



(12) 发明专利申请

(10) 申请公布号 CN 101986264 A

(43) 申请公布日 2011. 03. 16

(21) 申请号 201010559300. 5

(22) 申请日 2010. 11. 25

(71) 申请人 中国人民解放军国防科学技术大学
地址 410073 湖南省长沙市开福区砚瓦池正街 47 号中国人民解放军国防科学技术大学计算机学院微电子与微处理器研究所

(72) 发明人 陈书明 杨惠 万江华 孙永节
孙锁林 胡封林 余再祥 罗恒
吴虎成 胡春媚 汪东 唐涛

(74) 专利代理机构 湖南兆弘专利事务所 43008
代理人 周长清

(51) Int. Cl.
G06F 9/302(2006. 01)

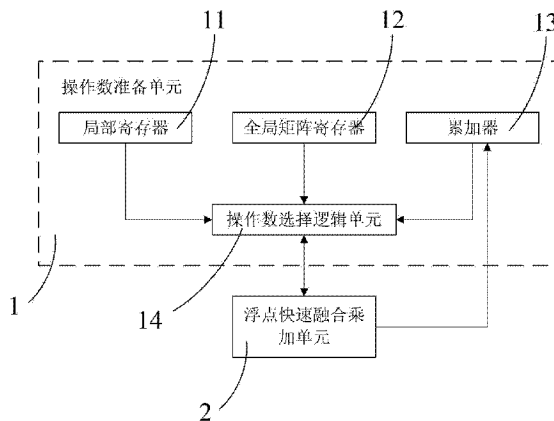
权利要求书 1 页 说明书 5 页 附图 5 页

(54) 发明名称

用于 SIMD 向量微处理器的多功能浮点乘加运算装置

(57) 摘要

本发明公开了一种用于 SIMD 向量微处理器的多功能浮点乘加运算装置, 它包括挂载于向量处理阵列(5)的每个向量处理单元(51)中的操作数准备单元(1)和浮点快速融合乘加单元(2), 操作数准备单元(1)将输入的源操作数处理得到指数、尾数和符号位分离的操作数并输出至浮点快速融合乘加单元(2), 浮点快速融合乘加单元(2)用于对所述操作数进行浮点乘加运算获取运算结果并将运算结果写回所述操作数准备单元(1)。本发明支持高效浮点矩阵运算和快速浮点求除法求倒数运算, 具有浮点运算性能好、开销小、功能多、编码少、速度快的优点。



1. 一种用于 SIMD 向量微处理器的多功能浮点乘加运算装置,其特征在于:它包括挂载于向量处理阵列(5)的每个向量处理单元(51)中的操作数准备单元(1)和浮点快速融合乘加单元(2),所述操作数准备单元(1)将输入的源操作数处理得到指数、尾数和符号位分离的操作数并输出至浮点快速融合乘加单元(2),所述浮点快速融合乘加单元(2)用于对所述操作数进行浮点乘加运算获取运算结果并将运算结果写回所述操作数准备单元(1)。

2. 根据权利要求1所述的用于 SIMD 向量微处理器的多功能浮点乘加运算装置,其特征在于:所述操作数准备单元(1)包括操作数选择逻辑单元(14)和分别与操作数选择逻辑单元(14)相连的局部寄存器(11)、全局矩阵寄存器(12)、累加器(13),所述局部寄存器(11)用于存放操作数或者写回的运算结果,可被本向量处理单元(51)访问;所述全局矩阵寄存器(12)用于存放操作数或者写回的运算结果,支持行列访问,可被全局向量处理单元(51)访问;所述累加器(13)用于存放操作数或者写回的运算结果;所述操作数选择逻辑单元(14)的数量与向量处理阵列(5)中向量处理单元(51)的数量相同,所述操作数选择逻辑单元(14)获取指数、尾数和符号位分离的操作数并输出至浮点快速融合乘加单元(2)。

3. 根据权利要求2所述的用于 SIMD 向量微处理器的多功能浮点乘加运算装置,其特征在于:所述操作数选择逻辑单元(14)包括数据选择单元(144)和分别与数据选择单元(144)相连的操作数例外判断单元(141)、倒数近似值查找表(142)、循环缓冲单元(143)、数据分离单元(145),所述操作数例外判断单元(141)分别与局部寄存器(11)、全局矩阵寄存器(12)、累加器(13)相连并对输入的操作数进行例外判断,所述倒数近似值查找表(142)用于为浮点除法运算的尾数输出倒数近似值,所述循环缓冲单元(143)用于存储除法或者求倒数运算过程的每次迭代后的中间结果,所述数据选择单元(144)与一个用于维护除法状态的状态机(146)相连,所述数据选择单元(144)根据执行运算类型选择进入数据分离单元(145)的操作数,所述数据分离单元(145)将操作数分离为尾数位、指数位和符号位并输出至浮点快速融合乘加单元(2)。

