

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022 年 8 月 18 日 (18.08.2022)



(10) 国际公布号
WO 2022/170809 A1

- (51) 国际专利分类号:
G06F 7/575 (2006.01) **G06F 7/523** (2006.01)
- (21) 国际申请号: PCT/CN2021/131745
- (22) 国际申请日: 2021 年 11 月 19 日 (19.11.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110178984.2 2021年2月9日 (09.02.2021) CN
- (71) 申请人: 南方科技大学(SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY) [CN/CN]; 中国广东省深圳市南山区学苑大道 1088 号, Guangdong 518055 (CN)。
- (72) 发明人: 毛伟(MAO, Wei); 中国广东省深圳市南山区学苑大道1088号, Guangdong 518055 (CN)。
余浩(YU, Hao); 中国广东省深圳市南山区学

苑大道 1088 号, Guangdong 518055 (CN)。 谢歆昂(XIE, Xinang); 中国广东省深圳市南山区学苑大道 1088 号, Guangdong 518055 (CN)。 李凯(LI, Kai); 中国广东省深圳市南山区学苑大道 1088 号, Guangdong 518055 (CN)。 李博宇(LI, Boyu); 中国广东省深圳市南山区学苑大道 1088 号, Guangdong 518055 (CN)。 杜来民(DU, Laimin); 中国广东省深圳市南山区学苑大道 1088 号, Guangdong 518055 (CN)。 代柳瑶(DAI, Liuyao); 中国广东省深圳市南山区学苑大道 1088 号, Guangdong 518055 (CN)。

(74) 代理人: 深圳市君胜知识产权代理事务所(普通合伙)(JOHNSON INTELLECTUAL PROPERTY AGENCY(SHENZHEN)); 中国广东省深圳市南山区粤海街道高新区社区高新南七道 20 号深圳国家工程实验室大楼 A503, Guangdong 518000 (CN)。

(54) **Title:** RECONFIGURABLE FLOATING POINT MULTIPLY-ACCUMULATE OPERATION UNIT AND METHOD SUITABLE FOR MULTI-PRECISION CALCULATION

(54) 发明名称: 一种适用于多精度计算的可重构浮点乘加运算单元及方法

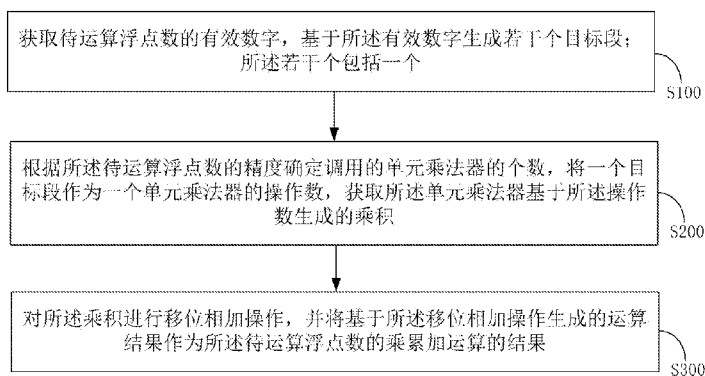


图 1

(57) **Abstract:** Disclosed in the present invention are a reconfigurable floating point multiply-accumulate operation unit and method suitable for multi-precision calculation. A uniform method is used to divide mantissas of floating points of different precision to obtain a plurality of bit segments; different numbers of same-type unit multipliers are called to implement multiplication operations of the plurality of bit segments in one period, and corresponding products are outputted; then, a shift-add operation is performed on the products to obtain a multiply-accumulate operation result of floating-point numbers. In the present invention, the problem of bit redundancy is avoided by employing a uniform mantissa division scheme, the hardware utilization rate is increased by employing a uniform unit multiplier, and the multiply-accumulate operation of half-precision floating-point numbers, the multiply-accumulate operation of single-precision dot product floating-point numbers, and the multiply-accumulate operation of double-precision floating-point numbers can be achieved. The problems in the prior art of bit redundancy, low hardware utilization rate and the like of an operation method supporting a multi-precision floating point multiplication operation are solved.

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本发明公开了一种适用于多精度计算的可重构浮点乘加运算单元及方法, 通过采用统一的方法对不同精度的浮点的尾数进行划分, 得到多个比特段, 并调用不同数量的同一类单元乘法器在一个周期内实现多个比特段的乘法运算并输出对应的乘积, 然后对所述乘积进行移位相加操作后即可得到浮点数的乘累加运算结果。本发明采用统一的尾数划分方案避免了比特冗余的问题, 采用统一的单元乘法器提高了硬件利用率, 还可以实现半精度浮点数的乘累加运算、单精度点积浮点数的乘累加运算和双精度浮点数的乘累加运算。解决了现有技术中支持多精度浮点乘法运算的运算方法会产生比特冗余、硬件利用率低等情况的问题。

一种适用于多精度计算的可重构浮点乘加运算单元及方法

技术领域

本发明涉及数字电路领域，尤其涉及的是一种适用于多精度计算的可重构浮点乘加运算单元及方法。

背景技术

随着科学计算和机器学习训练等的高速发展与广泛应用，能够支持浮点数据处理的乘法单元应运而生。常规定点乘法器的输入比特数固定，不能满足多精度计算的要求，因此出现了支持多精度浮点乘法运算的方法。然而现有的支持多精度浮点乘法运算的运算方法由于需要多种尾数划分方案以及需要将产生的乘积用补零方法分离为两个并行部分，因此存在精度损失以及比特冗余、硬件利用率低等问题。

因此，现有技术还有待改进和发展。

发明内容

本发明要解决的技术问题在于，针对现有技术的上述缺陷，提供一种适用于多精度计算的可重构浮点乘加运算单元及方法，旨在解决现有技术中支持多精度浮点乘法运算的运算方法会产生比特冗余、硬件利用率低等情况的问题。

本发明解决问题所采用的技术方案如下：

第一方面，本发明实施例提供一种适用于多精度计算的可重构浮点乘加运算方法，其中，所述方法包括：

获取待运算浮点数的有效数字，基于所述有效数字生成若干个目标段；所述若干个包括一个；

根据所述待运算浮点数的精度确定调用的单元乘法器的个数，将一个目标段作为一个单元乘法器的操作数，获取所述单元乘法器基于所述操作数生成的乘积；

对所述乘积进行移位相加操作，并将基于所述移位相加操作生成的运算结果作为所述待运算浮点数的乘累加运算的结果。

在一种实施方式中，所述获取待运算浮点数的有效数字，基于所述有效数字生成若干个目标段；所述若干个包括一个，包括：

在所述待运算浮点数的尾数部分添加1比特整数；

将添加完毕以后得到的浮点数的有效位数上的数字作为所述待运算浮点数的有效数字；

当所述有效数字的比特位数大于所述单元乘法器的比特位数时，根据所述单元乘法器的比特位数对所述有效数字进行划分，划分后生成若干个目标段；所述若干个包括一个。

在一种实施方式中，所述根据所述待运算浮点数的精度确定调用的单元乘法器的个数，将一个目标段作为一个单元乘法器的操作数，获取所述单元乘法器基于所述操作数生成的乘积包括：

