(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
2024 年 5 月 2 日 (02.05.2024)

WIPO | PCT

(10) 国际公布号
WO 2024/088231 A1

(51) 国际专利分类号:
G06F 5/10 (2006.01)

(21) 国际申请号: PCT/CN2023/126090

(22) 国际申请日: 2023 年 10 月 24 日 (24.10.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202211338060.5 2022年10月28日 (28.10.2022) CN

(71) 申请人: 华 为 技 术 有 限 公 司 (**HUAWEI TECHNOLOGIES CO., LTD.**) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 中国科学技术大学(**UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA**) [CN/CN]; 中 国 安 徽 省 合 肥 市 包 河 区 金 寨 路 96号, Anhui 230026 (CN)。

(72) 发明人: 梁华文(**LIANG, Huawen**); 中国安徽省合肥市包河区金寨路96号, Anhui 230026 (CN)。 肖世海(**XIAO, Shihai**); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 袁伟(**YUAN, Wei**); 中国安徽省合肥市包河区金寨路96号, Anhui 230026 (CN)。 田腾(**TIAN, Teng**); 中国安徽省合肥市包河区金寨路96号, Anhui 230026 (CN)。 金西(**JIN, Xi**); 中国安徽省合肥市包河区金寨路96号, Anhui 230026 (CN)。

(74) 代理人: 北 京 市 金 杜 律 师 事 务 所 (**KING & WOOD MALLESONS**); 中 国 北 京 市 朝 阳 区 东 三 环 中 路 1 号 环 球 金 融 中 心 办 公 楼 东 楼 20层, Beijing 100020 (CN)。

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) **Title:** SIGNAL PROCESSING METHOD AND APPARATUS, AND DEVICE, MEDIUM AND CHIP

(54) 发明名称: 信号处理方法、装置、设备、介质及芯片

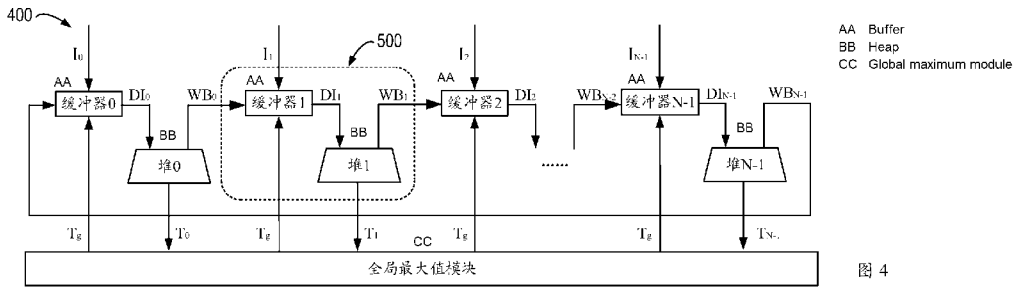


图 4

(57) **Abstract:** Provided in the embodiments of the present disclosure are a signal processing method and apparatus, and a device, a medium and a chip. The method comprises: inputting a plurality of signals into a multi-sorting module, wherein the multi-sorting module comprises N sorting modules which are connected in series in a ring, each sorting module comprising a Top-K module having K elements, and the multi-sorting module comprises a first sorting module and a second sorting module, the first sorting module comprising a first Top-K module, and a signal, which is output by the first sorting module and is discarded by the first Top-K module, or a replaced top end element of the first Top-K module serving as an input signal of the second sorting module; and updating elements in a plurality of Top-K modules in the multi-sorting module on the basis of the plurality of signals, and obtaining M pieces of top-ranked data among data of the plurality of signals, wherein the value of M is the product value of N and K. By means of the embodiments of the present disclosure, system performance can be improved, the signal processing efficiency can be improved, storage resources and processing resources can be saved on, and the data storage amount, the data transmission amount, and the data processing amount and processing time can all be reduced.

WO 2024/088231 A1

[见续页]

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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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: 本公开的实施例提供了一种信号处理方法、装置、设备、介质及芯片。该方法包括：向多排序模块输入多个信号，多排序模块包括环形串联的N个排序模块，每个排序模块包括具有K个元素的Top-K模块，多排序模块包括第一排序模块和第二排序模块，第一排序模块包括第一Top-K模块，第一排序模块输出的被第一Top-K模块舍弃的信号或被替换出的第一Top-K模块的顶端元素作为第二排序模块的输入信号。基于多个信号，更新多排序模块中多个Top-K模块中的元素，以及获得多个信号中数据排序靠前的M个数据，M的值为N与K的乘积值。实施本公开的实施例，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，以及减少数据存储量、数据传输量、数据处理量和处理时间。

信号处理方法、装置、设备、介质及芯片

本申请要求于 2022 年 10 月 28 日提交中国国家知识产权局，申请号为 202211338060.5，发明名称为“信号处理方法、装置、设备、介质及芯片”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本公开涉及计算机领域，更具体地，涉及一种信号处理方法、装置、设备、介质及芯片。

背景技术

Top-K 算法通常用于在多个数据的无序序列中找出前 K 个最大的元素或者最小的元素，可以用堆结构来实现。堆是基于完全二叉树的数据结构所构建的一组数据，常见的堆包括最大堆和最小堆。最大堆的性质包括其节点满足子节点的键值或者索引总是大于它的父节点，其堆顶的根节点是堆中所有元素中的最大值；最小堆的性质包括其节点满足子节点的键值或者索引总是小于它的父节点，其堆顶的根节点是堆中所有元素中的最小值。堆广泛应用于优先队列，以及在多个数据的无序序列中找出前 K 个最大的元素或者最小的元素，即 Top-K 算法。Top-K 算法常应用于相似性检索、推荐系统以及数据库等场景中。

发明内容

本申请的实施例提供了一种信号处理方法以及相关电子设备，用于优化实现在无序的大规模数据中快速找出前面一些最大的元素或者最小的元素。

在本申请的第一方面，提供了一种信号处理方法，该方法包括：向多排序模块输入多个信号，多排序模块包括环形串联的 N 个排序模块，N 个排序模块中的每个排序模块包括具有 K 个元素的 Top-K 模块，多排序模块包括第一排序模块和第二排序模块，第一排序模块包括第一 Top-K 模块，第一排序模块输出的被第一 Top-K 模块舍弃的信号或被替换出的第一 Top-K 模块的顶端元素作为第二排序模块的输入信号。基于多个信号，更新多排序模块中多个 Top-K 模块中的元素。以及获得多个信号中数据排序靠前的 M 个数据，其中 M 的值为 N 与 K 的乘积值，N、K、M 为正整数。

实施本方面的方法，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

结合第一方面，在一些实施例中，第一排序模块中包括第一缓冲器和第一 Top-K 模块，第二排序模块中包括第二缓冲器和第二 Top-K 模块，第一排序模块的第一缓冲器的输出作为第一排序模块的第一 Top-K 模块的输入，第一排序模块的第一 Top-K 模块的输出作为第二排序模块的第二缓冲器的输入，第二排序模块的第二缓冲器的输出作为第二排序模块的第二 Top-K 模块的输入。各个排序模块之间的缓冲器与 Top-K 模块交替连接以及多个排序模块环形串联的结构可以提高处理效率。

结合第一方面，在一些实施例中，多个信号包括第一外部信号和第二外部信号，第一外部信号为第一排序模块的输入，第二外部信号为第二排序模块的输入，并且该方法还包括：第一排序模块接收第一外部信号，第一排序模块基于第一外部信号获得第一写回信号，第一排序模块向第二排序模块发送第一写回信号。第二排序模块接收第二外部信号，第二排序模块接收第一写回信号，第二排序模块基于第二外部信号和/或第一写回信号获得第二写回信号，第二排序模块向第三排序模块发送第二写回信号。第一排序模块与第二排序模块串联可以提高处理效率，第一排序模块不需要的数据可以由第二排序模块继续处理，避免重复输入和处理。

结合第一方面，在一些实施例中，第一缓冲器中包括第一先入先出队列（first input first output，FIFO）、第一控制逻辑模块和第一多选器，第一多选器用于接收第一外部信号，第一先入先出队列用于多次接收并缓存多个写回信号，第一先入先出队列还用于在被多个写回信号填满后输出第一缓存信号，以及第二缓冲器中包括第二先入先出队列、第二控制逻辑模块和第二多选器，第二多选器用于接收第二外部信号，第二先入先出队列用于多次接收并缓存第一排序模块的第一 Top-K 模块输出的多个写回信号，多个写回信号包括第一写回信号，第二先入先出队列还用于在被多个写回信号填满后输出第二缓存信号。控制逻辑和多选器用于基于先入先出队列输出的缓存信号、输入的外部信号和全局堆顶最值来选择进入堆的输入信号。先入先出队列用于缓存上个排序模块输出的写回数据。设置该先入先出队列缓存写回数据的目的是由于全局

最值模块输出的堆顶值相对输入信号会存在几个周期的延迟，于是写回信号输入到缓冲器中的先入先出队列后，直到先入先出队列填充满才被读取，这时的缓存信号为多个周期前的输入，相对全局堆顶最值会延后，有利于控制逻辑能够发挥出较好筛选性能，避免写回信号在多个堆中传递而不会被筛选掉。

结合第一方面，在一些实施例中，第二排序模块基于第二外部信号和/或第一写回信号获得第二写回信号包括：第二控制逻辑模块基于第二外部信号和/或第二缓存信号确定第二输入信号，第二多选器向第二 Top-K 模块输出第二输入信号，第二 Top-K 模块接收来自于第二多选器的第二输入信号，第二 Top-K 模块基于第二输入信号获得第二写回信号。以及第二排序模块向第三排序模块发送第二写回信号，包括：第二 Top-K 模块向第三排序模块的第三缓冲器发送第二写回信号。Top-K 模块与缓冲器交替相连，可以提高信号处理效率。

结合第一方面，在一些实施例中，第二控制逻辑模块基于第二外部信号和/或第二缓存信号确定第二输入信号，包括：判断第二先入先出队列是否输出第二缓存信号。如果第二先入先出队列输出了第二缓存信号，那么判断第二缓存信号是否有效。如果第二缓存信号有效，则将第二缓存信号的数据与全局顶端最值进行比较。如果第二缓存信号的数据未超过全局顶端最值，则第二控制逻辑模块确定第二缓存信号作为第二输入信号，如果第二缓存信号的数据超过或等于全局顶端最值，则第二控制逻辑模块确定第二外部信号作为第二输入信号，并且第二缓存信号被舍弃。以及如果第二缓存信号无效，则第二控制逻辑模块确定第二外部信号作为第二输入信号，并且第二缓存信号被舍弃。通过多个堆之间传递数据以及全局最值（最大值或最小值）的约束，可以减少排序时所需比较的次数，节省处理资源的消耗。此外，基于该多堆结构可以较好的利用了高速缓存器等片上存储资源，提高了不同片上资源的利用效率。

结合第一方面，在一些实施例中，如果第二先入先出队列没有输出第二缓存信号，那么判断第二外部信号是否有效。如果第二外部信号有效，则第二控制逻辑模块确定第二外部信号作为第二输入信号。以及如果第二外部信号无效，则第二控制逻辑模块读取非空的第二先入先出队列中缓存的第三缓存信号作为第二输入信号。即先入先出队列中溢出的缓存信号优先于外部信号被处理。当先入先出队列中没有溢出的缓存信号时，将有效的外部信号作为堆的输入信号，如果外部信号无效，为了保证 FIFO 中的所有数据都进入排序，避免堆模块闲置，可以读取非空的先入先出队列中未溢出的缓存的写回信号作为堆的输入信号。

结合第一方面，在一些实施例中，第一写回信号包括第一写回信号的数据和第一写回使能信号，第一写回使能信号指示第一写回信号是否有效。第二外部信号包括第二外部信号的数据和第二外部使能信号，第二外部使能信号指示第二外部信号是否有效。第二缓存信号包括第二缓存信号的数据和第二缓存使能信号，第二缓存使能信号指示第二缓存信号是否有效。使用写回使能信号和外部使能信号可以更方便地控制信号的输入。

结合第一方面，在一些实施例中，Top-K 模块为最大堆，全局顶端最值为多个最大堆的堆顶元素的数据之中的最大值，第二缓存信号的数据未超过全局顶端最值，包括：第二缓存信号的数据小于最大值。第二缓存信号的数据超过或等于全局顶端最值，包括：第二缓存信号的数据大于或等于最大值。最大堆的结构能够帮助求出多个输入数据中预定数目的最小值。

结合第一方面，在一些实施例中，Top-K 模块为最小堆，全局顶端最值为多个最小堆的堆顶元素的数据之中的最小值，第二缓存信号的数据未超过全局顶端最值，包括：第二缓存信号的数据大于最小值。第二缓存信号的数据超过或等于全局顶端最值，包括：第二缓存信号的数据小于或等于最小值。最小堆的结构能够帮助求出多个输入数据中预定数目的最大值。

结合第一方面，在一些实施例中，第二 Top-K 模块基于第二输入信号获得第二写回信号，包括：判断第二输入信号是否有效，如果第二输入信号有效，则将第二输入信号的数据与第二 Top-K 模块的当前的顶端元素的数据进行比较。如果第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，则将第二输入信号的数据替换掉第二 Top-K 模块的当前顶端元素的数据以成为第二 Top-K 模块中新的元素，以及将被替换出的第二 Top-K 模块的当前顶端元素的数据作为第二写回信号的数据，并且使能第二写回信号有效。以及如果第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，则将第二输入信号的数据作为第二写回信号的数据，并且使能第二写回信号有效。通过利用 Top-K 模块内数据的有序性以及 Top-K 模块顶端数据的约束，将本 Top-K 模块不需要的数据传递给下一个排序模块处理或者及时丢弃，可以减少排序时所需比较的次数，节省处理资源和存储资源。

结合第一方面，在一些实施例中，还包括：如果第二输入信号无效，则将第二输入信号的数据作为第二写回信号的数据，并且使能第二写回信号无效。被确定无用的数据可以发送给下一个排序模块的 FIFO，

由下一个排序模块的控制逻辑来决策丢弃该数据。

结合第一方面，在一些实施例中，该方法还包括：第二 Top-K 模块更新第二 Top-K 模块的新的顶端元素，以及第二 Top-K 模块将第二 Top-K 模块的顶端元素发送给全局最值模块。全局最值模块用于获取多个 Top-K 模块的顶端元素，全局最值模块还用于选出多个 Top-K 模块的顶端元素的数据之中的最值，全局最值模块还用于将最值发送给多个缓冲器中的控制逻辑模块。由于被替换进入 Top-K 模块的元素不一定是 Top-K 模块内最值，所以在下一次输入信号输入堆之前，Top-K 模块需要更新出新的顶端元素，以保证下次处理的准确性。

结合第一方面，在一些实施例中，第二 Top-K 模块为最大堆，第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据小于最大堆的当前堆顶元素的数据。第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据大于或等于最大堆的当前堆顶元素的数据。最大堆的结构能够帮助求出多个输入数据中预定数目的最小值。

结合第一方面，在一些实施例中，第二 Top-K 模块为最小堆，第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据大于最小堆的当前堆顶元素的数据。第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据小于或等于最小堆的当前堆顶元素的数据。最小堆的结构能够帮助求出多个输入数据中预定数目的最大值。

结合第一方面，在一些实施例中，在第一时间周期内，第二缓冲器基于第二外部信号和/或第二缓存信号确定第二输入信号，以及在第二时间周期内，第二 Top-K 模块基于第二输入信号获得第二写回信号。在一个迭代周期内，一个排序模块可以完整的完成一次处理工作，一个迭代周期可以分成两个时间周期，例如第一时间周期和第二时间周期。在一些实施例中，第一时间周期与第二时间周期的时间长度相等。缓冲器判断选择进入堆的数据的处理周期与堆处理输入数据并输出写回数据的处理周期交替地执行，每个排序模块可以并行处理数据，这能够提高处理效率，缩减等待时间和处理时间，充分利用处理资源。

结合第一方面，在一些实施例中，多排序模块为多堆结构，Top-K 模块为堆，多堆结构包括多个堆，多个堆中的每个堆中包括 K 个元素，多堆结构输出 M 个数据，M 个数据为多堆结构的多个堆中的所有元素的集合。多堆结构的工作并行度可以增加到 N，实现了更高的数据吞吐量，提高信号处理效率，并且多堆结构中的每个堆的大小减小为 M/N，提高了单个堆的性能，可以节省存储资源、计算资源和处理资源，以及减少数据存储量、数据传输量、数据处理量和处理时间。

结合第一方面，在一些实施例中，如果堆为最大堆，那么 M 个数据为多个信号中排序数据最小的前 M 个数据，最大堆的堆顶元素为最大堆中数据最大的元素；如果堆为最小堆，那么 M 个数据为多个信号中排序数据最大的前 M 个数据，最小堆的堆顶元素为最小堆中数据最小的元素。最大堆有助于筛选最小值，最小堆有助于筛选最大值。

在本申请的第二方面，提供了一种信号处理装置，包括：输入模块，被配置为向多排序模块输入多个信号，多排序模块包括环形串联的 N 个排序模块，N 个排序模块中的每个排序模块包括具有 K 个元素的 Top-K 模块，多排序模块包括第一排序模块和第二排序模块，第一排序模块包括第一 Top-K 模块，第一排序模块输出的被第一 Top-K 模块舍弃的信号或被替换出的第一 Top-K 模块的顶端元素作为第二排序模块的输入信号。处理模块，被配置为基于多个信号，更新多排序模块中多个 Top-K 模块中的元素。以及输出模块，被配置为获得多个信号中数据排序靠前的 M 个数据，其中 M 的值为 N 与 K 的乘积值，N、K、M 为正整数。

实施本方面的信号处理装置，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

结合第二方面，在一些实施例中，第一排序模块中包括第一缓冲器和第一 Top-K 模块，第二排序模块中包括第二缓冲器和第二 Top-K 模块，第一排序模块的第一缓冲器的输出作为第一排序模块的第一 Top-K 模块的输入，第一排序模块的第一 Top-K 模块的输出作为第二排序模块的第二缓冲器的输入，第二排序模块的第二缓冲器的输出作为第二排序模块的第二 Top-K 模块的输入。各个排序模块之间的缓冲器与 Top-K 模块交替连接以及多个排序模块环形串联的结构可以提高处理效率。

结合第二方面，在一些实施例中，多个信号包括第一外部信号和第二外部信号，第一外部信号为第一排序模块的输入，第二外部信号为第二排序模块的输入，并且第一排序模块用于接收第一外部信号，第一排序模块还用于基于第一外部信号获得第一写回信号，第一排序模块还用于向第二排序模块发送第一写回信号。第二排序模块用于接收第二外部信号，第二排序模块还用于接收第一写回信号，第二排序模块还用

于基于第二外部信号和/或第一写回信号获得第二写回信号，第二排序模块还用于向第三排序模块发送第二写回信号。第一排序模块与第二排序模块串联可以提高处理效率，第一排序模块不需要的数据可以由第二排序模块继续处理，避免重复输入和处理。