4. 根据权利要求1或2或3所述的用于 SIMD 向量微处理器的多功能浮点乘加运算装置,其特征在于:所述浮点快速融合乘加单元(2)包括符号处理模块(21)、指数处理模块(22)、尾数处理模块(23)和写回模块(24),所述符号处理模块(21)、指数处理模块(22)、尾数处理模块(23)分别与写回模块(24)相连,所述写回模块(24)将运算结果或者中间运算结果写回所述局部寄存器(11)或者累加器(13)。

用于 SIMD 向量微处理器的多功能浮点乘加运算装置

技术领域

[0001] 本发明涉及微处理器领域,特别涉及一种用于 SIMD 向量微处理器的多功能浮点乘加运算装置。

背景技术

[0002] 图像处理、雷达信号处理和现代通信等应用领域,需要超高性能的信号处理器对大量数据进行处理,对数据计算的精度和实时性要求高。由于算法具有高的浮点计算密集性,涉及大量的浮点矩阵运算及浮点乘加/减或卷积运算,处理器的浮点数据处理能力显得越发重要。针对上述的应用特点,已有研究中提出了各种浮点乘加运算部件的运行机制和硬件实现结构,使其支持浮点乘加运算。如美国专利 US0,101,242 A1 多功能乘加浮点单元 MULTIPURPOSE MULTIPLY-ADD FUNCTION UNIT 保护了一个用于图形处理器中可配置多功能浮点乘加运算单元,提供对双精度算术算法的直接支持;再如北京龙芯中科技术服务有限公司,针对浮点乘加运算,申请的关于单双精度的兼容处理装置和方法,向量浮点乘加运算装置和方法,从各角度提升浮点处理能力。

[0003] 现有技术普遍存在的缺点是:第一,硬件复用率低,功能单一,利用浮点融合乘加结构,仅实现了浮点的乘/加/乘加运算功能;第二,利用浮点乘加运算装置的现有的结构和运算特点,实现求除法和求倒数运算,往往采用传统的不恢复除法 SRT 算法,或查表再配合软件编程迭代,收敛速度慢周期长,占用较多硬件资源;第三,不能支持现代数字信号处理算法中大规模的快速的矩阵运算,或矩阵运算的处理能力不足,编程代码量大。

[0004] 综上所述,尽量复用传统浮点乘加装置,提供一种可快速支持除法和倒数运算,且支持现代信号处理中强大矩阵运算能力的浮点乘加运算装置,是本发明需要解决的问题。

发明内容

[0005] 本发明针对上述现有技术的缺点,提供一种支持高效浮点矩阵运算和快速浮点求除法求倒数运算,浮点运算性能好、开销小、功能多、编码少、速度快的用于 SIMD 向量微处理器的多功能浮点乘加运算装置。

[0006] 为了解决上述技术问题,本发明采用的技术方案为:

一种用于 SIMD 向量微处理器的多功能浮点乘加运算装置,它包括挂载于向量处理阵列的每个向量处理单元中的操作数准备单元和浮点快速融合乘加单元,所述操作数准备单元将输入的源操作数处理得到指数、尾数和符号位分离的操作数并输出至浮点快速融合乘加单元,所述浮点快速融合乘加单元用于对所述操作数进行浮点乘加运算获取运算结果并将运算结果写回所述操作数准备单元。

[0007] 作为本发明上述技术方案为进一步改进:

所述操作数准备单元包括操作数选择逻辑单元和分别与操作数选择逻辑单元相连的局部寄存器、全局矩阵寄存器、累加器,所述局部寄存器用于存放操作数或者写回的运算结果,可被本向量处理单元访问;所述全局矩阵寄存器用于存放操作数或者写回的运算结果,

支持行列访问,可被全局向量处理单元访问;所述累加器用于存放操作数或者写回的运算结果;所述操作数选择逻辑单元的数量与向量处理阵列中向量处理单元的数量相同,所述操作数选择逻辑单元获取指数、尾数和符号位分离的操作数并输出至浮点快速融合乘加单元。

[0008] 所述操作数选择逻辑单元包括数据选择单元和分别与数据选择单元相连的操作数例外判断单元、倒数近似值查找表、循环缓冲单元、数据分离单元,所述操作数例外判断单元分别与局部寄存器、全局矩阵寄存器、累加器相连并对输入的操作数进行例外判断,所述倒数近似值查找表用于为浮点除法运算的尾数输出倒数近似值,所述循环缓冲单元用于存储除法或者求倒数运算过程的每次迭代后的中间结果,所述数据选择单元与一个用于维护除法状态的状态机相连,所述数据选择单元根据执行运算类型选择进入数据分离单元的操作数,所述数据分离单元将操作数分离为尾数位、指数位和符号位输出至浮点快速融合乘加单元。

[0009] 所述浮点快速融合乘加单元包括符号处理模块、指数处理模块、尾数处理模块和写回模块,所述符号处理模块、指数处理模块、尾数处理模块分别与写回模块相连,所述写回模块将运算结果或者中间运算结果写回所述局部寄存器或者累加器。

[0010] 本发明具有下述优点:

本发明用于 SIMD 向量微处理器的多功能浮点乘加运算装置包括挂载于向量处理阵列的每个向量处理单元中的操作数准备单元和浮点快速融合乘加单元,因此可以作为向量处理单元的一个功能部件,因此可以在可进行传统的浮点乘法、加法、乘加运算的基础上可以实现浮点矩阵运算,具有浮点运算功能丰富、硬件复用率高、硬件开销小的优点。

[0011] 操作数准备单元包括操作数选择逻辑单元和分别与操作数选择逻辑单元相连的局部寄存器、全局矩阵寄存器、累加器,因此通过局部寄存器、支持全局行列访问的全局矩阵寄存器、累加器,可以实现支持大规模的快速矩阵运算,矩阵运算能力强大,矩阵运算性能好,而且不需要可以进行操作数的调度和选择,从而编程代码量小;

操作数选择逻辑单元包括数据选择单元和分别与数据选择单元相连的操作数例外判断单元、倒数近似值查找表、循环缓冲单元、数据分离单元,本发明实施例结合其硬件结构采用快速算法,周期短速度快,可以实现快速的求倒数和浮点除法运算,占用资源小、运算周期短、运算速度快;

浮点快速融合乘加单元包括符号处理模块、指数处理模块、尾数处理模块和写回模块,调度流水线更加合理,并且可以通过硬件实现并行执行、合并执行,浮点运算效率高。

附图说明

[0012] 图 1 为本发明实施例的主体框架结构示意图;

图 2 为本发明实施例的结构示意图;

图 3 为具有本发明实施例的 SIMD 向量微处理器的局部结构示意图;

图 4 为本发明实施例进行浮点除法运算的流程示意图;

图 5 为本发明实施例进行浮点矩阵运算的执行示意图。

[0013] 图例说明:1、操作数准备单元;11、局部寄存器;12、全局矩阵寄存器;13、累加器;14、操作数选择逻辑单元;141、操作数例外判断单元;142、倒数近似值查找表;143、循环缓

冲单元 ;144、数据选择单元 ;145、数据分离单元 ;146、状态机 ;2、浮点快速融合乘加单元 ;21、符号处理模块 ;22、指数处理模块 ;23、尾数处理模块 ;24、写回模块 ;3、指令控制器 ;4、标量处理单元 ;5、向量处理阵列 ;51、向量处理单元 ;6、数据混洗网络 ;7、规约网络 ;8、多功能浮点乘加运算装置。

具体实施方式

[0014] 如图 1 所示,本发明实施例的用于 SIMD 向量微处理器的多功能浮点乘加运算装置包括挂载于向量处理阵列 5 的每个向量处理单元 51 中的操作数准备单元 1 和浮点快速融合乘加单元 2,操作数准备单元 1 将输入的源操作数处理得到指数、尾数和符号位分离的操作数并输出至浮点快速融合乘加单元 2,浮点快速融合乘加单元 2 用于对操作数进行浮点乘加运算获取运算结果并将运算结果写回操作数准备单元 1。

[0015] 操作数准备单元 1 包括操作数选择逻辑单元 14 和分别与操作数选择逻辑单元 14 相连的局部寄存器 11、全局矩阵寄存器 12、累加器 13,局部寄存器 11 用于存放操作数或者写回的运算结果,可被本向量处理单元 51 访问 ;全局矩阵寄存器 12 用于存放操作数或者写回的运算结果,支持行列访问,可被全局向量处理单元 51 访问 ;累加器 13 用于存放操作数或者写回的运算结果 ;操作数选择逻辑单元 14 的数量与向量处理阵列 5 中向量处理单元 51 的数量相同,操作数选择逻辑单元 14 将处理得到指数、尾数和符号位分离的操作数输出至浮点快速融合乘加单元 2。本实施例中,局部寄存器 11 只能被多功能浮点乘加运算装置 8 所属的向量处理单元 51 访问,而全局矩阵寄存器 12 具有全局性,可以被整个向量处理阵列 5 的 n 个向量处理单元 51 进行访问,且支持行列访问,访问位宽最大为 $n*B$,其中 n 表示向量宽度, B 表示处理器字长。

