

DOI:10.11784/tdxbz201601001

不跨层行操作并行 RCA 互连时延性能评估

陈乃金^{1,2}, 冯志勇¹

(1. 天津大学计算机科学与技术学院, 天津 300072; 2. 安徽工程大学计算机与信息学院, 芜湖 241000)

摘要: 针对三类行并行重构单元阵列互连时延性能评估问题, 提出了一种通过节点映射和运行机制来评测互连时延的方法, 基于前驱回溯不加旁节点不跨层时域映射算法, 对点到点、路由传输、行列总线等互连 RCA 进行了时延分析和计算. 实验结果表明, 与路由传输和行列总线互连相比, 点到点互连在最大不跨层互连、不跨层累加互连、考虑互连执行总时延等方面均是最小的, 从而表明了点到点重构单元阵列的互连时延优于路由器传输和行列总线互连.

关键词: 互连模式; 粗粒度可重构体系结构; 点到点互连; 路由传输互连; 行列总线互连

中图分类号: TP302 **文献标志码:** A **文章编号:** 0493-2137(2017)04-0429-08

Interconnect Delay Performance Evaluation for Non-Crossing Level and Row Operands Parallel RCA

Chen Naijin^{1,2}, Feng Zhiyong¹

(1. School of Computer Science and Technology, Tianjin University, Tianjin 300072, China;
2. College of Computer and Information Science, Anhui Polytechnic University, Wuhu 241000, China)

Abstract: In order to assess three types of row parallel reconfigurable cell array (RCA) interconnect delay performance, this paper presented a method for evaluating interconnect delay by means of node mapping and operation running mechanism. Based on preorder traversing backtracking no adding-bypass-node (PTBNA) non-crossing level temporal mapping algorithm, this paper analyzed and computed RCA interconnect delay of point to point (PP), router transmission (RT), and row column bus (RCB). Compared with RT and RCB, PP can get the least in maximum non-crossing level interconnect delay, non-crossing level accumulation interconnect delay and considering interconnect execution total delay. Thus PP-RCA interconnect delay is better than that of RT and RCB.

Keywords: interconnect mode; coarse grained reconfigurable architecture; point to point interconnect; router transmission interconnect; row column bus interconnect

粗粒度可重构体系结构(coarse grained reconfigurable architectures, CGRA)已经在嵌套循环编译^[1]、多媒体运算^[2]等方面获得了应用. CGRA 优化的编译性能是通过 RCA(reconfigurable cell array)来实现的, RCA 互连方式直接影响 CGRA 的编译性能^[3]. 因为行列

并行(row column parallel, RCP)执行架构可以获得运算操作数并行执行的最大化, 所以备受关注. 典型的 RCP 架构的互连方式有 RC 之间点到点(point to point, PP)、路由器传输 RT(router transmission)、行列总线(row column bus, RCB)互连 3 种情形. PP 类

收稿日期: 2016-01-01; 修回日期: 2016-04-13.

作者简介: 陈乃金(1972—), 男, 副教授, 86naijinchen@tongji.edu.cn.

通讯作者: 冯志勇, zvfeng@tju.edu.cn.

基金项目: 国家高技术研究发展计划(863 计划)资助项目(2013AA013204); 安徽省自然科学基金资助项目(1408085MF124); 安徽省高校自然科学基金重点资助项目(KJ2015A003); 安徽省高校优秀中青年骨干人才国内外访学研修重点项目(gxfzD2016102); 安徽工程大学国家自然科学基金预研基金资助项目.

Supported by the National High Technology Research and Development Program of China(No. 2013AA013204), the Natural Science Foundation of Anhui Province(No. 1408085MF124), the Key Program of University Natural Science Research Foundation of Anhui Province(No. KJ2015A003), the Key Program of University Outstanding Young and Middle Aged Backbone Talent Domestic and Foreign Visiting Scholar Research and Study of Anhui Province(No. gxfzD2016102) and the National Natural Science Pre-Research Foundation of China of Anhui Polytechnic University.

型典型架构有 MorphoSys^[4]、ADRES^[5]、REMUS 等^[6]; RT 类型典型架构有 LEAP 等^[7]; RCB 类型典型架构有 REMARC 等^[8]。在 CGRA 运算操作数行可并行执行层面, PP、RT、RCB 3 种互连方式 RCA 的硬件资源约束类似, 相关时域映射方案基本相同, 故对上述 3 种类型架构运算节点的互连时延分析显得尤为重要。

关于 PP 架构, 文献[5]介绍了所设计的 ADRES 可重构计算系统。文献[9]尝试把 MPEG-2 和 H.264 等多媒体计算关键循环映射到一个 8×8 的 ADRES 的计算阵列上去执行, 结果获得了 2.4 和 2.6 倍的性能加速。文献[10]设计了基于粗粒度 ADRES 架构的多输入多输出的正交频分复用 (multi-input multi-output based orthogonal frequency division multiplexing, MIMO-OFDM) 高速数据传输系统, 相比传统的方法, 获得了两倍性能改进。文献[6]介绍了所设计的 REMUS 可重构计算系统, 同时阐述了可重构计算处理器的架构和编译等内容。面向 REMUS 等统一的行流水架构, 文献[11]设计实现了二维可重构运算阵列的跨层数据传输的旁节点添加算法, 并给出其临界条件。

关于 RT 架构, 文献[7]介绍了所设计的 LEAP 可重构计算系统。基于 LEAP 架构, 文献[12]设计实现了一种循环任务图映射算法。

关于 RCB 架构, 文献[8]较为详细地介绍了 REMARC 的设计方案。文献[13]给出了总线和路由

基本传输时延的简单假设, 没有给出 CGRA 背景下较为具体的分析。

本文与之前已有的评估方法相比, 优势在于: 上述已有文献提出了 PP、RT、RCB 等互连结构, 但是在进行 RCA 计算性能评估时, 对块内数据传输的互连时延考虑不足或没有进行互连时延评估, 本文与已有的评估方法相比优势是较为详细地分析和比较了 PP、RT、RCB 3 种互连结构的互连时延, 同时指出了互连时延对应用在 RCA 上的计算性能具有较大的影响。