结合第二方面，在一些实施例中，第一缓冲器中包括第一先入先出队列、第一控制逻辑模块和第一多选器，第一多选器用于接收第一外部信号，第一先入先出队列用于多次接收并缓存多个写回信号，第一先入先出队列还用于在被多个写回信号填满后输出第一缓存信号，以及第二缓冲器中包括第二先入先出队列、第二控制逻辑模块和第二多选器，第二多选器用于接收第二外部信号，第二先入先出队列用于多次接收并缓存第一排序模块的第一 Top-K 模块输出的多个写回信号，多个写回信号包括第一写回信号，第二先入先出队列还用于在被多个写回信号填满后输出第二缓存信号。控制逻辑和多选器用于基于先入先出队列输出的缓存信号、输入的外部信号和全局堆顶最值来选择进入堆的输入信号。先入先出队列用于缓存上个排序模块输出的写回数据。设置该先入先出队列缓存写回数据的目的是由于全局最值模块输出的堆顶值相对输入信号会存在几个周期的延迟，于是写回信号输入到缓冲器中的先入先出队列后，直到先入先出队列填充满才被读取，这时的缓存信号为多个周期前的输入，相对全局堆顶最值会延后，有利于控制逻辑能够发挥出较好筛选性能，避免写回信号在多个堆中传递而不会被筛选掉。

结合第二方面，在一些实施例中，第二控制逻辑模块还用于基于第二外部信号和/或第二缓存信号确定第二输入信号，第二多选器还用于向第二 Top-K 模块输出第二输入信号，第二 Top-K 模块用于接收来自于第二多选器的第二输入信号，第二 Top-K 模块还用于基于第二输入信号获得第二写回信号。以及第二 Top-K 模块还用于向第三排序模块的第三缓冲器发送第二写回信号。Top-K 模块与缓冲器交替相连，可以提高信号处理效率。

结合第二方面，在一些实施例中，第二控制逻辑模块用于执行以下判断逻辑：判断第二先入先出队列是否输出第二缓存信号。如果第二先入先出队列输出了第二缓存信号，那么判断第二缓存信号是否有效。如果第二缓存信号有效，则将第二缓存信号的数据与全局顶端最值进行比较。如果第二缓存信号的数据未超过全局顶端最值，则第二控制逻辑模块确定第二缓存信号作为第二输入信号，如果第二缓存信号的数据超过或等于全局顶端最值，则第二控制逻辑模块确定第二外部信号作为第二输入信号，并且第二缓存信号被舍弃。以及如果第二缓存信号无效，则第二控制逻辑模块确定第二外部信号作为第二输入信号，并且第二缓存信号被舍弃。通过多个堆之间传递数据以及全局最值（最大值或最小值）的约束，可以减少排序时所需比较的次数，节省处理资源的消耗。此外，基于该多堆结构可以较好的利用了高速缓存器等片上存储资源，提高了不同片上资源的利用效率。

结合第二方面，在一些实施例中，第二控制逻辑模块还用于执行以下判断逻辑：如果第二先入先出队列没有输出第二缓存信号，那么判断第二外部信号是否有效。如果第二外部信号有效，则第二控制逻辑模块确定第二外部信号作为第二输入信号。以及如果第二外部信号无效，则第二控制逻辑模块读取非空的第二先入先出队列中缓存的第三缓存信号作为第二输入信号。即先入先出队列中溢出的缓存信号优先于外部信号被处理。当先入先出队列中没有溢出的缓存信号时，将有效的外部信号作为堆的输入信号，如果外部信号无效，为了保证 FIFO 中的所有数据都进入排序，避免堆模块闲置，可以读取非空的先入先出队列中未溢出的缓存的写回信号作为堆的输入信号。

结合第二方面，在一些实施例中，第一写回信号包括第一写回信号的数据和第一写回使能信号，第一写回使能信号指示第一写回信号是否有效。第二外部信号包括第二外部信号的数据和第二外部使能信号，第二外部使能信号指示第二外部信号是否有效。第二缓存信号包括第二缓存信号的数据和第二缓存使能信号，第二缓存使能信号指示第二缓存信号是否有效。使用写回使能信号和外部使能信号可以更方便地控制信号的输入。

结合第二方面，在一些实施例中，Top-K 模块为最大堆，全局顶端最值为多个最大堆的堆顶元素的数据之中的最大值，第二缓存信号的数据未超过全局顶端最值，包括：第二缓存信号的数据小于最大值。第二缓存信号的数据超过或等于全局顶端最值，包括：第二缓存信号的数据大于或等于最大值。最大堆的结构能够帮助求出多个输入数据中预定数目的最小值。

结合第二方面，在一些实施例中，Top-K 模块为最小堆，全局顶端最值为多个最小堆的堆顶元素的数据之中的最小值，第二缓存信号的数据未超过全局顶端最值，包括：第二缓存信号的数据大于最小值。第二缓存信号的数据超过或等于全局顶端最值，包括：第二缓存信号的数据小于或等于最小值。最小堆的结构能够帮助求出多个输入数据中预定数目的最大值。

结合第二方面，在一些实施例中，第二 Top-K 模块用于执行以下判断逻辑：第二 Top-K 模块基于第二输入信号获得第二写回信号，包括：判断第二输入信号是否有效，如果第二输入信号有效，则将第二输入信号的数据与第二 Top-K 模块的当前的顶端元素的数据进行比较。如果第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，则将第二输入信号的数据替换掉第二 Top-K 模块的当前顶端元素的数据以成为第二 Top-K 模块中新的元素，以及将被替换出的第二 Top-K 模块的当前顶端元素的数据作为第二写回信号的数据，并且使能第二写回信号有效。以及如果第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，则将第二输入信号的数据作为第二写回信号的数据，并且使能第二写回信号有效。通过利用 Top-K 模块内数据的有序性以及 Top-K 模块顶端数据的约束，将本 Top-K 模块不需要的数据传递给下一个排序模块处理或者及时丢弃，可以减少排序时所需比较的次数，节省处理资源和存储资源。

结合第二方面，在一些实施例中，第二 Top-K 模块还用于执行以下判断逻辑：如果第二输入信号无效，则将第二输入信号的数据作为第二写回信号的数据，并且使能第二写回信号无效。被确定无用的数据可以发送给下一个排序模块的 FIFO，由下一个排序模块的控制逻辑来决策丢弃该数据。

结合第二方面，在一些实施例中，第二 Top-K 模块还用于更新第二 Top-K 模块的新的顶端元素，以及第二 Top-K 模块还用于将第二 Top-K 模块的顶端元素发送给全局最值模块。全局最值模块用于获取多个 Top-K 模块的顶端元素，全局最值模块还用于选出多个 Top-K 模块的顶端元素的数据之中的最值，全局最值模块还用于将最值发送给多个缓冲器中的控制逻辑模块。由于被替换进入 Top-K 模块的元素不一定是 Top-K 模块内最值，所以在下一次输入信号输入堆之前，Top-K 模块需要更新出新的顶端元素，以保证下次处理的准确性。

结合第二方面，在一些实施例中，第二 Top-K 模块为最大堆，第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据小于最大堆的当前堆顶元素的数据。第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据大于或等于最大堆的当前堆顶元素的数据。最大堆的结构能够帮助求出多个输入数据中预定数目的最小值。

结合第二方面，在一些实施例中，第二 Top-K 模块为最小堆，第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据大于最小堆的当前堆顶元素的数据。第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据小于或等于最小堆的当前堆顶元素的数据。最小堆的结构能够帮助求出多个输入数据中预定数目的最大值。

结合第二方面，在一些实施例中，在第一时间周期内，第二缓冲器基于第二外部信号和/或第二缓存信号确定第二输入信号，以及在第二时间周期内，第二 Top-K 模块基于第二输入信号获得第二写回信号。在一个迭代周期内，一个排序模块可以完整的完成一次处理工作，一个迭代周期可以分成两个时间周期，例如第一时间周期和第二时间周期。在一些实施例中，第一时间周期与第二时间周期的时间长度相等。缓冲器判断选择进入堆的数据的处理周期与堆处理输入数据并输出写回数据的处理周期交替地执行，每个排序模块可以并行处理数据，这能够提高处理效率，缩减等待时间和处理时间，充分利用处理资源。

结合第二方面，在一些实施例中，多排序模块为多堆结构，Top-K 模块为堆，多堆结构包括多个堆，多个堆中的每个堆中包括 K 个元素，多堆结构输出 M 个数据，M 个数据为多堆结构的多个堆中的所有元素的集合。多堆结构的工作并行度可以增加到 N，实现了更高的数据吞吐量，提高信号处理效率，并且多堆结构中的每个堆的大小减小为 M/N，提高了单个堆的性能，可以节省存储资源、计算资源和处理资源，以及减少数据存储量、数据传输量、数据处理量和处理时间。

结合第二方面，在一些实施例中，如果堆为最大堆，那么 M 个数据为多个信号中排序数据最小的前 M 个数据，最大堆的堆顶元素为最大堆中数据最大的元素；如果堆为最小堆，那么 M 个数据为多个信号中排序数据最大的前 M 个数据，最小堆的堆顶元素为最小堆中数据最小的元素。最大堆有助于筛选最小值，最小堆有助于筛选最大值。

在本申请的第三方面，提供了一种电子设备，电子设备包括存储器以及耦合于存储器的处理器，存储器中存储有可执行指令，处理器用于调用可执行指令，使得电子设备执行以下根据上述第一方面或其任一实施例中的方法的操作。

在本申请的第四方面，提供了一种计算机可读存储介质，该计算机可读存储介质存储有计算机可执行指令，计算机可执行指令被处理器执行时实现根据上述第一方面或其任一实施例中的方法的操作。

在本申请的第五方面，提供了一种计算机程序产品，该计算机程序产品上包含计算机可执行指令，计算机可执行指令在被执行时实现根据上述第一方面或其任一实施例中的方法的操作。

在本申请的第六方面，提供了一种芯片或芯片系统，该芯片或芯片系统包括处理电路，其被配置为实现根据上述第一方面或其任一实施例中的方法的操作。

实施本申请以上一个或多个方面，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

附图说明

结合附图并参考以下详细说明，本公开各实施例的上述和其他特征、优点及方面将变得更加明显。在附图中，相同或相似的附图标注表示相同或相似的元素，其中：

图 1 示出了根据本公开的一些实施例的一种堆结构的示意图；

图 2 示出了根据本公开的一些实施例的信号处理方法流程的示意图；

图 3 示出了根据本公开的一些实施例的数据处理系统的功能模块的示意图；

图 4 示出了根据本公开的一些实施例的一种多堆结构的示意图；

图 5 示出了根据本公开的一些实施例的一种多堆结构的排序模块的示意图；

图 6 示出了根据本公开的一些实施例的一种多堆结构的排序模块的控制逻辑流程的示意图；

图 7 示出了根据本公开的一些实施例的一种堆模块的示意图；

图 8 示出了根据本公开的一些实施例的一种堆模块的堆控制逻辑流程的示意图；

图 9 示出了根据本公开的一些实施例的一种双堆结构的示意图；

图 10 示出了根据本公开的一些实施例的一种双堆结构的排序模块的示意图；

图 11 示出了根据本公开的一些实施例的一种双堆结构的排序模块的控制逻辑流程的示意图；

图 12 示出了根据本公开的一些实施例的一种堆模块的堆控制逻辑流程的示意图；

图 13 示出了根据本公开的一些实施例的电子设备结构的示意图。

具体实施方式

下面将参照附图更详细地描述本公开的实施例。虽然附图中显示了本公开的某些实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为限于这里阐述的实施例，相反提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

本申请以下实施例中所使用的术语只是为了描述特定实施例的目的，而并非旨在作为对本申请的限制。如在本申请的说明书和所附权利要求书中所使用的那样，单数表达形式“一个”、“一种”、“该”、“上述”、“该”和“这一”旨在也包括复数表达形式，除非其上下文中明确地有相反指示。还应当理解，本申请中使用的术语“和/或”是指并包含一个或多个所列出项目的任何或所有可能组合。

在本申请的实施例的描述中，术语“包括”及其类似用语应当理解为开放性包含，即“包括但不限于”。术语“基于”应当理解为“至少部分地基于”。术语“一个实施例”或“该实施例”应当理解为“至少一个实施例”。应理解，“第一”，“第二”，“第三”等表述只是为了表示多个对象可能是不同的，但是同时不排除两个对象之间是相同的。“第一”，“第二”，“第三”等表述不应当解释为对实施例的任何限制。术语“第一”、“第二”仅用于描述目的，而不能理解为暗示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征，在本申请实施例的描述中，除非另有说明，“多个”的含义是两个或两个以上。下文还可能包括其他明确的和隐含的定义。

在本申请中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本申请所描述的实施例可以与其它实施例相结合。

在大规模无序数据中求取找出前 K 个最大的元素或者最小的元素是个广泛的应用。示例性应用场景可以是，例如，在持续更新的海量视频中找出视频点击率最高的 100 个视频进行推荐；再例如，分布式系统中将任务分配给所监控到的中央处理器（CPU）利用率最低的 10 个设备；又例如，根据海量且不断更新的游戏玩家的数据，生成游戏分数最高的前 100 名游戏玩家列表。在这种情形下，可以使用最大堆或最小堆的堆结构来在大规模数据中求取动态更新的前 K 个最小或最大的数据元组。

由于堆满足的父节点与子节点的有序性，堆的更新操作，如进栈（pop）操作、出栈（push）操作，通常从根节点开始，限制了堆的并行性。对于一个大小为 N（N 为正整数）的堆，完成一次更新操作所需的比较次数为$O(logN)$。如图 1 所示的堆结构 100，在一种应用 Top-K 排序的堆并行方案中，将所有待处理数据分为两个部分，即数据库 101 和数据库 102，分别输入到大小为 K（K 为正整数）的堆 103 和堆 104 中完成 Top-K 排序，即数据库 101 输入堆 103 进行 Top-K 排序，数据库 102 输入堆 104 进行 Top-K 排序，再将堆 103 和堆 104 中的结果最终削减为一个 Top-K 的结果，即堆 105。在这里，堆 103 和堆 104 同时工作，并行度为 2。对于更高并行度的情况，可以通过使用更多的大小为 K 的堆完成第一次 Top-K 排序，再将多个第一次 Top-K 排序的结果逐步削减为最终 Top-K 的结果。这种方案在多个堆并行完成第一次堆化操作后，还有后续的提炼 K 个元素的步骤，会造成存储空间、计算资源、处理资源的浪费，并且这种方案中堆的大小为 K，当 K 较大时会给片上缓存带来较大的挑战，对于单个堆的性能也没有带来提升，因此往往硬件资源开销较大，并行度差。

本公开的一个或多个实施例提出了一种信号处理方法以及电子设备，用于优化实现在无序的大规模数据中快速找出前 M 个最大的元素或者最小的元素。该信号处理方法可以包括：向多排序模块输入多个信号，多排序模块包括环形串联的多个排序模块，多个排序模块中的每个排序模块包括 Top-K 模块，多排序模块包括第一排序模块和第二排序模块，第一排序模块的输出作为第二排序模块的输入。基于多个信号，更新多个 Top-K 模块中的元素，以及输出多排序模块中多个 Top-K 模块中的元素。

实施本公开的一个或多个实施例中的信号处理方法以及电子设备，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

参考图 2，图 2 示出了本申请实施例的信号处理方法流程 200 的示意图。示例性的，实施该信号处理方法的电子设备可以是智能手机、平板电脑、个人计算机、桌面型计算机、膝上型计算机、手持计算机、游戏机、智慧屏（智能电视）、无人机、增强现实（augmented reality，AR）设备、虚拟现实（virtual reality，VR）设备、人工智能（artificial intelligence，AI）设备、可穿戴式设备、车机（车载设备）、物联网（internet of things，IOT）设备或者其他类型的电子设备等等。本申请实施例对该电子设备的具体类型不做限定。

在框 201，向多排序模块输入多个信号。其中，多排序模块包括环形串联的多个排序模块，例如 N 个排序模块，多个排序模块中的每个排序模块包括具有 K 个元素的 Top-K 模块，多排序模块包括第一排序模块和第二排序模块，第一排序模块的输出作为第二排序模块的输入。例如，第一排序模块包括第一 Top-K 模块，第一排序模块输出的被第一 Top-K 模块舍弃的信号或被替换出的第一 Top-K 模块的顶端元素作为第二排序模块的输入信号。在本申请的实施例中，该多个信号是多排序模块之外的外部信号，该外部信号可以例如为网络数据信号，例如上述应用场景中提到的视频点击率数据信号，或者 CPU 利用率信号，或者游戏得分数据信号等。该外部信号可以包括数据元组以及使能信号。在一些实施例中，该多排序模块可以实现为图 4 所示的多堆结构 400，或者图 9 所示的双堆结构 900。Top-K 模块可以实现为堆，例如最大堆或最小堆。

在框 202，基于多个信号，更新多排序模块中多个 Top-K 模块中的元素。在一些实施例中，输入多排序模块中的外部信号在经过多排序模块的处理后，被筛选留下以及替换原有的 Top-K 模块中元素作为所需要的 Top-K 元素之一。

在框 203，获得多个信号中数据排序靠前的 M 个数据。其中 M 的值为 N 与 K 的乘积值，M = N × K，N、K、M 为正整数。在本申请实施例中，该多排序模块是一个由多个堆环形串联起来的结构，一个 Top-K 模块的输出信号会输入到下一个 Top-K 模块，由此将外部信号依次传输下去，直到进入某个 Top-K 模块或者被舍弃。该多排序模块中的经过排序完成的多个 Top-K 模块中的多个元素的集合即是所需要的 Top-K 元素。

实施上述实施例中的信号处理方法，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

在一些实施例中，在一些实施例中，多排序模块为多堆结构，Top-K 模块为堆，多堆结构包括多个堆，多个堆中的每个堆中包括 K 个元素，多堆结构输出 M 个数据，M 个数据为多堆结构的多个堆中的所有元素的集合。多堆结构的工作并行度可以增加到 N，实现了更高的数据吞吐量，提高信号处理效率，并且多堆结构中的每个堆的大小减小为 M/N，提高了单个堆的性能，可以节省存储资源、计算资源和处理资源，以及减少数据存储量、数据传输量、数据处理量和处理时间。

在一些实施例中，在一些实施例中，如果堆为最大堆，那么 M 个数据为多个信号中排序数据最小的前

M 个数据，最大堆的堆顶元素为最大堆中数据最大的元素；如果堆为最小堆，那么 M 个数据为多个信号中排序数据最大的前 M 个数据，最小堆的堆顶元素为最小堆中数据最小的元素。最大堆有助于筛选最小值，最小堆有助于筛选最大值。

在一些实施例中，第一排序模块中包括第一缓冲器和第一 Top-K 模块，第二排序模块中包括第二缓冲器和第二 Top-K 模块，第一排序模块的第一缓冲器的输出作为第一排序模块的第一 Top-K 模块的输入，第一排序模块的第一 Top-K 模块的输出作为第二排序模块的第二缓冲器的输入，第二排序模块的第二缓冲器的输出作为第二排序模块的第二 Top-K 模块的输入。各个排序模块之间的缓冲器与 Top-K 模块交替连接以及多个排序模块环形串联的结构可以提高处理效率。

在一些实施例中，多个信号包括第一外部信号和第二外部信号，第一外部信号为第一排序模块的输入，第二外部信号为第二排序模块的输入，并且该方法还包括：第一排序模块接收第一外部信号，第一排序模块基于第一外部信号获得第一写回信号，第一排序模块向第二排序模块发送第一写回信号。第二排序模块接收第二外部信号，第二排序模块接收第一写回信号，第二排序模块基于第二外部信号和/或第一写回信号获得第二写回信号，第二排序模块向第三排序模块发送第二写回信号。第一排序模块与第二排序模块串联可以提高处理效率，第一排序模块不需要的数据可以由第二排序模块继续处理，避免重复输入和处理。