[0016] 如图 2 所示,操作数选择逻辑单元 14 包括数据选择单元 144 和分别与数据选择单元 144 相连的操作数例外判断单元 141、倒数近似值查找表 142、循环缓冲单元 143、数据分离单元 145,操作数例外判断单元 141 分别与局部寄存器 11、全局矩阵寄存器 12、累加器 13 相连并对输入的操作数进行例外判断,倒数近似值查找表 142 用于为浮点除法运算的尾数输出倒数近似值,循环缓冲单元 143 用于存储除法或者求倒数运算过程的每次迭代后的中间结果,数据选择单元 144 与一个用于维护除法状态的状态机 146 相连,数据选择单元 144 根据执行运算类型选择进入数据分离单元 145 的操作数,数据分离单元 145 获取尾数位、指数位和符号位并输出至浮点快速融合乘加单元 2。本实施例中,倒数近似值查找表 142 的大小为 $2^7 \times 7$ bit。最终进入浮点快速融合乘加单元 2 的操作数来源有三个 :一是由操作数例外判断单元 141 对局部寄存器 11、全局矩阵寄存器 12、累加器 13 进行例外判断后的操作数 ;二是由操作数和倒数近似值查找表 142 进行查找后得到的精度为 7 位的中间结果 ;三是进行迭代运算过程中的写回结果。数据选择单元 144 保护两个数据选择器,数据选择单元 144 根据执行正常乘法、乘加操作,还是进行以乘代除的迭代运算,选择最终操作数。数据分离单元 145 则将操作数分离成为尾数位 M_a 、 M_b 、 M_c ,指数位 E_a 、 E_b 、 E_c ,符号位 S_a 、 S_b 、 S_c 。本实施例尽量复用了浮点快速融合乘加单元的硬件,使得在开销增加很少的前提下,能够实现单双精度浮点数除法、倒数运算,并大大提高浮点矩阵运算处理能力,与仅进行浮点乘加运算的结构相比,只增加了操作数准备单元中的部分逻辑,主要包括全局矩阵寄存器文件、维护除法操作的状态机、查找表,用于存放中间结果的循环缓冲单元。

[0017] 如图 2 所示,浮点快速融合乘加单元 2 包括符号处理模块 21、指数处理模块 22、尾数处理模块 23 和写回模块 24,符号处理模块 21、指数处理模块 22、尾数处理模块 23 分别与写回模块 24 相连,写回模块 24 将运算结果或者中间运算结果写回局部寄存器 11 或者累加器 13。本实施例在实际工作过程中,写回模块 24 将乘加单元运算的结果,作为除法或倒数运算的中间结果送到循环缓冲单元 143,或者作为乘加运算,矩阵运算的最终结果写回局部寄存器 11 或者累加器 13。

[0018] 如图 3 所示,具有本发明实施例的 SIMD 向量微处理器包括指令控制器 3、标量处理单元 4、向量处理阵列 5,标量处理单元 4 则主要面向串行应用的处理,向量处理阵列 5 包括 n 个向量处理单元 (VPE) 51, n 代表处理其的向量宽度,通常为 2 的幂,向量处理阵列 5 可面向计算密集型应用,配置为多组不同的 SIMD 宽度。标量处理单元 4、向量处理阵列 5 之间的数据交互可通过数据混洗网络 6 和规约网络 7 及数据存储体实现。指令控制器 3 完成指令的派发后,标量处理单元 4、向量处理阵列 5 分别对各自的指令进行译码,读取相应的操作数进行运算,并将运算的结果写入译码指定的目标存储中。每个向量处理单元 51 内部集成多个向量功能部件,使得基于超长指令字的指令集并行开发与基于向量的数据并行性开发同时进行,本发明实施例的多功能浮点乘加运算装置 8 则作为向量处理单元 51 内部的一个向量功能部件,因此整个 SIMD 向量微处理器共含有 n 个操作数准备单元 1 和 n 个浮点快速融合乘加单元 2,每一个多功能浮点乘加运算装置 8 的操作数准备单元 1 又包括 n 个操作数选择逻辑单元 14。多功能浮点乘加运算装置 8 主要实现的功能有:单/双精度浮点向量的乘和乘加/减运算、除法运算、求倒数运算、快速浮点矩阵运算。

[0019] 如图 4 所示,以源操作数 A 和 B (被除数 A 和除数 B,A 和 B 均可为矩阵或向量)进行浮点除法运算为例,本发明在进行浮点除法运算包括如下步骤:

1、将被除数 A 按行对应存放在局部寄存器 11 中,将除数 B 存放到全局矩阵寄存器 12 中,并假设操作数 0A 和操作数 0B 为除法运算中的一对操作数。

[0020] 2、0A 和 0B 操作数,经过操作数例外判断单元 141 进行例外判断和数据分离单元 145 进行符号指数尾数分离后,根据 0B 的尾数和倒数近似值查找表 142 进行倒数近似值的查找,获取精度为 7 位的倒数近似值。

[0021] 3、由控制除法执行状态的状态机 146 结合数据选择单元 144 通过指令译码获取的操作码,选择当前进入浮点快速乘加融合单元 2 的操作数,并记录迭代次数。

