摘 要

高性能计算是计算机科学的重要分支，主要指从体系结构、并行算法和并行软件等多个方面研究和利用高性能计算机的技术。数十年来，高性能计算被广泛地应用在大规模科学和工程计算、人工智能、工业仿真等关键领域，对基础科学发现、国民经济发展和国防工业建设具有极高的应用价值，已成为国家科技综合实力的重要体现。

图灵奖得主 Jim Gray 指出，可扩展性问题是高性能计算中的核心问题，也是未来信息技术领域十二项研究目标中排在首位的重要挑战。针对高性能计算中的并行软件设计和并行系统结构一软一硬两个层次，可扩展性挑战也表现在架构移植难、算法设计难、应用优化难等诸多方面。本文对大规模并行多层次不连续非线性可扩展理论开展研究，深入分析可扩展性发展规律，提出物理模型、并行算法以及性能优化多层次协同设计方法，在多种硬件并行规模、不同软件并行粒度、各级交叉并行应用上开展可扩展性优化设计，具体贡献包括：

1. **提出了大规模并行多层次不连续非线性可扩展理论。**基于高性能计算中的不连续非线性可扩展现象，通过对并行软件在同构多核到异构众核系统上可扩展性的分析，首次系统地对多个层次上的两种现象进行了深入的理论分析与内涵丰富，并提出了可扩展性的物理模型-并行算法-性能优化多层次协同设计理论。该可扩展理论考虑了从底层物理方法建模到上层应用性能优化完整的高性能计算研究链，为高性能计算领域可扩展性的研究，特别是大规模并行计算中的可扩展问题，提供了系统化、体系化的方法论层面指导。

2. **提出了百核量级的 Stencil 并行数值算法，利用新型向量化和分块技术设计并实现高可扩展单机多核 Stencil 计算。**在并行算法层次，提出了转置布局计算和时空计算折叠两种 Stencil 向量化策略，提高数据在 CPU 内的并行度；同时在性能优化层次，提出了高效寄存器数据重用算法和缓存分块优化算法，提高数据的访存效率。实验结果表明，通过并行算法-性能优化的双层协同设计思想，最高超过 state-of-the-art 方法的绝对性能 4.39 倍，有效地提升了 Stencil 并行数值算法的多核可扩展性能。

3. **提出了万核量级的分布式机器学习框架，利用新型聚类和回归技术设计**

并实现高可扩展的多机众核机器学习预测模型。本文在物理模型层次，提出了新型 Best Friend Graph 图数据结构及层次化的最小生成树网络模型，并设计了基于聚类的回归预测方法；在并行算法层次，提出了基于回溯的负载均衡算法和高效的并行通信算法，降低分布式系统的计算和通信开销。实验结果表明，通过物理模型-并行算法的双层协同设计思想，在保证聚类和回归方法准确性的同时，还能有效地将分布式机器学习框架的多机众核扩展性由已有工作的 1,536 核提升至 12,288 核。

4. **针对科学计算软件应用优化，提出了百万核量级的大规模扩展方法并设计实现一套大规模高可扩展国产核材料辐照损伤模拟的软件应用 OpenKMC。**在物理模型层次，设计优化了高可扩展的新型势函数模型和分组反应策略，支撑应用高效动力学蒙特卡罗模型建立；在并行算法层次，提出了适应于大规模应用的并行同步象限算法和高效自适应通信算法，提高应用的负载均衡和通信效率；在性能优化层次，提出了访存优化技术、高效局部性算法、Athread 线程级异构并行和从核向量化加速方法，通过主存级-缓存级-寄存器级多层次化访存特征提取优化和轻量级的进程级-线程级-数据级多层次并行性挖掘对异构众核体系结构算力进行充分利用。实验结果表明，通过物理模型-并行算法-性能优化的多层次可扩展性协同设计思想，实现神威・太湖之光千亿原子的 520 万核大规模模拟，并行效率高达 80%，成为核材料模拟新的里程碑。

关键词：大规模并行；可扩展性；科学计算应用；并行数值算法；机器学习框架

Abstract

High performance computing is an important branch of computing science, which mainly refers to the research and utilization of high-performance computer technology from the aspects of architecture, parallel algorithms and parallel software. For decades, high performance computing has been widely used in crucial fields such as large-scale scientific and engineering computing, artificial intelligence, and industrial simulation. It has extremely high application values for basic scientific discovery, national economic development and national defense industry construction, and has become an important manifestation of the country's comprehensive scientific and technological strength.