根据所述待运算浮点数的精度确定调用的单元乘法器的个数；

将一个目标段作为一个单元乘法器的一个操作数；

将所述操作数输入所述单元乘法器后生成若干行乘积。

在一种实施方式中，当所述单元乘法器为14比特乘法器时，所述根据所述待运算浮点数的精度和对数确定调用的单元乘法器的个数包括：

当所述待运算浮点数为半精度浮点数时， n 对待运算浮点数调用 n 个单元乘法器；

当所述待运算浮点数为单精度浮点数时， n 对待运算浮点数调用 $4n$ 个单元乘法器；

当所述待运算浮点数为双精度浮点数时， n 对待运算浮点数调用 $16n$ 个单元乘法器；

n 为大于0的整数。

在一种实施方式中，所述将所述操作数输入所述单元乘法器后生成若干行乘积包括：

将所述操作数输入所述单元乘法器中，通过无符号位布斯对所述操作数进行编码后生成若干行乘积。

在一种实施方式中，当所述待运算浮点数为双精度浮点数时，所述将一个目标段作为一个单元乘法器的一个操作数之前还包括：

当所述待运算浮点数应的目标段的比特位数不相等时，对比特位数最小的目标段进行补位操作。

在一种实施方式中，所述对所述乘积进行移位相加操作，并将基于所述移位相加操作生成的运算结果作为所述待运算浮点数的乘累加运算的结果包括：

将所述乘积输入预设的加法树中；

计算所述乘积的位移量，通过所述加法树根据所述位移量对所述乘积进行移位操作；

对所述移位操作后得到的数据进行求和操作后得到所述待运算浮点数的乘累加运算的结果。

在一种实施方式中，所述位移量包括内部位移量以及外部位移量中至少一种位移量；

所述内部位移量的计算方式为：将基于所述待运算浮点数划分出的段数的高低位之和作为所述段数对应的乘积的内部移位量；

所述外部位移量的计算方式为：将所述待运算浮点数的指数部分相加得到指数和，将得到的所有指数和的最大值作为参考值；将所述参考值与所述指数和作差得到指数差，将所述指数差作为所述待运算浮点数对应的乘积的外部移位量。

第二方面，本发明实施例还提供一种适用于多精度计算的可重构浮点乘加运算单元，其特征在于，所述运算单元包括：

划分模块，用于获取待运算浮点数的有效数字，基于所述有效数字生成若干个目标段；所述若干个包括一个；

单元乘法器，用于根据所述待运算浮点数的精度确定调用的单元乘法器的个数，将一个目标段作为一个单元乘法器的操作数，获取所述单元乘法器基于所述操作数生成的乘积；

加法树，用于对所述乘积进行移位相加操作，并将基于所述移位相加操作生成的运算结果作为所述待运算浮点数的乘累加运算的结果。

在一种实施方式中，所述运算单元中包含 $16n$ 个单元乘法器， n 为非负数。

本发明的有益效果：本发明实施例通过采用统一的尾数划分方案避免了比特冗余的问题，采用统一的单元乘法器提高了硬件利用率，还可以实现半精度浮点数的乘累加运算、单精度浮点数的乘累加运算和双精度浮点数的乘累加运算。解决了现有技术中支持多精度浮点乘法运算的运算方法会产生比特冗余、硬件利用率低等情况的问题。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明中记载的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1是本发明实施例提供的一种适用于多精度计算的可重构浮点乘加运算方法的流程示意图。

图2是本发明实施例提供的不同精度的浮点数的有效数字的划分方案的示意图。

图3是本发明实施例提供的14比特基本乘法器的工作原理示意图。

图4是本发明实施例提供的计算一对FP64时，16组乘积输入加法器树的计算图。

图5是本发明实施例提供的一种适用于多精度计算的可重构浮点乘加运算单元的内部基本模块图。

图6是本发明实施例提供的可以实现3种不同精度的浮点数的尾数乘累加运算的最小运算单元的参考图。

具体实施方式

为使本发明的目的、技术方案及优点更加清楚、明确，以下参照附图并举实施例对本发明进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

需要说明，若本发明实施例中有涉及方向性指示（诸如上、下、左、右、前、后……），则该方向性指示仅用于解释在某一特定姿态（如附图所示）下各部件之间的相对位置关系、运动情况等，如果该特定姿态发生改变时，则该方向性指示也相应地随之改变。

随着科学计算和机器学习训练等的高速发展与广泛应用，能够支持浮点数据处理的乘法单元应运而生。常规定点乘法器的输入比特数固定，不能满足多精度计算的要求，无法针对应用需求，最大化的利用硬件资源从而提高能效比和吞吐率。因此产生了多精度浮点乘法运算的方法。然而部分现有的支持多精度浮点乘法运算的方法在实现多精度乘法时，需要将产生的乘积用补零方法分离为两个并行部分，进而导致系统模块的利用率降低；还有部分现有的支持多精度浮点乘法运算的方法在实现多精度乘法时，需要采用不同的尾数划分方案，如使架构基于15位乘法器，其优化用以支持FP128精度的浮点乘法运算，但用于其他精度的浮点乘法运算时，将会产生了大量的比特冗余和硬件资源浪费。简言之，常规定点乘法器的输入比特数固定，不能满足多精度计算的要求，无法针对应用需求，最大化的利用硬件资源从而提高能效比和吞吐率；而现有的支持多精度浮点乘法运算的运算方法又存在精度损失以及比特冗余、硬件利用率低等问题。

基于现有技术的上述缺陷，本发明提供一种适用于多精度计算的可重构浮点乘加运算方法，通过采用统一的方法对不同精度的浮点的尾数进行划分，得到多个比特段，并调用不同数量的同一类单元乘法器实现多个比特段的乘法运算在一个周期内完成并输出对应的乘积，然后对所述乘积进行移位相加操作后即可得到浮点的尾数相乘的运算结果。本发明采用统一的尾数划分方案避免了比特冗余的问题，采用统一的单元乘法器提高了硬件利用率，还可以实现半精度浮点数的乘累加运算、单精度浮点数的乘累加运算

和双精度浮点数的乘累加运算。解决了现有技术中支持多精度浮点乘法运算的运算方法会产生比特冗余、硬件利用率低等情况的问题。

如图1所示，所述方法包括如下步骤：

步骤S100、获取待运算浮点数的有效数字，基于所述有效数字生成若干个目标段；所述若干个包括一个。

在浮点数的乘法运算中，乘法运算结果的指数部分为相乘的两个浮点数的指数部分之和，乘法运算结果的尾数部分为相乘的两个浮点数的尾数之积。本实施例主要是针对浮点数的乘法运算中生成乘法运算结果的尾数部分，即相乘的两个浮点数的尾数之积的方法进行优化。具体地，首先本实施例需要获取待运算浮点数的有效数字，所述有效数字指的是待运算浮点数的尾数中需要参与乘法运算的数据，只有首先确定了需要参与乘法运算的有效数字，才能进行后续的乘法运算。获取到所述有效数字以后，本实施例需要基于所述有效数字生成一个或者多个目标段，然后再将所述目标段作为单元乘法器的输入数据。