[0022] 4、将运算结果写回循环缓冲单元 143,进入下一次迭代,重复步骤 3,或者写回局部寄存器 11 中的目的寄存器。

[0023] 上述步骤中,由求倒数查找表可得到精度为 7 位的倒数近似值。由 IEEE-754 标准可知,单精度浮点的最终结果要求尾数满足 23 位精度,双精度浮点要求满足 52 位精度。因此查表得到的 7 位倒数近似值 L ,需进行迭代以满足精度要求。本发明采用的迭代方法是 Goldschmidt 算法,其初始函数如下:

$$D_j = L \times b$$

$$N_j = a \times L$$

迭代函数为:

$$R_{i-1} = 2 - D_{i-1}$$

$$D_i = D_{i-1} \times R_{i-1} = 1 - (1 - Lb)^{2^i}$$

$$N_i = N_{i-1} \times R_{i-1}$$

上述公式中, L 是指由查找表查表得到的 7 位倒数近似值, a 和 b 分别是由被除数 0A 和除数 0B 的尾数。每次迭代过程需要进行两次并行的乘法操作, 和一次求补操作, 这些操作数据不相关, 可全流水计算, 由浮点快速融合乘加单元 2 完成。每迭代一次精度增加一倍, 因此精度随着迭代次数的增加呈 14、28、56 增长。满足单精度需进行两次迭代, 满足双精度需进行三次迭代运算。

[0024] 本发明在进行浮点求倒数运算包括如下步骤:

1、实施 m 个浮点数 $a_1, a_2 \dots a_m$, 除以 m 个浮点数 $b_1, b_2 \dots b_m$, 或求 m 个 $b_1, b_2 \dots b_m$ 的倒数, 其中 $m \leq n$ 。获取除数的尾数, 通过 n 个操作数准备单元 1 中的倒数近似值查找表 142, 获得 m 个除数倒数的 7 位精度估计值。经过维护除法的状态机 146 和操作数选择判别后, 将该值对应输送到 m 个浮点快速融合乘加单元 2。

[0025] 2、在 n 个浮点快速融合乘加单元 2 中, 进行乘加或乘法操作, 完成一次迭代, 并将运算结果写回操作数准备单元 1, 作为中间运算结果存储在局部寄存器 11 中。迭代算法采用 Goldschmidt 快速除法算法, 以乘代除, 以二次速度逼近最终满足精度的结果, 迭代一次后, 精度达到 14 位, 迭代两次精度达到 28 位, 迭代三次后精度达到 56 位。

[0026] 3、每完成一次迭代, 结果写回操作数准备单元 1, 都通过维护除法操作的状态机 141 来提供控制信号, 从而支持除法的迭代运算。操作数来源可为: 局部寄存器 11 读出的操作数、倒数近似值查找表 142 中获取的数据、迭代运算中浮点快速融合乘加单元 2 产生的写回结果。根据操作类型译码信号和状态机 141 来控制操作数选择。

[0027] 如图 5 所示, 以实施矩阵 A 乘矩阵 B 加矩阵 C 为例, 本发明在进行浮点矩阵运算时, 首先, 将矩阵 A 和矩阵 C 中的元素, 按行分别加载到 n 个操作数准备单元 1 的局部寄存器 11 和向量累加器 13 中, 将矩阵 B 中的元素加载到全局矩阵寄存器 12 中, 经过操作数选择逻辑单元 14 判别后, 对应输送到 n 个浮点快速融合乘加单元 2; 然后, 将操作数准备单元 1 发送来的操作数, 在 n 个浮点快速融合乘加单元 2 中, 进行乘加或乘法操作, 并将运算结果写回操作数准备单元 1, 做为最终写回结果, 存放在存储矩阵 C 元素的局部寄存器 11 中。多功能浮点乘加运算装置 8 位于各个向量处理单元 51 内, 当进行浮点矩阵乘加操作时, 将矩阵 A 存入局部寄存器 11, 加载方式为在向量寄存器 11 的相同编码的寄存器中存放 A 矩阵的一行元素, 将矩阵 B 按列存入全局矩阵寄存器 12, 全局矩阵寄存器 12 支持行列访问, 将矩阵 C 存入向量累加器 13, 加载方式和 A 类似, 共同作为操作数的三个来源, 进入各个向量处理单元 51 并行做浮点乘法, 产生的结果由规约网络 7 进行叠加。因此, 同一周期内可完成矩阵 A 和矩阵 B 的一行一列的乘加运算, 计算最终矩阵乘累加结果的一行, 需 4 个周期, 整个矩阵乘加运算结果在 16 周期内产生, 从而可以快速实现浮点矩阵乘加。类似于浮点矩阵乘加, 本发明的操作数准备模块中提供的供数方式, 同样适用于矩阵加 / 减法 / 乘 / 除法等其他算数逻辑运算。

[0028] 以上仅是本发明的优选实施方式, 本发明的保护范围并不局限于上述实施例, 凡属于本发明思路下的技术方案均属于本发明的保护范围。应当指出, 对于本技术领域的普通技术人员来说, 在不脱离本发明原理前提下的若干改进和润饰, 应视为本发明的保护范围。