1 PP、RT 和 RCB 不跨层互连时延过程说明及理论分析

1.1 PP、RT、RCB 不跨层互连时延过程说明

PP、RT、RCB 3 种 RCP 互连模型如图 1 所示, 本文研究的 2 个前提条件如下: ①目标架构为一块行操作并行执行 RCA 架构, 一个运算节点占用一个 RC, 乘法时延为 2 cycle, 其他 ALU 运算时延和单个 RC 配置时延为 1 cycle, ②点到点 RC 数据传输互连时延为 1 cycle。

图 2 (a) ~ (c) 中 DT (data transmission) 表示数据传输; 图 2 (b) 中 DLE (data link establishment) 表示数据链路建立; 图 2 (c) 中 BR (bus request) 表示总线请求, BA (bus acknowledge) 表示总线应答, DTC (data

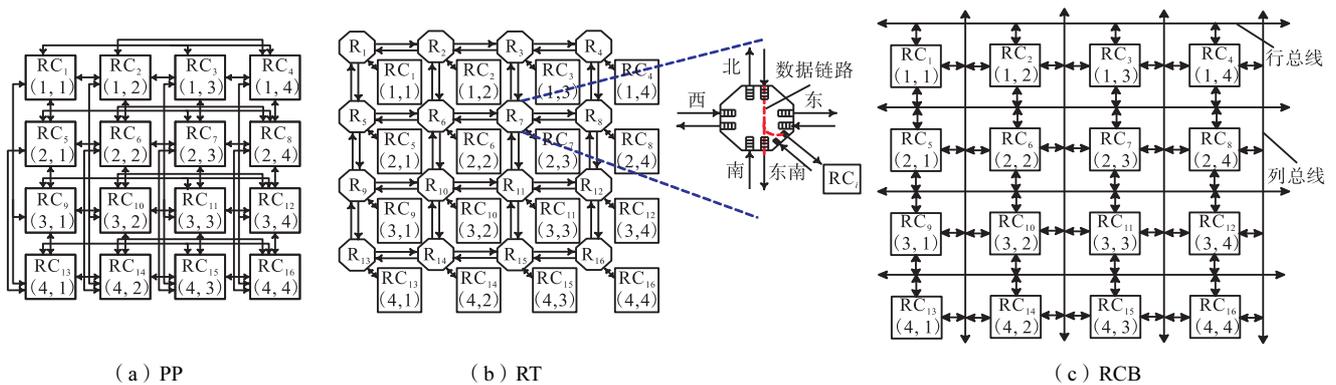


图 1 RCP 3 种互连模型
Fig.1 Three interconnect model of RCP

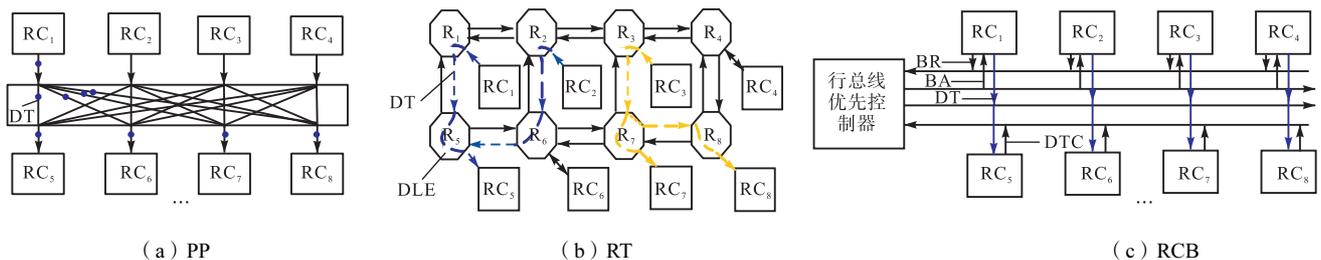


图 2 RCP 3 种不跨层数据传输互连模型
Fig.2 Three non-crossing level data transmission interconnect model of RCP

transmission completing) 表示数据传输完成。

PP、RT、RCB 不跨层互连时延产生的过程说明: 设 RC_1 为源数据传送的起点, 就 PP 而言, RC_1 和其他重构单元(如 RC_2 等)同时把数据传输到 RC_5 , 数据通路固定, 数据传输无冲突, 故需经过 DT 阶段; 就 RT 而言, 若把 RC_1 的数据传输到 RC_5 , 需要数据链路建立, 故需经过 DLE、DT 两个阶段. 就 RCB 而言, 由于要用行线, 若把 RC_1 的数据传输到 RC_5 , 需要总线请求、总线应答等环节, 故需经过 BR、BA、DT、DTC 4 个阶段。

1.2 理论分析

定义 1 最大不跨层互连时延.

一个任务数据流图 DFG (data flow graph) 按硬件约束映射到一块 RCA, 假设 V 为顶点集合, $\forall v_i, v_j, v_k \in V$, 其依赖关系是 v_k 为 v_i, v_j 的父节点, $\text{level}(v_i)$ 表示节点 v_i 的 DFG 层次, 且 $\text{level}(v_i) = \text{level}(v_j)$, $\text{level}(v_i) = \text{level}(v_k) + 1$, 若 v_i, v_j, v_k 被映射到同一块 RCA 后, 这种数据依赖存在且节点互连边均能够在硬件上实现正确连接, 则 RCA 块内运算节点间的层差为 1. 若第 i 块 RCA_i 成功映射了若干个 DFG 子图, 路径最长的 DFG 子图时延称为第 i 块 RCA 最大不跨层互连时延, 记为 $I_{\text{MAX-ID}_i}$. 假定一个 DFG 的映射需要 m 块 RCA, 则 m 块 RCA 最大不跨层互连时延的累加和称为 $I_{\text{MAX-ID}}$.

