Currently amended**) A computer cluster comprising:

a plurality of nodes comprising a hardware processor, wherein one or more of the nodes are configured to receive a command to start a cluster initialization process for the computer cluster, and wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that executes computational instructions~~, when executed, causes the hardware processor to interpret user instructions, to evaluate mathematical expressions, and to produce results of mathematical expression evaluation~~; and

a mechanism for the nodes to communicate results of mathematical expression evaluation with each other using a peer-to-peer architecture;

wherein the plurality of nodes comprises:

a first node comprising at least one hardware processor configured to access a memory comprising program code for a user interface and program code for a single-node kernel configured to interpret user code and distribute

computational instructions to at least one of a plurality of other nodes for execution;

a second node comprising at least one hardware processor with a plurality of processing cores, wherein the second node is configured to receive computational instructions from the first node, execute the instructions, and communicate a result of the computation to at least one other node;

a third node comprising at least one processor with a plurality of processing cores, wherein the third node is configured to receive the result of an executed computation from another node, execute a second computation based in part on the received result , and communicate the result of the second computation to the first node.

~~wherein at least one of the nodes is configured to return at least one result of mathematical expression evaluation to a user interface or a script~~.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/181,112	02/14/2014	Zvi Tannenbaum	ZTANN.002P1C3	6874

20995 7590 04/13/2018

KNOBBE MARTENS OLSON & BEAR LLP
2040 MAIN STREET
FOURTEENTH FLOOR
IRVINE, CA 92614

EXAMINER	
ASRES, HERMON	

ART UNIT	PAPER NUMBER
2449	

NOTIFICATION DATE	DELIVERY MODE
04/13/2018	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

jayna.cartee@knobbe.com
efiling@knobbe.com

Office Action Summary	Application No. 14/181,112	Applicant(s) TANNENBAUM ET AL.	
	Examiner HERMON ASRES	Art Unit 2449	AIA (First Inventor to File) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *08/08/2017*.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☒ Claim(s) *2-41* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *2-41* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☒ The drawing(s) filed on *02/14/2014* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date *See Continuation Sheet*.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

4) ☐ Other: _____.

Continuation of Attachment(s) 2). Information Disclosure Statement(s) (PTO/SB/08), Paper No(s)/Mail Date :08/08/2017, 09/22/2017, 12/01/2017.

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on August 08, 2017 has been entered.

Claims 2, 28, and 31 have been amended.

Claims 32-41 have been added.

Claims 2-41 are pending.

Response to Arguments

Double Patenting Rejection

The double patenting rejection will not be withdrawn because applicant in remarks filed on 08/08/2017 indicated that the double patenting rejection be held in abeyance until the other rejections are overcome.

35 U.S.C. 103 Rejections

Applicant's argument filed 08/08/2017 has been fully considered but they are deemed not persuasive in view of new grounds of rejection.

Double Patenting

The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory double patenting rejection is appropriate where the claims at issue are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the reference application or patent either is shown to

be commonly owned with this application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement. A terminal disclaimer must be signed in compliance with 37 CFR 1.321(b).

The USPTO internet Web site contains terminal disclaimer forms which may be used. Please visit http://www.uspto.gov/forms/. The filing date of the application will determine what form should be used. A web-based eTerminal Disclaimer may be filled out completely online using web-screens. An eTerminal Disclaimer that meets all requirements is auto-processed and approved immediately upon submission. For more information about eTerminal Disclaimers, refer to http://www.uspto.gov/patents/process/file/efs/guidance/eTD-info-I.jsp.

Claim 2, 28, and 31 are rejected on the ground of nonstatutory double patenting as being unpatentable over claims 1 and 8 of U.S. Patent No. 8,676,877 in view of Howard et al. (U.S. PGPub 2003/0195938). Although the conflicting claims are not identical, they are not patentably distinct from each other because: see table below.

Instant Application		US Patent 8,676,877	
Claim	**Limitation**	**Claim**	**Limitation**
2, 28, and 31	**A computer cluster comprising:**	**1 and 8**	**A system for performing an instruction received from a front end by executing commands on one or more special purpose microprocessors, the system comprising:**

	a plurality of nodes, comprising a hardware processor wherein one or more of the nodes are configured to receive a command to start a cluster initialization process for the computer cluster, and **wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel that when executed causes the hardware processor to interpret user instructions <u>to evaluate mathematical expressions and to produce results of mathematical expression evaluation</u> ; and a mechanism for the nodes to communicate <u>results of mathematical expression evaluation</u> with each other using a peer-to-peer architecture; wherein at least one of the nodes is configured <u>to return at least one result of mathematical expression</u> evaluation to a user interface or a script**		**a plurality of nodes, wherein each node is configured to access a computer-readable memory system comprising program code for a single-node kernel module, ……..** **wherein each cluster node module is stored in a computer-readable memory system and configured to communicate with a single-node kernel and with one or more other cluster node modules, to accept instructions, and to interpret at least some of the instructions such that the plurality of cluster node modules communicate with one another in order to act as a cluster in executing commands using one or more hardware processors**

