文章编号: 0253-2239(2010)02-0498-05

光互连与电互连的并行运算性能比较

李 倩 乔耀军 纪越峰

(北京邮电大学信息光子学与光通信教育部重点实验室, 北京 100876)

摘要 光互连在短距离传输中的应用是未来大规模计算领域的关键技术。针对并行运算系统对高速高效传输的应用需求,研究了不同结构和不同传输方式下并行运算系统的性能,并从并行运算性能这一角度比较了光互连系统与电互连系统的优劣。采用网孔结构和超立方结构两种模型,用加速比和效率两个指标来评价光互连系统和电互连系统在并行运算方面的性能差异。得到光互连系统及电互连系统在进行并行计算时的加速比和效率与计算规模之间的关系,分析了计算规模对计算速比和效率的影响。通过加速比和效率这两个量的比较得到结论,在大规模并行计算中,光互连系统有着电互连系统不可比拟的优势。

关键词 光通信;光互连;并行体系结构;快速傅里叶变换

中图分类号 O439 **文献标识码** A **doi**: 10.3788/AOS20103002.0498

Comparison of Parallel Computing Performance Between the Optical and Electrical Interconnects

Li Qian Qiao Yaojun Ji Yuefeng

(*Key Laboratory of Information Photonics and Optical Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing* 100876, *China*)

Abstract Optical interconnects for short-distance transmission are becoming more and more important along with the increasing demand on large-scale computing tasks. An analytic method to evaluate the computing performance of optical and electrical chip-to-chip interconnect systems from the parallel computing performance view is proposed. The mesh architecture model and hypercube architecture model are both investigated. The speedup and efficiency are selected to represent the parallel computing performance of an interconnect system. The relation between speedup and computing number, together with the relation between efficiency and computing number, are summarized and analyzed. The comparison shows that the parallel system based on optical chip-to-chip interconnects system has much higher speedup and efficiency than that based on electrical system.

Key words optical communications; optical interconnect; parallel architecture; fast Fourier transform (FFT)

1 引 言

光互连是以光波作为传递信息的载体的一种互连传输方式,其主要特色之一就是光学信息的并行传输,现已发展成为一门独立的网络通讯技术。光互连技术在高速传输下有着电互连技术不可比拟的优点,其最有潜质的优点就是功耗低[1,2]。近年来,光互连在短距离上的应用成为研究的主要热点。短距离传输中光互连的优势早已经被公认,尤其是在带宽上较之于电互连的显著提高,更是有着广泛的应用前景[3,4]。随着并行计算规模的不断增大,对

收稿日期: 2009-01-07; **收到修改稿日期**: 2009-05-18

基金项目: 国家 863 计划(2007AA01Z2a6)、国家 973 计划(2007CB310705)、国家自然科学基金(60711140087)、教育部博士点基金(200800130001)、教育部创新团队项目(IRT0609)和科技部国际合作计划(2006DFA11040)资助课题。

作者简介: 李 倩(1985—),女,硕士研究生,主要从事宽带通信网和高速芯片光互连等方面的研究。
E-mail: littlesen@126.com

导师简介: 乔耀军(1972—),男,博士,副教授,主要从事高速光通信系统与并行光互连等方面的研究。
E-mail: qiao@bupt.edu.cn

并行计算系统的研究和改进越来越受到重视。当处理器的计算速度提高到一定程度时，并行计算系统内部处理器间的数据传输速率就成了影响系统效能的主要因素。但是，能提供较大带宽的光互连并没有在计算机内部得到广泛应用。一个主要的原因就是目前还没有可以有效地量化地评价系统性能的分析方法。本文主要研究的只是光互连在传输速率上的优势所带来的系统性能的改进，对于光互连与电互连在互连密度和功耗等方面的差异，将在后续研究中进行深入讨论。