例如，该多排序模块的具体结构可以结合参考图 4 所示的示例多堆结构 400。结合图 4 和图 5 所示的实施例，多堆结构 400 包括环形串联的多个排序模块，例如排序模块 500（可以是第一排序模块或第二排序模块），每个排序模块具有对应的数据源输入I_k（即多个信号中的一个信号，或者称为外部信号，如第二外部信号），堆（如第一 Top-K 模块/第一堆或第二 Top-K 模块/第二堆），堆顶的数据值的输出T_k（即顶端元素/堆顶元素的信号），本排序模块输入WB_{k-1}（如第一写回信号）、下一个单排序模块的输入WB_k（如第二写回信号）、以及全局堆顶最大值T_g（即全局最值）。N 个单堆结构通过 N 个缓冲器交替相连，使得 N 个排序模块的输出和输入首尾相接连成环状。图 4 所示的可扩展高并行的多堆结构中可以具有 N 个单独的信号输入，可同时完成 N 路信号输入以找到这 N 路信号的大量数据输入中共同的 TOP-M 个数据值。或者，例如，该多堆结构也可以是图 9 实施例所示的双堆结构 900，具体可以参考后续图 9 描述的实施例，这里暂不赘述。

在一些实施例中，第一缓冲器中包括第一先入先出队列、第一控制逻辑模块和第一多选器，第一多选器用于接收第一外部信号，第一先入先出队列用于多次接收并缓存多个写回信号，第一先入先出队列还用于在被多个写回信号填满后输出第一缓存信号。第二排序模块（如图 5 中所示的排序模块 500 或者图 10 所示的缓冲器 1000）中包括第二缓冲器（参考图 5 所示的缓冲器 510 或者图 10 所示的缓冲器 1010），第二缓冲器（如图 5 的缓冲器 510 或者图 10 的缓冲器 1010）中包括第二先入先出队列（如图 5 的$FIFO$ 550 或者图 10 的$FIFO$ 1050）、第二控制逻辑模块（如图 5 的控制逻辑 520 或者图 10 的控制逻辑 1020）和第二多选器（如图 5 的多选器 530 或者图 10 的多选器 1030），第二多选器（如MUX_k）用于接收第二外部信号（如I_k），第二先入先出队列（如$FIFO_k$）用于多次接收并缓存第一排序模块的第一 Top-K 模块输出的多个写回信号（如WB_{k-1}），多个写回信号包括第一写回信号，第二先入先出队列还用于在被多个写回信号填满后输出第二缓存信号（如WB'_{k-1}）。控制逻辑（如C_k）和多选器（如MUX_k）用于基于先入先出队列（如$FIFO_k$）输出的缓存信号（或者称为缓存写回信号，如WB'_{k-1}）、输入的外部信号（如I_k）和全局堆顶最值（如T_g）来选择进入堆（如$heap_k$）的输入信号。先入先出队列用于缓存上个排序模块输出的写回数据。设置该先入先出队列缓存写回数据的目的是由于全局最值模块输出的堆顶值相对输入信号会存在几个周期的延迟，于是写回信号输入到缓冲器中的先入先出队列后，直到先入先出队列填充满才被读取，这时的缓存信号为多个周期前的输入，相对全局堆顶最值会延后，有利于控制逻辑能够发挥出较好筛选性能，避免写回信号在多个堆中传递而不会被筛选掉。

在一些实施例中，在一些实施例中，第二控制逻辑模块（如C_k）基于第二外部信号（如I_k）和/或第二缓存信号（如WB'_{k-1}）确定第二输入信号（如DI_k）。第二多选器（如MUX_k）向第二 Top-K 模块（如$heap_k$）输出第二输入信号（如DI_k）。第二 Top-K 模块（如$heap_k$）接收来自于第二多选器（如MUX_k）的第二输入信号（如DI_k）。第二 Top-K 模块（如$heap_k$）基于第二输入信号（如DI_k）获得第二写回信号（如WB_k）。第二 Top-K 模块（如$heap_k$）向第三排序模块的第三缓冲器发送第二写回信号（如WB_k）。Top-K 模块与缓冲器交替相连，可以提高信号处理效率。

在一些实施例中，第二控制逻辑执行以下判断逻辑：判断第二先入先出队列是否输出第二缓存信号。如果第二先入先出队列输出了第二缓存信号，那么判断第二缓存信号是否有效。如果第二缓存信号有效，

则将第二缓存信号的数据与全局顶端最值进行比较。如果第二缓存信号的数据未超过全局顶端最值，则第二控制逻辑模块确定第二缓存信号作为第二输入信号，如果第二缓存信号的数据超过或等于全局顶端最值，则第二控制逻辑模块确定第二外部信号作为第二输入信号，并且第二缓存信号被舍弃。以及如果第二缓存信号无效，则第二控制逻辑模块确定第二外部信号作为第二输入信号，并且第二缓存信号被舍弃。通过多个堆之间传递数据以及全局最值（最大值或最小值）的约束，可以减少排序时所需比较的次数，节省处理资源的消耗。此外，基于该多堆结构可以较好的利用了高速缓存器等片上存储资源，提高了不同片上资源的利用效率。例如，在图 5 所示的实施例中，单堆模块的堆顶最大值输出T_k，表示本排序模块中当前堆结构的堆顶。信号T_g表示从 N 个堆顶输入获取的最值，即全局最值（又可称为全局堆顶最值）。

在一些实施例中，第二控制逻辑还执行以下判断逻辑：如果第二先入先出队列没有输出第二缓存信号，那么判断第二外部信号是否有效。如果第二外部信号有效，则第二控制逻辑模块确定第二外部信号作为第二输入信号。以及如果第二外部信号无效，则第二控制逻辑模块读取非空的第二先入先出队列中缓存的第三缓存信号作为第二输入信号。即先入先出队列中溢出的缓存信号优先于外部信号被处理。当先入先出队列中没有溢出的缓存信号时，将有效的外部信号作为堆的输入信号，如果外部信号无效，为了保证 FIFO 中的所有数据都进入排序，避免堆模块闲置，可以读取非空的先入先出队列中未溢出的缓存的写回信号（如第三缓存信号）作为堆的输入信号。关于第二控制逻辑执行的判断逻辑可以结合参考图 6 所示的控制逻辑的工作流程 600 或者图 11 所示的控制逻辑的工作流程 1100 的实施例进行理解，此处不再赘述。

在一些实施例中，第一写回信号包括第一写回信号的数据和第一写回使能信号，第一写回使能信号指示第一写回信号是否有效。第二外部信号包括第二外部信号的数据和第二外部使能信号，第二外部使能信号指示第二外部信号是否有效。第二缓存信号包括第二缓存信号的数据和第二缓存使能信号，第二缓存使能信号指示第二缓存信号是否有效。使用写回使能信号和外部使能信号可以更方便地控制信号的输入。例如，在一些实施例中，数据源I_k表示外部输出的待进行排序的数据。数据源I_k（如第二外部信号）又具体的包括数据信号I_data_k（如第二外部信号的数据）和使能信号I_en_k（如第二外部使能信号）。使能信号I_en_k表示数据信号I_data_k是否有效。环状结构的上一个排序模块给本$FIFO_k$的输入信号WB_{k-1}又具体的包括数据信号WB_data_{k-1}（如第一写回信号的数据）和使能信号WB_en_{k-1}（如第一写回使能信号），使能信号WB_en_{k-1}表示数据信号WB_data_{k-1}是否有效。本排序模块中的$FIFO_k$的输出的缓存写回信号WB'_{k-1}（如第二缓存信号）又具体的包括数据信号$WB_data'_{k-1}$（如第二缓存信号的数据）和使能信号$WB_en'_{k-1}$（如第二缓存使能信号），使能信号$WB_en'_{k-1}$表示数据信号$WB_data'_{k-1}$是否有效。多堆结构中本单排序模块输出给下一个单排序模块，作为下一个单排序模块的输入的WB_k，类似的WB_k又具体的分为数据信号WB_data_k和使能信号WB_en_k，使能信号WB_en_k表示数据信号WB_data_k是否有效。在一些实施例中，外部信号是否有效依据交替输入的周期。

在一些实施例中，Top-K 模块为最大堆，全局顶端最值为多个最大堆的堆顶元素的数据之中的最大值，第二缓存信号的数据未超过全局顶端最值，包括：第二缓存信号的数据小于最大值。第二缓存信号的数据超过或等于全局顶端最值，包括：第二缓存信号的数据大于或等于最大值。最大堆的结构能够帮助求出多个输入数据中预定数目的最小值。在一些实施例中，Top-K 模块为最小堆，全局顶端最值为多个最小堆的堆顶元素的数据之中的最小值，第二缓存信号的数据未超过全局顶端最值，包括：第二缓存信号的数据大于最小值。第二缓存信号的数据超过或等于全局顶端最值，包括：第二缓存信号的数据小于或等于最小值。最小堆的结构能够帮助求出多个输入数据中预定数目的最大值。

在一些实施例中，第二 Top-K 模块可以执行以下判断逻辑：判断第二输入信号是否有效，如果第二输入信号有效，则将第二输入信号的数据与第二 Top-K 模块的当前的顶端元素的数据进行比较。如果第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，则将第二输入信号的数据替换掉第二 Top-K 模块的当前顶端元素的数据以成为第二 Top-K 模块中新的元素，以及将被替换出的第二 Top-K 模块的当前顶端元素的数据作为第二写回信号的数据，并且使能第二写回信号有效。以及如果第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，则将第二输入信号的数据作为第二写回信号的数据，并且使能第二写回信号有效。通过利用 Top-K 模块内数据的有序性以及 Top-K 模块顶端数据的约束，将本 Top-K 模块不需要的数据传递给下一个排序模块处理或者及时丢弃，可以减少排序时所需比较的次数，节省处理资源和存储资源。在一些实施例中，第二 Top-K 模块还执行以下判断逻辑：如果第二输入信号无效，则将第二输入信号的数据作为第二写回信号的数据，并且使能第二写回信号无效。被确定无用的数据可以发送给下一个排序模块的 FIFO，由下一个排序模块的控制逻辑来决策丢弃该数据。

在一些实施例中，第二 Top-K 模块更新第二 Top-K 模块的新的顶端元素。例如堆结构执行堆化操作以更新堆结构的新的堆顶元素，其中，堆化操作指的是对堆中的元素进行排序更新。)。第二 Top-K 模块将第二 Top-K 模块的顶端元素发送给全局最值模块。全局最值模块用于获取多个 Top-K 模块的顶端元素，全局最值模块还用于选出多个 Top-K 模块的顶端元素的数据之中的最值，全局最值模块还用于将最值发送给多个缓冲器中的控制逻辑模块。由于被替换进入 Top-K 模块的元素不一定是 Top-K 模块内最值，所以在下一次输入信号输入堆之前，Top-K 模块需要更新出新的顶端元素，以保证下次处理的准确性。

参考图 7 所示的堆模块 700，堆数据一般为层次结构，具有堆顶，如果是最大堆则堆顶为堆的最大值，如果是最小堆则堆顶为堆的最小值，可以以最大堆为例，求 N 个数据输入通道内数据的共同的 K 个最大值。堆模块 700 有输入信号DI_k（如第二输入信号）作为堆模块（$heap_k$）（如第二 Top-K 模块）的输入，具体的输入信号DI_k又分为数据信号DI_data_k和使能信号DI_en_k，使能信号DI_en_k表示数据信号DI_data_k是否有效。堆模块 700 的输出为数据信号WB_data_k和使能信号WB_en_k，使能信号WB_en_k表示数据信号WB_data_k是否有效。关于第二堆的结构以及执行的判断逻辑可以结合参考图 7-图 8 所示的堆模块 700 以及堆控制逻辑的工作流程 800 或者图 12 所示的控制逻辑的工作流程 1200 的实施例进行理解，此处不再赘述。

在一些实施例中，第二 Top-K 模块为最大堆，第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据小于最大堆的当前堆顶元素的数据。第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据大于或等于最大堆的当前堆顶元素的数据。最大堆的结构能够帮助求出多个输入数据中预定数目的最小值。在一些实施例中，第二 Top-K 模块为最小堆，第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据大于最小堆的当前堆顶元素的数据。第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据小于或等于最小堆的当前堆顶元素的数据。最小堆的结构能够帮助求出多个输入数据中预定数目的最大值。

在一些实施例中，在第一时间周期内，第二缓冲器基于第二外部信号和/或第二缓存信号确定第二输入信号，以及在第二时间周期内，第二 Top-K 模块基于第二输入信号获得第二写回信号。在一个迭代周期内，一个排序模块可以完整的完成一次处理工作，一个迭代周期可以分成两个时间周期，例如第一时间周期和第二时间周期。在一些实施例中，第一时间周期与第二时间周期的时间长度相等。缓冲器判断选择进入堆的数据的处理周期与堆处理输入数据并输出写回数据的处理周期交替地执行，每个排序模块可以并行处理数据，这能够提高处理效率，缩减等待时间和处理时间，充分利用处理资源。

以上实施例仅仅是为了更加清楚的说明本申请的技术方案，并不构成对本申请的限定。在其他实施例中还可以包括更多或更少的流程或步骤，或者各实施例进行组合或拆分，得到更多的实现方式，本申请实施例对此不作限制。

结合上述实施例可以看出，本公开的一些实施例可以实现一种针对大规模数据排序的可扩展高并行的多堆结构，其并行度可以增加到 N，实现了更高的数据吞吐量，并且对于求取 M 个 Top 数据的多堆结构，其中每个堆的大小减小为 M/N，提高了单个堆的性能，可以节省存储资源、计算资源和处理资源，以及减少处理时间，更快的实现得到所需的排序靠前的多个数据。

此可扩展的环形多输入的多堆结构，其整体结构可以包含首尾串联的 N 个排序模块以及一个全局最值（最大值或最小值）模块，每个排序模块可以至少包括一个单堆结构、一个缓冲器、一个数据源等。其中，缓冲器接收来自上一个排序模块的写回数据以及外部数据源的数据，并且接收全局最值（最大值或最小值），并向相应的堆输出数据。每个缓冲器至少包含一个先入先出队列、一个多选器以及相应的控制逻辑，先入先出队列用于缓冲前一个排序模块的堆输出的写回数据，多选器及其控制逻辑用于选择输入给自身排序模块的堆的输入信号。每个排序模块的堆接收来自身的多选器的输出，并且生成写回数据输出到下一个排序模块的缓冲器，以及将堆顶输入到全局最值模块。所有的堆与缓冲器交替连接构成一个环形结构。由于缓冲器隔离了相邻的前一个堆的输出，所有的堆可以同步工作或者异步工作，输入数据通过堆的写回数据以及缓冲器，在堆间单向传递直至进入某个堆或者被比较全局最值时扔掉，全局最值（最大值或最小值）可以约束数据的传递，每个堆中存储的数据不会被浪费。因此，本公开的一些实施例的多堆结构可以利用堆内数据排布的有序性质，并通过堆之间的串联，以及全局最值（最大值或最小值）的约束，减少了排序所需的比较次数，使得硬件资源的消耗较少以及节省了存储资源。即，相较于图 1 所示的排序结构，处理资源开销降低，并提高了吞吐量，可以实现更大数据量的 Top 排序以及多个多堆结构的并发工作，并且基于

该多堆结构，每个堆中存储的数据不会被浪费，可以较好的利用片上存储资源（如，静态随机存取存储器（static random access memory，SRAM），提高不同片上资源的利用效率。

下面介绍本申请的实施例提供的数据处理系统 300 的功能模块（或者称为逻辑单元）。参考图 3，图 3 示出了本申请一些实施例提供的数据处理系统 300 的功能模块的示意图。

本申请实施例可以对数据处理系统 300 进行功能模块的划分，例如，可以将数据处理系统 300 的各个功能划分各个功能模块，也可以将数据处理系统 300 的两个或两个以上的功能集成在一个功能模块中，本实施例不作限制。上述集成的模块既可以采用硬件或软件的形式实现，也可以采用软硬件结合的形式实现。

示例性的，如图 3 所示，数据处理系统 300 中包括存储模块 301、控制模块 302、数据处理模块 303、数据消费模块 304，其中，数据处理模块 303 中还包括前处理模块 305、排序模块 306、后处理模块 307。

该数据处理系统 300 可以是基于专用集成电路芯片的专用加速器系统以及基于可编程逻辑器件的硬件系统。对于专用加速器系统，数据处理模块 303 可以包括应用专用集成电路芯片（ASIC）。对于可编程硬件系统，数据处理模块 303 包括可编程逻辑器件（FPGA）。存储子系统由 ASIC 或 FPGA 上的片上 SRAM 缓冲、随机存储器（DDR 或 HBM）以及非易失性存储器（SSD 或 HHD）等组成。数据消费模块 304 可以是本地通用计算机系统也可以是远程客户端应用。本公开的应用场景可以是针对需要数据排序的计算机系统，例如云计算平台加速相似性搜索、推荐系统以及数据库系统中的 Top 排序计算。

本公开的实施例可以是在专用加速器系统或者可编程硬件系统的软硬件实现。其中控制模块 302 控制存储模块 301 的数据的输入输出以及数据处理模块 303 中其他模块的控制。前处理模块 305 和后处理模块 307 根据应用场景的不同可以有不同的实现。前处理模块 305 可以针对片外存储系统输入的数据进行处理，以得到排序所需的数据；排序模块 306 用于执行排序；后处理模块 307 可以根据排序的结果再进一步处理以得到可以输出给数据消费模块 304 的结果。例如，排序模块 306 可以执行前述方法 200 所描述的相应一个或多个实施例，排序模块 306 中包含片上存储单元，排序的中间结果在排序过程中存储在片上，排序结果输出到后处理模块 307。在专用加速器系统或者可编程硬件系统上，前处理模块 305、排序模块 306 以及后处理模块 307 可以均在片上实现，其中控制模块 302 可使用通用处理器或者在加速器芯片上进行实现。

实施本申请实施例的数据处理系统 300 中的功能模块，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

数据处理系统 300 中的各个功能模块可以通过软件、硬件或者两者结合的方式实现。可以理解的，关于数据处理系统 300 包括的各个功能单元的具体实现可参考本申请其他各实施例中的详细描述，这里不多赘述。

需要注意的是，本申请实施例的数据处理系统 300 中不限定各个功能模块的类型、数量和关联关系，在一些实施例中，数据处理系统 300 中可以包括更多或更少的功能模块等。

需要说明的是，本申请实施例中对模块或单元的划分是示意性的，仅仅为一种逻辑功能划分，实际实现时也可以有另外的划分方式。另外，在实施例中的各功能单元可以集成在一个单元中，也可以是单独物理存在，也可以两个或两个以上单元集成为一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现，或者软硬件结合的形式实现。

为了更清楚的理解本公开的内容，下面介绍本公开针对大规模数据的排序提出的一种可扩展高并行的多输入的多堆结构，以及一种双堆结构。

本公开的一些实施例提供了一种可扩展高并行的环形多输入的多堆结构，这里具体以最大堆为例进行说明，最小堆的情况反之亦然。假设该多堆结构可以求取多个数据中的 M 个最小值，该多堆结构中并行的堆数目为 N，可扩展的环形多输入多堆结构可以由 N 个堆、N 个缓冲器、N 个数据源、N 个控制逻辑以及一个全局最大值模块组成。N 个堆的结构可以均为最大堆，最大堆用于筛选最小值，每个堆可以包括 K 个的元素，完成一系列输入和排序之后，最后输出的这 N 个堆中所包括的 K 个元素的集合即为所求取的 M 个最小值，即 M 的值为 K 与 N 的乘积值，其中 K、M、N 为正整数。