定义 2 不跨层累加互连时延.

假定一个 DFG 被划分映射到 m 块 RCA, 第 i 块 RCA 累加不跨层互连延迟记为 $I_{\text{ACC-ID}_i}$, 则 m 块不跨层互连时延的累加和, 称为不跨层累加互连时延, 记为 $I_{\text{ACC-ID}}$. $I_{\text{MAX-ID}}$ 越大会导致 $I_{\text{ACC-ID}}$ 增大.

定义 3 异步不跨层互连累加时延.

若第 i 块 RCA 中存在双目或三目数据来源不同的现象, 这时产生了异步延迟, 第 i 块 RCA 所有异步延迟之和称为累加异步不跨层互连延迟 $I_{\text{ASY-ID}_i}$, m 块 RCA 所有 DFG 映射子图的异步不跨层互连时延累加和, 称为异步不跨层互连累加时延, 记为 $I_{\text{ASY-ID}}$. 一般而言, 双目或三目运算的异步延迟分别为 1 或 2 cycle.

定义 4 路由共享路径.

任意 $v_i, v_j, v_k \in V$ 且 v_i 是 v_j 和 v_k 的父节点映射到 RCA_i . $\text{level}(v_i) + 1 = \text{level}(v_j) = \text{level}(v_k)$, 若有两条数据传递路径 $v_i \rightarrow R_i \rightarrow R_j \rightarrow v_j$ 和 $v_i \rightarrow R_i \rightarrow R_j \rightarrow v_k$, 则称 $R_i \rightarrow R_j$ 为节点 v_j 和 v_k 的路由共享路径. R_i, R_j 为路由器, 因 v_i 和 R_i, R_j 和 v_j 的连接线很短, 故时延近似为 0. 由图 2(b) 可知, RC_3 的数据传至 RC_7 和 RC_8 , $R_3 \rightarrow R_7$ 为路由共享路径, 其时延只统计 1 次.

定义 5 总线共享路径.

任意 $v_i, v_j, v_k \in V$ 且 v_i 是 v_j 和 v_k 的父节点映射到 RCA_i . $\text{level}(v_i) + 1 = \text{level}(v_j) = \text{level}(v_k)$, 则 v_i 把数据传输到 v_j, v_k , DT、BR、BA、DTC 各需 1 次, 称为总线共享路径. 由图 2(c) 可知, RC_3 的数据传至 RC_7 和 RC_8 , 由于 RC_7 和 RC_8 的来源数据均为 RC_3 , 其为总线共享路径时延, 完成 1 次数据传输需要 DT、BR、BA、DTC 各 1 次, 即时延为 4 cycle.

下面从理论层面分析 PP、RT、RCB 的互连时延, 若节点存在异步执行, 由定义 3 可知, 每存在一个异步, 双目或三目运算要加上取数据的 1 cycle 或 2 cycle 互连时延, 一个循环 DFG 在 RCA 总时延计算方法与文献[11]一致, 但文献[11]对互连时延考虑不足.

PP-RCA 表示点点互连 RCA, 其互连时延为 DT; RT-RCA 表示路由互连 RCA, 其互连时延为 DT、DLE; BUS-RCA 表示总线互连 RCA, 其互连时延为 BR、BA、DT、DTC.

1) PP-RCA

定理 1 单块多子图.

设 k 个子图均被映射到第 i 块 RCA_i , $RCA_{a \times b}$, $G_k = \langle V_k, E_k \rangle, i \in [1, k], G_k \xrightarrow{f_i} RCA_i$, 则

$$I_{\text{MAX-ID}_i} = \max_{u, v \in V_k} |\text{level}(f(u)) - \text{level}(f(v))| \quad (1)$$

$$I_{\text{ACC-ID}_i} = \sum_{k=1}^n |\{f(e_k) | e_k \in E_k\}| \quad (2)$$

证明: 不妨设有 k 个子图被成功映射到 RCA_i , 则第 1 个子图的 $I_{\text{MAX-ID}_1} = \max_{u_1, v_1 \in V_k} |\text{level}(f(u_1)) - \text{level}(f(v_1))|$; \dots ; 第 k 个子图的 $I_{\text{MAX-ID}_k} = \max_{u_k, v_k \in V_k} |\text{level}(f(u_k)) - \text{level}(f(v_k))|$. 故可得 $I_{\text{MAX-ID}_i} = \max_{u, v \in V_k} \{I_{\text{MAX-ID}_1}, I_{\text{MAX-ID}_2}, \dots, I_{\text{MAX-ID}_k}\} = \max_{u, v \in V_k} |\text{level}(f(u)) - \text{level}(f(v))|$.

设 G_1 有 $|E_1|$ 条边, 且被成功映射到 RCA_i 的边数为 $|\{f(e_1) | e_1 \in E_1\}|$; G_2 有 $|E_2|$ 条边, 且被成功映射到 RCA_i 的边数为 $|\{f(e_2) | e_2 \in E_2\}|$; \dots ; G_n 有 $|E_n|$ 条边, 且被成功映射到 RCA_i 的边数为 $|\{f(e_n) | e_n \in E_n\}|$. 故

$$I_{\text{ACC-ID}_i} = \sum_{k=1}^n |\{f(e_k) | e_k \in E_k\}|.$$

定理 2 多块多子图.

