Name	Kenjiro Taura, Professor	Location	Hongo	Field	Parallel processing, programming languages

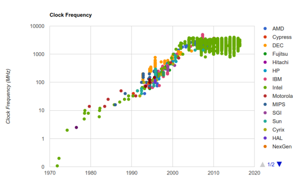
High Performance Computing and High Level Programming

Research Introduction: http://www.eidos.ic.i.u-tokyo.ac.jp/Research

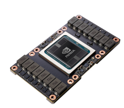
1 Background

Recent advances in intelligent information processing technologies such as machine learning are built on high performance computing. More traditionally, computer simulations are indispensable tools in all fields of science and technology (understanding the function of biomolecules, predicting climate change, designing new materials, understanding the formation of galaxies, etc.) and they have been crucially dependent on high performance computing too.

In the past, computers were getting faster without significantly rewriting software, thanks to the increasing serial processing performance of CPUs, particularly to frequencies. Recently, however, frequencies have ceased increasing (the end of Dennard scaling) and software performance can now improve only by parallel processing. What's more, it is commonly believed that the exponential doubling of transistor densities (so-called Moore's Law) will cease around 2025 too, after which performance improvement are believed to come mainly from domain-specific processors and languages and innovations in memory and networking technologies.



http://cpudb.stanford.edu/



http://www.nvidia.com/

2 Research Themes

Whether they are computers of today or emerging computers in future, application performance improvement no longer comes solely from hardware but critically depends on system software such as programming languages and operating systems. Particularly important is system software that automates processes to extract hardware performance such as parallelization, vectorization, load balancing, data placement, etc. thereby reducing the burden on the programmer.

To this end, we work at every layer of system software including compilers, runtime systems, operating systems, algorithms and performance analysis tools, with the central focus being on designing and implementing new parallel programming systems and domain-specific languages. Some of the ongoing projects are below.

Load Balancing and Distributed Shared Memory in Massively Parallel Environments This research aims at making programming on massively parallel machines such as supercomputers much easier. The ultimate goal is to automatically distribute data and allow programs to transparently access them as if all data are in a single machine in a range of environments from multicore nodes to large scale networks of them.

Performance Analysis and Optimization of Deep Learning Frameworks Deep learning frameworks are languages or libraries specifically designed for deep learning. We are analyzing and optimizing the performance of existing frameworks on various processors (CPUs and GPUs) and also developing a new execution engine.

High Performance String Data Processing Systems Being focused on text data processing, we are working on frameworks that achieve both ease of programming and high performance/parallelization. Given the grammar of the string to analyze, the system automatically generates a program (a parser) that analyzes it with vectorization and parallelization. We are building a system that analyzes large data with high performance on top of it.

Pattern-based Domain-specific Compilers Parallel patterns, which are abstract models structurizing parallel computing, are known to be useful for high-level parallel programming. Their library implementation, however, often faces with the limitation of their applicability to actual applications and the overhead derived from their abstraction. To alleviate these drawbacks, we are studying on domain-specific compilers, which apply domain-specific transformations and optimizations to pattern-based parallel programs.

Performance Analysis Tools of Parallel Programs It is all too common that parallel programs do not show satisfactory performance in the first attempt. We are working on tools that analyze what is going on at the bottom of it and help programmers achieve high performance.

Virtual Machine Monitors for Next Generation Non-Volatile Memory A new high performance memory technology is emerging that can be accessed just like ordinary memory (by load/store instructions of CPUs) and is non-volatile just like HDD/SSD (keeping data after power off). Using this technology, we are working on virtual machine monitors that provide virtual machines with a large memory efficiently with low cost.

We also welcome students who are interested in applying large scale and high performance computing toward socially important outcomes.

Corporate Data Mining We are working with Advanced-i Ltd. and Tsuruoka Laboratory on large-scale corporate data mining; we extract similarity and other relations among companies, so as to support a good business succession.

Medical Text Mining We are working with St. Luke Hospital on text mining from patient records so as to detect and prevent medical incidents.

We also welcome students interested in Digital Annealer, a device implementing quantum annealing with digital circuits.

3 Goal of Student Development

I want students to become individuals who "can explore by themselves" and "think deeply by themselves." In this era when answers to many questions and necessary pieces of information are just a google away, it has never been more important to be "roseau pensant." Whether you are a fluent programmer or not is only a part of your research excellence. I want those who are already good at programming to learn and master broader subjects new to you.