该多堆结构包括 N 个堆，其中第i个堆$heap_i$包括输入数据DI_i与写回数据WB_i以及堆顶数据T_i的接口，其中$i = 0,1,\dots,N-1$。在一些实施例中，第i个堆可每两个周期处理一个输入数据DI_i，并且在输入数据DI_i有效的一个周期后返回一个写回数据WB_i。其中，输入数据DI_i与写回数据WB_i均可以包括进行排序的元组以及使能信号。例如，输入数据DI_i包括元组DI_Data_i以及使能信号DI_en_i，写回数据WB_i包括元组WB_Data_i以及使能信号WB_en_i，其中元组可以表示为$(ID, Value)$，包含用于比较的数值$Value$以及用于对应数值的编号ID或者相应的键值等。在一些实施例中，输入数据DI_i的使能信号DI_en_i为真表示输入数据

DI_i有效，在输入数据DI_i有效一周期后输出写回数据WB_i，并伴随着写回数据WB_i的使能信号WB_en_i有效。在一些实施例中，当输入数据DI_i大于等于堆顶T_i时，写回数据WB_i为输入数据DI_i，堆顶保持不变。否则，写回数据WB_i为T_i，堆顶被替换为DI_i。

该多堆结构还包括 N 个缓冲器，第i个缓冲器$buffer_i$可以包括先入先出队列$FIFO_i$和多选器MUX_i以及相应的控制逻辑C_i，其中$i = 0,1,\dots,N-1$。$FIFO_i$用于缓冲来自前一个堆$heap_{i-1}$的写回数据WB_{i-1}。当$FIFO_i$被填满到一定程度后，可以输出被读取数据WB'_{i-1}。多选器MUX_i其中一个输入I_1^i连接着对应的数据源I_i，另一个输入I_2^i连着$FIFO_i$的写回数据WB'_{i-1}。多选器MUX_i的选择信号S_i由控制逻辑C_i给出。多选器MUX_i的选择输出连着堆$heap_i$的输入数据DI_i。在一些实施例中，对于多选器控制逻辑C_i，当$FIFO_i$的输出数据WB'_{i-1}有效时，比较WB'_{i-1}和全局最大值T_g，当WB'_{i-1}的值小于T_g时，选择$FIFO_i$的输出数据WB'_{i-1}作为多选器的输出。当$FIFO_i$的输出数据WB'_{i-1}无效时，如果数据源I_i有效则选择数据源I_i作为多选器的输出，如果数据源I_i无效可以选择读取非空的$FIFO_i$中缓存的数据（如第三缓存信号）作为多选器的输出。在一些实施例中，因为堆每两个周期处理一个数据，所以缓冲器可以相应的每两个周期发生一次判断。

该多堆结构还包括 N 个数据源，第i个数据源I_i可以使用先入先出存储器等实现，用于缓冲外部信号的输入数据元组并提供相应的输入数据使能。全局最大值模块用于接收所有堆的堆顶数据T_i，并产生所有堆的堆顶数据中的最大值，即全局最大值T_g，T_g基于实现形式的不同会相比输入存在几个周期的延迟。

在一些实施例中，在没有数据输入时，所有堆的输入数据的使能无效，所有堆都不工作。初始时，N 个数据源使能有效，可以每两个周期分别向 N 个缓冲器发送数据。例如对于第i个缓冲器$buffer_i$，由于缓冲器$buffer_i$中的$FIFO_i$未被填充到可以读取数据，控制逻辑C_i直接选择数据源I_i作为相应堆$heap_i$的输入。如此重复多个周期，直到存在缓冲器$buffer_i$中的$FIFO_i$被填充到可以向多选器MUX_i输出数据WB'_{i-1}，相应的控制逻辑C_i可以进行判断。当$FIFO_i$的输出数据WB'_{i-1}有效时，比较WB'_{i-1}和全局最大值T_g，如果$FIFO_i$输出数据WB'_{i-1}小于全局最大值模块的输出T_g，则多选器MUX_i选择$FIFO_i$的输出数据WB'_{i-1}作为堆$heap_i$的输入数据DI_i。当$FIFO_i$的输出数据WB'_{i-1}无效时，如果数据源I_i使能有效，则选择数据源I_i作为堆$heap_i$的输入数据DI_i，如果数据源I_i使能无效，控制逻辑C_i可以读取非空的$FIFO_i$中缓存的其他数据（如第三缓存信号）作为堆$heap_i$的输入数据DI_i，直到$FIFO_i$为空。当所有数据源输入完成，并且所有缓冲器中的$FIFO$为空时，待堆执行完所有排序操作后，最后输出的这 N 个堆中（每个堆包括 K 个的元素）所包括的 K × N 个元素的集合即为所求取的 M 个最小值，排序完成。实现该多堆结构，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

下面结合图 4 所示的多堆结构 400、图 5 所示的排序模块 500、排序模块的控制逻辑流程 600、堆模块 700、堆模块的堆控制逻辑流程 800 来更具体地说明本公开提供的一些实施例。

图 4 示出了本公开提出的一种可扩展的高并行的多输入的环形多堆结构 400 的示意图。本实施例以堆为最大堆来进行示例说明，最小堆的实施例可以反向推之。如图 4 所示，多堆结构 400 中包含 N 个如虚线框所示的单个排序模块 500 以及一个全局最大值模块，这 N 个排序模块可以串连成一个环形结构，单个排序模块中可以包括一个缓冲器和一个单堆结构。单个排序模块 500 的具体结构可以参考图 5 所示的实施例。其中，N 个单堆结构通过 N 个缓冲器相连接，N 个单堆结构与 N 个缓冲器交替排列，首尾相接连成环状，一个缓冲器的输入可以是与之相连的前一个堆的输出，一个缓冲器的输出可以是与之相连的下一个堆的输入。全局最大值模块用于接收所有堆的堆顶数据T_i，并产生所有堆的堆顶数据中的最大值，即全局最大值T_g，T_g基于实现形式的不同会相比输入存在几个周期的延迟。可以看出，多堆结构 400 中所包括的 N 个排序模块中的第k个排序模块可以具有单独的数据源输入I_k，堆顶的数据值的输出T_k，以及全局堆顶最大值输入T_g，其中$k = 0,1,\dots,N-1$。图 4 所示的可扩展的高并行的多堆结构 400，可以具有 N 个单独的数据源输入，可同时完成 N 路数据输入，以及多堆结构 400 中的 N 个堆的每个堆的结构可以均为最大堆，最大堆用于筛选最小值，每个堆可以包括 K 个的元素。将很多数据输入该多堆结构 400，在完成一系列输入和排序之后，最后输出的这 N 个堆中所包括的 K × N 个元素的集合即为所求取的 M 个最小值，即找到该输入的很多数据中的 TOP-M 个数据。实现该多堆结构 400，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

参考图 5 中所示的单个排序模块 500，第k个排序模块 500 的外部接口包括数据源I_k、上一个排序模块给本排序模块的输入WB_{k-1}、本排序模块给下一个排序模块的输入WB_k、本单堆模块的堆顶最大值输出T_k、全局最大值输入信号T_g等，其中以下小写的 k 表示如图 4 结构中的第 k 个单排序模块（区别于在本公开中的大写的 K 表示 TOP-K 排序中需要获得的 K 个最大或者最小数据）。具体地，数据源I_k表示外部输出的待

进行排序的数据。数据源I_k又具体的分为数据信号I_data_k和使能信号I_en_k。使能信号I_en_k表示数据信号I_data_k是否有效。环状结构的上一个单排序模块给本单排序模块的输入WB_{k-1}又具体的分为数据信号WB_data_{k-1}和使能信号WB_en_{k-1}，使能信号WB_en_{k-1}表示数据信号WB_data_{k-1}是否有效。多堆结构中本单排序模块给下一个单排序模块的输出WB_k可以作为下一个单排序模块的输入，类似的WB_k又具体的分为数据信号WB_data_k和使能信号WB_en_k，使能信号WB_en_k表示数据信号WB_data_k是否有效。本单堆模块的堆顶最大值输出T_k，表示本排序模块中当前堆结构的堆顶。输入信号T_g表示从 N 个排序模块的堆顶输出的 N 个堆的最大值中的最大值，即全局最大值。

如图 5 所示的单个排序模块 500，以第k个模块为例，包含第k堆模块（$heap_k$）520 和对应的缓冲器模块（$buffer_k$）510。缓冲器模块 510 包含控制逻辑模块（C_k）540、多路选择器模块（MUX_k）530 和先入先出模块($FIFO_k$)550等。多路选择器 530 分别接收本排序模块对应的数据源I_k，和$FIFO_k$ 550 的输出WB'_{k-1}。$FIFO_k$ 550 接收上一个排序模块给本排序模块的写回信号WB_{k-1}，并输出缓冲过的数据WB'_{k-1}。多路选择器 530 的选择信号S_k来自控制逻辑模块 540 的输出。多路选择器 530 的输出DI_k作为第 k 堆模块（$heap_k$）520 的输入，具体的DI_k信号又分为数据信号DI_data_k和使能信号DI_en_k，使能信号DI_en_k表示数据信号DI_data_k是否有效。第 k 堆模块($heap_k$)520 的输出为WB_k，包含数据信号WB_data_k和使能信号WB_en_k。控制逻辑模块 540 的输入是WB'_{k-1}、I_k和T_g，输出为多路选择器 530 的选择信号S_k。

以第k个堆的缓冲器 510 中的控制逻辑（C_k）540 为例，其图 5 中的控制逻辑的工作流程 600 可以如图 6 所示。在一些实施例中，控制逻辑模块 540 可以每两个周期开始一次判断。如图 6 所示，在框 601，判断$FIFO_k$是否有数据输出，如果有数据输出，则进入框 602，如果没有数据输出，则进入框 603。此判断的目的在于，因为全局最大值模块输出的T_g相对输入会存在几个周期的延迟，WB_{k-1}输入到缓冲器中的$FIFO_k$后，直到$FIFO_k$填充满才被读出，这时WB'_{k-1}为多个周期前的输入，相对T_g会延后，有利于控制逻辑能够发挥出较好筛选性能，避免WB_{k-1}在多个堆中传递而不会被筛选掉。在框 602，判断$FIFO_k$输出的数据WB'_{k-1}对应的使能$WB_en'_{k-1}$是否有效，即是否为真（true），有效（即为真）则进入框 604，无效（即为假）则进入框 606。在框 604，判断$WB_data'_{k-1}$是否小于T_g，如果是则进入框 605，否则进入框 606。在框 605，输出选择信号S_k使多选器MUX_k选择WB'_{k-1}作为多选器MUX_k的输出，如上文描述DI_k信号又分为数据信号DI_data_k和使能信号DI_en_k，对应的使能信号DI_en_k为真，DI_data_k为$WB_data'_{k-1}$的数据值，此时数据源I_k不被选择，需要保持数据源I_k。在框 606，输出选择信号S_k使多选器MUX_k选择数据源I_k作为多选器MUX_k的输出，如上文描述DI_k信号又分为数据信号DI_data_k和使能信号DI_en_k，DI_k信号的数据信号和使能信号都取数据源I_k的数据信号和使能信号输出，WB'_{k-1}的数据则被忽略。在框 603，判断I_en_k是否有效，即是否为真，有效则进入框 606，无效则进入框 607。在框 607，读取非空的$FIFO_k$中缓存的其他数据作为DI_k，此处的目的是为了在没有外部输入数据时，将$FIFO_k$中并未溢出的数据提取出来，保证$FIFO_k$中所有数据都进入排序，避免堆模块闲置。在另一些可选的实施例中，在提取非空的$FIFO_k$中的数据之后，还可以对该写回数据执行如框 602 和/或框 604 中的判断过程，例如与T_g进行比较，再决定是否将该写回数据作为DI_k。

以上图 4、图 5 和图 6 所示的多堆结构 400、排序模块 500 及其控制逻辑的工作流程 600 相关的实施例仅是为了更加清楚的说明本申请的技术方案而示出的一些示例，并不构成对本申请的限定。在其他实施例中还可以包括更多或更少的模块、结构、流程或步骤，或者各实施例进行组合或拆分，得到更多的实现方式，本申请实施例对此不作限制。

以第k个单排序模块 500 中的堆模块（$heap_k$）520 为例描述的结构可以如图 7 所示的堆模块 700。堆模块 700 可以包括堆数据和堆控制逻辑。堆数据一般为层次结构，具有堆顶，如果是最大堆则堆顶为堆的最大值，如果是最小堆则堆顶为堆的最小值，本实施例的堆模块是以最大堆为例，求输入数据中的 K 个最小值。

堆模块 700 具有输入信号DI_k作为堆模块（$heap_k$）520 的输入，具体的输入信号DI_k又分为数据信号DI_data_k和使能信号DI_en_k，使能信号DI_en_k表示数据信号DI_data_k是否有效。堆模块 700 的输出为WB_k，包括数据信号WB_data_k和使能信号WB_en_k，使能信号WB_en_k表示数据信号WB_data_k是否有效。另外，堆模块 700 还输出堆顶T_k给全局最大值模块。

堆模块 700 中的堆控制逻辑流程 800 如图 8 所示。在一些实施例中，堆模块 700 可以每两个周期开始处理一个数据输入。在框 801，判断输入信号DI_k对应的使能信号DI_en_k是否有效，即是否为真（true），有效（即为真）则进入框 802，无效（即为假）则进入框 805。在框 802，判断DI_data_k是否小于当前堆顶值T_k，如果是则进入框 803，否则进入框 804。在框 803，设置WB_en_k为真，DI_data_k与当前的堆顶数据T_k进

行替换，DI_data_k的值变成新的堆顶数据T_k的值，赋予WB_data_k数据的值为刚才被替换出去的堆顶数据T_k的值，同时堆逻辑开始执行堆的堆化操作。所谓堆的堆化操作，是指当堆的堆顶发生变化后为了维护堆的性质，而执行的堆内数据特定的操作。在框 804，设置WB_en_k为真，堆顶数据T_k保持不变，赋予WB_data_k数据的值为输入数据DI_data_k的值，同时堆逻辑开始执行堆的堆化操作。在框 805，DI_en_k为假，输入数据DI_data_k无效，设置WB_en_k为假，WB_data_k数据无效，堆继续完成堆化操作。可以理解的是，进行堆化操作只需要在下一次向单堆结构输入信号之前更新出新的堆顶元素即可，无需将堆中每个元素都按顺序排序好。可以理解的是，堆模块会将WB_en_k和WB_data_k发送给下一个排序模块的缓冲器，由下一个缓冲器的控制逻辑决定保留还是舍弃该WB_data_k。

需要说明的是实现如图 7 和图 8 中所述的 Top-K 排序功能并不一定需要由堆这种数据结构实现，上述实施例中以堆作为示例进行了说明，但是 Top-K 模块并不仅限于堆结构来实现，其他可以实现类似的 Top-K 排序功能的任何结构或模块都在本公开的保护范围之内。例如，框 803 中当前堆顶可选地为其他备选数据结构中多个数据的最大值，框 804 中，可选地为替换出其他备选数据结构中多个数据的最大值。

在一些实施例中，单个堆每两个周期处理一个数据，如流程 800，对应的单个缓冲器每两个周期发生一次控制逻辑的判断流程，如流程 600。因此对于包括 N 个排序模块的多堆结构 400，N 个排序模块可以以一个迭代周期并行执行排序活动，N 个缓冲器、N 个堆可以分别并行处理数据。在一个示例中，对于单个排序模块，一个迭代周期可以包括两个时钟周期（非限制性的，也可以是四个时钟周期，具体可以根据实际情况设定），在一个时钟周期内，缓冲器选择并向堆输出输入数据DI_k，在下一个周期里，堆模块对输入数据DI_k进行数据处理，并向下一个排序模块的$FIFO$输出WB_k。

此外，可以理解的是，在一些实施例中，多堆结构开始工作后，在没有数据输入时，所有堆的输入使能信号无效，所有堆都不工作。当有数据输入时，向所有缓冲器输入数据，由于此时所有缓冲器中$FIFO$为空，直接选择外部输入数据作为堆的输入。堆$heap_k$接收数据进行处理后向下一个堆$heap_{k+1}$的缓冲器输出写回数据。在输入外部数据多个周期后，直到缓冲器$buffer_k$中的$FIFO$被填满，$buffer_k$开始按照图 6 中的控制逻辑流程 600 进行工作，筛选堆的写回数据并选择输入数据到堆$heap_k$。根据以上实施例，外部的输入数据可以在相邻的堆之间单向传递，直至进入某个堆中，或者被扔掉。全局最大值模块产生的全局最大值T_g可以约束所有堆中的值，保证堆之间存在着相应的阈值上限，使得所有堆中存放的数据满足排序的要求。当所有数据输入完成后，$buffer_k$继续工作直到相应$buffer_k$中的$FIFO$被清空，最终完成排序。

根据本公开的上述一个或多个实施例所实现的用于排序的多输入的多堆结构，数据平均吞吐量提高，并且具有接近线性的扩放性的增益效果。本公开的实施例通过利用堆内数据的有序性，并且多个堆之间通过传递数据以及全局最值约束，减少了排序时所需比较的次数，使得硬件处理资源的消耗较少，并且基于堆结构较好的利用了片上存储资源，提高了不同片上资源的利用效率。此外，本方案排序的比较次数较少。例如，对于并行度为N、需要求取 K 个靠前值的实现 Top-K 排序的多堆结构，其中每个堆的大小为$\frac{K}{N}$，层数为$\left\lceil \log_2 \frac{K}{N} \right\rceil$，即每个堆每周期会进行$O\left(\log_2 \frac{K}{N}\right)$次比较，同时堆与堆之间传递数据，每周期比较$N$次，因此总的比较次数为$O\left(\mathrm{N}(\log_2 \frac{K}{N} + 1)\right)$。

图 4 至图 8 示出的实施例是本申请提供的一种实现方式，仅仅是为了更加清楚的说明本申请的技术方案，并不对本申请的其他实施例构成限定。实施本公开的上述一些实施例中的针对大规模数据的用于排序的可扩展高并行的多堆结构，其并行度可以增加到 N，实现了更高的数据吞吐量，并且对于求取 M 个 Top 数据的多堆结构，其中每个堆的大小减小为 M/N，提高了单个堆的性能，可以节省存储资源、计算资源和处理资源，以及减少处理时间，更快的得到靠前的 Top 数据。

此可扩展的环形多输入的多堆结构，其整体结构可以包含 N 个单堆、N 个缓冲器、相应的$FIFO$以及一个全局最大/小值模块，N 个缓冲器包含$FIFO$、多路器MUX以及相应的控制逻辑C。每个堆接收来自身的缓冲器模块的输出，并且自身写回数据连接到下一个堆的缓冲器，还将堆顶输入到全局最大/小值模块。缓冲器接收来自上一个堆的写回数据以及外部数据源，并且接收全局最大/小值，并向相应的堆输出数据。所有的堆与缓冲器交替相连构成一个环形结构。由于缓冲器隔离了相邻的前一个堆的数据输出，所有堆可以同步工作或者异步工作。输入数据通过堆的写回值以及缓冲器，在堆间单向传递直至进入某个堆或者被全局最大/小值模块扔掉，全局最大/小值模块可以约束数据的传递，使得每个堆中存储的数据不会被浪费。因此，本公开的一些实施例的多堆结构可以利用堆内数据排布的有序性质，并通过堆之间的串联，以及全局最大/小值的约束，减少了排序所需的比较次数，节省处理资源的消耗以及节省了存储资源，并提高了吞吐量，可以实现更大数据量的求取靠前数据的排序以及多个多堆结构的并发工作。