所述步骤S100包括如下步骤：

步骤S110、在所述待运算浮点数的尾数部分添加1比特整数；

步骤S120、将添加完毕以后得到的浮点数的有效位数上的数字作为所述待运算浮点数的有效数字；

步骤S130、当所述有效数字的比特位数大于所述单元乘法器的比特位数时，根据所述单元乘法器的比特位数对所述有效数字进行划分，划分后生成若干个目标段；所述若干个包括一个。

具体地，为了获取到所述待运算浮点数的有效数字，本实施例首先需要在所述待运算浮点数的尾数部分添加1比特整数，然后将添加完毕以后得到的浮点数的有效位上的数字作为所述待运算浮点数的有效数字。举例说明，对于半精度浮点数（浮点16比特，FP16）的尾数部分加上1比特整数后，其有效位数为11位；对于单精度浮点数（浮点32

比特数，FP32）的尾数部分加上1比特整数后，其有效位数为24位；对于双精度浮点数（浮点64比特数，FP64）的尾数部分加上1比特整数后，其有效位数为53位。获取到有效数字以后，本实施例还需要根据所述有效数字得到目标段，并将所述目标段作为后续单元乘法器的输入数据。

具体地，本实施例需要比较所述有效数字的比特位数与所述单元乘法器的比特位数，并最终判断应该对所述有效数字进行何种处理，进而生成目标段。当所述有效数字的比特位数小于或者等于所述单元乘法器的比特位数时，可以直接将所述有效数字作为目标段输入所述单元乘法器中。举例说明，如图2所示，当所述单元乘法器为14位基本单元乘法器时，16比特浮点数的有效数字只有11位，因此不需要对16比特浮点数的有效数字进行划分，可以直接将其作为一个目标段。

当所述有效数字的比特位数大于所述单元乘法器的比特位数时，很明显无法将所述有效数字直接输入所述单元乘法器中，因此需要对所述有效数字进行划分，再将划分后生成的若干个目标段输入单元乘法器中。举例说明，当所述单元乘法器为14位基本单元乘法器时，32比特浮点数的有效数字为24位，因此需要对该有效数字进行划分，生成2个12比特的目标段。同理可得，64比特浮点数的有效数字为53位，因此也需要对该有效数字进行划分，进而生成14：13：13：13的4个目标段。

获取到目标段以后，需要将所述目标段输入单元乘法器，因此如图1所示，所述方法还包括如下步骤：

步骤S200、根据所述待运算浮点数的精度确定调用的单元乘法器的个数，将一个目标段作为一个单元乘法器的操作数，获取所述单元乘法器基于所述操作数生成的乘积。

本实施例首先需要根据所述待运算浮点数的精度确定调用的单位乘法的个数。然后将得到的一个目标段作为一个单元乘法器的一个操作数，可以理解的是一个单元乘法器需要两个操作数才能进行乘法运算，一个操作数作为乘数，另一个操作数作为被乘数。然后获取所述单元乘法器基于所述操作数生成的乘积。在乘法中，如果乘数是两位或两

位以上的数，乘的时候，就要用乘数的每一位去乘被乘数，每次乘得的积，叫做乘积，或叫做不完全积。

所述步骤S200具体包括如下步骤：

步骤210、根据所述待运算浮点数的精度确定调用的单元乘法器的个数；

步骤220、将一个目标段作为一个单元乘法器的一个操作数；

步骤230、将所述操作数输入所述单元乘法器后生成若干行乘积。

本实施例中首先需要确定调用的单元乘法器的个数。在一种实现方式中，当所述单元乘法器为14比特乘法器时，所述根据所述待运算浮点数的精度和对数确定调用的单元乘法器的个数包括：当所述待运算浮点数为半精度浮点数时， n 对待运算浮点数调用 n 个单元乘法器；当所述待运算浮点数为单精度浮点数时， n 对待运算浮点数调用 $4n$ 个单元乘法器；当所述待运算浮点数为双精度浮点数时， n 对待运算浮点数调用 $16n$ 个单元乘法器； n 为大于0的整数。

举例说明，假设所述单元乘法器为14比特基本单元乘法器，当需要同时计算16对半精度浮点数的乘累加运算结果时，需要调用16个14比特基本单元乘法器，理由如下，因为本实施例是基于单元乘法器的比特位数对所述有效数字进行划分，半精度浮点数的有效数字可以直接作为一个目标段，因此同时计算16对半精度浮点数的乘累加运算结果时，每一对半精度浮点数需要调用1个14比特基本单元乘法器，一共需要调用16个14比特基本单元乘法器。同理，当需要同时计算4对单精度浮点数的乘累加运算结果时，也需要调用16个14比特基本单元乘法器。因为单精度浮点数的有效数字需要划分后才能输入单元乘法器中，其划分结果是生成2个目标段，则1对单精度浮点数对应的4个目标段之间有 $2*2=4$ 种乘法组合方式，即需要4个单元乘法器，则4对单精度浮点数的乘累加运算就需要 $4*4=16$ 个单元乘法器。同理当需要同时计算1对双精度浮点数的乘累加运算结果时，也需要调用16个14比特基本单元乘法器，理由如下，因为双精度浮点数的有效数字划分后生成4个目标段，则1对双精度浮点数对应的8个目标段之间有 $4*4=16$ 种乘法组

合方式，因此一共需要16个14比特基本单元乘法器。

此外，由于对有效数字进行划分后有可能生成比特位数不相等的目标段，因此在一种实现方式中，将一个目标段作为一个单元乘法器的一个操作数之前还包括：当所述待运算浮点数应的目标段的比特位数不相等时，对比特位数最小的目标段进行补位操作，所述补位操作可以以补零的方式实现。举例说明，当所述单元乘法器为14比特基本单元乘法器时，双精度浮点数对应的有效数字为53位，划分后生成的14：13：13：13的4个目标端，则需要对13比特的目标段进行补零操作。

然后将一个目标段作为一个单元乘法器的一个操作数，之后获取所述单元乘法器生成的若干行乘积。具体地，所述操作数输入到单元乘法器以后，在所述单元乘法器里会通过无符号位布斯(booth)对所述操作数进行编码后生成若干行乘积（如图3所示）。