为了恰当地评价光互连和电互连系统，建立了基本的处理器结构模型来衡量光互连及电互连条件下并行计算的效能，模型概要将在第一部分进行介绍。第二部分主要分析并行快速傅里叶变换(FFT)的性能，以 FFT 为例，考虑到 FFT 的并行度很高，在蝶式计算中，数据通信有固定的模式，在并行处理中有一定的代表性，对 FFT 的性能指标的衡量可以直观地反映光互连与电互连系统对并行问题的处理能力[5]。模型参数主要包括通信时间和计算时间，模型输出的则是光互连系统和电互连系统的加速比与效率，由此可以方便地比较二者的性能，这将在第三部分详细讨论。在这一模型中，光互连的优势得到了直观地体现。

2　模型概要

计算机中的各个处理器并不是完全一致的，处理器间通信的带宽也不尽相同。简单起见，假设它们是同构的，不同之处仅在于处理器间的通信带宽。

众所周知，独享处理器资源时，串行程序的执行时间近似等于程序指令执行花费的 CPU 时间。但是，并行程序相对复杂，其执行时间等于从并行程序开始执行，到所有进程执行完毕，时钟走过的时间。对各个进程，执行时间可以进一步分解为计算 CPU 时间、通信 CPU 时间、同步开销时间和同步导致的进程空闲时间[3]。

在讨论 FFT 的计算时间时，选取基 2FFT 的运算法则，在这一法则下 N 点 FFT 需要 $N\mathrm{lb}N$ 次的运算。那么串行执行时间就是单次运算时间与运算次数的乘积，而并行执行时间则是运算时间与通信及开销时间的和。

用加速比和效率两个指标来评价并行性相对于串行性的优势。在处理器资源独享的前提下，假设某个串行运算在某台并行计算机单处理器上的执行时间为 T_s，而该运算并行化后，P 个进程在 P 个处理器并行执行所需要的时间为 T_p，则该并行运算在该并行计算机上的加速比 S_p 可定义为

$$S_p = \frac{T_s}{T_p}. \tag{1}$$

效率定义为

$$E_p = \frac{S_p}{P}. \tag{2}$$

这里需要说明的是 T_1 指处理器个数为 1 时并行运算的执行时间，通常情形下 T_1 大于 T_s，因为并行运算往往引入一些冗余的控制和管理开销[3]。

加速比和效率是衡量一个并行运算性能的基本的评价方法，加速比表示了并行化之后的效率提升情况，是并行求解问题获得相应益处的一种度量。效率又将加速比平均到了每个处理器上，是处理器被有效利用的时间的度量。加速比着眼于并行算法和计算机本身的可扩展性，效率则可以理解为每个处理器对加速比的贡献。显然，执行最慢的进程将决定并行运算的性能[4]。这正解释了为何假设各处理器同构。若处理器不完全同构，总的执行时间取决于速度最慢的处理器，对加速比和效率这两个评价指标来说，各处理器的差异无法体现，相当于整体速度都被放慢，在分析过程中实质是一样的。

考虑两种处理器结构。一种是网孔型结构，这是一种带宽受限的结构，其中任意不相邻的两个处理器都不能直接通信，而是要经过位于它们之间的处理器的传递，起传递作用的处理器必须要等到一个数据被完整地接收后才能将其继续向后传递。另外一种是超立方结构，这是一种全带宽结构，其中起传递作用的处理器不需要等待一个数据被完整地接收就可以将其继续向后传递[5]。显然，后者在数据交换时的速度会更快一些。在前一种结构中，假设现在要在两个处理器间传送 m 个数据，这两个处理器的间隔为 x 个处理器，每个数据长 y bit，处理器间传输带宽为 B bit/s，传输的其它开销时间为 t_s，那么整个传输需要的时间 $T_1 = t_s + mxy/B$。把运算启动时间等时间消耗都归到 t_s 中。而同样情况下在后一种结构中传输只需要 $T_2 = t_s + my/B$。

3　并行 FFT 性能分析

考虑将 n 点 FFT 映射到 P 个处理器中，如图 1 所示。假设 $n = 2^r$，$P = 2^d$ 且 d 为偶数。那么每个处理器将会存储并计算 n/P 个数据。根据基 2 的运算法则，整个运算总共需要 $r = \mathrm{lb}n$ 轮迭代。图 1 是四处理器十六点 FFT 运算的模型。从中发现，迭代过