Turing Award winner Jim Gray pointed out that scalability issues are at the heart of high performance computing, and it is also an important challenge that ranks first among the twelve research goals in the future information technology field. Parallel software design and parallel system architecture are two typical aspects of scalability, and the scalability challenges are reflected in the difficulty of architecture porting, algorithm design, and application optimization. This thesis conducts research on the theory of massively parallel multi-level discontinuous nonlinear scalability, deeply analyzes the development law of scalability, and proposes physical models, parallel algorithms and performance optimization multi-level collaborative design methods. The scalability optimization design is carried out at various hardware parallel scales, different software parallel granularities, and multi-level interdisciplinary parallel applications. The specific contributions include:

1. **A massively parallel multi-level discontinuous nonlinear scalable theory is proposed.** Through the analysis of the scalability on parallel software from homogeneous multi-core systems to heterogeneous many-core systems, it is found that there are two typical phenomena of scalability, namely discontinuity and nonlinearity, in the physical model, parallel algorithm and performance optimization of applications with different parallel granularity. The two phenomena at multiple levels are systematically analyzed, and the multi-level collaborative design theory of scalability is first proposed.

It provides methodological guidance for the research of scalability in the field of high-performance computing, especially the scalability problem in massively parallel computing.

2. **A hundred-core-level Stencil parallel numerical algorithm is proposed, using novel vectorization and tilling technology to achieve highly scalable single-machine multi-core Stencil calculations.** At the parallel algorithm level, two Stencil vectorization strategies, transposed layout calculation and spacial-temporal calculation folding, are proposed to improve the parallelism of data in the CPU; at the same time, at the performance optimization level, an efficient register data reuse algorithm and tilling optimization algorithm are designed to improve data access efficiency. The experimental results show that based on the collaborative design idea of parallel algorithm and performance optimization, the absolute performance is up to 4.39 times higher than the state-of-the-art method, and the multi-core scalable performance of the Stencil parallel numerical algorithm is effectively improved.

3. **A ten-thousand-core-level distributed machine learning framework is proposed, which uses novel clustering and regression techniques to achieve a highly scalable multi-machine many-core machine learning prediction model.** At the theoretical modeling level, this thesis proposes a new Best Friend graph data structure and a hierarchical minimum spanning tree network model, and designs a regression prediction method based on clustering; at the parallel algorithm level, a backtracking-based load balancing algorithm and an efficient parallel communication algorithms are proposed to reduce the computational and communication overhead of distributed systems. The experimental results show that through the physical model and parallel algorithm collaborative design idea, it ensures the accuracy of the clustering and regression methods, and effectively improve the scalability on multi-machine and many-core distributed machine learning framework from the existing work of 1,536 cores to 12,288 cores.

4. **Aiming at the optimization of scientific computing software applications, a million-core-level large-scale scaling method is proposed, and a large-scale and highly scalable domestic software application OpenKMC is designed to simulate the radiation damage of nuclear materials.** At the physical model level, a highly

scalable new potential function model and grouped reaction strategy are optimized to support the establishment of an efficient kinetic Monte Carlo model; at the parallel algorithm level, a parallel synchronous quadrant algorithm suitable for large-scale applications and efficient adaptive communication algorithm are proposed to improve load balancing and communication efficiency; at the performance optimization level, memory access optimization technology, efficient locality algorithm, Athread thread-level heterogeneous parallel strategy and vectorized acceleration method are proposed. Hierarchical memory-cache-register data access features are extracted and lightweight process-thread-data parallelism is tapped to optimize the utilization of the computing power on heterogeneous many-core architecture. The experimental results show that, through the multi-level collaborative design idea of physical model, parallel algorithm, and performance optimization, our OpenKMC achieves high accuracy and good scalability of applying hundred-billionatom simulation on 5.2 million cores with a performance of over 80.1% parallel efficiency, which becomes a new milestone in nuclear material simulation.

Keywords: Massively Parallelism; Scalability; Scientific Computing Applications; Parallel Numerical Algorithms; Machine Learning Frameworks

目 录