获取到乘积以后，为了获得浮点数的乘累加运算的结果，如图1所示，所述方法还包括如下步骤：

步骤S300、对所述乘积进行移位相加操作，并将基于所述移位相加操作生成的运算结果作为所述待运算浮点数的乘累加运算的结果。

本实施例获取到单元乘法器输出的乘积以后，为了获取到准确的乘法运算结果，还需要对得到的乘积进行移位相加操作以后，再将移位相加操作后的运算结果作为所述待运算浮点数的乘累加运算的结果。

在一种实现方式中，所述步骤S300具体包括如下步骤：

步骤S310、将所述乘积输入预设的加法树中；

步骤S320、计算所述乘积的位移量，通过所述加法树根据所述位移量对所述乘积进行移位操作；

步骤S330、对所述移位操作后得到的数据进行求和操作后得到所述待运算浮点数的乘累加运算的结果。

本实施例针对生成目标段的方案以及单元乘法器的使用情况，预先设置了一个加法

树，以实现数据的无损处理。具体地，获取到乘积以后，将所述乘积输入到所述加法树中，然后在所述加法树中计算出所述乘积的位移量，再根据所述位移量对所述乘积进行移位操作。具体地，所述加法树中计算出的位移量包括内部位移量以及外部位移量中至少一种位移量，即可以仅包括内部位移量，也可以仅包括外部位移量，还可以同时包括内部位移量和外部位移量，位移量的最终值需要具体根据待运算浮点数的精度以及计算对数确定。其中，所述内部位移量的计算方式为将基于所述待运算浮点数划分出的段数的高低位之和作为所述段数对应的乘积的内部移位置量。所述外部位移量的计算方式为：将所述待运算浮点数的指数部分相加得到指数和，将得到的所有指数和的最大值作为参考值，然后将所述参考值与所述指数和作差得到指数差，最后将所述指数差作为所述待运算浮点数对应的乘积的外部移位置量。

简言之，当同时计算多对浮点数的时候，则浮点数本身指数部分的数字需要作为外部位移量。当对所述有效数字进行划分后，由于调用了不同的单元乘法器进行运算，虽然每一个单元乘法器输出乘积的比特位数相同，但是需要将划分出的段数的高低位之和作为内部位移量，并进行相应的高低位移位后再累加才可生成正确的浮点数的乘累加运算结果。

举例说明，同时计算16对半精度浮点数的乘累加运算结果，需要调用16个14比特基本单元乘法器，由于半精度浮点数的有效数字并未进行划分，但是需要同时计算多对浮点数的乘累加运算结果，因此在这种情况下乘积的位移量只有外部位移量，首先查询16对半精度浮点数的16个指数和，将16个指数和中的最大的和作为参考值，将所述参考值分别减去16个不同的指数和，以此得到外部位移量。

当同时计算4对单精度浮点数的乘累加运算结果时，由于单精度浮点数的有效数字是经过划分以后才输入单元乘法器中的，且需要同时计算多对浮点数的乘累加运算结果，因此这种情况下乘积的位移量既包含内部位移量也包含外部位移量。首先查询4对单精度浮点数的4个指数和，将4个指数和中的最大的和作为参考值，然后将所述参考值

分别减去4个不同的指数和，以此得到外部位移量。此外，还需要将基于所述待运算浮点数划分出的段数的高低位之作为所述段数对应的乘积的内部移位量，例如 $a_0 \times b_0$ 的段数和位 $0+0=0$ ，则内部位移量为0； $a_1 \times b_0$ 的段数和位 $1+0=1$ ，则内部位移量为向左移位 $1 \times 14=14$ 位； $a_3 \times b_1$ 的段数和位 $3+1=4$ ，则内部位移量为向左移位 $4 \times 14=56$ 位。

当计算1对双精度浮点数的乘累加运算结果时，由于双精度浮点数的有效数字是经过划分以后才输入单元乘法器中的，但是只需要计算1对浮点数的乘累加运算结果，因此这种情况下乘积的位移量只包含内部位移量，例如 $a_1 \times b_0$ 结果移位量为 $1 \times 14\text{bit}$ 、 $a_1 \times b_1$ 及 $a_2 \times b_0$ 结果移位量均为 $2 \times 14\text{bit}$ （如图4所示）。

计算出乘积的位移量以后，根据所述位移量对所述乘积进行移位操作，然后对所述移位操作后得到的数据进行求和操作以后，即可得到所述待运算浮点数的乘累加运算的结果。

基于上述实施例，本发明还提供一种适用于多精度计算的可重构浮点乘加运算单元，如图5所示，所述运算单元包括：

划分模块01，用于获取待运算浮点数的有效数字，基于所述有效数字生成若干个目标段；所述若干个包括一个；

单元乘法器02，用于根据所述待运算浮点数的精度确定调用的单元乘法器的个数，将一个目标段作为一个单元乘法器的操作数，获取所述单元乘法器基于所述操作数生成的乘积；

加法树03，用于对所述乘积进行移位相加操作，并将基于所述移位相加操作生成的运算结果作为所述待运算浮点数的乘累加运算的结果。

在一种实现方式中，为了使所述运算单元可以实现半精度浮点数、单精度浮点数以及双精度浮点数的乘累加运算，所述运算单元中包含 $16n$ 个单元乘法器， n 为非负数。举例说明，如图6所示，当所述运算单元中包含16个14比特基本单元乘法器时，每个单元乘法器可以实现1组半精度浮点数乘法运算，因此可以同时实现16对半精度浮点数的乘

累加运算。每4个单元乘法器可以实现1组单精度浮点数乘法运算，因此还可以同时实现4对单精度浮点数的乘累加运算。16个单元乘法器可以实现1组双精度浮点数乘法运算，因此还可以实现1对双精度浮点数的乘累加运算。图6是本发明提供的可以实现3种不同精度的浮点数的乘累加运算的最小运算单元的参考图。因此本发明实施例可在无限制硬件资源的情况下在一个时钟周期内至少完成多对半精度浮点数的乘累加运算、多对单精度浮点数的乘累加运算或者1对双精度浮点数的乘累加运算。与固定的FP32和FP64乘加单元相比，本发明提供的运算单元可分别提高4倍和16倍的最大吞吐率。

综上所述，本发明公开了一种适用于多精度计算的可重构浮点乘加运算单元及方法，通过采用统一的方法对不同精度的浮点的尾数进行划分，得到多个比特段，并调用不同数量的同一类单元乘法器实现多个比特段的乘法运算在一个周期内完成并输出对应的乘积，然后对所述乘积进行移位相加操作后即可得到浮点数的乘累加运算结果。本发明采用统一的尾数划分方案避免了比特冗余的问题，采用统一的单元乘法器提高了硬件利用率，还可以实现半精度浮点数的乘累加运算、单精度浮点数的乘累加运算和双精度浮点数的乘累加运算。解决了现有技术中支持多精度浮点乘法运算的运算方法会产生比特冗余、硬件利用率低等情况的问题。