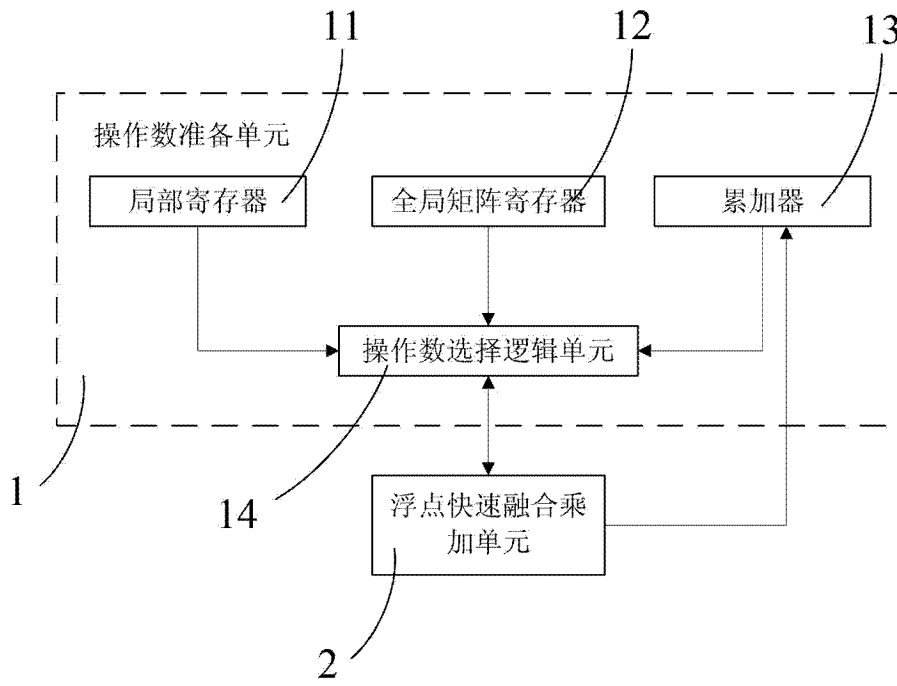


图 1

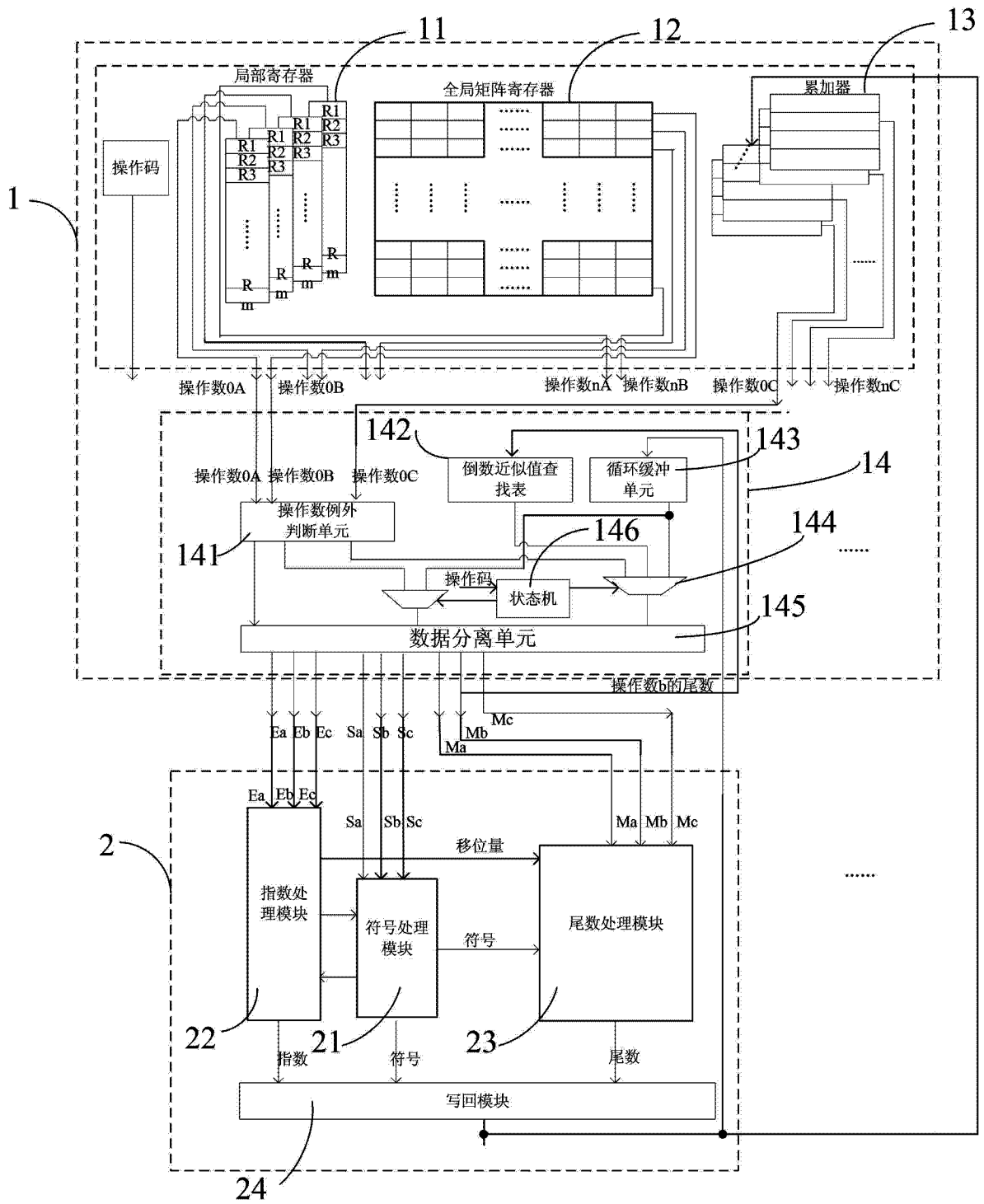