本公开的一些实施例还提供了一种针对双周期单输入交替的双堆结构，这里具体以最大堆为例进行说明，最小堆的情况反之亦然。该双堆结构可以包括两个堆结构、两个缓冲器以及数据源。其中两个堆为正常的双周期堆结构$heap_i$，包括输入数据DI_i与写回数据WB_i以及堆顶数据T_i的接口，其中$i=0,1$，分别代表两个堆。每个堆结构可每两个周期处理一个输入数据DI_i，并且在输入数据有效的一个周期后返回一个写回数据WB_i。在一些实施例中，输入数据DI_i可以包括元组DI_Data_i以及使能信号DI_en_i，写回数据WB_i可以包括元组WB_Data_i以及使能信号WB_en_i，其中元组可以表示为$(ID, Value)$，包含用于比较的数值$Value$以及用于对应数值的编号ID或者相应的键值等。在一些实施例中，输入数据DI_i的使能信号DI_en_i为真表示输入数据DI_i有效，在输入数据DI_i有效一周期后输出写回数据WB_i，并伴随着写回数据WB_i的使能信号WB_en_i有效。在一些实施例中，当输入数据DI_i大于等于堆顶T_i时，写回数据WB_i为DI_i，堆顶保持不变。否则，写回数据WB_i为T_i，堆顶被替换为DI_i。

该双堆结构还包括两个缓冲器，第i个缓冲器$buffer_i$可以包括先入先出队列$FIFO_i$和多选器MUX_i以及相应的控制逻辑C_i，其中$i=0,1$。$FIFO_i$用于缓冲来自另外一个堆的写回数据WB_{i-1}。当$FIFO_i$被填满到一定程度后，可以输出被读取数据WB'_{i-1}。多选器MUX_i其中一个输入I_1^i连接着对应的数据源I，另一个输入I_2^i连着$FIFO_i$的写回数据WB'_{i-1}。多选器MUX_i的选择信号S_i由控制逻辑C_i给出。多选器MUX_i的选择输出连着堆$heap_i$的输入数据DI_i。在一些实施例中，对于多选器控制逻辑C_i，当$FIFO_i$的输出数据WB'_{i-1}有效时，比较WB'_{i-1}和当前堆的堆顶T_i，当WB'_{i-1}的值小于T_i时，选择$FIFO_i$的输出数据WB'_{i-1}作为多选器的输出。当$FIFO_i$的输出数据WB'_{i-1}无效时，如果数据源I_i有效则选择数据源I_i作为多选器的输出，如果数据源I_i无效可以选择读取非空的$FIFO_i$中的另一数据作为多选器的输出。在一些实施例中，因为堆每两个周期处理一个数据，所以缓冲器可以相应的每两个周期发生一次判断。

该双堆结构还可以包括数据源I，数据源I可以使用先入先出存储器等实现，用于缓冲外部信号的输入数据元组并提供相应的输入数据使能。在一些实施例中，数据源I可以轮流为双堆结构提供输入数据，使能信号对两个缓冲器交替有效。当数据源为空时，两个堆的输入数据使能信号无效，两个堆不工作。在初始时，数据源I向$buffer_0$提供输入数据并且使能信号有效，控制逻辑C_0选择数据源I作为堆的输入数据DI_0。在一个周期后堆$heap_0$向$buffer_1$输出写回数据WB_0。由于$buffer_1$中的$FIFO_1$未被填充到可以输出数据，控制逻辑C_1选择数据源I向$buffer_1$提供的使能信号有效的输入数据作为堆的输入数据DI_1。一个周期后，堆$heap_1$向$buffer_0$输出写回数据WB_1，由于$buffer_0$中的$FIFO_0$未被填充到可以输出数据，控制逻辑C_0选择有效的数据源I作为堆的输入数据DI_0。如此重复直到多个周期后，有一个缓冲器中的$FIFO$被填充到可以输出数据，例如为$buffer_0$，相应的控制逻辑C_0进行判断，如果$FIFO_0$输出数据WB'_1小于堆$heap_0$的堆顶T_0，则多选器选择$FIFO_0$的输出WB'_1作为堆$heap_0$的输入数据DI_0。如果$FIFO_0$输出数据WB'_1大于堆$heap_0$的堆顶T_0，则判断数据源I是否有效，如果数据源I有效则选择数据源作为堆$heap_0$的输入数据DI_0，如果数据源I无效则可以读取非空的$FIFO_0$中的其他缓存数据作为堆$heap_0$的输入数据DI_0。堆$heap_0$在数据输入后的一个周期输出写回数据WB_0并将WB_0发送给$buffer_1$，相应的控制逻辑C_1可以执行与上述控制逻辑C_0类似的判断过程。之后，双堆结构交替输入数据并完成排序。当数据源I的使能信号无效时，控制逻辑可以读取$FIFO$中的数据，直到$FIFO$为空。当数据源I输入完成且所有缓冲器中的$FIFO$为空时，待堆执行完所有排序操作后，最后输出的这两个堆中所包括的元素的集合即为所求取的 M 个最小值，排序完成。实现该双堆结构，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

示例性的，可以结合参考图 9 至图 11 示出的本公开提供的针对双周期单输入交替的双堆结构的一些实施例来更好的理解。以最大堆为例，实施例中针对双周期单输入交替的双堆结构可以包括两个堆结构、两个缓冲器以及一个数据源输入。参考图 9 示出的双堆结构 900 中，双堆结构中的两个堆$heap_0$和$heap_1$通过缓冲器$buffer_0$和$buffer_1$相连，也构成一个环形的循环结构，但整体只有一个单一的输入。其中每个缓冲器模块包含了多选器以及控制逻辑。可以看出，本质上，双堆结构 900 是多堆结构 400 的 N=2 时的一个特殊的实例。其中，数据源I分别连接到缓冲器 0（$buffer_0$）和缓冲器 1（$buffer_1$），用于将外部数据输入进入堆结构。该数据源I可以使向$buffer_0$和$buffer_1$的输入信号交替有效，例如第一个周期使能向$buffer_0$的输入信号有效，向$buffer_1$的输入信号无效，第二个周期使能向$buffer_1$的输入信号有效，向$buffer_0$的输入信号无效，第三个周期又使能向$buffer_0$的输入信号有效，向$buffer_1$的输入信号无效，如此循环。$buffer_0$可以接收$heap_1$的写回数据WB_1、$heap_0$的堆顶数据T_0和数据源I，并根据控制逻辑C_0进行判断，选择输出一个数据信号作为$heap_0$的输入数据DI_0。$heap_0$接收输入数据DI_0，在一个周期后，$heap_0$向

$buffer_1$输出写回信号WB_0，以及向$buffer_0$输出堆顶数据T_0。类似地，$buffer_1$可以接收$heap_0$的写回数据WB_0、$heap_1$的堆顶数据T_1和数据源I，并根据控制逻辑C_1进行判断，选择输出一个数据信号作为$heap_1$的输入数据DI_1。$heap_1$接收输入数据DI_1，在一个周期后，$heap_1$向$buffer_0$输出写回信号WB_1，以及向$buffer_1$输出堆顶数据T_1。实现该双堆结构 900，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

双堆结构 900 中的单个排序模块 1000 结构如图 10 所示，以堆 0 及缓冲器 0 组成的排序模块，即堆$heap_0$ 1020 及其$buffer_0$ 1010 为例示意。堆$heap_0$ 1020 的写回数据WB_0连接$buffer_1$以及控制逻辑C_1，堆$heap_0$ 1020 的输入数据DI_0与多选器（MUX_0）1030 的输出相连，写回信号WB_1作为缓冲器 1010 内部先入先出队列（$FIFO_0$）1050 的输入，$FIFO_0$ 1050 的输出WB_1'作为多选器MUX_0 1030 的一个输入。$buffer_0$ 1010 与数据源输入数据I相连，数据源I连接到内部多选器 1030，作为多选器 1030 的一个输入，并且使能信号I_en连接到控制逻辑C_0 1040。控制逻辑（C_0）1040 接收WB_1'、I_en以及$heap_0$的堆顶T_0，并进行判断，选择输出一个数据信号作为$heap_0$的输入数据DI_0。$heap_0$接收输入数据DI_0，在一个周期后，堆$heap_0$ 1020 向$buffer_1$输出写回数据WB_0，以及向控制逻辑（C_0）1040 输出堆顶数据T_0。堆$heap_1$及其$buffer_1$的连接关系与双堆结构 900 所示的堆$heap_0$及其$buffer_0$的连接关系类似。

在一些实施例中，堆$heap_0$的缓冲器$buffer_0$的控制逻辑C_0的判断流程 1100 可以如图 11 所示。在一些实施例中，控制逻辑C_0可以每两个周期开始一次判断。如图 11 所示，在框 1101，判断$FIFO_0$是否有数据输出，如果有数据输出，则进入框 1102，如果没有数据输出，则进入框 1103。此判断的目的在于，因为输出的堆顶T_0相对输入会存在几个周期的延迟，WB_1输入到缓冲器中的$FIFO_0$后，直到$FIFO_0$填充满才被读出，这时WB_1'为多个周期前的输入，相对T_0会延后，有利于控制逻辑能够发挥出较好筛选性能，避免WB_1在双堆中传递而不会被筛选掉。在框 1102，判断$FIFO_0$输出的数据WB_1'对应的使能WB_en_1'是否有效，即是否为真（true），有效（即为真）则进入框 1104，无效（即为假）则进入框 1106。在框 1104，判断WB_data_1'是否小于T_0，如果是则进入框 1105，否则进入框 1106。在框 1105，输出选择信号S_0使多选器MUX_0选择WB_1'作为多选器MUX_0的输出，如上文描述DI_0信号又分为数据信号DI_data_0和使能信号DI_en_0，对应的使能信号DI_en_0为真，DI_data_0为WB_data_1'的数据值，此时数据源I不被选择，需要保持数据源I。在框 1106，输出选择信号S_0使多选器MUX_0选择数据源I作为多选器MUX_0的输出，如上文描述DI_0信号又分为数据信号DI_data_0和使能信号DI_en_0，DI_0信号的数据信号和使能信号都取数据源I的数据信号和使能信号输出，WB_1'的数据则被忽略。在框 1103，判断I_en是否有效，即是否为真，有效则进入框 1106，无效则进入框 1107。在框 1107，读取非空的$FIFO_0$中缓存的数据作为DI_0，此处的目的是为了在没有外部输入数据时，将$FIFO_0$中并未溢出的数据提取出来，保证$FIFO_0$中所有数据都进入排序，避免堆模块闲置。在另一些可选的实施例中，在提取非空的$FIFO_0$中的数据之后，还可以对该写回数据执行如框 1102 和/或框 1104 中的判断过程，例如与T_0进行比较，再决定是否将该写回数据作为DI_0。

以第 0 个堆模块（$heap_0$）1020 为例描述的结构可以如图 7 所示的堆模块 700。堆模块可以包括堆数据和堆控制逻辑。堆数据一般为层次结构，具有堆顶，如果是最大堆则堆顶为堆的最大值，如果是最小堆则堆顶为堆的最小值，本实施例的堆模块是以最大堆为例，求输入数据中的 K 个最小值。堆模块$heap_0$具有输入信号DI_0作为堆模块$heap_0$的输入，具体的输入信号DI_0又分为数据信号DI_data_0和使能信号DI_en_0，使能信号DI_en_0表示数据信号DI_data_0是否有效。堆模块$heap_0$的输出为WB_0，包括数据信号WB_data_0和使能信号WB_en_0，使能信号WB_en_0表示数据信号WB_data_0是否有效。另外，堆模块$heap_0$还输出堆顶T_0给控制逻辑C_0。

堆模块$heap_0$中的堆控制逻辑流程 1200 如图 12 所示。在一些实施例中，堆模块$heap_0$可以每两个周期开始处理一个数据输入。在框 1201，判断输入信号DI_0对应的使能信号DI_en_0是否有效，即是否为真（true），有效（即为真）则进入框 1202，无效（即为假）则进入框 1205。在框 1202，判断DI_data_0是否小于当前堆顶值T_0，如果是则进入框 1203，否则进入框 1204。在框 1203，设置WB_en_0为真，DI_data_0与当前的堆顶数据T_0进行替换，DI_data_0的值变成新的堆顶数据T_0的值，赋予WB_data_0数据的值为刚才被替换出去的堆顶数据T_0的值，同时堆逻辑开始执行堆的堆化操作。所谓堆的堆化操作，是指当堆的堆顶发生变化后为了维护堆的性质，而执行的堆内数据特定的操作。在框 1204，设置WB_en_0为真，堆顶数据T_0保持不变，赋予WB_data_0数据的值为输入数据DI_data_0的值，同时堆逻辑开始执行堆的堆化操作。在框 1205，DI_en_0为假，输入数据DI_data_0无效，设置WB_en_0为假，WB_data_0数据无效，堆继续完成堆化操作。可以理解的是，进行堆化操作只需要在下一次向堆结构输入信号之前更新出新的堆顶元素即可，无需将堆中每个元

素都按顺序排序好。可以理解的是，堆模块会将WB_en_0和WB_data_0发送给下一个排序模块的缓冲器，由下一个缓冲器的控制逻辑决定保留还是舍弃该WB_data_0。

需要说明的是如图 12 中所述的功能并不一定需要由堆这种数据结构实现，只要实现类似的功能，均能实现。例如，框 1203 中当前堆顶可选地为其他备选数据结构中多个数据的最大值，框 1204 中，可选地为替换出其他备选数据结构中多个数据的最大值。

在一些实施例中，单个堆每两个周期处理一个数据，如流程 1200，对应的单个缓冲器每两个周期发生一次控制逻辑的判断流程，如流程 1100。因此对于包括两个排序模块的双堆结构 900，两个排序模块可以以一个迭代周期并行执行排序活动，两个缓冲器、两个堆可以分别并行处理数据。在一个示例中，对于单个排序模块，一个迭代周期可以包括两个时钟周期（非限制性的，也可以是四个时钟周期，具体可以根据实际情况设定），在一个时钟周期内，缓冲器选择并向堆输出输入数据DI，在下一个周期里，堆模块对输入数据DI进行数据处理，并向下一个排序模块的$FIFO$输出写回数据WB。

此外，可以理解的是，在一些实施例中，双堆结构开始工作后，在没有数据输入时，两个堆的输入使能信号无效，两个堆都不工作。当有数据输入时，两个堆交替有输入数据。在一些实施例中，第一个周期，数据源I向$buffer_0$输入数据，并且该输入信号的使能信号有效，因为此时$FIFO_0$为空，所以直接将I作为$heap_0$的DI_0；第二个周期，$heap_0$向$buffer_1$输出写回数据WB_0，数据源I向$buffer_1$输入数据，并且该输入信号的使能信号有效，由于$FIFO_1$为空，直接将I作为$heap_1$的输入DI_1；第三个周期，$heap_1$向$buffer_0$输出写回数据WB_1，因为此时$FIFO_0$非满，直接将I作为$heap_0$的DI_0；第四个周期，$heap_0$向$buffer_1$输出写回数据WB_0，由于$FIFO_1$非满，直接将I作为$heap_1$的输入DI_1。如此输入数据多个周期，直到$buffer_0$中的$FIFO_0$被填满，控制逻辑C_0根据$FIFO_0$的输出WB_1'和$heap_0$堆顶T_0进行判断，如果WB_1'有效且小于T_0，则控制多选器MUX_0的选择信号S_0选择WB_1'输入到$heap_0$，否则，判断I的使能I_en是否有效，如果有效选择I输入到$heap_0$，否则读取非空的$FIFO_0$中缓存的其他写回数据WB_1'，将该WB_1'输入到$heap_0$，或者将该WB_1'输入到多选器MUX_0。之后两个缓冲器根据相应的$FIFO$输出和数据源输入I，重复以上操作，向堆中输入数据。当所有双堆结构的外部数据输入完毕，并且所有$FIFO$被清空，排序完成。

因为对于最大堆来说，一个堆的写回数据，其大小必定大于当前的堆顶，如果写回值大于另外一个堆的堆顶，必然不会进入到另外一个堆中，通过简单的控制逻辑，完成堆两个之间的数据传递，并且共用一个数据输入，其接口与单个堆结构相同。相比于多堆结构 400，双堆结构 900 的硬件结构更简单，吞吐量近似，所消耗的资源更少，并且也可以作为一个单输入的堆来使用。

图 9 至图 12 示出的实施例是本申请提供的一种实现方式，仅仅是为了更加清楚的说明本申请的技术方案，并不对本申请的其他实施例构成限定。实施本公开的上述实施例中的针对双周期单输入交替的双堆结构，两个堆交替工作，并且相互写回数据，通过两个缓冲器以及相应的控制逻辑使得两个堆结构能够交替从一个数据输入以及另一个堆的写回数据中得到数据输入，保持堆的数据排布有序的性质，以较低的处理资源开销提高吞吐量到了接近传统堆结构的上限，并且具有与传统的堆硬件结构相同的功能抽象以及接口，能够轻易的替换到已有的堆结构硬件设计中。

下面介绍本申请的实施例提供的电子设备 1300 的结构。

本申请实施例对电子设备 1300 的具体类型不做限定。示例性的，实施本公开的一个或多个实施例的电子设备 1300 可以是智能手机（mobile phone）、平板电脑（Pad）、个人计算机（personal computer，PC）、台式电脑、便携电脑、桌面型计算机、膝上型计算机、手持计算机、笔记本电脑、上网本、游戏机、电子会议白板、智慧屏（智能电视）、无人机、蜂窝电话、个人数字助理（personal digital assistant，PDA）、增强现实（augmented reality，AR）设备、虚拟现实（virtual reality，VR）设备、人工智能（artificial intelligence，AI）设备、可穿戴式设备、车机（车载设备）、物联网（internet of things，IOT）设备、以及工业控制（industrial control）中的设备、无人驾驶（self driving）中的设备、远程医疗（remote medical）中的设备、智能电网（smart grid）中的设备、运输安全（transportation safety）中的设备、智慧城市（smart city）中的设备、智慧家庭（smart home）中的设备或者其他类型的电子设备等等。电子设备 1300 可以搭载各种类型的操作系统，本实施例对此不作限制。

图 13 示例性示出了本申请实施例提供的电子设备 1300 的结构。

如图 13 所示，电子设备 1300 可包括：一个或多个终端设备处理器 1310、通信接口 1320、存储器 1330 等，这些部件可通过总线 1340 或者其他方式连接，以通过总线 1340 连接为例。其中：

处理器 1310 是电子设备 1300 的控制中心，通过各种接口和线路连接电子设备 1300 的各个部分，可

用于读取和执行计算机可读指令。具体的，处理器 1310 可用于调用存储于存储器 1330 中的程序或数据，并执行该程序包含的指令，实现电子设备 1300 的各种功能。在一些实施例中，处理器 1310 可包括一个或多个处理单元。处理器 1310 还可以集成应用处理器和调制解调处理器。其中，应用处理器主要用于处理操作系统、用户界面和应用程序等；调制解调处理器主要用于处理无线通信。可以理解的是，上述调制解调处理器也可以不集成到处理器 1310 中。在本申请的一些实施例中，处理器可用于执行本申请上述任一实施例中提供的方法和功能。