应当理解的是，本发明的应用不限于上述的举例，对本领域普通技术人员来说，可以根据上述说明加以改进或变换，所有这些改进和变换都应属于本发明所附权利要求的保护范围。

权利要求书

1.一种适用于多精度计算的可重构浮点乘加运算方法，其特征在于，所述方法包括：

获取待运算浮点数的有效数字，基于所述有效数字生成若干个目标段；所述若干个包括一个；

根据所述待运算浮点数的精度确定调用的单元乘法器的个数，将一个目标段作为一个单元乘法器的操作数，获取所述单元乘法器基于所述操作数生成的乘积；

对所述乘积进行移位相加操作，并将基于所述移位相加操作生成的运算结果作为所述待运算浮点数的乘累加运算的结果。

2.根据权利要求1所述的一种适用于多精度计算的可重构浮点乘加运算方法，其特征在于，所述获取待运算浮点数的有效数字，基于所述有效数字生成若干个目标段；所述若干个包括一个包括：

在所述待运算浮点数的尾数部分添加1比特整数；

将添加完毕以后得到的浮点数的有效位数上的数字作为所述待运算浮点数的有效数字；

当所述有效数字的比特位数大于所述单元乘法器的比特位数时，根据所述单元乘法器的比特位数对所述有效数字进行划分，划分后生成若干个目标段；所述若干个包括一个。

3.根据权利要求1所述的一种适用于多精度计算的可重构浮点乘加运算方法，其特征在于，所述根据所述待运算浮点数的精度确定调用的单元乘法器的个数，将一个目标段作为一个单元乘法器的操作数，获取所述单元乘法器基于所述操作数生成的乘积包括：

根据所述待运算浮点数的精度确定调用的单元乘法器的个数；

将一个目标段作为一个单元乘法器的一个操作数；

将所述操作数输入所述单元乘法器后生成若干行乘积。

4.根据权利要求3所述的一种适用于多精度计算的可重构浮点乘加运算方法，其特征在于，当所述单元乘法器为14比特乘法器时，所述根据所述待运算浮点数的精度和对数

确定调用的单元乘法器的个数包括：

当所述待运算浮点数为半精度浮点数时， n 对待运算浮点数调用 n 个单元乘法器；

当所述待运算浮点数为单精度浮点数时， n 对待运算浮点数调用 $4n$ 个单元乘法器；

当所述待运算浮点数为双精度浮点数时， n 对待运算浮点数调用 $16n$ 个单元乘法器；

n 为大于0的整数。

5.根据权利要求3所述的一种适用于多精度计算的可重构浮点乘加运算方法，其特征在于，所述将所述操作数输入所述单元乘法器后生成若干行乘积包括：

将所述操作数输入所述单元乘法器中，通过无符号位布斯对所述操作数进行编码后生成若干行乘积。

6.根据权利要求3所述的一种适用于多精度计算的可重构浮点乘加运算方法，其特征在于，当所述待运算浮点数为双精度浮点数时，所述将一个目标段作为一个单元乘法器的一个操作数之前还包括：

当所述待运算浮点数应的目标段的比特位数不相等时，对比特位数最小的目标段进行补位操作。

7.根据权利要求1所述的一种适用于多精度计算的可重构浮点乘加运算方法，其特征在于，所述对所述乘积进行移位相加操作，并将基于所述移位相加操作生成的运算结果作为所述待运算浮点数的乘累加运算的结果包括：

将所述乘积输入预设的加法树中；

计算所述乘积的位移量，通过所述加法树根据所述位移量对所述乘积进行移位操作；

对所述移位操作后得到的数据进行求和操作后得到所述待运算浮点数的乘累加运算的结果。

8.根据权利要求7所述的一种适用于多精度计算的可重构浮点乘加运算方法，其特征在于，所述位移量包括内部位移量以及外部位移量中至少一种位移量；

所述内部位移量的计算方式为：将基于所述待运算浮点数划分出的段数的高低位之和作为所述段数对应的乘积的内部移位量；

所述外部位移量的计算方式为：将所述待运算浮点数的指数部分相加得到指数和，将得到的所有指数和的最大值作为参考值；将所述参考值与所述指数和作差得到指数差，将所述指数差作为所述待运算浮点数对应的乘积的外部移位量。