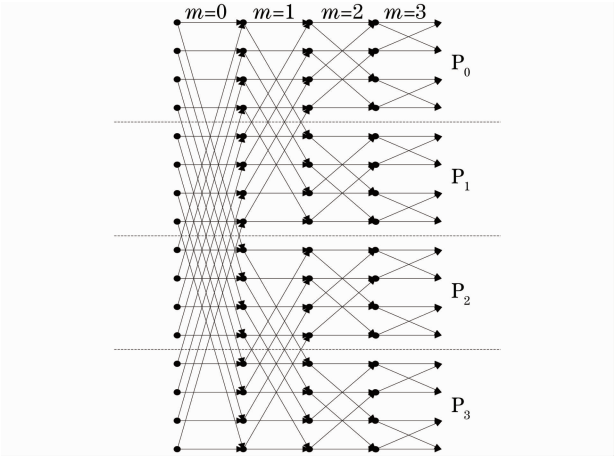

图 1 FFT 运算结构模型

Fig. 1 FFT computing scheme

程可以被分为两个阶段：由于 n/P 个数据存储在同一处理器中，那么，在最后的 $r-d=\mathrm{lb}(n/P)$ 轮迭代中不需要处理器之间交换数据，各个处理器独自进行运算；在前 $d=\mathrm{lb}P$ 轮迭代中，待计算的数据位于不同的处理器中，所以在每一轮运算中每个处理器都要和其它处理器交换 n/P 个数据。

为了研究在不同传输模式下并行 FFT 的性能，将分别对处理器的网孔结构模型和超立方结构模型进行分析。由于各处理器是同构的，那么每个处理器单独进行一次 FFT 运算的时间也是一样的。假设单处理器进行一次 FFT 运算的时间为 t_c，包括运算启动时间及同步开销等在内的开销时间的和为 t_e，相邻的两个处理器之间传输一个数据的时间为 t_w。每个处理器会存储并计算 n/P 个数据。根据前面的分析，很容易得到串行运算条件下 n 点 FFT 的运算时间为

$$T_s=t_c n\mathrm{lb}n. \tag{3}$$

在 P 个处理器的网孔结构模型中，处理器被排列成 $\sqrt{P}\times\sqrt{P}$ 阵列，如图 2 所示。不是任意的两个处理器都可以直接通信，位于通信源和通信端处理器之间的处理器将会起到传递的作用。考虑网孔结构中的某一个处理器 P_i。总共需要有 $r=\mathrm{lb}n$ 轮迭代计算，在前 $d=\mathrm{lb}P$ 轮迭代中，处理器 P_i 需要与其他处理器进行数据通信。在一半即 $d/2$ 轮的数据通信中，P_i 与跟它处在同一行的处理器进行通信，在另外一半中与跟它处在同一列的处理器进行通信。在 $d/2$ 轮的迭代中，行或列间的通信距离从 1 个单位到 $\sqrt{P}/2$ 个单位，成倍地增加。在第 $k(k=0,1,\cdots\cdots,d/2-1)$ 轮迭代中，开销时间是 t_e，通信时间是 $t_w(n/p)2^k$。将所有的迭代都考虑进来，在一行或一列中，数据交换时间为 $\sum_{k=0}^{d/2-1}[t_e+t_w(n/p)2^k]$。于是，在全部需要数据通信的 d 轮迭代中，总的数据交换所需的时间为

$$T_w=\sum_{k=0}^{d/2-1}[t_e+t_w(n/p)2^k]+\sum_{k=0}^{d/2-1}[t_e+t_w(n/p)2^k]=t_e\mathrm{lb}P+2t_w n/\sqrt{P}. \tag{4}$$

