論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Performance Optimizations on a Many-core Processor
著者(和文)	林新華
Author(English)	Lin Xinhua (James)
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:乙第4151号, 授与年月日:2018年2月28日, 学位の種別:論文博士, 審査員:松岡 聡,渡辺 治,脇田 建,遠藤 敏夫,横田 理央
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:乙第4151号, Conferred date:2018/2/28, Degree Type:Thesis doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（論文博士）

論　文　要　旨（英文）　（800語程度）

（Summary）

報告番号	乙 第　　号	氏　名	林 新華

（要　旨）

The landscape of computing is historically changing from single-core to many-core. Due to the end of Moore's Law in the recent decade, instead of improving the clock frequency, more cores are adding to a single chip, resulting in the emerging of a new breed of many-core processor architecture. Moreover, as the memory performance has been lagging behind the processor performance for a couple of decades, the increasing flops-to-byte ratio of many-core processors makes most application memory-bound. As a result, one promising solution to address this historic challenge is exploring on-chip data locality with an efficient inter-core communication. The inter-core communication has two basic schemas: the load/store and message-passing. The former, such as the cache-coherence protocol, is easy-to-program but less scalable; the latter, such as the register communication, is scalable but hard-to-program.

Despite of its strong scalability, the register communication brings the significant increasing programming challenge to many-core processors. Taking the China's home-grown 260-core SW26010 processor as an example, the register-level communication (RLC) adopted in SW26010 has two major constraints. First, RLC can only support row/column-based communications, which requires designing algorithms to only communicate among the cores in the same row/column. Second, RLC uses an anonymous producer-consumer protocol, which requires orchestrating message sequences manually so as to ensure a correct sending and receiving order.

This study, divided in three steps, aims to tackle the key programming challenges of the register communication. Taking the SW26010 processor as a research case, we first identified the programming challenges through a comprehensive evaluation of the processor, and then developed a systematic optimization solution. The research findings are envisioned to make a breakthrough in the performance optimizations and to inform researchers who face the same challenges.

The purpose of the first step in this study was to illuminate the uncharted area of the SW26010 processor. The inadequate public information of this processor's micro-architecture prevents global researchers from improving application performance on the TaihuLight supercomputer. To address this issue, we developed the micro-benchmark suite *swCandle*, mostly written in assemble language, to evaluate the key micro-architectural features of the SW26010. The benchmark revealed some unanticipated findings on

the processor micro-architecture beyond the publicly available data. For instance, the broadcast mode of RLC has the same latency as the peer-to-peer mode. According to this finding, we speculated that the broadcast mode might be implemented by default while the P2P mode might be implemented as a special case of the broadcast mode; the implementations could be similar to the mask operations in vector processing. These findings provide important information for performance optimizations in the following two steps.

Based on the findings reveled in the first step, we conducted the second one in which we optimized two compute-bound kernels. The first kernel is the direct N-body simulation. Due to the lack of efficient hardware support, the reciprocal square root (*rsqrt*) operations turned out to be the performance bottleneck of N-body on the SW26010. We applied the computation-oriented optimizations and achieved about 25% efficiency in a single core-group of the SW26010. The second kernel is double-precision general matrix-multiplication (DGEMM). We designed a novel algorithm for RLC and applied several on-chip communication-oriented optimizations. These endeavors improved the efficiency to up to 88.7% in a single core-group of the SW26010.

In contrast to the compute-bound kernels, due to the limited memory bandwidth of the SW26010, the single memory-bound kernel -- such as Sparse matrix-vector multiplication (SpMV -- cannot perform well on the processor, despite comprehensive optimizations. However, we anticipated, the overall performance of an algorithm can be effectively improved by overlapping multiple memory-bound kernels within the algorithm, and, thus, can provide a promising optimization approach for those multi-kernel memory-bound algorithms to achieve better performance on the SW26010. The aim of the third step in this study is to optimize the memory-bound algorithm Preconditioned Conjugate Gradient (PCG). First, in order to minimize the all_reduce communication cost of PCG, we developed a new algorithm RNPCG, a non-blocking PCG leveraging the on-chip register communication. Second, we optimized three key kernels of the PCG, including proposing a localized version of the Diagonal-based Incomplete Cholesky (LDIC) preconditioner. Third, to scale the RNPCG on TaihuLight, we designed the three-level non-blocking all_reduce operations. With these three steps, we implemented the RNPCG in the computational fluid dynamics software OpenFOAM. The experimental results on TaihuLight show that 1) compared with the default implementations of OpenFOAM, the RNPCG and the LDIC on a single-core group of SW26010 can achieve a maximum speedup of 8.9X and 3.1X, respectively; 2) the scalable RNPCG can outperform the standard PCG both in the strong and the weak scaling up to 66,560 cores.