It is noted that (U.S. Patent No. 8,676,877) teaches substantially the same limitation as Instant application's independent claims while failing to teach the limitation underlined in the table above. However Howard teaches this in paragraph [0017] "…transmitting an algorithm computation request and associated data from a requesting host to a home node of a computing system wherein the request includes a requested number (N) of processing nodes to be applied to computation of the request. …distributes the <u>computation request (Note: this is the mathematical expression)</u> from the home node to a plurality of processing nodes wherein the plurality of processing

nodes includes N processing nodes coupled to the home node and wherein the distribution is in a hierarchical ordering. Then the method broadcasts the associated data from the home node to all of the plurality of processing nodes". Howard in paragraph [0019] also teaches "…agglomerates a final computation result from partial computation results received from the plurality of processing nodes wherein the agglomeration is performed in the reverse order of the hierarchical ordering, and returns the final computation result from the home node to the requesting host". Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine these teaching because this enables each remote host to communicate and parallel process algorithm code and data within the cluster but without direct communication with individual cluster nodes and/or detailed knowledge of cluster topology.

Claim Rejections - 35 USC § 103

The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 2, and 4-41 are rejected under 35 U.S.C. 103(a) as being unpatentable over Block et al. (U.S. PGPub 2005/0021751) hereinafter Block in

view of Singh (US Patent 8,601,101) hereinafter Singh and further in view of Howard et al. (U.S. PGPub 2003/0195938) hereinafter Howard.

As per claim 2, Block teaches a computer cluster ***(Block, ….cluster data port services within a cluster infrastructure. See paragraph [0025])*** comprising:

a plurality of nodes, comprising a hardware processor wherein one or more of the nodes are configured to receive a command to start a cluster initialization process for the computer cluster, and wherein each of the nodes is configured to access a non-transitory computer-readable medium comprising program code for a single-node kernel ***(Block, …..multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044])*** that when executed causes the hardware processor to interpret user instructions; and a mechanism for the nodes to communicate with each other. ***(Block, establishment of multiple network connections between a source node, a target node and one or more backup nodes in such a manner that a cluster data port is effectively utilized as single data port. See paragraph [0025]).***

Block doesn't explicitly teach that the nodes communicate with each other using a peer-to-peer architecture and also that one of the nodes returns a result to a user interface or a script. In analogous art Singh teaches nodes that communicate with each other using a peer-to-peer architecture. ***(Singh, …….each node in a cluster including a peer-to-peer communication channel compiling and maintaining its own cluster***

membership are disclosed……. each cluster node may be coupled to every other node and commonly-accessible storage through a network. …..Based on the cluster membership information, the joining node may request a peer-to-peer connection with each potential cluster member. See Column 2 line 26-38). Singh also teaches one of the nodes returns a result to a user interface or a script ***(Singh, The cluster configuration service 1510A may include a user interface to allow a system administrator to perform various node management functions through administrative console 1500. When the administrator makes changes to the node 240A configuration, the cluster configuration service 1510A may record all of the modifications that are made. See Column 14 line 61-66).***

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to take the teaching of Singh and apply them on the teaching of Block because when each node on the cluster communication channel is in peer-to-peer connection, each node can detect the connection failure based on the socket communication with the failed node and can update its cluster membership data without the need for a central entity to pass an updated membership list among nodes. ***(Singh, See column 13 line 23-28).***

Block-Singh doesn't explicitly teach that the clusters evaluate mathematical expressions and produce results of Mathematica expressions. Block-Singh also doesn't teach return results of mathematical expression evaluation to a user interface. In analogous art Howard teaches clusters evaluate mathematical expressions and produce results of Mathematica expressions ***(Howard, …transmitting an algorithm***

computation request and associated data from a requesting host to a home node of a computing system wherein the request includes a requested number (N) of processing nodes to be applied to computation of the request. …distributes the computation request (Note: this is the mathematical expression) from the home node to a plurality of processing nodes wherein the plurality of processing nodes includes N processing nodes coupled to the home node and wherein the distribution is in a hierarchical ordering. Then the method broadcasts the associated data from the home node to all of the plurality of processing nodes. See paragraph [0017]). Howard also teaches return results of mathematical expression evaluation to a user interface ***(Howard, …agglomerates a final computation result from partial computation results received from the plurality of processing nodes wherein the agglomeration is performed in the reverse order of the hierarchical ordering, and returns the final computation result from the home node to the requesting host. See paragraph [0017]. …returning the partial result to the requesting host as a final result in response to further direction to complete processing of the complex algorithm. See paragraph [0019]).***