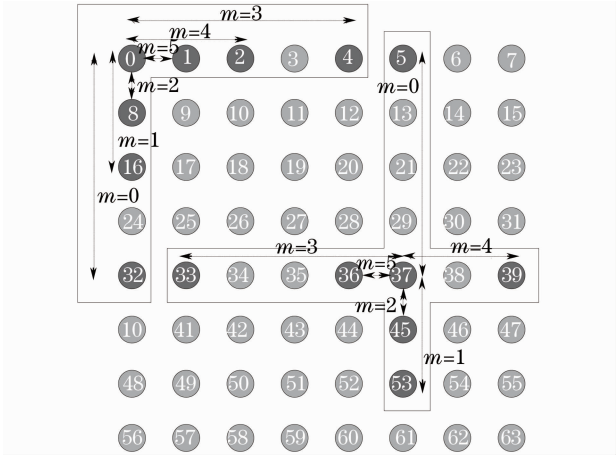

图 2 网孔结构模型

Fig. 2 Mesh architecture

假设单处理器一次运算的时间为 t_c，那么 r 轮迭代的运算时间为 $t_c(n/P)\mathrm{lb}n$，于是总的执行时间为

$$T_p=t_c(n/P)\mathrm{lb}n+t_e\mathrm{lb}P+2t_w n/\sqrt{P}. \tag{5}$$

根据以上结论分析加速比和效率。加速比 S 定义为串行运算在单处理器上的执行时间与该运算并行化后在多处理器上的执行时间的比值，它可以描述并行化的效能。效率 E 定义为加速比与处理器数量 P 的比值，它可以描述并行化的程度。但实际中，任何一个串行运算都不能被完全并行化，因为并行化总要有额外的开销。于是加速比为

$$S=\frac{T_s}{T_p}=\frac{t_c n\mathrm{lb}n}{t_c(n/P)\mathrm{lb}n+t_e\mathrm{lb}P+2t_w n/\sqrt{P}}. \tag{6}$$

效率为

$$E=\frac{S}{P}=\frac{t_c n\mathrm{lb}n}{t_c n\mathrm{lb}n+t_e P\mathrm{lb}P+2t_w n\sqrt{P}}. \tag{7}$$

用同样的方法来分析超立方结构，如图 3 所示。不同之处仅在于通信时间，在前 $d=\mathrm{lb}P$ 轮迭代中，通信时间是 $t_w(n/P)\mathrm{lb}P$，于是总数据交换时间为

$$T_w=t_e\mathrm{lb}P+t_w(n/P)\mathrm{lb}P, \tag{8}$$

总执行时间为

$$T_p=t_c(n/P)\mathrm{lb}n+t_e\mathrm{lb}P+t_w(n/P)\mathrm{lb}P. \tag{9}$$

同样的分析，加速比为

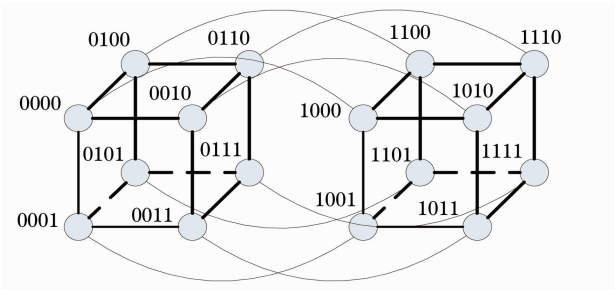

图 3　超立方结构模型

Fig. 3　Hypercube architecture

$$S=\frac{T_s}{T_p}=\frac{t_c n\mathrm{lb}n}{t_c(n/P)\mathrm{lb}n+t_e\mathrm{lb}P+t_w(n/P)\mathrm{lb}P},\quad(10)$$

效率为

$$E=\frac{S}{P}=\frac{t_c n\mathrm{lb}n}{t_c n\mathrm{lb}n+t_e P\mathrm{lb}P+t_w n\mathrm{lb}_2 P}.\quad(11)$$