9.一种适用于多精度计算的可重构浮点乘加运算单元，其特征在于，所述运算单元包括：

划分模块，用于获取待运算浮点数的有效数字，基于所述有效数字生成若干个目标段；所述若干个包括一个；

单元乘法器，用于根据所述待运算浮点数的精度确定调用的单元乘法器的个数，将一个目标段作为一个单元乘法器的操作数，获取所述单元乘法器基于所述操作数生成的乘积；

加法树，用于对所述乘积进行移位相加操作，并将基于所述移位相加操作生成的运算结果作为所述待运算浮点数的乘累加运算的结果。

10.根据权利要求9所述的一种适用于多精度计算的可重构浮点乘加运算单元，其特征在于，所述运算单元中包含 $16n$ 个单元乘法器， n 为非负数。

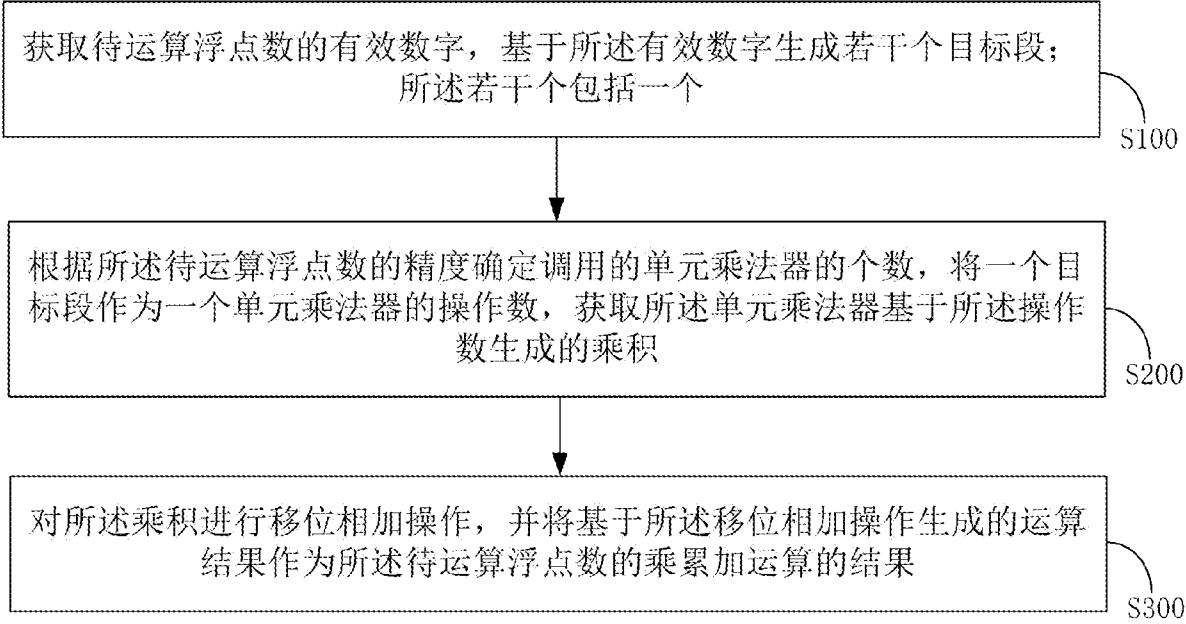


图 1

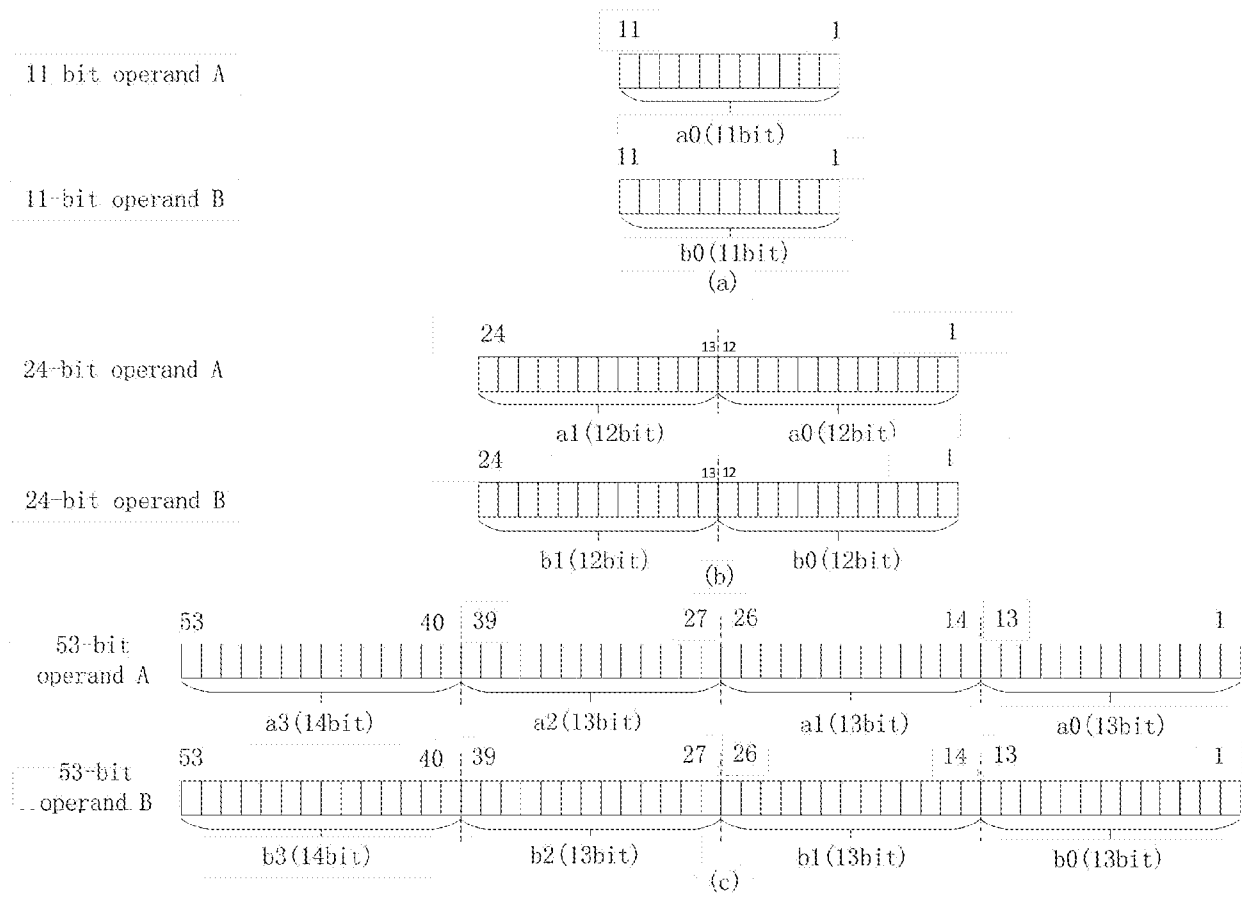


图 2

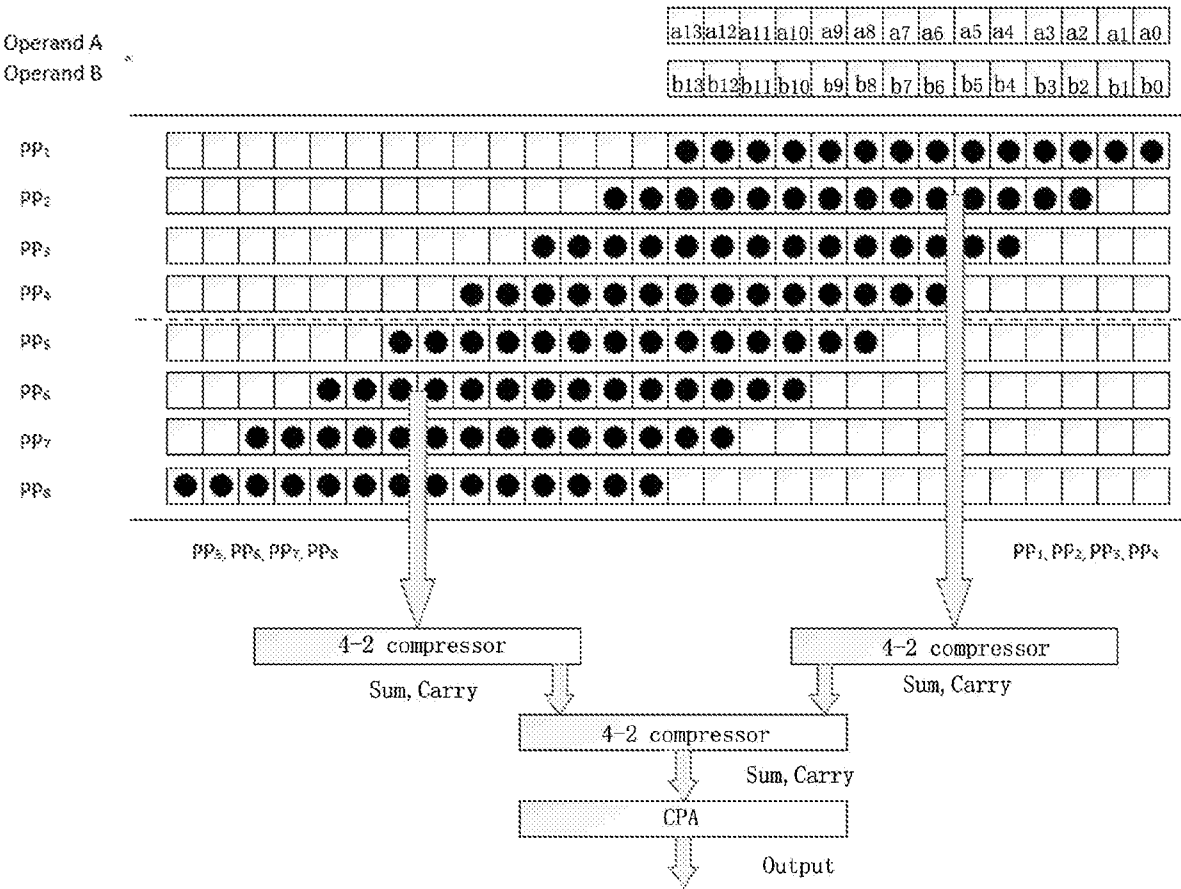


图 3

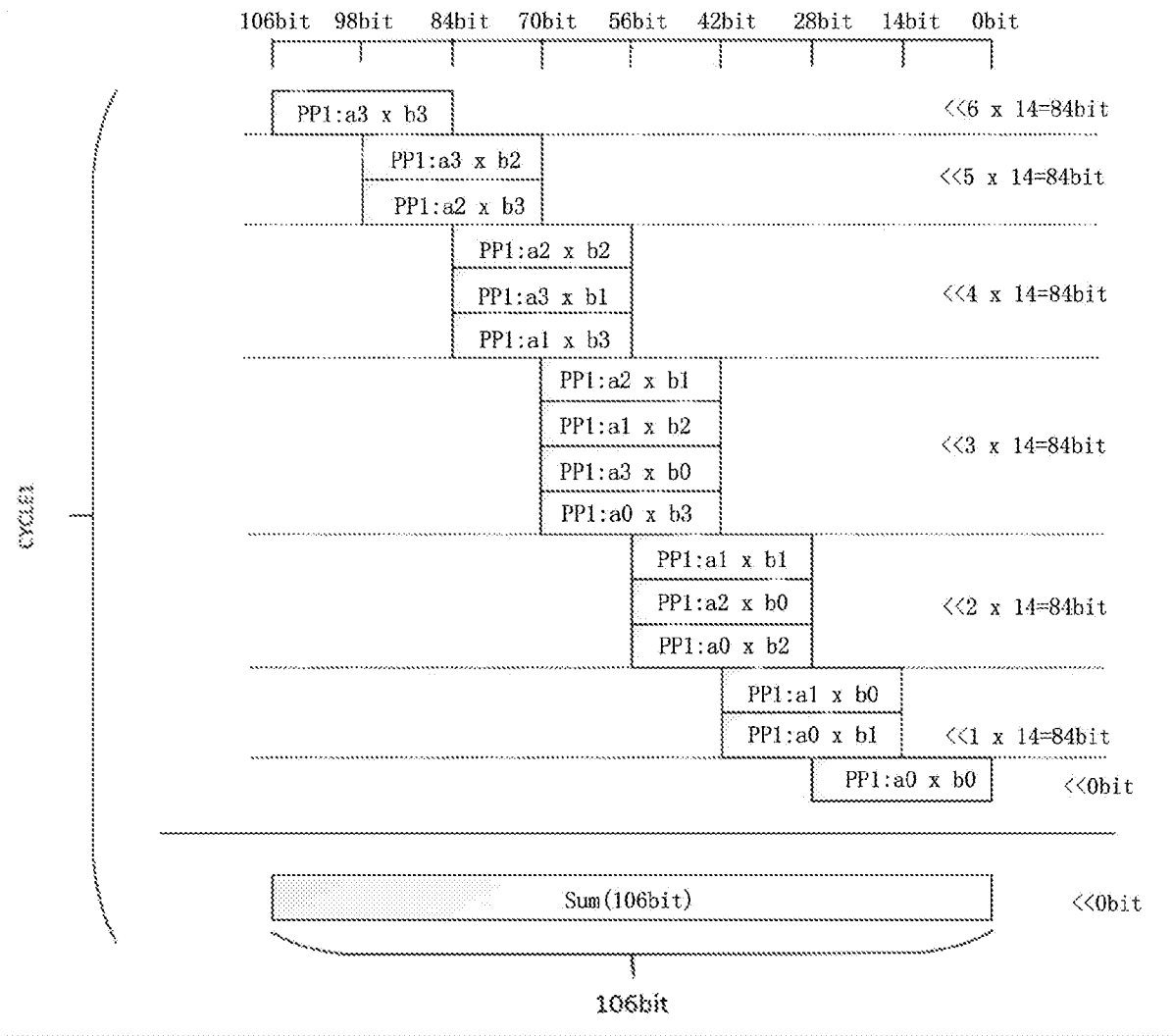


图 4

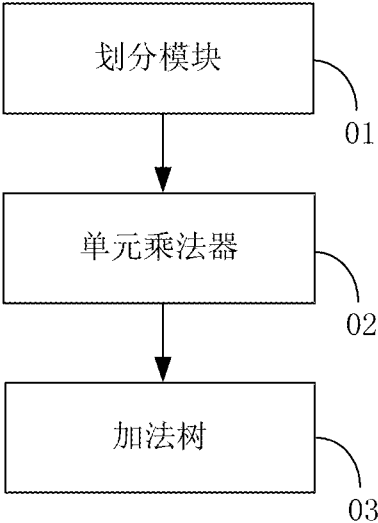


图 5

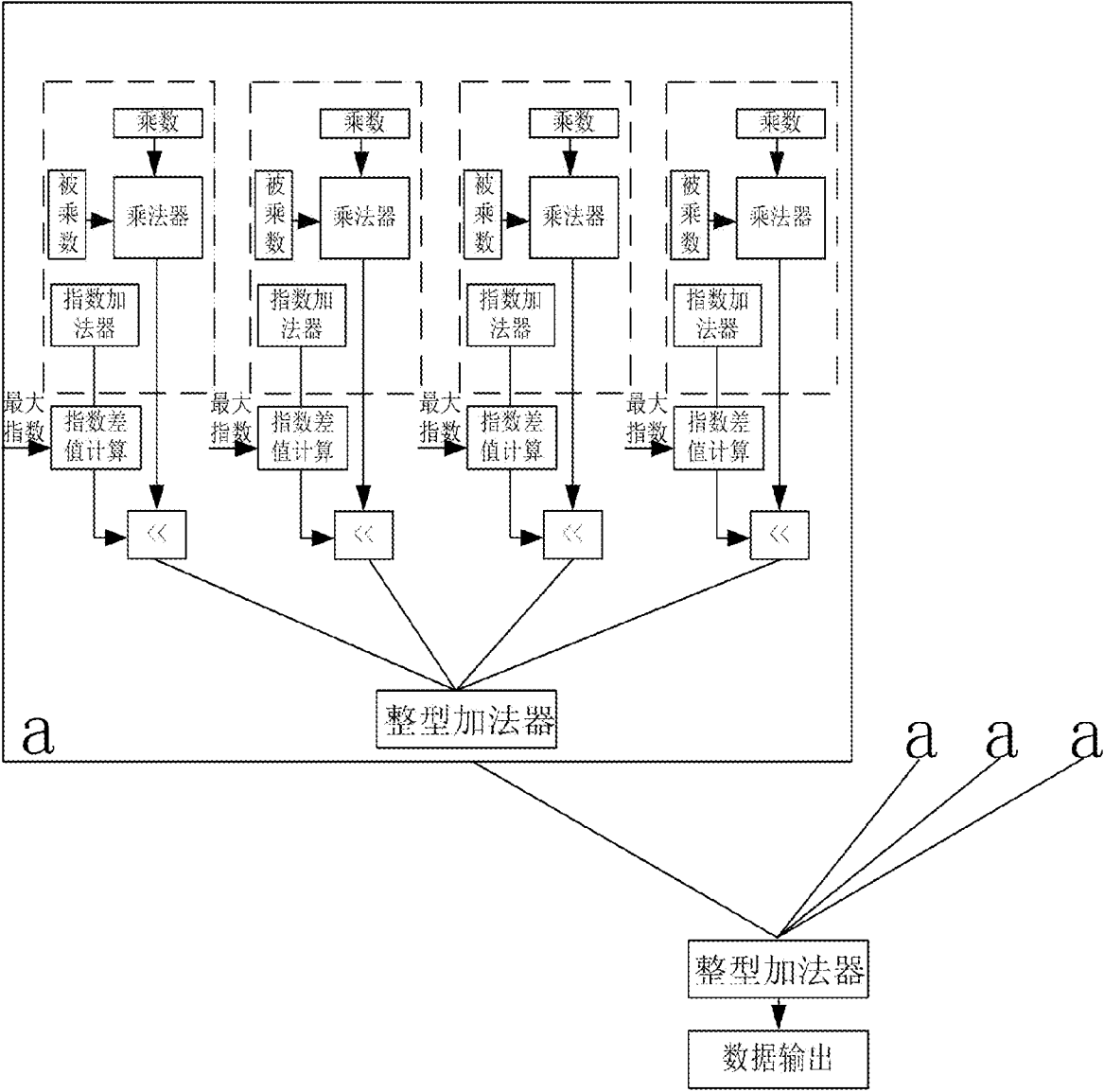


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/131745

A. CLASSIFICATION OF SUBJECT MATTER

G06F 7/575(2006.01)i; G06F 7/523(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI; SIPOABS; DWPI; USTXT; WOTXT; EPTXT; IEEE: 多精度, 可重构, 浮点, 乘加, 运算, 有效数字, 尾数, 段, 乘法器, 加法树, Multiple-Precision, Reconfigurable, floating point, multiply-add, operation, significant digit number, mantissas, segments, multiplier, adder tree

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 112860220 A (SOUTH UNIVERSITY OF SCIENCE AND TECHNOLOGY) 28 May 2021 (2021-05-28) claims 1-10, description paragraphs [0002]-[0090]	1-10
X	Wei Mao et al. "A Reconfigurable Multiple-Precision Floating-Point Dot Product Unit for High-Performance Computing" 2021 Design, Automation & Test in Europe Conference & Exhibition (DATE), 05 February 2021 (2021-02-05), page 1793 line 1 to page 1798 last line	1-10
A	CN 111492343 A (QUALCOMM INC.) 04 August 2020 (2020-08-04) entire document	1-10