存储器 1330 与处理器 1310 耦合，用于存储各种软件程序和/或多组指令。具体地，存储器 1330 可包括高速随机存取的存储器，并且也可包括非易失性存储器，例如一个或多个磁盘存储设备、闪存设备或其他非易失性固态存储设备。存储器 1330 可以存储操作系统，例如嵌入式操作系统。存储器 1330 还可以存储网络通信程序，该网络通信程序可用于与一个或多个附加设备，一个或多个终端设备，一个或多个网络设备进行通信。存储器 1330 还可以存储终端设备接口程序，该终端设备接口程序可以通过图形化的操作界面将应用程序的内容形象逼真的显示出来，并通过菜单、对话框以及按键等输入控件接收终端设备对应用程序的控制操作。在本申请的一些实施例中，存储器 1330 可用于存储本申请的一个或多个实施例提供的方法在电子设备 1300 侧的实现程序。

电子设备 1300 还可包括输入输出模块。输入输出模块可用于实现电子设备 1300 和其他终端设备/外部环境之间的交互，可主要包括音频输入输出模块、按键输入模块以及显示器等。具体地，输入输出模块还可包括：摄像头、触摸屏以及传感器等等。其中，输入输出模块均通过接口与处理器 1310 进行通信。

需要说明的是，图 13 所示的电子设备 1300 仅仅是本申请实施例的一种实现方式，实际应用中，电子设备 1300 还可以包括更多或更少的部件，可以组合两个或多个的部件，或者组合某些部件，或者拆分某些部件，或者不同的部件布置。图示的部件可以以硬件、软件或软件和硬件的组合实现，在此不作限定。

以上实施例仅仅是为了更加清楚的说明本申请的技术方案，并不构成对本申请的限定。在其他实施例中还可以包括更多或更少的流程或步骤，或者各实施例进行组合或拆分，得到更多的实现方式，本申请实施例对此不作限制。

下面介绍本申请实施例提供的一种信号处理装置。该信号处理装置可包括：输入模块，被配置为为向多排序模块输入多个信号，多排序模块包括环形串联的 N 个排序模块，N 个排序模块中的每个排序模块包括具有 K 个元素的 Top-K 模块，多排序模块包括第一排序模块和第二排序模块，第一排序模块包括第一 Top-K 模块，第一排序模块输出的被第一 Top-K 模块舍弃的信号或被替换出的第一 Top-K 模块的顶端元素作为第二排序模块的输入信号。处理模块，被配置为基于多个信号，更新多排序模块中多个 Top-K 模块中的元素。以及输出模块，被配置为获得多个信号中数据排序靠前的 M 个数据，其中 M 的值为 N 与 K 的乘积值，N、K、M 为正整数。

实施本实施例的信号处理装置，能够改进系统性能，提高信号处理效率，节省存储资源和处理资源，减少数据存储量、数据传输量、数据处理量和处理时间。

在一些实施例中，在一些实施例中，多排序模块为多堆结构，Top-K 模块为堆，多堆结构包括多个堆，多个堆中的每个堆中包括 K 个元素，多堆结构输出 M 个数据，M 个数据为多堆结构的多个堆中的所有元素的集合。多堆结构的工作并行度可以增加到 N，实现了更高的数据吞吐量，提高信号处理效率，并且多堆结构中的每个堆的大小减小为 M/N，提高了单个堆的性能，可以节省存储资源、计算资源和处理资源，以及减少数据存储量、数据传输量、数据处理量和处理时间。

在一些实施例中，如果堆为最大堆，那么 M 个数据为多个信号中排序数据最小的前 M 个数据，最大堆的堆顶元素为最大堆中数据最大的元素；如果堆为最小堆，那么 M 个数据为多个信号中排序数据最大的前 M 个数据，最小堆的堆顶元素为最小堆中数据最小的元素。最大堆有助于筛选最小值，最小堆有助于筛选最大值。

在一些实施例中，第一排序模块中包括第一缓冲器和第一 Top-K 模块，第二排序模块中包括第二缓冲器和第二 Top-K 模块，第一排序模块的第一缓冲器的输出作为第一排序模块的第一 Top-K 模块的输入，第一排序模块的第一 Top-K 模块的输出作为第二排序模块的第二缓冲器的输入，第二排序模块的第二缓冲器的输出作为第二排序模块的第二 Top-K 模块的输入。各个排序模块之间的缓冲器与 Top-K 模块交替连接以及多个排序模块环形串联的结构可以提高处理效率。例如，该多排序模块的具体结构可以结合参考图 4 所示的示例多堆结构 400 或者图 9 所示的双堆结构 900，单个排序模块的具体结构可以结合参考图 5 所示的示例排序模块 500 或者图 10 所示的排序模块 1000，这里不再赘述。

在一些实施例中，多个信号包括第一外部信号和第二外部信号，第一外部信号为第一排序模块的输入，第二外部信号为第二排序模块的输入，并且第一排序模块用于接收第一外部信号，第一排序模块还用于基于第一外部信号获得第一写回信号，第一排序模块还用于向第二排序模块发送第一写回信号。第二排序模块用于接收第二外部信号，第二排序模块还用于接收第一写回信号，第二排序模块还用于基于第二外部信号和/或第一写回信号获得第二写回信号，第二排序模块还用于向第三排序模块发送第二写回信号。第一排序模块与第二排序模块串联可以提高处理效率，第一排序模块不需要的数据可以由第二排序模块继续处理，避免重复输入和处理。

在一些实施例中，第一缓冲器中包括第一先入先出队列、第一控制逻辑模块和第一多选器，第一多选器用于接收第一外部信号，第一先入先出队列用于多次接收并缓存多个写回信号，第一先入先出队列还用于在被多个写回信号填满后输出第一缓存信号，以及第二缓冲器中包括第二先入先出队列、第二控制逻辑模块和第二多选器，第二多选器用于接收第二外部信号，第二先入先出队列用于多次接收并缓存第一排序模块的第一 Top-K 模块输出的多个写回信号，多个写回信号包括第一写回信号，第二先入先出队列还用于在被多个写回信号填满后输出第二缓存信号。控制逻辑和多选器用于基于先入先出队列输出的缓存信号、输入的外部信号和全局堆顶最值来选择进入堆的输入信号。先入先出队列用于缓存上个排序模块输出的写回数据。设置该先入先出队列缓存写回数据的目的是由于全局最值模块输出的堆顶值相对输入信号会存在几个周期的延迟，于是写回信号输入到缓冲器中的先入先出队列后，直到先入先出队列填充满才被读取，这时的缓存信号为多个周期前的输入，相对全局堆顶最值会延后，有利于控制逻辑能够发挥出较好筛选性能，避免写回信号在多个堆中传递而不会被筛选掉。

在一些实施例中，第二控制逻辑模块还用于基于第二外部信号和/或第二缓存信号确定第二输入信号，第二多选器还用于向第二 Top-K 模块输出第二输入信号，第二 Top-K 模块用于接收来自于第二多选器的第二输入信号，第二 Top-K 模块还用于基于第二输入信号获得第二写回信号。以及第二 Top-K 模块还用于向第三排序模块的第三缓冲器发送第二写回信号。Top-K 模块与缓冲器交替相连，可以提高信号处理效率。

在一些实施例中，第二控制逻辑模块用于执行以下判断逻辑：判断第二先入先出队列是否输出第二缓存信号。如果第二先入先出队列输出了第二缓存信号，那么判断第二缓存信号是否有效。如果第二缓存信号有效，则将第二缓存信号的数据与全局顶端最值进行比较。如果第二缓存信号的数据未超过全局顶端最值，则第二控制逻辑模块确定第二缓存信号作为第二输入信号，如果第二缓存信号的数据超过或等于全局顶端最值，则第二控制逻辑模块确定第二外部信号作为第二输入信号，并且第二缓存信号被舍弃。以及如果第二缓存信号无效，则第二控制逻辑模块确定第二外部信号作为第二输入信号，并且第二缓存信号被舍弃。通过多个堆之间传递数据以及全局最值（最大值或最小值）的约束，可以减少排序时所需比较的次数，节省处理资源的消耗。此外，基于该多堆结构可以较好的利用了高速缓存器等片上存储资源，提高了不同片上资源的利用效率。例如，在图 5 中，单堆模块的堆顶最大值输出T_k，表示本排序模块中当前堆结构的堆顶。信号T_g表示从 N 个堆顶输入获取的最值，即全局最值（又可称为全局堆顶最值）。

在一些实施例中，第二控制逻辑模块还用于执行以下判断逻辑：如果第二先入先出队列没有输出第二缓存信号，那么判断第二外部信号是否有效。如果第二外部信号有效，则第二控制逻辑模块确定第二外部信号作为第二输入信号。以及如果第二外部信号无效，则第二控制逻辑模块读取非空的第二先入先出队列中缓存的第三缓存信号作为第二输入信号。即先入先出队列中溢出的缓存信号优先于外部信号被处理。当先入先出队列中没有溢出的缓存信号时，将有效的外部信号作为堆的输入信号，如果外部信号无效，为了保证 FIFO 中的所有数据都进入排序，避免堆模块闲置，可以读取非空的先入先出队列中未溢出的缓存的写回信号作为堆的输入信号。关于第二控制逻辑模块执行的判断逻辑可以结合参考图 6 所示的控制逻辑的工作流程 600 或者图 11 所示的控制逻辑的工作流程 1100 的实施例进行理解，此处不再赘述。

在一些实施例中，第一写回信号包括第一写回信号的数据和第一写回使能信号，第一写回使能信号指示第一写回信号是否有效。第二外部信号包括第二外部信号的数据和第二外部使能信号，第二外部使能信号指示第二外部信号是否有效。第二缓存信号包括第二缓存信号的数据和第二缓存使能信号，第二缓存使能信号指示第二缓存信号是否有效。使用写回使能信号和外部使能信号可以更方便地控制信号的输入。

在一些实施例中，Top-K 模块为最大堆，全局顶端最值为多个最大堆的堆顶元素的数据之中的最大值，第二缓存信号的数据未超过全局顶端最值，包括：第二缓存信号的数据小于最大值。第二缓存信号的数据超过或等于全局顶端最值，包括：第二缓存信号的数据大于或等于最大值。最大堆的结构能够帮助求出多个输入数据中预定数目的最小值。在一些实施例中，Top-K 模块为最小堆，全局顶端最值为多个最小堆的

堆顶元素的数据之中的最小值，第二缓存信号的数据未超过全局顶端最值，包括：第二缓存信号的数据大于最小值。第二缓存信号的数据超过或等于全局顶端最值，包括：第二缓存信号的数据小于或等于最小值。最小堆的结构能够帮助求出多个输入数据中预定数目的最大值。

在一些实施例中，第二 Top-K 模块用于执行以下判断逻辑：第二 Top-K 模块基于第二输入信号获得第二写回信号，包括：判断第二输入信号是否有效，如果第二输入信号有效，则将第二输入信号的数据与第二 Top-K 模块的当前的顶端元素的数据进行比较。如果第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，则将第二输入信号的数据替换掉第二 Top-K 模块的当前顶端元素的数据以成为第二 Top-K 模块中新的元素，以及将被替换出的第二 Top-K 模块的当前顶端元素的数据作为第二写回信号的数据，并且使能第二写回信号有效。以及如果第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，则将第二输入信号的数据作为第二写回信号的数据，并且使能第二写回信号有效。通过利用 Top-K 模块内数据的有序性以及 Top-K 模块顶端数据的约束，将本 Top-K 模块不需要的数据传递给下一个排序模块处理或者及时丢弃，可以减少排序时所需比较的次数，节省处理资源和存储资源。在一些实施例中，第二 Top-K 模块还用于执行以下判断逻辑：如果第二输入信号无效，则将第二输入信号的数据作为第二写回信号的数据，并且使能第二写回信号无效。被确定无用的数据可以发送给下一个排序模块的 FIFO，由下一个排序模块的控制逻辑来决策丢弃该数据。

在一些实施例中，第二 Top-K 模块还用于更新第二 Top-K 模块的新的顶端元素，例如第二 Top-K 模块可以是堆结构，该堆结构可以执行堆化操作以更新堆结构的新的堆顶元素，堆化操作指的是对堆中的元素进行排序更新。第二 Top-K 模块还用于将第二 Top-K 模块的顶端元素发送给全局最值模块。全局最值模块用于获取多个 Top-K 模块的顶端元素，全局最值模块还用于选出多个 Top-K 模块的顶端元素的数据之中的最值，全局最值模块还用于将最值发送给多个缓冲器中的控制逻辑模块。由于被替换进入 Top-K 模块的元素不一定是 Top-K 模块内最值，所以在下一次输入信号输入堆之前，Top-K 模块需要更新出新的顶端元素，以保证下次处理的准确性。关于第二 Top-K 模块的结构以及执行的判断逻辑可以结合参考图 7-图 8 所示的堆模块 700 以及堆控制逻辑的工作流程 800 或者图 12 所示的控制逻辑的工作流程 1200 的实施例进行理解，此处不再赘述。

在一些实施例中，第二 Top-K 模块为最大堆，第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据小于最大堆的当前堆顶元素的数据。第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据大于或等于最大堆的当前堆顶元素的数据。最大堆的结构能够帮助求出多个输入数据中预定数目的最小值。在一些实施例中，第二 Top-K 模块为最小堆，第二输入信号的数据未超过第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据大于最小堆的当前堆顶元素的数据。第二输入信号的数据超过或等于第二 Top-K 模块的当前顶端元素的数据，包括：第二输入信号的数据小于或等于最小堆的当前堆顶元素的数据。最小堆的结构能够帮助求出多个输入数据中预定数目的最大值。

在一些实施例中，在第一时间周期内，第二缓冲器基于第二外部信号和/或第二缓存信号确定第二输入信号，以及在第二时间周期内，第二 Top-K 模块基于第二输入信号获得第二写回信号。在一个迭代周期内，一个排序模块可以完整的完成一次处理工作，一个迭代周期可以分成两个时间周期，例如第一时间周期和第二时间周期。在一些实施例中，第一时间周期与第二时间周期的时间长度相等。缓冲器判断选择进入堆的数据的处理周期与堆处理输入数据并输出写回数据的处理周期交替地执行，每个排序模块可以并行处理数据，这能够提高处理效率，缩减等待时间和处理时间，充分利用处理资源。

需要说明的是，以上实施例描述的信号处理装置仅仅是为了更加清楚的说明本申请的技术方案的一种实现方式，并不构成对本申请其他实施例的限定，实际应用中，不限定该信号处理装置的各个功能模块的类型、数量和关联关系。在一些实施例中，该信号处理装置中可以包括更多或更少的功能模块等。该信号处理装置中的各个功能模块可以通过软件、硬件或者两者结合的方式实现。可以理解的，关于该信号处理装置包括的各个功能单元的具体实现可参考本申请其他各实施例中的详细描述，这里不多赘述。

需要说明的是，本申请实施例中对模块或单元的划分是示意性的，仅仅为一种逻辑功能划分，实际实现时也可以有另外的划分方式。另外，在实施例中的各功能单元可以集成在一个单元中，也可以是单独物理存在，也可以两个或两个以上单元集成为一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现，或者软硬件结合的形式实现。

本申请的实施例还提供了一种芯片，该芯片可以包括输入接口、输出接口和处理电路。在本公开的实

施例中，可以由输入接口和输出接口完成信令或数据的交互，由处理电路完成信令或数据信息的生成以及处理。

本申请的实施例还提供了一种芯片系统，包括处理器，用于支持计算设备以实现上述任一实施例中所涉及的功能。在一种可能的设计中，芯片系统还可以包括存储器，用于存储必要的程序指令和数据，当处理器运行该程序指令时，使得安装该芯片系统的设备实现上述任一实施例中所涉及的方法。示例性地，该芯片系统可以由一个或多个芯片构成，也可以包含芯片和其他分立器件。

本申请的实施例还提供了一种处理器，用于与存储器耦合，存储器存储有指令，当处理器运行所述指令时，使得处理器执行上述任一实施例中涉及的方法和功能。

本申请的实施例还提供了一种计算机可读存储介质，其上存储有计算机指令或程序代码，当处理器运行所述指令或所述程序代码时，使得处理器执行上述任一实施例中涉及的方法和功能。计算机可读介质可以是包含或存储用于或有关于指令执行系统、装置或设备的程序的任何有形介质。计算机可读介质可以是计算机可读信号介质或计算机可读存储介质。计算机可读介质可以包括但不限于电子的、磁的、光学的、电磁的、红外的或半导体系统、装置或设备，或其任意合适的组合。该计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。计算机可读存储介质的更详细示例包括带有一根或多根导线的电气连接、磁性介质（例如，磁盘、软盘、硬盘、磁带、磁存储设备）、光介质（例如光存储设备、DVD）、半导体介质（例如固态硬盘）、随机存储存取器（RAM）、只读存储器（ROM）、可擦除可编程只读存储器（EPROM 或闪存），或其任意合适的组合等。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。本申请的实施例还提供了有形地存储在非暂时性计算机可读存储介质上的至少一个计算机程序产品。该计算机程序产品包括一个或多个计算机可执行指令，例如包括在程序模块中的指令，其在目标的真实或虚拟处理器上的设备中执行，以执行上述各实施例中任一实施例中涉及的过程、方法和功能。在计算机上加载和执行该计算机程序指令时，全部或部分地产生按照本申请实施例的流程或功能。该计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。该计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，该计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线）或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。

根据本申请实施例提供的方法，本申请还提供一种计算机程序产品，该计算机程序产品包括：计算机程序代码，当该计算机程序代码在计算机上运行时，使得该计算机执行上述实施例中的过程、方法和功能。通常，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、库、对象、类、组件、数据结构等。在各种实施例中，可以根据需要在程序模块之间组合或分割程序模块的功能。用于程序模块的机器可执行指令可以在本地或分布式设备内执行。在分布式设备中，程序模块可以位于本地和远程存储介质中。

用于实现本公开的方法的计算机程序代码可以用一种或多种编程语言编写。这些计算机程序代码可以提供给通用计算机、专用计算机或其他可编程的数据处理装置的处理器，使得程序代码在被计算机或其他可编程的数据处理装置执行的时候，引起在流程图和/或框图中规定的功能/操作被实施。程序代码可以完全在计算机上、部分在计算机上、作为独立的软件包、部分在计算机上且部分在远程计算机上或完全在远程计算机或服务器上执行。在本公开的上下文中，计算机程序代码或者相关数据可以由任意适当载体承载，以使得设备、装置或者处理器能够执行上文描述的各种处理和操作。载体的示例包括信号、计算机可读介质、等等。信号的示例可以包括电、光、无线电、声音或其它形式的传播信号，诸如载波、红外信号等。

通常，本申请的各种实施例可以以硬件或专用电路、软件、逻辑或其任何组合来实现。一些方面可以用硬件实现，而其他方面可以用固件或软件实现，其可以由控制器，微处理器或其他计算设备执行。虽然本公开的实施例的各个方面被示出并描述为框图，流程图或使用一些其他图示表示，但是应当理解，本文描述的框，装置、系统、技术或方法可以实现为，如非限制性示例，硬件、软件、固件、专用电路或逻辑、通用硬件或控制器或其他计算设备，或其某种组合。

以上已经描述了本公开的各实现，上述说明是示例性的，并非穷尽的，并且也不限于所公开的各实现。在不偏离所说明的各实现的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择，旨在很好地解释各实现的原理、实际应用或对市场中的技术的改进，或者使本技术领域的其他普通技术人员能理解本文公开的各个实现方式。

以上的具体实施方式，对本申请实施例的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上仅为本申请实施例的具体实施方式而已，并不用于限定本申请实施例的保护范围，凡在本申请实施例的技术方案的基础之上，所做的任何修改、等同替换、改进等，均应包括在本申请实施例的保护范围之内。