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to take the teaching of Howard and apply them on the teaching of Block-Singh as doing so would enable a cluster of nodes to act as a single machine to one or more remote host computers outside of the inter-cluster network and also enables each remote host to communicate and parallel process algorithm code and

data within the cluster but without direct communication with individual cluster nodes and/or detailed knowledge of cluster topology. ***(Howard, see paragraph [0068]).***

As per claim 4, the modified Block-Singh-Howard teaches wherein the mechanism comprises a message passing interface. ***(Block, a single data port instance may also support target to source message sends. See paragraph [0050]).***

As per claim 5, the modified Block-Singh-Howard teaches wherein each of the nodes comprises one or more cluster node modules. ***(Block, cluster data port services within a cluster infrastructure to provide reliable and efficient communications between nodes. See paragraph [0010]).***

As per claim 6, the modified Block-Singh-Howard teaches wherein each single-node kernel is stored in a non-transitory computer-readable medium and configured to accept and execute a request. ***(Block, The routines executed to implement ……operating system or a specific application, component, program, object, module or sequence of instructions, will also be referred to herein as "computer program code," or simply "program code." See paragraph [0039], [0049]).***

As per claim 7, the modified Block-Singh-Howard teaches wherein each of the nodes comprises one or more cluster node modules, wherein each of the cluster node

modules comprises instructions stored in a non-transitory computer- readable medium and wherein the instructions when executed by the hardware processor cause the cluster node module to communicate with the single-node kernel and with one or more other cluster node modules. ***(Block, …..multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044], [0039]).***

As per claim 8, the modified Block-Singh-Howard teaches wherein the plurality of cluster node modules act as a cluster. ***(Block, utilize cluster data port services within a cluster infrastructure to provide reliable and efficient communications between nodes in a clustered computer system. See paragraph [0010]).***

As per claim 9, the modified Block-Singh-Howard teaches wherein the plurality of cluster node modules communicate with one another to act as a cluster. ***(Block, Nodes 12, 14 and 16 are typically coupled together via a clustering network 18. See paragraph [0028]).***

As per claim 10, the modified Block-Singh-Howard teaches wherein the computer cluster includes the user interface. ***(Block, Cluster manager application 60 that provides the user interface whereby a user such as a systems administrator can manage clustering operations in the system. See paragraph [0037]).***

As per claim 11, the modified Block-Singh-Howard teaches wherein each cluster node module accepts instructions from the user interface and interprets one or more of the instructions. ***(Block, Cluster manager application 60 that provides the user interface whereby a user such as a systems administrator can manage clustering operations in the system. See paragraph [0037], [0039]).***

As per claim 12, the modified Block-Singh-Howard teaches wherein the cluster initialization process comprises establishing communication among two or more of the nodes. ***(Bloack, communications between nodes in a clustered computer system. See paragraph [0010]).***

As per claim 13, the modified Block-Singh-Howard teaches wherein the cluster initialization process comprises configuring access by one or more of the nodes to a computer-readable medium comprising program code for the single-node kernel. ***(Block, The cluster data port services described hereinafter, on the other hand, provide an abstracted transport service that encapsulates and manages the establishment of multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044]).***

As per claim 14, the modified Block-Singh-Howard teaches wherein the cluster initialization process comprises establishing message-passing support among the nodes in the cluster. ***(Block, data port services may also be configured to provide synchronous and asynchronous caller send models, as well as support message encryption. As such, the herein-described services may be used to provide a general messaging service that allows a variety of operating system or kernel components to make use of the services in a clustered environment. See paragraph [0045]).***

As per claim 15, the modified Block-Singh-Howard teaches wherein the cluster initialization process comprises launching cluster node modules by the cluster. ***(Block, see paragraph [0011], [0012]).***

As per claim 16, the modified Block-Singh-Howard teaches wherein the cluster initialization process comprises assigning a processor identification number to each of the cluster node modules. ***(Block, Data port internal object 72 also has a data port address table 78, which stores remote IP address, node ID pairs relating to the available IP addresses through which a particular node (identified by the node ID) is coupled to the clustering network. See paragraph [0059]).***