A	CN 109739555 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 10 May 2019 (2019-05-10) entire document	1-10
A	CN 107273090 A (INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 20 October 2017 (2017-10-20) entire document	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

11 January 2022

Date of mailing of the international search report

27 January 2022

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088, China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/131745

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112860220	A	28 May 2021	None			
CN	111492343	A	04 August 2020	WO	2019125835	A1	27 June 2019
				EP	3729257	A1	28 October 2020
				US	2019196785	A1	27 June 2019
				US	10346133	B1	09 July 2019
				BR	112020012183	A2	24 November 2020
				CA	3083043	A1	27 June 2019
				CA	3083043	C	29 December 2020
				AU	2018388451	A1	11 June 2020
				AU	2018388451	B2	25 March 2021
				IN	202047021398	A	24 July 2020
				CN	111492343	B	12 March 2021
				BR	112020012183	B1	14 September 2021
CN	109739555	A	10 May 2019	US	2021326118	A1	21 October 2021
				WO	2020140766	A1	09 July 2020
CN	107273090	A	20 October 2017	CN	107273090	B	31 July 2020

国际检索报告

国际申请号

PCT/CN2021/131745

A. 主题的分类

G06F 7/575 (2006.01) i; G06F 7/523 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNABS; CNTXT; CNKI; SIPOABS; DWPI; USTXT; WOTXT; EPTXT; IEEE: 多精度, 可重构, 浮点, 乘加, 运算, 有效数字, 尾数, 段, 乘法器, 加法树, Multiple-Precision, Reconfigurable, floating point, multiply-add, operation, significant digit number, mantissas, segments, multiplier, adder tree