设有一个 DFG 被映射到 m 个 RCA 且每块 RCA 有 n_k 个子图 ($k = 1, 2, 3, \dots, m$), RCA_1 有 n_1 个子图被成功映射; RCA_2 有 n_2 个子图被成功映射; \dots ; RCA_m 有 n_m 个子图被成功映射. 则 m 块 RCA 的最大不跨层互连时延为

$$I_{\text{MAX-ID}} = \sum_{k=1}^m I_{\text{MAX-ID}k} \quad (3)$$

m 块 RCA 的累加不跨层互连时延为

$$I_{\text{ACC-ID}} = \sum_{k=1}^m I_{\text{ACC-ID}k} = \sum_{k=1}^m \sum_{i=1}^{n_k} | \{ f(e_i) | e \in E_{n_k} \} | \quad (4)$$

证明:由定理 1 可知 RCA_1 的最大不跨层互连时延为 $I_{\text{MAX-ID}1}$, 累加不跨层互连时延为 $I_{\text{ACC-ID}1}$; RCA_2 的最大不跨层互连时延为 $I_{\text{MAX-ID}2}$, 累加不跨层互连时延为 $I_{\text{ACC-ID}2}$; \dots ; RCA_m 的最大不跨层互连时延为 $I_{\text{MAX-ID}m}$, 累加不跨层互连时延为 $I_{\text{ACC-ID}m}$; 所以 m 块

RCA 的最大不跨层互连时延为 $I_{\text{MAX-ID}} = \sum_{k=1}^m I_{\text{MAX-ID}k}$, 累加不跨层互连时延为 $I_{\text{ACC-ID}} = \sum_{k=1}^m I_{\text{ACC-ID}k} = \sum_{k=1}^m \sum_{i=1}^{n_k} | \{ f(e_i) | e \in E_{n_k} \} |$.

2) RT-RCA

设 $\text{RCA}_{a \times b}$, $G = (V, E)$, $G_k \xrightarrow{f_i} \text{RCA}_{a \times b}$, $|E| = e$, $|V| = v$.

引理 1 单块 RCA 一对多两层子图.

设一个 DFG 子图的关系为 $1:m$, $m \leq b$, $\forall v_i, v_j, v_k \in V$, 3 个节点具有的依赖关系是 v_k 为 v_i, v_j 的父节点, 且被映射到同一块 RCA 后, $\text{level}(v_i)$ 表示节点 v_i 的在 DFG 中的层次, 且 $\text{level}(v_i) = \text{level}(v_j)$, $\text{level}(v_i) = \text{level}(v_k) + 1$, 层差 $\theta = 1$, $|E| = e = m$, $|V| = v = m + 1$. 则 $I_{\text{MAX-ID}} = 3m\theta = 3m$, $I_{\text{ACC-ID}} = (2m + 1)\theta = 2m + 1$.

证明:当 $m = 1$ 时, $I_{\text{MAX-ID}} = 3 \times 1 = 3$, $I_{\text{ACC-ID}} = 2 \times 1 + 1 = 3$, 结论成立.

设 $m = k$ 时结论成立, 即 $I_{\text{MAX-ID}} = 3k$, $I_{\text{ACC-ID}} = 2k + 1$. 当 $m = k + 1$ 时, 由于每增加一个点, $I_{\text{MAX-ID}}$ 增加 3, $I_{\text{ACC-ID}}$ 增加 2, 故 $I_{\text{MAX-ID}} = 3(k + 1) = 3k + 3$, $I_{\text{ACC-ID}} = 2(k + 1) + 1 = (2k + 2) + 1 = 2k + 3$, 所以 $m = k + 1$ 结论成立. 综上所述, 结论成立.

引理 2 单块 RCA 多对一两层子图.

设子图关系为 $c:1$, 多对一的关系主要体现在双目和三日运算, c 只能取 2 和 3, 由引理 1 可知层差 $\theta = 1$, 则①当子图关系为 $2:1$ 时, $I_{\text{MAX-ID}} = 6\theta = 6$, $I_{\text{ACC-ID}} = 9\theta = 9$; ②当子图关系为 $3:1$ 时, $I_{\text{MAX-ID}} = 9\theta = 9$, $I_{\text{ACC-ID}} = 17\theta = 17$.

定理 3 单块 RCA_i 多对多及多层子图.

设 k 个子图均被映射到第 i 块 RCA_i , $\text{RCA}_{a \times b}$, $G_k = (V_k, E_k)$, $G_k \xrightarrow{f_i} \text{RCA}_i$, k 个子图的叠加包含以

下 3 种情况.

(1) k 个仅包含 2 个及以上 $1:m$ 的两层子图叠加, 若 k 个子图最大层差为 θ_{max} , 由引理 1 可知

$$I_{\text{MAX-ID}i} = \max_{u, v \in V_k} \{ I_{\text{MAX-ID}1}, I_{\text{MAX-ID}2}, \dots, I_{\text{MAX-ID}k} \} = 3m\theta_{\text{max}} \quad (5)$$

$$I_{\text{ACC-ID}i} = k(2m + 1)\theta_{\text{max}} \quad (6)$$

(2) k 个 $1:m$ 和 $2:1$ 或 $3:1$ 的两层子图叠加.

设 $\theta_{\text{max}1}$ 为 $2:1$ 的层差, $\theta_{\text{max}2}$ 为 $3:1$ 的层差, $\theta_{\text{max}3}$ 为 $1:m$ 的层差, 则由引理 1 和引理 2 可知

$$I_{\text{MAX-ID}i} = 6\theta_{\text{max}1} + 9\theta_{\text{max}2} + 3m\theta_{\text{max}3} \quad (7)$$

$$I_{\text{ACC-ID}i} = k[9\theta_{\text{max}1} + 17\theta_{\text{max}2} + (2m + 1)\theta_{\text{max}3}] \quad (8)$$

(3) 设 k_1 个包含 2 个及以上 $1:m$ 的两层子图叠加, k_2 个 $1:m$ 和 $2:1$ 或 $3:1$ 的两层子图叠加, $k = k_1 + k_2$, 则

$$I_{\text{MAX-ID}i} = 6\theta_{\text{max}1} + 9\theta_{\text{max}2} + 3m\theta_{\text{max}3} \quad (9)$$

$$I_{\text{ACC-ID}i} = k_1(2m + 1)\theta_{\text{max}} + k_2[9\theta_{\text{max}1} + 17\theta_{\text{max}2} + (2m + 1)\theta_{\text{max}3}] \quad (10)$$

定理 4 多块多层次子图.