As per claim 17, the modified Block-Singh-Howard teaches wherein the cluster initialization process comprises: launching cluster node modules by the cluster; and

after launching the cluster node modules, configuring access by one or more of the nodes to a non-transitory computer-readable medium comprising program code for the single-node kernel. ***(Block, data port services may also be configured to provide synchronous and asynchronous caller send models, as well as support message encryption. As such, the herein-described services may be used to provide a general messaging service that allows a variety of operating system or kernel components to make use of the services in a clustered environment. See paragraph [0045]).***

As per claim 18, the modified Block-Singh-Howard teaches wherein the cluster initialization process comprises: launching cluster node modules by the cluster; after launching the cluster node modules, establishing communication among two or more of the nodes; ***(Block, a cluster data port consistent with the invention may be used to facilitate the transfer of large volumes of data between a source node and specified target nodes in a clustering environment. See paragraph [0012])*** and after establishing communication among two or more of the nodes, assigning a processor identification number to each of the cluster node modules. ***(Block, Data port internal object 72 also has a data port address table 78, which stores remote IP address, node ID pairs relating to the available IP addresses through which a particular node (identified by the node ID) is coupled to the clustering network. See paragraph [0059]).***

As per claim 19, the modified Block-Singh-Howard teaches wherein one or more of the nodes are configured to communicate at least some of the user instructions. ***(Block, establishment of multiple network connections between a source node, a target node and one or more backup nodes in such a manner that a cluster data port is effectively utilized as single data port. See paragraph [0025], [0044]).***

As per claim 20, the modified Block-Singh-Howard teaches wherein one or more of the nodes are configured to communicate at least some of the user instructions to one or more single-node kernel. ***(Block, …..multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044]).***

As per claim 21, the modified Block-Singh-Howard teaches wherein one or more of the nodes are configured to accept user instructions via one or more of the nodes. ***(Block, see paragraph [0025], [0044]).***

As per claim 22, the modified Block-Singh-Howard teaches wherein one or more of the nodes are configured to communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other. ***(Block, see paragraph [0025], [0044]).***

As per claim 23, the modified Block-Singh-Howard teaches wherein one or more of the nodes are configured to transmit at least some of the user instructions that originate from a user interface. ***(Singh, The cluster configuration service 1510A may include a user interface to allow a system administrator to perform various node management functions through administrative console 1500. When the administrator makes changes to the node 240A configuration, the cluster configuration service 1510A may record all of the modifications that are made. See Column 14 line 61-66).***

As per claim 24, the modified Block-Singh-Howard teaches wherein one or more of the nodes are configured to parallelize at least some of the user instructions before communicating at least some of the user instructions to one or more single-node kernels. ***(Block, …..multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044]).***

As per claim 25, the modified Block-Singh-Howard teaches wherein one or more of the nodes are configured to accept user instructions before communicating at least some of the user instructions to one or more single-node kernels. ***(Singh, The cluster configuration service 1510A may include a user interface to allow a system administrator to perform various node management functions through***

administrative console 1500. When the administrator makes changes to the node 240A configuration, the cluster configuration service 1510A may record all of the modifications that are made. See Column 14 line 61-66).

As per claim 26, the modified Block-Singh-Howard teaches wherein one or more of the nodes are configured to: accept user instructions; after accepting user instructions, communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other; ***(Block, see paragraph [0025], [0044])*** and after communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels. ***(Block, …..multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044]). (Also see Singh Column 14 line 61-66).***

As per claim 27, the modified Block-Singh-Howard teaches wherein the plurality of nodes are configured to communicate with one another to interpret and translate commands for execution by a plurality of single-node kernels. ***(Block, …..multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044]). (Also see Singh Column 14 line 61-66).***

As per claim 28, Block teaches a computer cluster comprising: a plurality of nodes, ***(Block, ….cluster data port services within a cluster infrastructure. See paragraph [0025])*** wherein one or more of the nodes are configured to receive:

a command to start a cluster initialization process for the computer cluster, wherein the cluster initialization process comprises establishing communication among two or more of the nodes; ***(Block, establishment of multiple network connections between a source node, a target node and one or more backup nodes in such a manner that a cluster data port is effectively utilized as single data port. See paragraph [0025]).***

and a mechanism for the nodes to communicate with each other using asynchronous calls; wherein each of the nodes is configured to access a non-transitory computer- readable medium comprising program code for a single-node kernel configured to interpret user instructions; and wherein at least one of the nodes is configured to return a result to the user interface or the script. ***(Block, …..multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044]).***

Block doesn't explicitly teach receiving an instruction from a user interface or a script. In analogous art Singh teaches receiving instruction from a user interface in a computer cluster. ***(Singh, The cluster configuration service 1510A may include a user interface to allow a system administrator to perform various node management functions through administrative console 1500. When the***

administrator makes changes to the node 240A configuration, the cluster configuration service 1510A may record all of the modifications that are made. See Column 14 line 61-66).