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 112860220 A (南方科技大学) 2021年5月28日 (2021 - 05 - 28) 权利要求1-10项, 说明书第[0002]-[0090]段	1-10
X	Wei Mao 等. "A Reconfigurable Multiple-Precision Floating-Point Dot Product Unit for High-Performance Computing" 2021 Design, Automation & Test in Europe Conference & Exhibition (DATE), 2021年2月5日 (2021 - 02 - 05), 第1793页第1行至第1798页倒数第1行	1-10
A	CN 111492343 A (高通股份有限公司) 2020年8月4日 (2020 - 08 - 04) 全文	1-10
A	CN 109739555 A (腾讯科技深圳有限公司) 2019年5月10日 (2019 - 05 - 10) 全文	1-10
A	CN 107273090 A (中国科学院计算技术研究所) 2017年10月20日 (2017 - 10 - 20) 全文	1-10

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2022年1月11日

国际检索报告邮寄日期

2022年1月27日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

徐生芹

电话号码 (86-512)88995786

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/131745

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	112860220	A	2021年5月28日	无			
CN	111492343	A	2020年8月4日	WO	2019125835	A1	2019年6月27日
				EP	3729257	A1	2020年10月28日
				US	2019196785	A1	2019年6月27日
				US	10346133	B1	2019年7月9日
				BR	112020012183	A2	2020年11月24日
				CA	3083043	A1	2019年6月27日
				CA	3083043	C	2020年12月29日
				AU	2018388451	A1	2020年6月11日
				AU	2018388451	B2	2021年3月25日
				IN	202047021398	A	2020年7月24日
				CN	111492343	B	2021年3月12日
				BR	112020012183	B1	2021年9月14日
CN	109739555	A	2019年5月10日	US	2021326118	A1	2021年10月21日
				WO	2020140766	A1	2020年7月9日
CN	107273090	A	2017年10月20日	CN	107273090	B	2020年7月31日