设有一个 DFG 被映射到 m 个 RCA 且每块 RCA 有 n_k 个子图 ($k = 1, 2, 3, \dots, m$), $I_{\text{MAX-ID}k}$ 为第 k 块 RCA 最大不跨层时延, $I_{\text{ACC-ID}k}$ 为第 k 块 RCA 累加不跨层时延.

$$I_{\text{MAX-ID}} = \sum_{k=1}^m I_{\text{MAX-ID}k} \quad (11)$$

$$I_{\text{ACC-ID}} = \sum_{k=1}^m I_{\text{ACC-ID}k} \quad (12)$$

3) BUS-RCA

设 $\text{RCA}_{a \times b}$, $G = (V, E)$, $G \xrightarrow{f} \text{RCA}_{a \times b}$, $|E| = e$, $|V| = v$.

引理 3 单块 RCA 一对多两层子图.

设一个 DFG 子图的关系为 $1:m$, $m \leq b$, 则 $I_{\text{MAX-ID}} = I_{\text{ACC-ID}} = 3 + m$.

引理 4 单块 RCA 多对一两层子图.

设子图关系为 $c:1$, 多对 1 的关系主要体现在双目和三日运算, c 只能取 2 和 3, 则①当子图关系为 $2:1$ 时, $I_{\text{MAX-ID}} = I_{\text{ACC-ID}} = 9$; ②当子图关系为 $3:1$ 时, $I_{\text{MAX-ID}} = I_{\text{ACC-ID}} = 14$.

定理 5 单块 RCA_i 多对多及多层子图.

设 k 个子图均被映射到第 i 块 RCA_i , $\text{RCA}_{a \times b}$, $G_k = (V_k, E_k)$, $G_k \xrightarrow{f_i} \text{RCA}_i$, k 个子图的叠加包含以下 3 种情况.

(1) k 个仅包含 2 个及以上 $1:m$ 的两层子图叠加, 若 k 个子图最大层差为 θ_{max} , 则由引理 3 可知

$$I_{\text{MAX-ID}i} = 3 + m \quad (13)$$

$$I_{\text{ACC-ID}i} = k(3 + m) \quad (14)$$

(2) k 个 $1:m$ 和 $2:1$ 或 $3:1$ 的两层子图叠加。设 $\theta_{\max 1}$ 为 $2:1$ 的层差, $\theta_{\max 2}$ 为 $3:1$ 的层差, $\theta_{\max 3}$ 为 $1:m$ 的层差, 则由引理 1 和引理 2 可知

$$I_{\text{MAX-ID}i} = 9\theta_{\max 1} + 14\theta_{\max 2} + (3+m)\theta_{\max 3} \quad (15)$$

$$I_{\text{ACC-ID}i} = k[9\theta_{\max 1} + 14\theta_{\max 2} + (3+m)\theta_{\max 3}] \quad (16)$$

(3) 设 k_1 个包含 2 个及以上 $1:m$ 的两层子图叠加, k_2 个 $1:m$ 和 $2:1$ 或 $3:1$ 的两层子图叠加, $k = k_1 + k_2$, 有

$$I_{\text{MAX-ID}i} = 9\theta_{\max 1} + 14\theta_{\max 2} + (3+m)\theta_{\max 3} \quad (17)$$

$$I_{\text{ACC-ID}i} = k_1(3+m)\theta_{\max 3} + k_2[9\theta_{\max 1} + 14\theta_{\max 2} + (3+m)\theta_{\max 3}] \quad (18)$$

定理 6 多块多层子图。

设有一个 DFG 被映射到 m 个 RCA 且每块 RCA 有 n_k 个子图 ($k = 1, 2, 3, \dots, m$), $I_{\text{MAX-ID}k}$ 为第 k 块 RCA 最大不跨层互连时延, $I_{\text{ACC-ID}k}$ 为第 k 块 RCA 累加不跨层互连时延, 有

$$I_{\text{MAX-ID}} = \sum_{k=1}^m I_{\text{MAX-ID}k} \quad (19)$$

$$I_{\text{ACC-ID}} = \sum_{k=1}^m I_{\text{ACC-ID}k} \quad (20)$$

2 实验及分析

本节从实验的角度给予 PP、RT、RCB 互连的

$I_{\text{MAX-ID}}$ 和 $I_{\text{ACC-ID}}$ 分析。采用课题组自主研发的时域划分映射工具软件中面向 RCP 架构的不跨层映射算法 (preorder traversing backtracking no adding_bypass_node, PTBNA) 获得的多个基准的时域映射结果来计算 PP、RT、RCB 的 $I_{\text{MAX-ID}}$ 和 $I_{\text{ACC-ID}}$, 相应互连时延的计算方法按第 1.2 节中定理 1~6 以及引理 1~4 的计算方法来确定, A_{RPU} 选取 49 RC ($\text{RCA}_{7 \times 7}$)、56 RC ($\text{RCA}_{7 \times 8}$)、64 RC ($\text{RCA}_{8 \times 8}$) 3 个值, 对于不同的 A_{RPU} 值采用的基准程序集 (见表 1) 对 PP、RT、RCB 的互连时延进行了求解, 具体结果如表 2 和表 3 所示。

表 1 基准程序集

Tab.1 Benchmark assembly

基准	原始输入次数	原始输出次数	运算节点总数	加法	减法	乘法
FFT4	10	4	12	4	4	4
FFT8	20	8	36	12	12	12
EWF	20	5	34	28	—	6
EWF3	60	15	102	84	—	18
EWF6	120	30	204	168	—	36
FDCT	32	8	42	13	13	16
FDCT3	96	24	126	39	39	48
FDCT6	192	48	252	78	78	96
MATRIX4	128	16	112	48	—	64