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to take the teaching of Singh and apply them on the teaching of Block because it would allow a system administrator to perform various node management functions through administrative console. ***(Singh, … allow a system administrator to perform various node management functions through administrative console 1500. When the administrator makes changes to the node 240A configuration, the cluster configuration service 1510A may record all of the modifications that are made. See column 14 line 61-66).***

Block-Singh doesn't explicitly teach that the clusters evaluate mathematical expressions and produce results of Mathematica expressions. Block-Singh also doesn't teach return results of mathematical expression evaluation to a user interface. In analogous art Howard teaches clusters evaluate mathematical expressions and produce results of Mathematica expressions ***(Howard, …transmitting an algorithm computation request and associated data from a requesting host to a home node of a computing system wherein the request includes a requested number (N) of processing nodes to be applied to computation of the request. …distributes the computation request (Note: this is the mathematical expression) from the home node to a plurality of processing nodes wherein the plurality of processing nodes includes N processing nodes coupled to the home node and wherein the***

distribution is in a hierarchical ordering. Then the method broadcasts the associated data from the home node to all of the plurality of processing nodes. See paragraph [0017]). Howard also teaches return results of mathematical expression evaluation to a user interface ***(Howard, …agglomerates a final computation result from partial computation results received from the plurality of processing nodes wherein the agglomeration is performed in the reverse order of the hierarchical ordering, and returns the final computation result from the home node to the requesting host. See paragraph [0017]. …returning the partial result to the requesting host as a final result in response to further direction to complete processing of the complex algorithm. See paragraph [0019]).***

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to take the teaching of Howard and apply them on the teaching of Block-Singh as doing so would enable a cluster of nodes to act as a single machine to one or more remote host computers outside of the inter-cluster network and also enables each remote host to communicate and parallel process algorithm code and data within the cluster but without direct communication with individual cluster nodes and/or detailed knowledge of cluster topology. ***(Howard, see paragraph [0068]).***

As per claim 29, the modified Block-Singh-Howard teaches wherein the asynchronous calls comprise a first command to create a first packet containing: an expression to be sent as payload; and a target node where the expression should be sent; ***(Block, provide synchronous and asynchronous caller send models, as well***

as support message encryption. As such, the herein-described services may be used to provide a general messaging service that allows a variety of operating system or kernel components to make use of the services in a clustered environment. See paragraph [0045]) wherein the first command is configured to be called from within a single-node kernel; wherein the single-node kernel is configured to send the first packet to a local cluster node module; and wherein the local cluster node module is configured to forward the expression to the target node ***(Block, …..multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044]). (Also see Singh Column 14 line 61-66).***

As per claim 30,

[Rejection rational for claim 29 is applicable].

As per claim 31,

[Rejection rational for claim 2 and 28 is applicable].

As per claim 32, the modified Block-Singh-Howard teaches the computer cluster of claim 5, wherein each of the plurality of nodes implements asynchronous calls that enable the single-node kernel to perform computation tasks while the cluster node modules are simultaneously communicating with one another. ***(Block, …cluster data***

port services may also be configured to provide synchronous and asynchronous caller send models, as well as support message encryption. As such, the herein-described services may be used to provide a general messaging service that allows a variety of operating system or kernel components to make use of the services in a clustered environment. See paragraph [0045]).

As per claim 33, the modified Block-Singh-Howard teaches the computer cluster of claim 5, wherein intercommunication among the plurality of single-node kernels during thread execution is enabled by the plurality of cluster node modules, and wherein the computer cluster is configured to permit exchange of information between nodes during the course of a parallel computation ***(Block, The cluster data port services described hereinafter, on the other hand, provide an abstracted transport service that encapsulates and manages the establishment of multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating system or kernel component) like a single data port or data pipe. See paragraph [0044]).***

As per claim 34, the modified Block-Singh-Howard teaches the computer cluster of claim 5, wherein each of the single-node kernels are configured to call a send command involving creating a packet containing: an expression to be sent as payload; ***(Block, ...multiple logical or TCP connections to a designated target node with designated backup target nodes, which appears to the user (i.e., an operating***