图 2

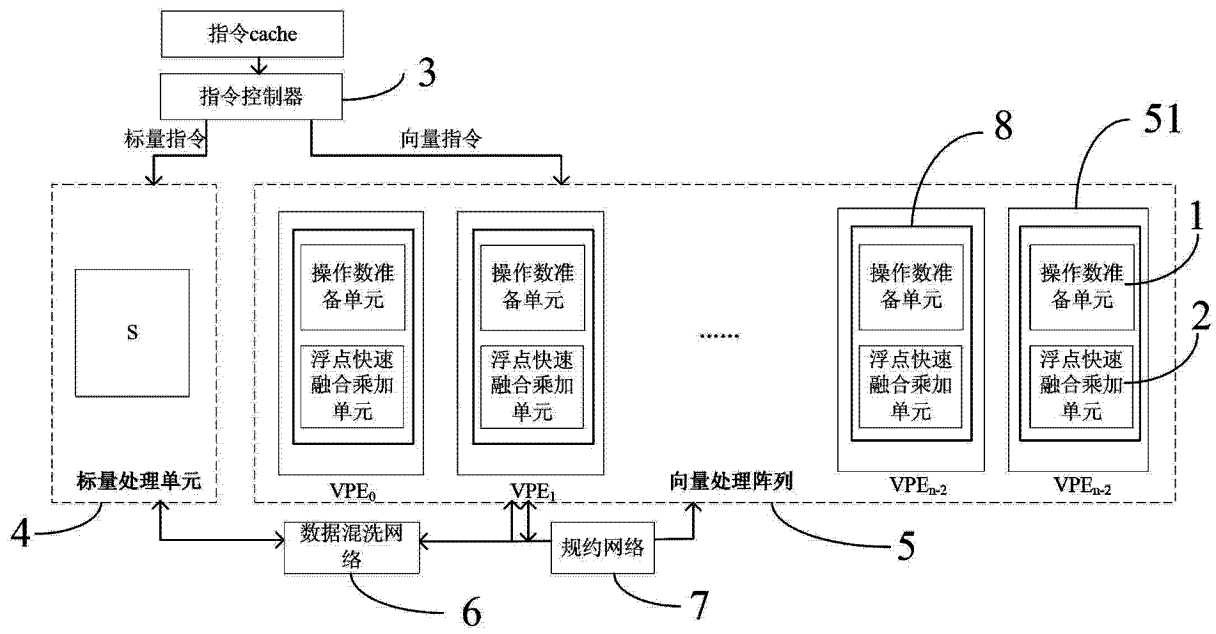


图 3