2.1 PP和RT比较

从表 2 和表 3 可以看出, 相比于 RT 互连, PP 获得了较小的 $I_{\text{MAX-ID}}$ 和 $I_{\text{ACC-ID}}$ 。

表 2 $A_{\text{RPU}} = 49、56、64$ 时 PP 和 RT 的 $I_{\text{MAX-ID}}$ 比较

Tab.2 Comparison of the $I_{\text{MAX-ID}}$ of PP and RT when $A_{\text{RPU}} = 49, 56, 64$

划分基准	$\text{RCA}_{7 \times 7}$			$\text{RCA}_{7 \times 8}$			$\text{RCA}_{8 \times 8}$		
	RT	PP	$\Delta/\%$	RT	PP	$\Delta/\%$	RT	PP	$\Delta/\%$
FFT4	8	2	-75.0	8	2	-75.0	8	2	-75.0
FFT8	11	3	-72.7	11	3	-72.7	11	3	-72.7
EWF	35	11	-68.6	35	11	-68.6	35	11	-68.6
EWF3	35	11	-68.6	35	11	-68.6	38	12	-68.4
EWF6	59	19	-67.8	53	17	-67.9	59	19	-67.8
FDCT	34	7	-79.4	41	5	-87.8	26	5	-80.8
FDCT3	39	9	-76.9	26	8	-69.2	41	8	-80.5
FDCT6	78	16	-79.5	84	15	-82.1	61	12	-80.3
MATRIX4	21	5	-76.2	24	6	-75.0	24	5	-79.2
平均 $\Delta/\%$			-73.9			-74.1			-74.8

表 3 $A_{\text{RPU}} = 49、56、64$ 时 PP 和 RT 的 $I_{\text{ACC-ID}}$ 比较

Tab.3 Comparison of the $I_{\text{ACC-ID}}$ of PP and RT when $A_{\text{RPU}} = 49, 56, 64$

划分基准	$\text{RCA}_{7 \times 7}$			$\text{RCA}_{7 \times 8}$			$\text{RCA}_{8 \times 8}$		
	RT	PP	$\Delta/\%$	RT	PP	$\Delta/\%$	RT	PP	$\Delta/\%$
FFT4	18	6	-66.7	18	6	-66.7	18	6	-66.7
FFT8	45	15	-66.7	48	16	-66.7	48	16	-66.7
EWF	72	24	-66.7	72	24	-66.7	72	24	-66.7
EWF3	216	72	-66.7	216	72	-66.7	210	70	-66.7
EWF6	408	136	-66.7	393	131	-66.7	405	135	-66.7
FDCT	162	42	-74.1	158	44	-72.2	158	44	-72.2
FDCT3	309	92	-70.2	362	102	-71.8	336	94	-72.0
FDCT6	670	193	-71.2	497	188	-62.2	636	184	-71.1
MATRIX4	156	40	-74.4	162	42	-74.1	165	43	-73.9
平均 $\Delta/\%$			-69.3			-68.2			-69.2

2.2 PP 和 RCB 比较

方式 RCA 的互连时延, 相比 RCB, PP 获得了 I_{MAX-ID} 和 I_{ACC-ID} 总体优化.

表 4 和表 5 列出了 $A_{RPU} = 49、56、64$ 时 PTBNA 映射算法所得到的 PP 和 RCB 两种不同互连

表 4 $A_{RPU} = 49、56、64$ 时 PP 和 RCB 的 I_{MAX-ID} 比较

Tab.4 Comparison of the I_{MAX-ID} of PP and RCB when $A_{RPU} = 49, 56, 64$

划分基准	RCA _{7×7}			RCA _{7×8}			RCA _{8×8}		
	RCB	PP	Δ%	RCB	PP	Δ%	RCB	PP	Δ%
FFT4	9	2	-77.8	9	2	-77.8	9	2	-77.8
FFT8	13	3	-76.9	13	3	-76.9	12	3	-75.0
EWf	45	11	-75.6	45	11	-75.6	45	11	-75.6
EWf3	45	11	-75.6	45	11	-75.6	49	12	-75.5
EWf6	77	19	-75.3	69	17	-75.4	77	19	-75.3
FDCT	51	7	-86.3	42	5	-88.1	46	5	-89.1
FDCT3	49	9	-81.6	59	8	-86.4	59	8	-86.4
FDCT6	97	16	-83.5	125	15	-88.0	87	12	-86.2
MATRIX4	22	5	-77.3	26	6	-76.9	26	5	-80.8
平均 Δ%			-78.9			-78.9			-80.2

表 5 $A_{RPU} = 49、56、64$ 时 PP 和 RCB 的 I_{ACC-ID} 比较

Tab.5 Comparison of the I_{ACC-ID} of PP and RCB when $A_{RPU} = 49, 56, 64$

划分基准	RCA _{7×7}			RCA _{7×8}			RCA _{8×8}		
	RCB	PP	Δ%	RCB	PP	Δ%	RCB	PP	Δ%
FFT4	8	6	-25.0	18	6	-66.7	18	6	-66.7
FFT8	51	15	-70.6	52	16	-69.2	52	16	-69.2
EWf	96	24	-75.0	96	24	-75.0	96	24	-75.0
EWf3	279	72	-74.2	279	72	-74.2	271	70	-74.2
EWf6	529	136	-74.3	512	131	-74.4	521	135	-74.1
FDCT	144	42	-70.8	148	44	-70.3	148	44	-70.3
FDCT3	306	92	-69.9	346	102	-70.5	334	94	-71.9
FDCT6	648	193	-70.2	462	188	-59.3	609	184	-69.8
MATRIX4	172	40	-76.7	180	42	-76.7	184	43	-76.6
平均 Δ%			-67.4			-70.7			-72.0

2.3 PP、RT、RCB 互连的 RCA 性能评估

了 PP-RCA、RT-RCA、RCB-RCA 考虑 I_{ACC-ID} 执行总时延比较. 由表 6~9 可知 PP-RCA 均能获得 RCA 总的计算总时延的减少.