system or kernel component) like a single data port or data pipe. See paragraph [0044]) and where the expression should be sent; wherein, upon reception of the send command by a local cluster node module associated with the single-node kernel, the local cluster node module is configured to decode the packet and forward a payload to the cluster node module specified in the packet; ***(Block, …message M1 is transmitted between the sockets on the source and target nodes. Upon receipt of the message, the socket reports to the data port of the target node that the record is complete (i.e., the message has been received and stored in a record of an IOCP queue). See paragraph [0084])*** wherein each of the single-node kernels is configured to call a receipt command involving creating a packet specifying which message to test for completion and to then wait for a reply expression to evaluate; wherein, upon reception of the receipt command by the local cluster node module associated with the single-node kernel, the local cluster node module is configured to decode the packet and use a message specifier to search for any matching expressions listed as completed in a received message queue; ***(Block, The data port notifies the user on the target node that such data has been received via a callback function rcvddata(M1,SN), where SN is the sequence number of the message. The user then processes the received message, and initiates the communication of an acknowledgment to the source node via a sendack(SN,RC) call to the data port, with RC storing a return code appropriate for the message. See paragraph [0084])*** wherein, upon determining that such a completed expression is found, the local cluster node module is configured to send the completed expression to a local single- node

kernel associated with the cluster node module in response to the receipt command, and wherein each of the single-node kernels is configured to receive the completed expression and to update variables used by the single-node kernel during evaluation of expressions ***(Block, Once the acknowledgment is received by the socket, the socket notifies the data port that N acknowledgments have been received (to account for the possibility that multiple acknowledgments may be received). In response to such notification, the data port issues a callback function to the user to free the buffer allocated for the message M1, and to forward the return code RC generated by the target node. See paragraph [0085]).***

As per claim 35, the modified Block-Singh doesn't explicitly teach the computer cluster of claim 2 wherein the plurality of nodes are configured to permit exchange of information between nodes during the course of parallel computation. In analogous art Howard teaches the computer cluster of claim 2 wherein the plurality of nodes are configured to permit exchange of information between nodes during the course of parallel computation ***(Howard, …parallel computing by providing methods and systems for configuration of and operation of a Howard Cascade (HC). Algorithm computation requests are transmitted to a home node of a HC and distributed (cascaded) through nodes of the HC to expand the algorithm computation request to a desired number of processing nodes. See paragraph [0013]).***

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to take the teaching of Howard and apply them on the teaching

of Block-Singh as doing so would enable a cluster of nodes to act as a single machine to one or more remote host computers outside of the inter-cluster network and also enables each remote host to communicate and parallel process algorithm code and data within the cluster but without direct communication with individual cluster nodes and/or detailed knowledge of cluster topology. ***(Howard, see paragraph [0068]).***

As per claim 36, the modified Block-Singh-Howard teaches the computer cluster of claim 2, wherein each of the plurality of nodes comprises instructions executable by the hardware processor and configured to implement asynchronous behavior, wherein the instructions comprise: a first instruction to asynchronously send a payload to another node; ***(Block, The cluster data port services may rely on sockets asynchronous I/O completion ports (IOCPs) to minimize the number of tasks required to support the service, given the potential for multiple logical or TCP connections being established between nodes. The services may also provide the structure to support asynchronous sends, receive callbacks, and improve efficiency for load balancing sends across multiple TCP connections. Enhanced socket support coupled with cluster data port services may also insure that ordered message delivery is maintained in spite of load balanced messaging across multiple connections simultaneously. See paragraph [0051])*** a second instruction to asynchronously receive a payload from another node; and a third instruction to search for a payload matching a message specifier. ***(Block, …cluster data port services may also be configured to provide synchronous and***

asynchronous caller send models, as well as support message encryption. As such, the herein-described services may be used to provide a general messaging service that allows a variety of operating system or kernel components to make use of the services in a clustered environment. See paragraph [0045]).

As per claim 37,

[Rejection rational for claim 2 and 28 is applicable].

As per claim 38, the modified Block-Singh doesn't explicitly teach the computer cluster node of claim 37, wherein the one or more non-transitory memory devices comprise program code for performing a parallel fast Fourier transform, wherein the program code causes the hardware processor to evaluate a command to perform a Fourier transform on an array comprising a first data portion that is stored on the computer cluster node and a second data portion that is not stored on the computer cluster node. In analogous art Howard teaches the computer cluster node of claim 37, wherein the one or more non-transitory memory devices comprise program code for performing a parallel fast Fourier transform, wherein the program code causes the hardware processor to evaluate a command to perform a Fourier transform on an array comprising a first data portion that is stored on the computer cluster node and a second data portion that is not stored on the computer cluster node ***(Howard, Decomposition of Parallel Processing Data for Two-Dimensional Convolution Using Fast Fourier Transforms, see paragraph [0437], a two-dimensional convolution using Fast***

Fourier Transforms may be partitioned on a HC. The partitioning can handle an arbitrary number of parallel processing nodes and arbitrarily large `images` ***see paragraph [0438], [0439]).***

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to take the teaching of Howard and apply them on the teaching of Block-Singh as doing so would enable a cluster of nodes to act as a single machine to one or more remote host computers outside of the inter-cluster network and also enables each remote host to communicate and parallel process algorithm code and data within the cluster but without direct communication with individual cluster nodes and/or detailed knowledge of cluster topology. ***(Howard, see paragraph [0068]).***

As per claim 39,

[Rejection rational for claim 35 is applicable].