4　光互连与电互连系统的性能比较

为了衡量光互连与电互连系统的优劣以及运算规模对 FFT 运算性能的影响，选取合适的参数进行仿真。所用到的是 1024 个处理器组成的系统。

分别在光互连与电互连的条件下分析加速比和效率与运算规模间的关系。各时间参数采用当前处理器的平均工作能力，通过调研资料设置如下仿真参数：开销时间 t_e 在平均意义下取 80 ns，单次 FFT 运算时间 t_c 为 20 ns。光互连系统每通道的传送速率为 6.25 Gb/s，电互连系统每通道的传送速率为 1 Gb/s，分别考虑光互连系统与电互连系统在 2，4，8 通道情况下的加速比和效率性能。

研究在固定的系统中进行不同规模 FFT 运算时加速比和效率的变化情况。FFT 运算规模与加速比的关系如图 4 所示，可以看出，光互连系统能为大规模的并行运算提供较大的加速比。以网孔模型为例，在 8 通道系统中计算 16384 点 FFT 时，电互连系统的加速比为 496.1，而光互连的加速比为 778.8，提高了 57%。运算规模与效率的关系如图 5 所示，同样以网孔模型为例，在 8 通道系统中计算 16384 点 FFT 时，电互连系统的效率为 0.48，而光互连的效率为 0.76，提高了 58%。随着运算规模的增大，系统的效率越来越高，说明并行化程度越来越高，而光互连系统与电互连系统相比，在并行化程度上始终占据着优势。

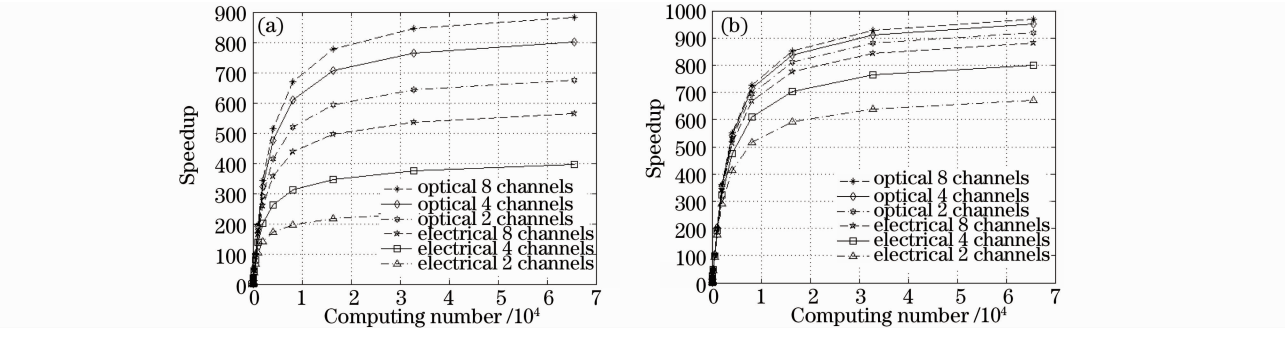

图 4　加速比与计算规模间的关系。(a)网孔结构；(b)超立方结构

Fig. 4　Relation between speedup and computing number with mesh structure (a) and hypercube structure (b)