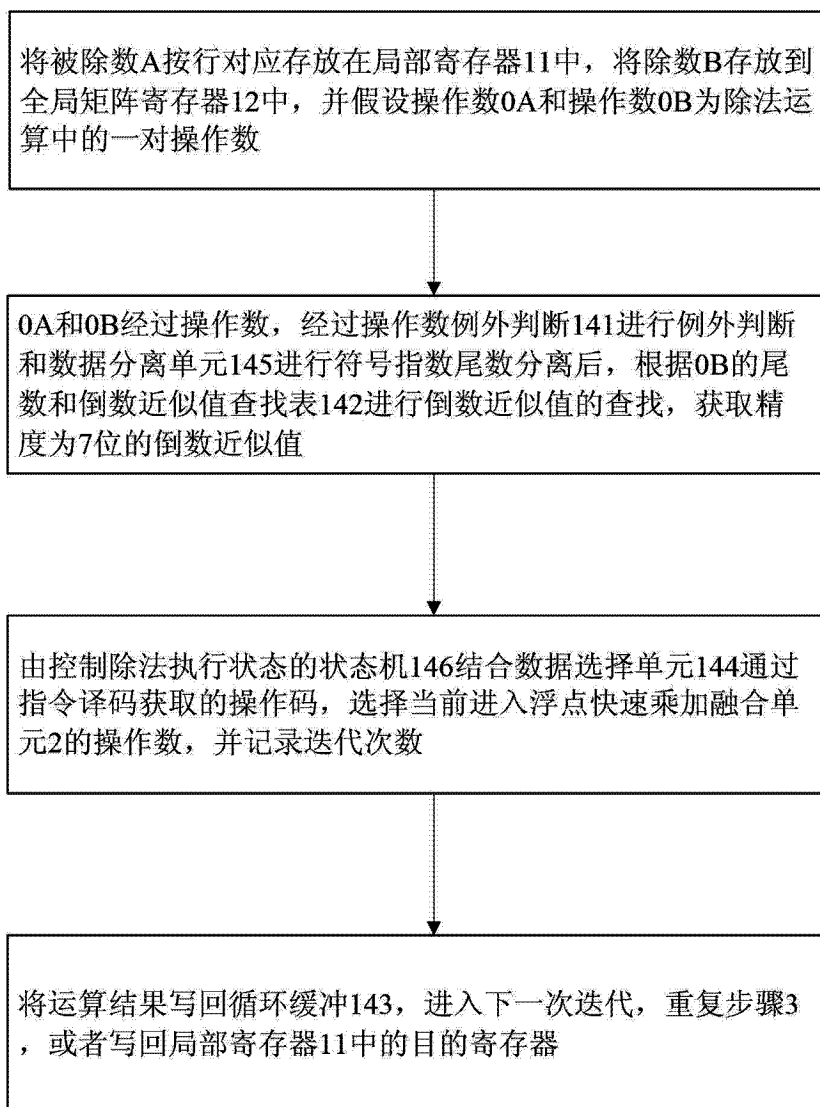


图 4

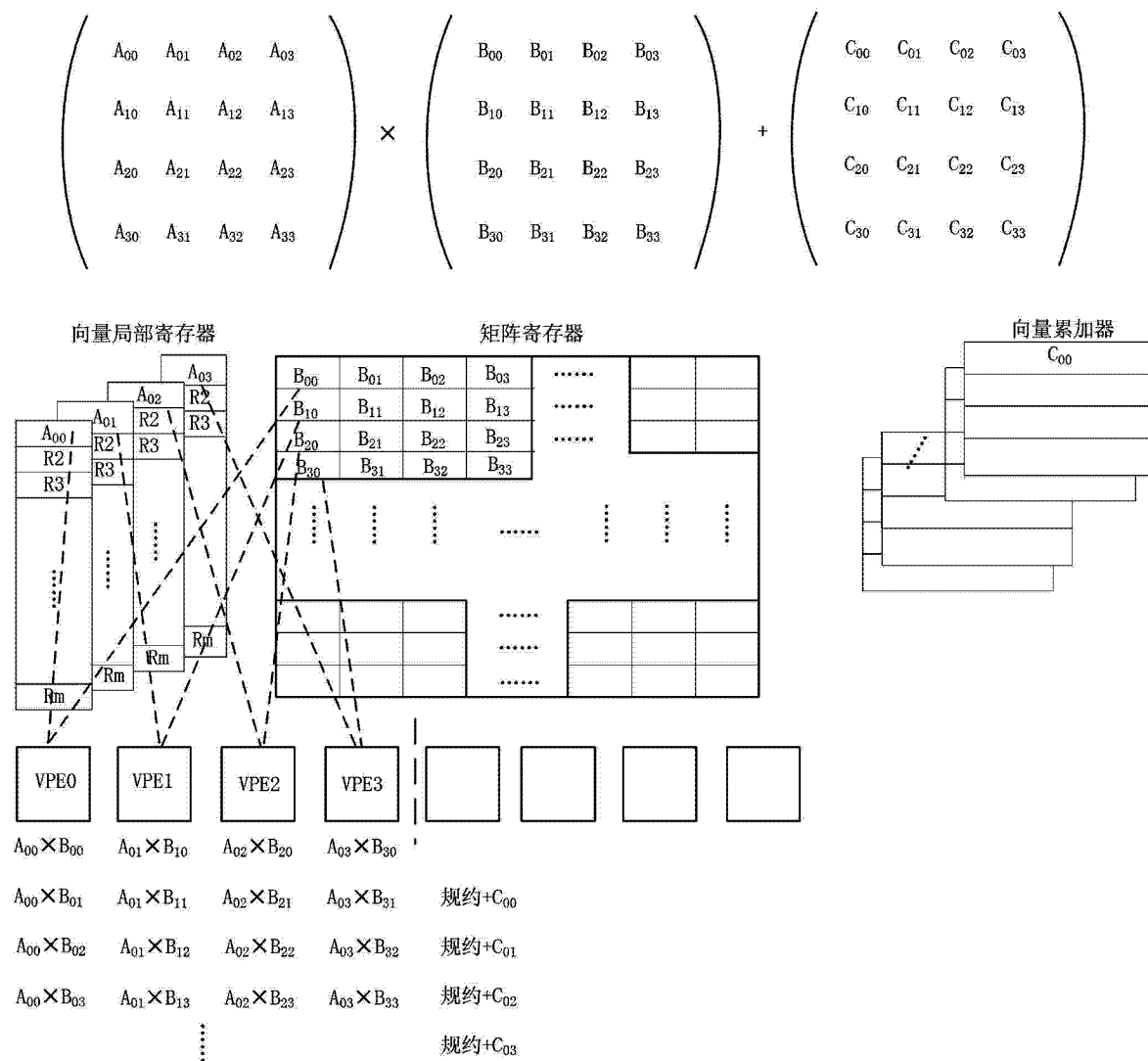


图 5