As per claim 40,

[Rejection rational for claim 36 is applicable].

As per claim 41, the modified Block-Singh doesn't explicitly teach the computer cluster node of claim 37 wherein the hardware processor comprises a special purpose microprocessor. In analogous art Howard teaches the computer cluster node of claim 37 wherein the hardware processor comprises a special purpose microprocessor ***(Howard, …a first processor 420 connected to a second processor 422 via a***

single communication channel 424. (Note: this are special purpose processors) Multiple parallel channels of expansion, on the other hand, consist of multiple computer processors and multiple communication channels, as shown in FIG. 33. Processor 430 and processor 432 of processing node 112 are connected to processor 34 and processor 436 of processing node 114 by two communication channels 438 and 440, respectively. Other parallel channel configurations follow and may be implemented such that a HC moves data as efficiently as desired. See paragraph [0167]).

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to take the teaching of Howard and apply them on the teaching of Block-Singh as doing so would enable a cluster of nodes to act as a single machine to one or more remote host computers outside of the inter-cluster network and also enables each remote host to communicate and parallel process algorithm code and data within the cluster but without direct communication with individual cluster nodes and/or detailed knowledge of cluster topology. ***(Howard, see paragraph [0068]).***

Claim 3 is rejected under 35 U.S.C. 103(a) as being unpatentable over Block et al. (US PGPub 2005/0021751) hereinafter Block in view of Singh (US Patent 8,601,101) hereinafter Singh and further in view of Gray (USPGPub 2007/0073705) and further in view of Howard et al. (U.S. PGPub 2003/0195938) herein Howard.

As per claim 3, Block-Singh-Howard teaches the computer cluster of claim 2 but doesn't teach a Mathematica kernel. In analogous art Gray teaches a Mathematica kernel. ***(Gray, The front end may include an interactive document referred to as a notebook similar to those often used with MATHEMATICA.RTM. software systems. A notebook may include input to be sent to the kernel 104 and output received from the kernel, as well as text, graphics, palettes, etc. see paragraph [0031]).***

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to take the teaching of Gray and apply them on the teaching of Block-Singh-Howard because the kernel and the front end may be implemented on a same computing system or on different computing systems that are communicatively coupled to one another. ***(Gray, see paragraph [0030]).***

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HERMON ASRES whose telephone number is (571)272-4257. The examiner can normally be reached on Monday - Friday 7:30 am to 5:00 pm est.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, VIVEK SRIVASTAVA can be reached on 571-272-7304. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/HERMON ASRES/
Examiner, Art Unit 2449

Notice of References Cited	Application/Control No. 14/181,112	Applicant(s)/Patent Under Reexamination TANNENBAUM ET AL.	
	Examiner HERMON ASRES	Art Unit 2449	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-2005/0021751 A1	01-2005	Block, Timothy Roy	G06F9/54	709/225
*	B	US-2007/0073705 A1	03-2007	Gray; Theodore W.	G06F9/4443	1/1
*	C	US-8,601,101 B1	12-2013	Singh; Servesh Pratap	H04L69/40	370/254
*	D	US-6,968,359 B1	11-2005	Miller; Robert	G06F9/5061	709/205
*	E	US-2007/0094532 A1	04-2007	SENGUPTA; Sabyasachi	G06F11/3632	714/5.1
*	F	US-2003/0195938 A1	10-2003	Howard, Kevin David	G06F8/45	709/208
	G	US-				
	H	US-				
	I	US-				
	J	US-				
	K	US-				
	L	US-				
	M	US-				

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

PTO/SB/08 Equivalent

INFORMATION DISCLOSURE STATEMENT BY APPLICANT	Application No.	14/181112
	Filing Date	February 14, 2014
	First Named Inventor	Zvi Tannenbaum
	Art Unit	2449
(Multiple sheets used when necessary)	Examiner	Asres, Hermon
SHEET 1 OF 1	Attorney Docket No.	ZTANN.002P1C3

U.S. PATENT DOCUMENTS					
Examiner Initials	Cite No.	Document Number *Number - Kind Code (if known)* Example: 1,234,567 B1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear

FOREIGN PATENT DOCUMENTS						
Examiner Initials	Cite No.	Foreign Patent Document *Country Code-Number-Kind Code* Example: JP 1234567 A1	Publication Date MM-DD-YYYY	Name	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear	T[1]