应注意的是，尽管上面分别结合附图描述了本申请的实施例，但是上面的实施例并不是彼此独立的，它们也可以结合以得到其他的实施例。本申请的实施例中的方式、情况、类别以及实施例的划分仅是为了描述的方便，不应构成特别的限定，各种方式、类别、情况以及实施例中的特征在符合逻辑的情况下，可以相互结合。本申请的各实施方式可以任意进行组合，以实现不同的技术效果。本申请的实施例不再对各种结合进行罗列。

此外，尽管在附图中以特定顺序描述了本公开的方法的操作，但是这并非要求或者暗示必须按照该特定顺序来执行这些操作，或是必须执行全部所示的操作才能实现期望的结果。相反，流程图中描绘的步骤可以改变执行顺序。附加地或备选地，可以省略某些步骤，将多个步骤组合为一个步骤执行，和/或将一个步骤分解为多个步骤执行。还应当注意，根据本公开的两个或更多装置的特征和功能可以在一个装置中具体化。反之，上文描述的一个装置的特征和功能可以进一步划分为由多个装置来具体化。

上述实施例中所用，根据上下文，术语“当…时”可以被解释为意思是“如果…”或“在…后”或“响应于确定…”或“响应于检测到…”。类似地，根据上下文，短语“在确定…时”或“如果检测到（所陈述的条件或事件）”可以被解释为意思是“如果确定…”或“响应于确定…”或“在检测到（所陈述的条件或事件）时”或“响应于检测到（所陈述的条件或事件）”。

权利要求书

1. 一种信号处理方法，其特征在于，所述方法包括：

向多排序模块输入多个信号，所述多排序模块包括环形串联的 N 个排序模块，所述 N 个排序模块中的每个排序模块包括具有 K 个元素的 Top-K 模块，所述多排序模块包括第一排序模块和第二排序模块，所述第一排序模块包括第一 Top-K 模块，所述第一排序模块输出的被所述第一 Top-K 模块舍弃的信号或被替换出的所述第一 Top-K 模块的顶端元素作为所述第二排序模块的输入信号；

基于所述多个信号，更新所述多排序模块中多个所述 Top-K 模块中的元素；以及

获得所述多个信号中数据排序靠前的 M 个数据，其中 M 的值为 N 与 K 的乘积值，N、K、M 为正整数。

2. 根据权利要求 1 所述的方法，其特征在于，所述第一排序模块中包括第一缓冲器和所述第一 Top-K 模块，所述第二排序模块中包括第二缓冲器和第二 Top-K 模块，所述第一排序模块的所述第一缓冲器的输出作为所述第一排序模块的所述第一 Top-K 模块的输入，所述第一排序模块的所述第一 Top-K 模块的输出作为所述第二排序模块的所述第二缓冲器的输入，所述第二排序模块的所述第二缓冲器的输出作为所述第二排序模块的所述第二 Top-K 模块的输入。

3. 根据权利要求 2 所述的方法，其特征在于，所述多个信号包括第二外部信号，所述第二外部信号为所述第二排序模块的外部输入信号，并且所述方法还包括：

所述第二排序模块接收第二外部信号；

所述第二排序模块接收来自于所述第一排序模块的所述第一写回信号；

所述第二排序模块根据包括所述第二外部信号、所述第一写回信号和全局顶端最值中的至少一项获得第二写回信号，其中所述全局顶端最值为多个所述 Top-K 模块的顶端元素之中的最值；

所述第二排序模块向第三排序模块发送所述第二写回信号。

4. 根据权利要求 3 所述的方法，其特征在于，以及所述第二缓冲器中包括第二先入先出队列、第二控制逻辑模块和第二多选器，所述第二多选器用于接收所述第二外部信号，所述第二先入先出队列用于多次接收并缓存所述第一排序模块的所述第一 Top-K 模块输出的多个写回信号，所述多个写回信号包括所述第一写回信号，所述第二先入先出队列还用于在被所述多个写回信号填满后输出第二缓存信号。

5. 根据权利要求 4 所述的方法，其特征在于，所述第二排序模块基于所述第二外部信号和/或所述第一写回信号获得第二写回信号包括：

所述第二控制逻辑模块基于所述第二外部信号和/或所述第二缓存信号确定第二输入信号；

所述第二多选器向所述第二 Top-K 模块输出所述第二输入信号；

所述第二 Top-K 模块接收来自于所述第二多选器的所述第二输入信号；

所述第二 Top-K 模块基于所述第二输入信号获得所述第二写回信号；以及

所述第二排序模块向第三排序模块发送所述第二写回信号，包括：

所述第二 Top-K 模块向所述第三排序模块的第三缓冲器发送所述第二写回信号。

6. 根据权利要求 5 所述的方法，其特征在于，所述第二控制逻辑模块基于所述第二外部信号和/或所述第二缓存信号确定第二输入信号，包括：

判断所述第二先入先出队列是否输出所述第二缓存信号；

如果所述第二先入先出队列输出了所述第二缓存信号，那么判断所述第二缓存信号是否有效；

如果所述第二缓存信号有效，则将所述第二缓存信号的数据与所述全局顶端最值进行比较；

如果所述第二缓存信号的数据未超过所述全局顶端最值，则所述第二控制逻辑模块确定所述第二缓存信号作为所述第二输入信号；

如果所述第二缓存信号的数据超过或等于所述全局顶端最值，则所述第二控制逻辑模块确定所述第二外部信号作为所述第二输入信号，并且所述第二缓存信号被舍弃；以及

如果所述第二缓存信号无效，则所述第二控制逻辑模块确定所述第二外部信号作为所述第二输入信号，并且所述第二缓存信号被舍弃。

7. 根据权利要求 6 所述的方法，其特征在于，还包括：

如果所述第二先入先出队列没有输出所述第二缓存信号，那么判断所述第二外部信号是否有效；

如果所述第二外部信号有效，则所述第二控制逻辑模块确定所述第二外部信号作为所述第二输入信号；

以及

如果所述第二外部信号无效，则所述第二控制逻辑模块读取非空的所述第二先入先出队列中缓存的第三缓存信号作为所述第二输入信号。

8. 根据权利要求 4-7 任一项所述的方法，其特征在于，其中，所述第一写回信号包括所述第一写回信号的数据和第一写回使能信号，所述第一写回使能信号指示所述第一写回信号是否有效；所述第二外部信号包括所述第二外部信号的数据和第二外部使能信号，所述第二外部使能信号指示所述第二外部信号是否有效；所述第二缓存信号包括所述第二缓存信号的数据和第二缓存使能信号，所述第二缓存使能信号指示所述第二缓存信号是否有效。

9. 根据权利要求 6 所述的方法，其特征在于，所述 Top-K 模块为最大堆，所述全局顶端最值为多个所述最大堆的堆顶元素的数据之中的最大值，

所述第二缓存信号的数据未超过所述全局顶端最值，包括：

所述第二缓存信号的数据小于所述最大值；

所述第二缓存信号的数据超过或等于所述全局顶端最值，包括：

所述第二缓存信号的数据大于或等于所述最大值。

10. 根据权利要求 6 所述的方法，其特征在于，所述 Top-K 模块为最小堆，所述全局顶端最值为多个所述最小堆的堆顶元素的数据之中的最小值，

所述第二缓存信号的数据未超过所述全局顶端最值，包括：

所述第二缓存信号的数据大于所述最小值；

所述第二缓存信号的数据超过或等于所述全局顶端最值，包括：

所述第二缓存信号的数据小于或等于所述最小值。

11. 根据权利要求 5 所述的方法，其特征在于，所述第二 Top-K 模块基于所述第二输入信号获得所述第二写回信号，包括：

判断所述第二输入信号是否有效；

如果所述第二输入信号有效，则将所述第二输入信号的数据与所述第二 Top-K 模块的当前的顶端元素的数据进行比较；

如果所述第二输入信号的数据未超过所述第二 Top-K 模块的当前顶端元素的数据，则将所述第二输入信号的数据替换掉所述第二 Top-K 模块的当前顶端元素的数据以成为所述第二 Top-K 模块中新的元素，以及将被替换出的所述第二 Top-K 模块的当前顶端元素的数据作为所述第二写回信号的数据，并且使能所述第二写回信号有效；以及

如果所述第二输入信号的数据超过或等于所述第二 Top-K 模块的当前顶端元素的数据，则将所述第二输入信号的数据作为所述第二写回信号的数据，并且使能所述第二写回信号有效。

12. 根据权利要求 11 所述的方法，其特征在于，还包括：

如果所述第二输入信号无效，则将所述第二输入信号的数据作为所述第二写回信号的数据，并且使能所述第二写回信号无效。

13. 根据权利要求 11 或 12 所述的方法，其特征在于，所述方法还包括：

所述第二 Top-K 模块更新所述第二 Top-K 模块的新的顶端元素；以及

所述第二 Top-K 模块将所述第二 Top-K 模块的顶端元素发送给全局最值模块，所述全局最值模块用于获取所述多个 Top-K 模块的顶端元素，所述全局最值模块还用于选出所述多个 Top-K 模块的顶端元素的数据之中的最值，所述全局最值模块还用于将所述最值发送给多个缓冲器中的控制逻辑模块。

14. 根据权利要求 11 所述的方法，其特征在于，所述第二 Top-K 模块为最大堆，

所述第二输入信号的数据未超过所述第二 Top-K 模块的当前顶端元素的数据，包括：

所述第二输入信号的数据小于所述最大堆的当前堆顶元素的数据；

所述第二输入信号的数据超过或等于所述第二 Top-K 模块的当前顶端元素的数据，包括：

所述第二输入信号的数据大于或等于所述最大堆的当前堆顶元素的数据。

15. 根据权利要求 11 所述的方法，其特征在于，所述第二 Top-K 模块为最小堆，

所述第二输入信号的数据未超过所述第二 Top-K 模块的当前顶端元素的数据，包括：

所述第二输入信号的数据大于所述最小堆的当前堆顶元素的数据；

所述第二输入信号的数据超过或等于所述第二 Top-K 模块的当前顶端元素的数据，包括：

所述第二输入信号的数据小于或等于所述最小堆的当前堆顶元素的数据。

16. 根据权利要求 5 所述的方法，其特征在于，包括:

在第一时间周期内，所述第二缓冲器基于所述第二外部信号和/或所述第二缓存信号确定第二输入信号；以及

在第二时间周期内，所述第二 Top-K 模块基于所述第二输入信号获得所述第二写回信号。

17. 根据权利要求 1-16 中任一项所述的方法，其特征在于，所述多排序模块为多堆结构，所述 Top-K 模块为堆，所述多堆结构包括多个所述堆，多个所述堆中的每个所述堆中包括 K 个元素，所述多堆结构输出所述 M 个数据，所述 M 个数据为所述多堆结构的多个所述堆中的所有元素的集合。

18. 根据权利要求 17 所述的方法，其特征在于，如果所述堆为最大堆，那么所述 M 个数据为所述多个信号中排序数据最小的前 M 个数据，所述最大堆的堆顶元素为所述最大堆中数据最大的元素；如果所述堆为最小堆，那么所述 M 个数据为所述多个信号中排序数据最大的前 M 个数据，所述最小堆的堆顶元素为所述最小堆中数据最小的元素。

19. 一种信号处理装置，其特征在于，包括:

输入模块，被配置为向多排序模块输入多个信号，所述多排序模块包括环形串联的多个排序模块，所述多个排序模块中的每个排序模块包括 Top-K 模块，所述多排序模块包括第一排序模块和第二排序模块，所述第一排序模块的输出作为所述第二排序模块的输入；

处理模块，被配置为基于所述多个信号，更新多个所述 Top-K 模块中的元素；以及

输出模块，被配置为输出所述多排序模块中多个所述 Top-K 模块中的元素。

20. 一种电子设备，其特征在于，所述电子设备包括存储器以及耦合于所述存储器的处理器，所述存储器中存储有可执行指令，所述处理器用于调用所述可执行指令，使得所述电子设备执行根据权利要求 1 至 18 中任一项所述的方法。

21. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有计算机可执行指令，所述计算机可执行指令被处理器执行时实现根据权利要求 1 至 18 中任一项所述的方法。

22. 一种芯片，包括处理电路，被配置为执行根据权利要求 1 至 18 中任一项所述的方法。

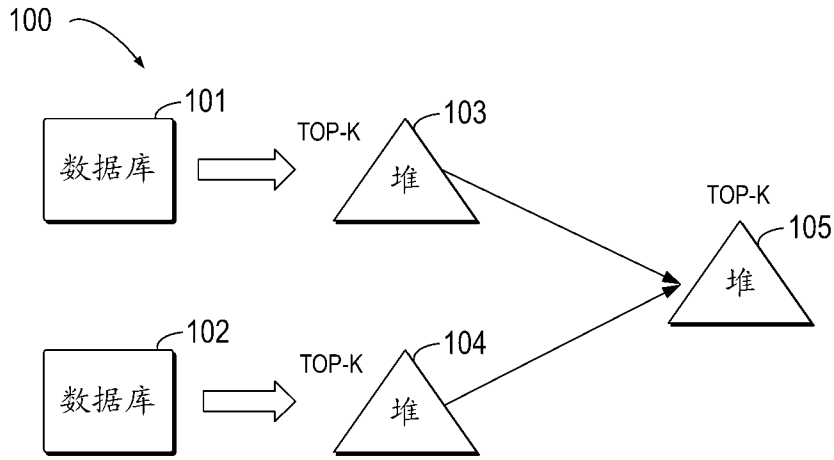

100
101
数据库
TOP-K
103
堆
TOP-K
105
堆
102
数据库
TOP-K
104
堆

图 1

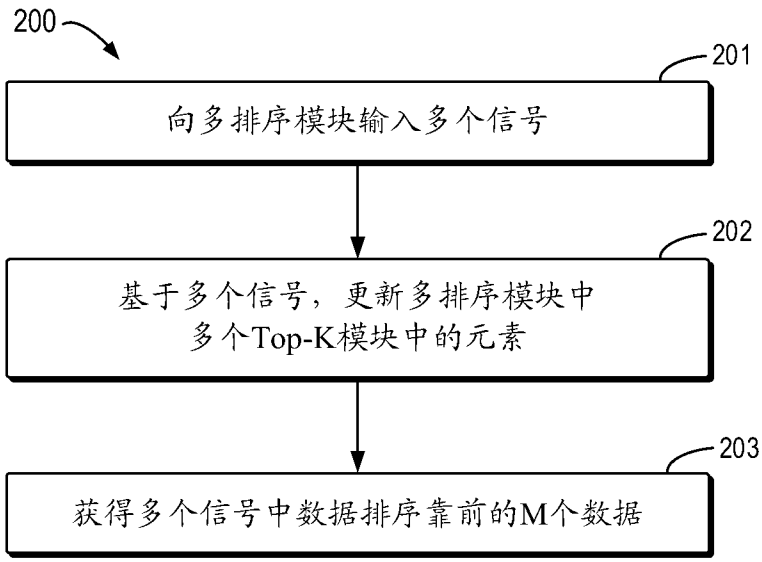

200
201
向多排序模块输入多个信号
202
基于多个信号，更新多排序模块中
多个Top-K模块中的元素
203
获得多个信号中数据排序靠前的M个数据

图 2

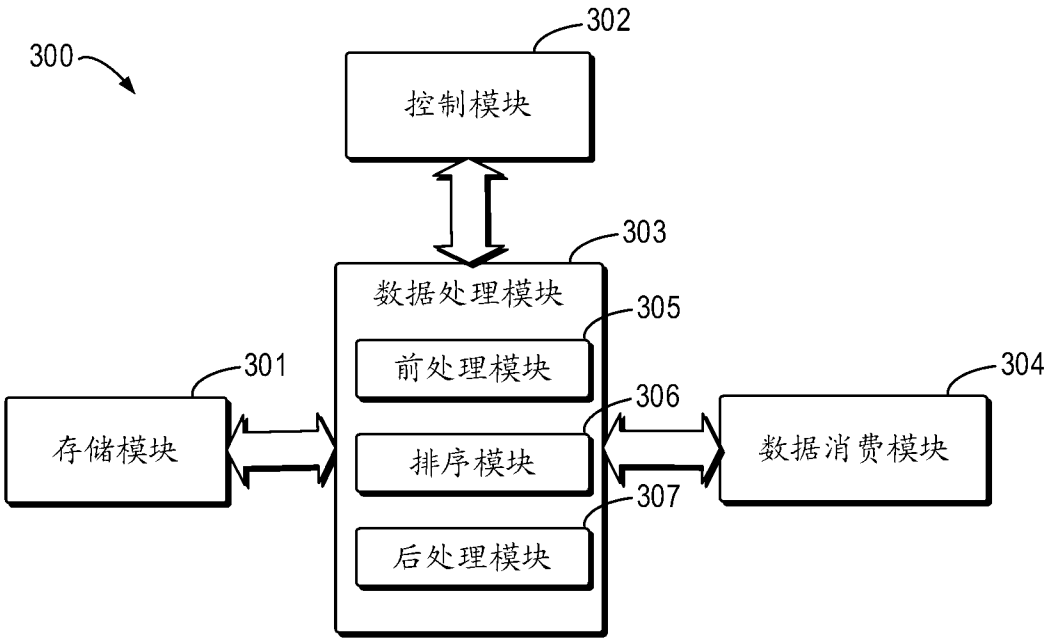

300
302
控制模块
303
数据处理模块
305
前处理模块
301
存储模块
306
排序模块
304
数据消费模块
307
后处理模块

图 3

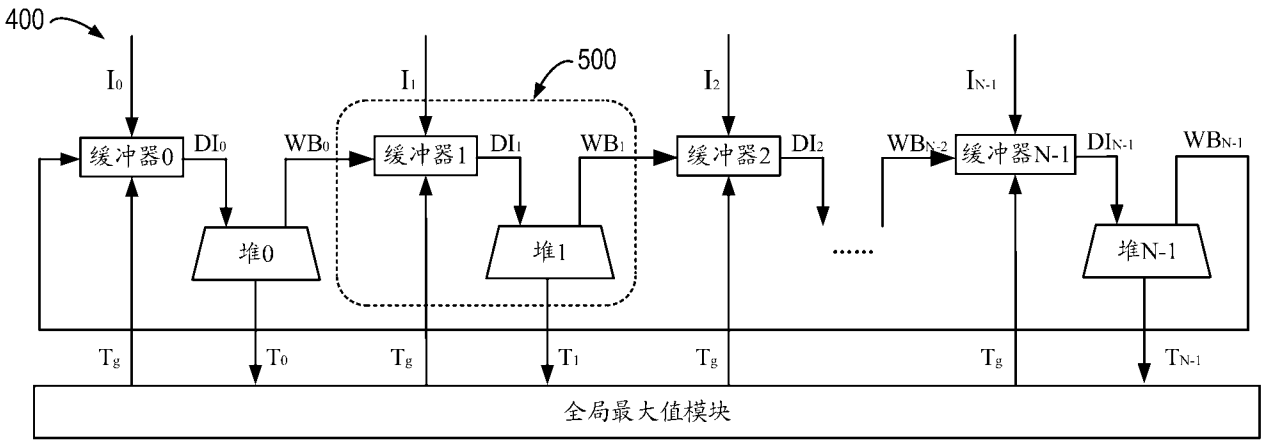

400
500
I_0
I_1
I_2
I_{N-1}
缓冲器0
缓冲器1
缓冲器2
缓冲器N-1
DI_0
DI_1
DI_2
DI_{N-1}
WB_0
WB_1
WB_{N-2}
WB_{N-1}
堆0
堆1
堆N-1
.....
T_g
T_0
T_1
T_{N-1}
全局最大值模块