表 6 和表 7 给出了 PP-RCA、RT-RCA、RCB-RCA 考虑 I_{MAX-ID} 执行总时延比较. 表 8 和表 9 给出

表 6 $A_{RPU} = 49、56、64$ 时 PP-RCA 和 RT-RCA 包含 I_{MAX-ID} 总时延的比较

Tab.6 Comparison of the total delay including I_{MAX-ID} of PP-RCA and RT-RCA when $A_{RPU} = 49, 56, 64$

划分基准	RCA _{7×7}		Δ%	RCA _{7×8}		Δ%	RCA _{8×8}		Δ%
	RT	PP		RT	PP		RT	PP	
FFT4	71	65	-8.5	71	65	-8.5	71	65	-8.5
FFT8	156	148	-5.1	137	129	-5.8	137	129	-5.8
EWf	184	160	-13.0	184	160	-13.0	185	161	-13.0
EWf3	311	287	-7.7	310	286	-7.7	314	288	-8.3
EWf6	568	528	-7.0	558	522	-6.5	559	519	-7.2
FDCT	168	141	-16.1	151	115	-23.8	136	115	-15.4
FDCT3	366	336	-8.2	338	320	-5.3	355	322	-9.3
FDCT6	688	626	-9.0	692	623	-10.0	644	595	-7.6
MATRIX4	339	323	-4.7	339	321	-5.3	338	319	-5.6
平均 Δ%			-8.8			-9.5			-9.0

表7 $A_{RPU} = 49、56、64$ 时 PP-RCA 和 RCB-RCA 包含 I_{MAX-ID} 总时延比较Tab.7 Comparison of the total delay including I_{MAX-ID} of PP-RCA and RCB-RCA when $A_{RPU} = 49, 56, 64$

划分基准	RCA _{7×7}			RCA _{7×8}			RCA _{8×8}		
	RCB	PP	$\Delta/\%$	RCB	PP	$\Delta/\%$	RCB	PP	$\Delta/\%$
FFT4	72	65	-9.7	72	65	-9.7	72	65	-9.7
FFT8	158	148	-6.3	139	129	-7.2	138	129	-6.5
EWf	194	160	-17.5	194	160	-17.5	195	161	-17.4
EWf3	321	287	-10.6	320	286	-10.6	325	288	-11.4
EWf6	586	528	-9.9	574	522	-9.1	577	519	-10.1
FDCT	185	141	-23.8	152	115	-24.3	156	115	-26.3
FDCT3	376	336	-10.6	371	360	-3.0	373	322	-13.7
FDCT6	707	626	-11.5	733	623	-15.0	670	595	-11.2
MATRIX4	340	323	-5.0	341	321	-5.9	340	319	-6.2
平均 $\Delta/\%$			-11.7			-11.4			-12.5

表8 $A_{RPU} = 49、56、64$ 时 PP-RCA 和 RT-RCA 包含 I_{ACC-ID} 总时延比较Tab.8 Comparison of the total delay including I_{ACC-ID} of PP-RCA and RT-RCA when $A_{RPU} = 49, 56, 64$

划分基准	RCA _{7×7}			RCA _{7×8}			RCA _{8×8}		
	RT	PP	$\Delta/\%$	RT	PP	$\Delta/\%$	RT	PP	$\Delta/\%$
FFT4	81	69	-14.8	81	69	-14.8	81	69	-14.8
FFT8	190	160	-15.8	174	142	-18.4	174	142	-18.4
EWf	221	173	-21.7	221	173	-21.7	222	174	-21.6
EWf3	492	348	-29.3	491	347	-29.3	486	346	-28.8
EWf6	917	645	-29.7	898	636	-29.2	905	635	-29.8
FDCT	296	176	-40.5	268	154	-42.5	268	154	-42.5
FDCT3	636	419	-34.1	674	414	-38.6	650	408	-37.2
FDCT6	1 280	803	-37.3	1 105	796	-28.0	1 219	767	-37.1
MATRIX4	474	358	-24.5	477	357	-25.2	479	357	-25.5
平均 $\Delta/\%$			-27.5			-27.5			-28.4

表9 $A_{RPU} = 49、56、64$ 时 PP-RCA 和 RCB-RCA 包含 I_{ACC-ID} 总时延比较Tab.9 Comparison of the total delay including I_{ACC-ID} of PP-RCA and RCB-RCA when $A_{RPU} = 49, 56, 64$

划分基准	RCA _{7×7}			RCA _{7×8}			RCA _{8×8}		
	RCB	PP	$\Delta/\%$	RCB	PP	$\Delta/\%$	RCB	PP	$\Delta/\%$
FFT4	81	69	-14.8	81	69	-14.8	81	69	-14.8
FFT8	196	160	-18.4	178	142	-20.2	178	142	-20.2
EWf	245	173	-29.4	245	173	-29.4	246	174	-29.3
EWf3	555	348	-37.3	554	347	-37.4	547	346	-36.7
EWf6	1 038	645	-37.9	1 017	636	-37.5	1 021	635	-37.8
FDCT	278	176	-36.7	258	154	-40.3	258	154	-40.3
FDCT3	633	419	-33.8	658	414	-37.1	648	408	-37.0
FDCT6	1 258	803	-36.2	1 070	796	-25.6	1 192	767	-35.7
MATRIX4	490	358	-26.9	495	357	-27.9	498	357	-28.3
平均 $\Delta/\%$			-30.2			-30.0			-31.1

2.4 互连时延与应用最终在 RCA 上的计算性能关系讨论

由表 2~表 9 可知, RT-RCA 和 RCB-RCA 的 I_{MAX-ID} 和 I_{ACC-ID} 均较大, 当 $A_{RPU} = 49$ 时, 以 FDCT6 为例. ①RT-RCA: $I_{MAX-ID} = 78$ cycle, RCA 上的计算总时延 $T = 688$ cycle, 所占的百分比 $\Delta = 11.3\%$; $I_{ACC-ID} = 670$ cycle, $T = 1 280$ cycle, $\Delta = 52.3\%$. ②RCB-RCA: $I_{MAX-ID} = 97$ cycle, $T = 707$ cycle, $\Delta = 13.7\%$; $I_{ACC-ID} = 648$ cycle, $T = 1 258$ cycle, $\Delta =$

51.5%. 由上可知, 互连时延对应用最终在 RCA 上的计算性能具有重要的影响.