NON PATENT LITERATURE DOCUMENTS			
Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[1]
	1	English translation of MATSUMURA, et al., "Construction of distributed computing system for large-scale matrices intended or reduction of communication blocks," The Special Interest Group Technical Reports of IPSJ, Information Processing Society of Japan, March 5, 1998, Vol. 98, No. 18, pp. 19-24.	T

27161567

Examiner Signature	/HERMON ASRES/	Date Considered	02/19/2018

*__Examiner:__ Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

T[1] - Place a check mark in this area when an English language Translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.A/

PTO/SB/08 Equivalent

INFORMATION DISCLOSURE STATEMENT BY APPLICANT	Application No.	14/181112
	Filing Date	February 14, 2014
	First Named Inventor	Tannenbaum, Zvi
	Art Unit	2449
(Multiple sheets used when necessary)	Examiner	Asres, Hermon
SHEET 1 OF 1	Attorney Docket No.	ZTANN.002P1C3

U.S. PATENT DOCUMENTS

Examiner Initials	Cite No.	Document Number *Number - Kind Code (if known)* Example: 1,234,567 B1	Publication Date MM-DD-YYYY	Name of Patentee or Applicant	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear
	1	2004/0195572	10-07-2004	Kato et al.	

FOREIGN PATENT DOCUMENTS

Examiner Initials	Cite No.	Foreign Patent Document *Country Code-Number-Kind Code* Example: JP 1234567 A1	Publication Date MM-DD-YYYY	Name of Patentee or Applicant	Pages, Columns, Lines Where Relevant Passages or Relevant Figures Appear	T[1]
	2	JP 11-328134 w/ English Abstract	11-30-1999	Hitachi Ltd.		
	3	JP 2004-061359 w/ English Abstract	02-26-2004	Ricoh Co. Ltd.		

NON PATENT LITERATURE DOCUMENTS

Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[1]
	4	KONISHI, et al., "Performance Evaluation of Parallel Computing Tool for MATLAB on PC Grid Environment," IPSJ SIG Technical Reports, August 5, 2005, Vol. 2005, No. 84. 9 pages.	
	5	MATSUMURA, et al. "Construction of distributed computing system for large-scale matrices intended or reduction of communication blocks," August 5, 1998, Vol. 98, No. 18. 9 pages.	
	6	TATEBE, et al., "Efficient Implementation of MPI Using Remote Memory Write," Transactions of Information Processing Society of Japan, May 1999, Vol. 40, No. 5. 14 pages.	

Examiner Signature /HERMON ASRES/	Date Considered 02/19/2018

***Examiner:** Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

T[1] - Place a check mark in this area when an English language Translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.A/

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	6	((Zvi) near2 (Tannenbaum)).INV.	USPAT	OR	OFF	2015/07/24 20:25
S2	7	((Dean) near2 (Dauger)).INV.	USPAT	OR	OFF	2015/07/24 20:25
S3	177	(computer with cluster) same nodes same kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:26
S4	24	S3 and interpret	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:26
S5	28950	709/223.ccls.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:26
S6	5291	H04L41/0213.cpc.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:27
S7	1692	H04L29/08072.cpc.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:27
S8	27776	H04L29/06.cpc.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:27
S9	6793	H04L41/22.cpc.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S10	8156	H04L41/12.cpc.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S11	0	S3 and S6	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28

S12	0	S3 and S7	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S13	15	S3 and S8	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S14	0	S3 and S9	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S15	2	S3 and S10	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/24 20:28
S16	99	("20020049859" | "20030005266" | "20030051062" | "20030135621" | "20030195931" | "20030195938" | "20040110209" | "20040157203" | "20040252710" | "20050021751" | "20050038852" | "20050060237" | "20050076105" | "20050108394" | "20050180095" | "20060053216" | "20060106931" | "20070044099" | "20070073705" | "20080250347" | "20080281997" | "5423046" | "5881315" | "5930768" | "6108699" | "6202080" | "6546403" | "6578068" | "6751698" | "6782537" | "6968335" | "7093004" | "7136924" | "7174381" | "7249357" | "7334232" | "7437460" | "7472193" | "7554949" | "7937455" | "8082289" | "8140612" | "8402083" | "8676877").PN.	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/25 12:54
S17	177	(computer with cluster) same nodes same kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/25 13:19
S18	10	(computer with cluster) same nodes same kernel and mathematica	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/25 13:20
S19	177	(computer with cluster) same nodes same kernel	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/07/25 13:23
S20	10	("8402080").PN.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2015/11/29 23:38
S21	2	("8849889").PN.	US-PGPUB; USPAT;	OR	OFF	2015/11/29 23:38