

メニーコア環境におけるノード内通信の高速化  
および効率化を実現するタスクモデルの研究

2017 年度

島 田 明 男

# 主 論 文 要 旨

No.1

報告番号	甲 乙 第	号	氏 名	島田明男
主論文題名：				
メニーコア環境におけるノード内通信の高速化および効率化を実現するタスクモデルの研究				
(内容の要旨)				
<p>近年、電力効率の観点から、コア単体の処理性能を向上させるよりも、コア数を増加させるCPUアーキテクチャが一般的になっており、High-performance Computing (HPC) システムを構成するノード1台あたりのコア数は飛躍的に増加してきている。一方、1コアあたりのメモリ量は減少する傾向にある。このようなメニーコア環境では、同一ノード内で動作する並列プロセスの数が従来のマルチコア環境よりも増加する。よって、ノード内通信（同一ノード内の並列プロセス間の通信）の発生回数が、従来よりも多くなる。ノード内通信の性能が、並列アプリケーションの性能に従来よりも大きな影響を与えることになり、より高速なノード内通信が求められる。また、1コアあたりのメモリ量は減少する傾向にあるため、ノード内通信に掛かるメモリ消費量の低減も求められる。しかし、従来のマルチコア環境向けのノード内通信では、並列プロセスのアドレス空間をまたいでデータ転送するために、通信遅延の増加やノード内通信に掛かるメモリ消費量の増加をまねいてしまう問題がある。</p> <p>そこで、本論文は、メニーコア環境において高速かつ効率的なノード内通信を実現するため、Partitioned Virtual Address Space (PVAS) と呼ぶ新たなタスクモデルを提案する。PVAS タスクモデルは、並列処理をおこなうノード内の並列プロセス群を同一アドレス空間で動作させることを可能にする。並列プロセスを同一アドレス空間で動作させることで、アドレス空間をまたいでデータ転送することなく、ノード内通信を実行することが可能になり、高速かつメモリ消費量の少ないメニーコア環境向けのノード内通信を実現することができる。PVAS タスクモデルをLinux カーネルのメモリ管理を対象に実装した。</p> <p>本論文では、PVAS タスクモデルを Message Passing Interface (MPI) の通信に適用し、PVAS タスクモデルの検証をおこなう。PVAS タスクモデルをMPI 通信に適用し、高速かつメモリ消費量の少ない、よりメニーコア環境に適したMPI ノード内通信を実現する。PVAS の適用は、MPI における連続データの送受信と不連続データの送受信を対象にしておこなう。実装は、MPI の Open Source Software (OSS) 実装の一つである Open MPI に、PVAS を利用するメニーコア環境向けのノード内通信モジュールを組み込むことでおこなう。これを、マイクロベンチマークとミニアプリケーションにより評価し、マルチコア環境向けのMPI ノード内通信モジュールを用いた場合と比較する。マイクロベンチマークによる評価では、連続データの送受信において、通信遅延が大きく改善することを示す。また、不連続データの送受信においても、小さなデータが多数不連続にメモリ上に配置される通信パターンを除き、通信遅延が大きく改善することを示す。連続データを送受信するミニアプリケーションによる評価では、最大で約18%、実行性能が改善することを示す。不連続データを送受信するミニアプリケーションによる評価では、最大で約21%、実行性能が改善することを示す。また、マイクロベンチマークによってMPI ノード内通信のメモリ消費量を測定し、最大で約18%、メモリ消費量を削減可能であることを示す。これらの評価により、本論文で提案するPVAS タスクモデルによって、メニーコア環境において高速かつ効率的なノード内通信を実現可能であることを示す。</p>				

## Thesis Abstract

No. 1

Registration Number	<input checked="" type="checkbox"/> "KOU" <input type="checkbox"/> "OTSU"	Name	Akio Shimada
No.	*Office use only		
Thesis Title			
A Study on Task Models for High-performance and Efficient Intra-node Communication in Many-core Environments			
Thesis Summary			
<p>Having reached the evolutionary limits of single-core performance in terms of power-efficiency, increasing the number of cores being incorporated into processors is becoming popular in recent years. Thus, the number of cores in High-performance Computing (HPC) system is growing rapidly. Meanwhile, per-core memory size is becoming smaller. In such many-core environments, the number of parallel processes running in a node becomes larger. Therefore, the number of times of intra-node communications (communications among parallel processes in the same node) taking places in a node becomes larger than that in multi-core environments. Then, the performance of intra-node communication gives a greater impact on parallel application performance. As a result, high-performance intra-node communication is required. Moreover, reducing memory footprint for intra-node communication is required because per-core memory size becomes smaller in many-core environments. However, in case of intra-node communication for multi-core environments, transmitting data crossing address space boundaries among parallel processes results in large communication latency and large memory footprint.</p> <p>In this thesis, Partitioned Virtual Address Space (PVAS), which is a new task model for achieving high-performance and efficient intra-node communication in many-core environments, is proposed. PVAS task model enables parallel processes in the same node to run in the same address space. Parallel processes running in the same address space can perform intra-node communication without crossing address space boundaries. As a result, the intra-node communication for many-core environments, which is high-performance and whose memory usage is small, can be achieved. PVAS task model was implemented by modifying the memory management of Linux kernel.</p> <p>In this thesis, the effectiveness of PVAS task model is verified by applying it to Message Passing Interface (MPI) communication. MPI intra-node communication for many-core environments, which is high-performance and whose memory usage is small, can be achieved by utilizing PVAS task model. PVAS task model is applied to MPI intra-node communications for transmitting both contiguous and non-contiguous data. Intra-node communication module for many-core environments, which utilizes PVAS task model, has been implemented into Open MPI (Open MPI is one of the open source MPI implementations) and compared to existing intra-node communication module for multi-core environments by micro-benchmarks and mini-applications. Micro-benchmark results show that PVAS task model accelerates MPI intra-node communication for transmitting contiguous data and accelerates MPI intra-node communication for transmitting non-contiguous data excluding several communication patterns, which sends a large number of small data discontinuously stored on the memory. PVAS task model improves the performance of mini-application performing contiguous data transfer by up to 18%, and the performance of</p>			

mini-application performing non-contiguous data transfer by up to 21%. Moreover, micro-benchmark results show that PVAS task model can reduce memory footprint for MPI intra-node communication by up to 18%. These results shows that PVAS task model achieves high-performance and high-efficient intra-node communication in many-core environments.