2.5 小结

面向统一行并行 RCP 架构, 相比较 RT 和 RCB 互连 RCA, PP-RCA 获得最小的 I_{MAX-ID} 和 I_{ACC-ID} , PP-RCA 总体性能均优于 RT-RCA 和 RCB-RCA. PP-RCA 适用于运算性能较高, 并且要求 I_{MAX-ID} 和 I_{ACC-ID} 均较小的场合. 若 DFG 每个节点出度均很大, 两层近邻 RC 单元间传输数据线增多, PP-RCA、RT-

RCA、RCB-RCA 可能会出现互连线动态运行功耗的增加的情形,但是这种情况出现很少.由于 PP-RCA 互连 RC 之间均是用数据线直接相连,如果有一根数据线出现故障,整个 RCA 的计算结果就会出错,但路由型 RT-RCA 可以通过容错路由机制等来避免,故 PP-RCA 容错性能不如 RT-RCA 等架构.

3 结 语

基于不跨层行操作并行 RCA 互连时延性能评估,本文分析和比较了 PP、RT、RCB 互连方式的 RCA 架构,相比较 RT、RCB 互连 RCA 互连延迟,结果发现 PP 互连的 RCA 架构在减少 $I_{\text{MAX-ID}}$ 、 $I_{\text{ACC-ID}}$ 等方面优势明显,同时比较了 PP-RCA、RT-RCA、RCB-RCA 总的执行性能,结果表明 PP-RCA 优势明显,在减少互连延迟和 RCA 总时延方面,PP-RCA 架构具有可行性.

参考文献:

- [1] Kim Y, Lee J, Mai T, et al. Improving performance of nested loops on reconfigurable array processors [J]. *ACM Transactions on Architecture and Code Optimization*, 2012, 8(4): 1-32.
- [2] Zhao X, Erdogan A T, Arslan T. High-efficiency customized coarse-grained dynamically reconfigurable architecture for JPEG2000[J]. *IEEE Transactions on Very Large Scale Integration Systems*, 2013, 21(12): 2343-2348.
- [3] Cardoso J M P, Diniz C D, Weinhardt M. Compiling for reconfigurable computing: A survey [J]. *ACM Computing Surveys*, 2010, 42(4): 1301-1365.
- [4] Singh H, Lee M H, Lu G M, et al. MorphoSys: An integrated reconfigurable system for data parallel and computation intensive applications[J]. *IEEE Transactions on Computers*, 2000, 49(5): 465-481.
- [5] Mei B, Vernalde S, Verkest D, et al. ADRES: An architecture with tightly coupled VLIW processor and coarse-grained reconfigurable matrix [C]// *Proceedings of 13th International Conference on Field Programmable Logic and Application*. Lisbon, Portugal, 2003: 61-70.
- [6] 魏少军, 刘雷波, 尹首一. 可重构计算处理器技术 [J]. *中国科学: 信息科学*, 2012, 42(12): 1559-1576.
Wei Shaojun, Liu Leibo, Yin Shouyi. Key techniques of reconfigurable computing processor[J]. *Scientia Sinica: Informationis*, 2012, 42(12): 1559-1576 (in Chinese).
- [7] 窦 勇, 邬贵明, 徐进辉, 等. 支持循环流水的粗粒度可重构阵列体系结构[J]. *中国科学: 信息科学*, 2008, 38(4): 579-591.
Dou Yong, Wu Guiming, Xu Jinhui, et al. A coarse-grained reconfigurable computing architecture with loop self-pipelining[J]. *Scientia Sinica: Informationis*, 2008, 38(4): 579-591 (in Chinese).
- [8] Miyamori T, Olukotun K, Budiu M, et al. REMARC: Reconfigurable multimedia array coprocessor[J]. *IEICE Transactions on Information and Systems*, 1999, E82-D(2): 389-397.
- [9] Berekovic M, Kanstein A, Mei B, et al. Mapping of nomadic multimedia applications on the ADRES reconfigurable array processor[J]. *Microprocessors and Microsystems*, 2009, 33(4): 290-294.
- [10] Janakiraman N, Nirmalkumar P, Akram S M. Coarse grained ADRES based MIMO-OFDM transceiver with new radix-25 pipeline FFT/IFFT processor[J]. *Circuits, Systems, and Signal Processing*, 2015, 34(3): 851-873.
- [11] 陈乃金, 冯志勇, 江建慧. 用于二维 RCA 跨层数据传输的旁节点无冗余添加算法[J]. *通信学报*, 2015, 36(4): 1-17.
Chen Naijin, Feng Zhiyong, Jiang Jianhui. Bypass node non-redundant adding algorithm for crossing-level data transmission in two-dimension reconfigurable cell array[J]. *Journal on Communications*, 2015, 36(4): 1-17 (in Chinese).
- [12] 王大伟, 窦 勇, 李思昆. 核心循环到粗粒度可重构体系结构的流水化映射[J]. *计算机学报*, 2009, 32(6): 1089-1099.
Wang Dawei, Dou Yong, Li Sikun. Loop kernel pipelining mapping onto coarse-grained reconfigurable architectures[J]. *Chinese Journal of Computers*, 2009, 32(6): 1089-1099 (in Chinese).
- [13] 孙 康. 可重构计算相关技术研究[D]. 杭州: 浙江大学计算机科学与技术学院, 2007.
Sun Kang. Research on Reconfigurable Computing Technologies[D]. Hangzhou: College of Computer Science and Technology, Zhejiang University, 2007 (in Chinese).

(责任编辑: 王晓燕)