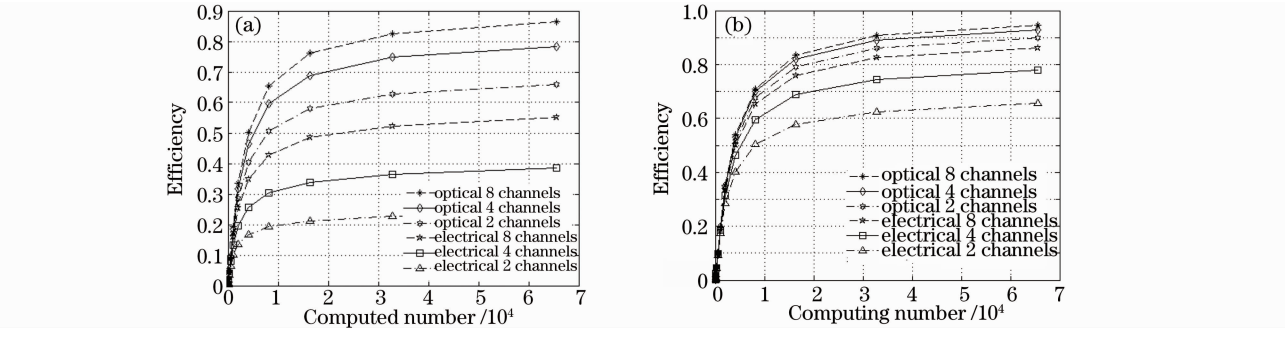

图 5　效率与计算规模间的关系。(a) 网孔结构；(b) 超立方结构

Fig. 5　Relation between efficiency and computing number with mesh structure (a) and hypercube structure (b)

4 结 论

讨论了一种评价并行计算效能的方法，并基于此方法比较了光互连系统与电互连系统在并行运算方面的效能，得到了其加速比与效率的仿真结果。比较显示，同等条件下，不论是加速比还是效率，光互连系统都有着电互连系统无法比拟的优势，在网孔模型中各方面性能的提高都超过了50%。同时，并行运算的规模会对运算的效能产生一定影响，通过以上结论，也可以找到改善并行运算效能的方法。

致谢 感谢顾仁涛的悉心指导，他在学习和工作中对我的无私帮助让我受益匪浅。

参 考 文 献

1 C. Hoyeol, P. Kapur, K. C. Saraswat. Power comparison between high-speed electrical and optical interconnects for interchip communication [J]. *J. Lightwave Technol.*, 2004, **22**(9): 2021～2033

2 Li Hongjian, Dou Yusheng, Tang Hong *et al*.. Parallel computer simulation of photochemical reactions [J]. *Chinese J. Lasers*, 2009, **36**(2): 356～361
李鸿健，豆育升，唐 红 等. 激光诱导化学反应的并行计算机模拟[J]. 中国激光，2009，**36**(2)：356～361

3 Tang Jianxiong, Jin Yaohui, Gao Yu *et al*.. Microring resonator optical switch for ultralow-latency interconnections [J]. *Chinese J. Lasers*, 2008, **35**(s2): 200～203
唐健雄，金耀辉，高 煜 等. 微环共振器光开关在高速互连中的应用[J]. 中国激光，2008，**35**(s2)：200～203

4 J. E. Roth, S. Palermo, N. C. Helman *et al*.. An optical interconnect transceiver at 1550 nm using low-voltage electro absorption modulators directly integrated to CMOS [J]. *J. Lightwave Technol.*, 2007, **25**(12): 3739～3747

5 Tang Yuan, Zhang Yunquan, Sun Jiachang. Analysis on extensibility of parallel FFT and feasibility as index of performance of new super computing [J]. *Development and Application of High-Performance Computing*, 2007, **4**: 16～26
唐 渊，张云泉，孙家昶. 并行FFT可扩展性分析及作为新超级计算性能评测指标的可行性分析研究[J]. 高性能计算发展与应用，2007，**4**：16～26

6 Zhang Linbo, Chi Xuebin, Mo Zeyao *et al*.. Introduction to Parallel Computing [M]. Beijing: Tsinghua University Press, 2006. 210～212
张林波，迟学斌，莫则尧 等. 并行计算导论[M]. 北京：清华大学出版社，2006. 210～212

7 A. Grama, A. Gupta, G. Karypiis *et al*.. Introduction to Parallel Computing [M]. Zhang Wu, Mao Guoyong, Cheng Haiying *et al*.. Transl., Beijing: China Machine Press, 2005. 77～81
格兰马，古普塔，卡瑞皮斯 等. 并行计算导论[M]. 张 武，毛国勇，程海英 等译. 北京：机械工业出版社，2005. 77～81

8 Gu Rentao, Qiao Yaojun, Ji Yuefeng. Optical or electrical interconnects: quantitative comparison from parallel computing performance view, Global Telecommunications Conference, 2008, New Orleans, LO