图 4

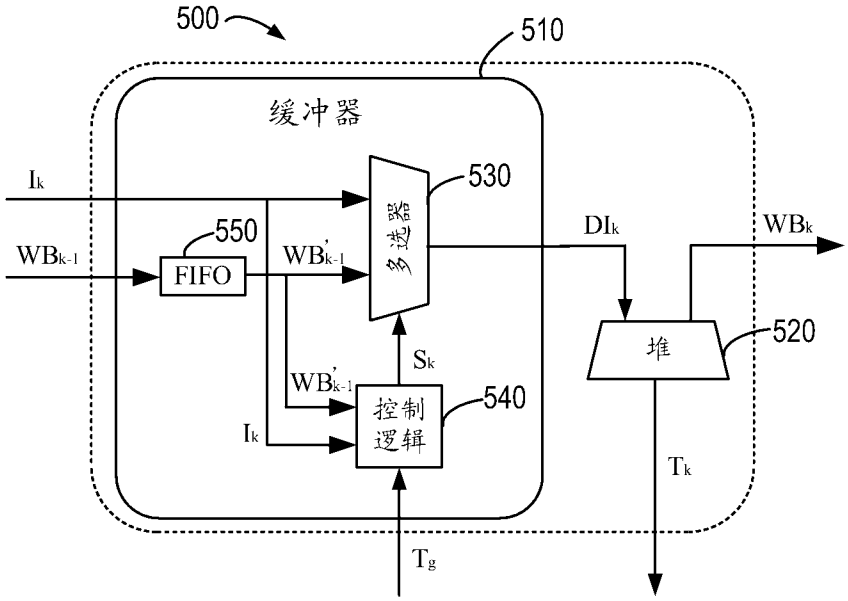

500
510
缓冲器
Ik
WBk-1
550
FIFO
WB'k-1
530
多选器
DIk
WBk
520
堆
Sk
WB'k-1
Ik
控制逻辑
540
Tk
Tg

图 5

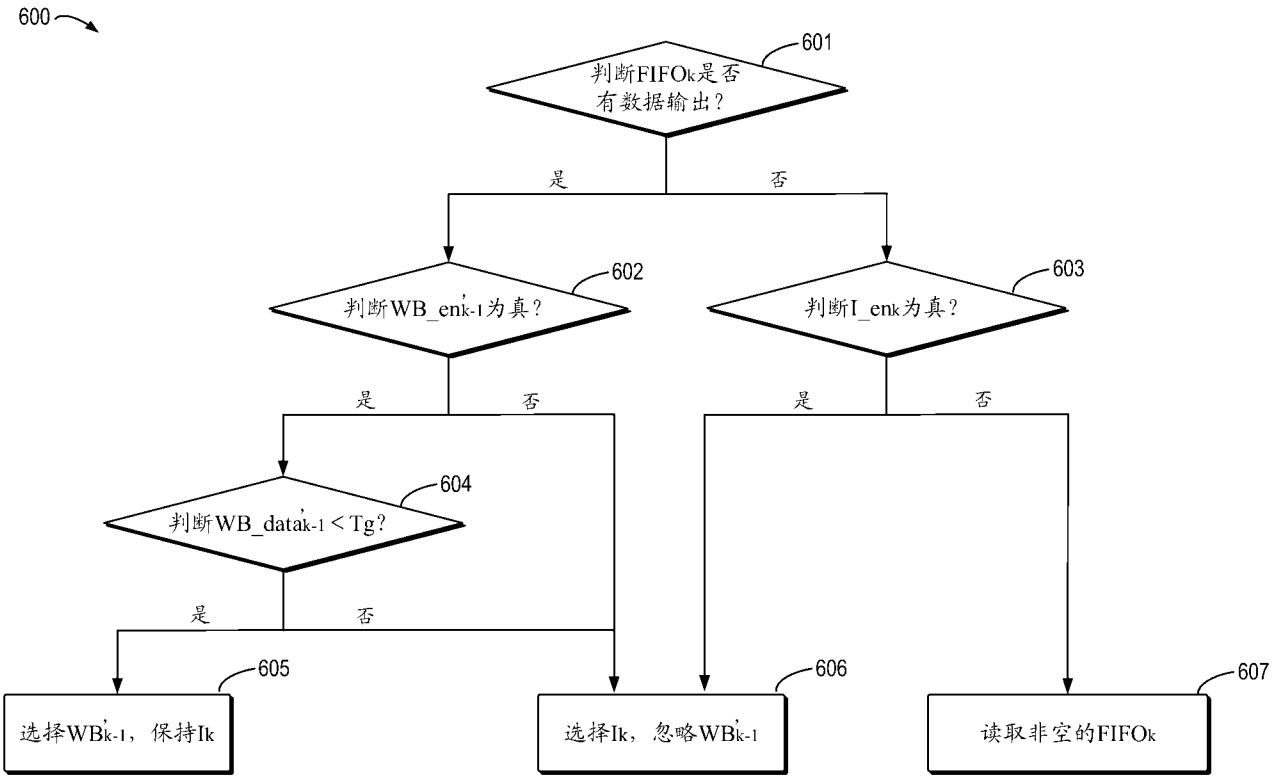

600
601
判断FIFOk是否有数据输出？
是
否
602
判断WB_en'k-1为真？
603
判断I_enk为真？
是
否
是
否
604
判断WB_data'k-1 < Tg?
是
否
605
选择WB'k-1，保持Ik
606
选择Ik，忽略WB'k-1
607
读取非空的FIFOk

图 6

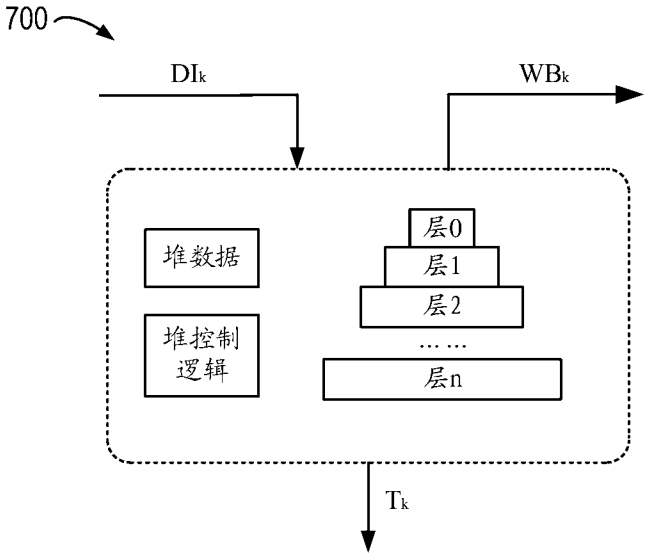

700
DIk
WBk
堆数据
堆控制逻辑
层0
层1
层2
……
层n
Tk

图 7

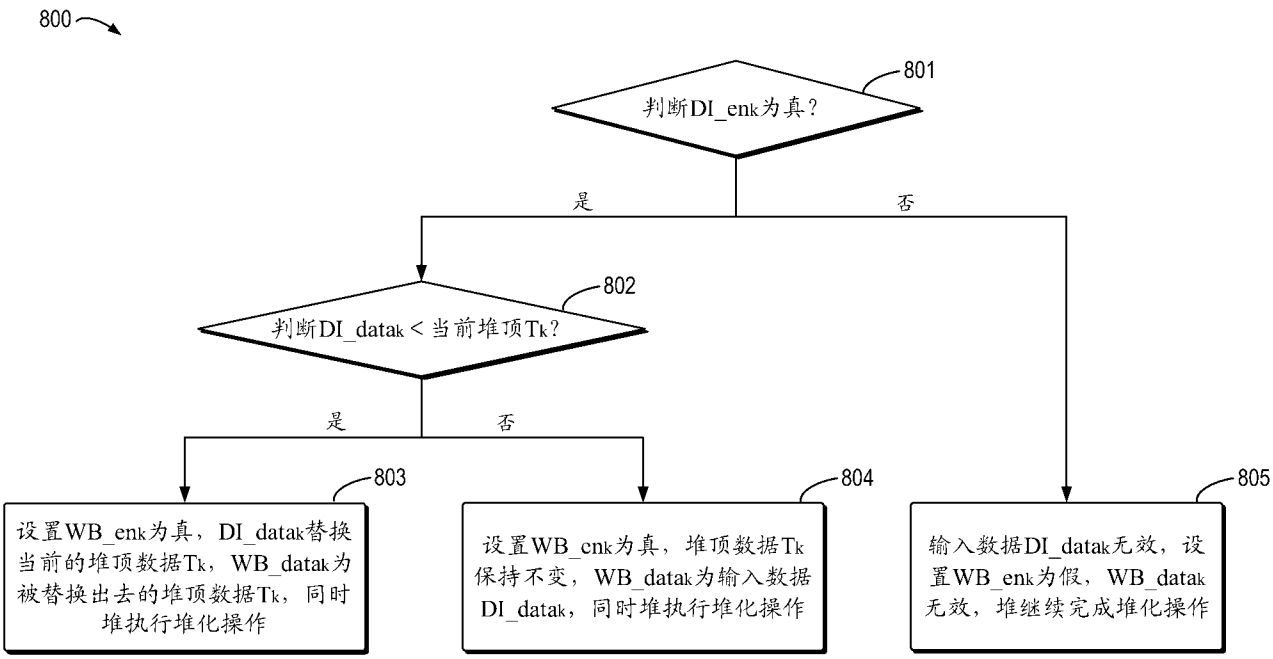

800
801
判断DI_enk为真?
是
否
802
判断DI_datak < 当前堆顶Tk?
是
否
803
设置WB_enk为真，DI_datak替换当前的堆顶数据Tk，WB_datak为被替换出去的堆顶数据Tk，同时堆执行堆化操作
804
设置WB_cnk为真，堆顶数据Tk保持不变，WB_datak为输入数据DI_datak，同时堆执行堆化操作
805
输入数据DI_datak无效，设置WB_enk为假，WB_datak无效，堆继续完成堆化操作

图 8

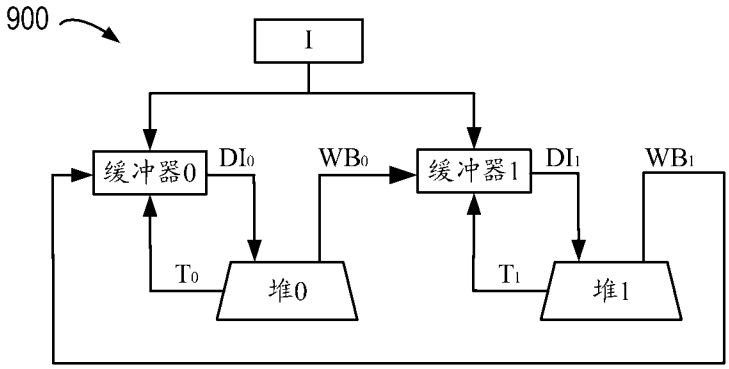

900
I
缓冲器0
DI0
WB0
缓冲器1
DI1
WB1
T0
堆0
T1
堆1

图 9

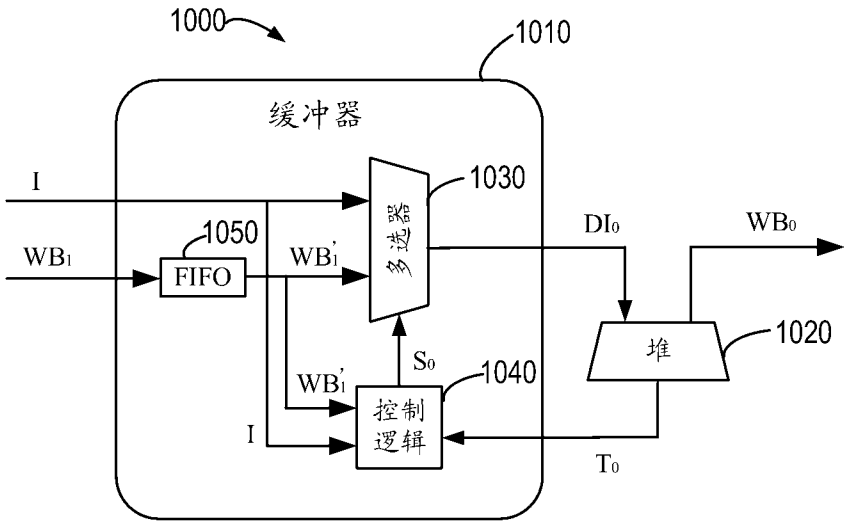

1000
1010
缓冲器
I
WB1
1050
FIFO
WB1'
1030
多选器
DI0
WB0
1020
堆
S0
WB1'
1040
控制
逻辑
I
T0

图 10

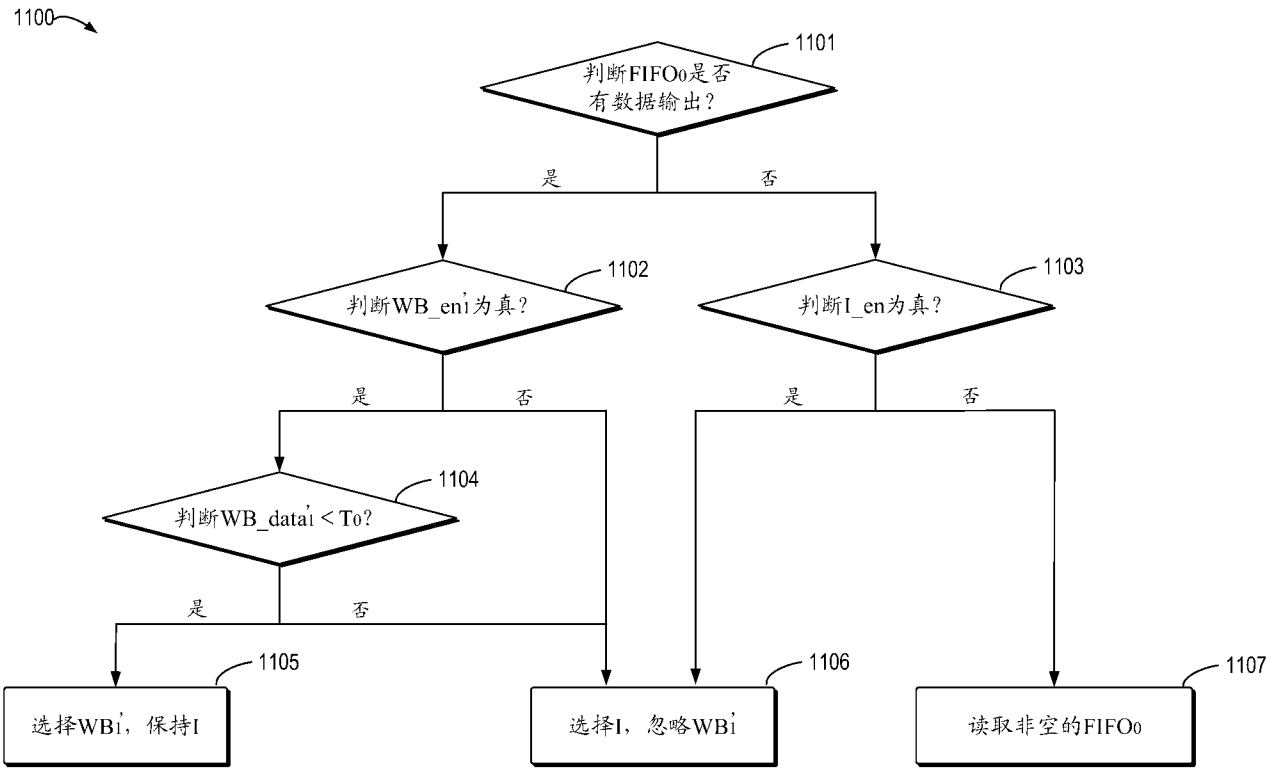

1100
1101
判断FIFO0是否有数据输出？
是
否
1102
判断WB_en'1为真？
1103
判断I_en为真？
是
否
是
否
1104
判断WB_data'1 < T0?
是
否
1105
选择WB1'，保持I
1106
选择I，忽略WB1'
1107
读取非空的FIFO0

图 11

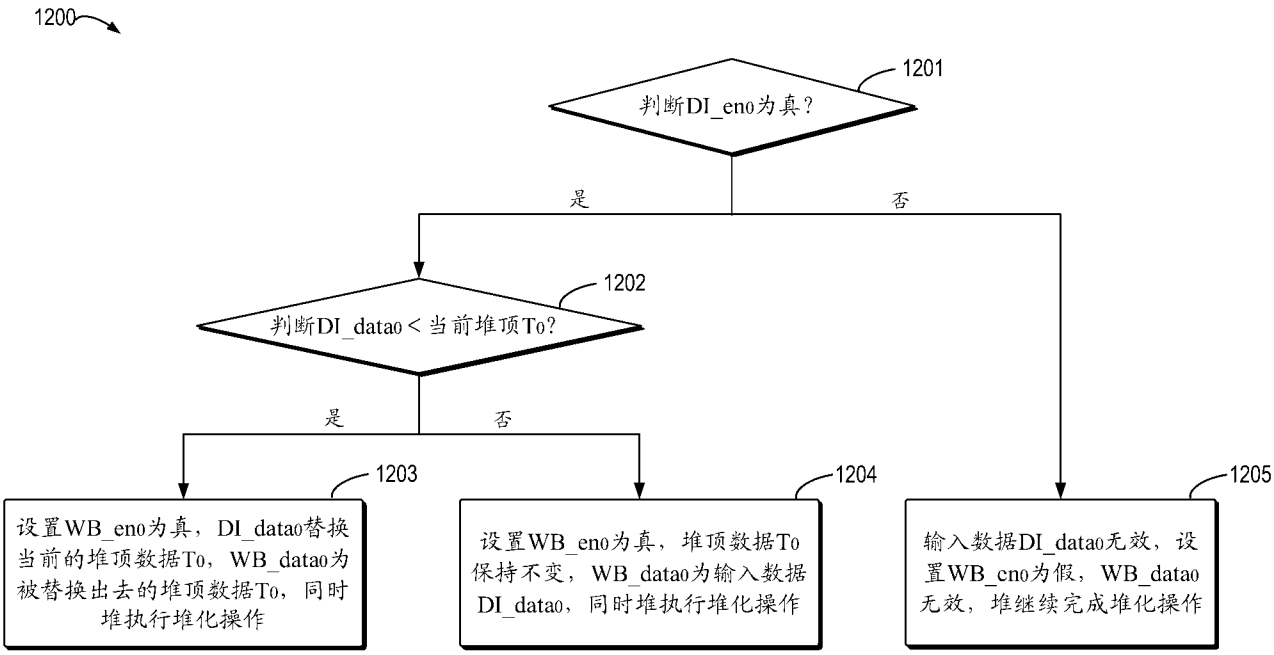

1200
1201
判断DI_en0为真？
是
否
1202
判断DI_data0 < 当前堆顶T0?
是
否
1203
设置WB_en0为真，DI_data0替换当前的堆顶数据T0，WB_data0为被替换出去的堆顶数据T0，同时堆执行堆化操作
1204
设置WB_en0为真，堆顶数据T0保持不变，WB_data0为输入数据DI_data0，同时堆执行堆化操作
1205
输入数据DI_data0无效，设置WB_cn0为假，WB_data0无效，堆继续完成堆化操作

图 12

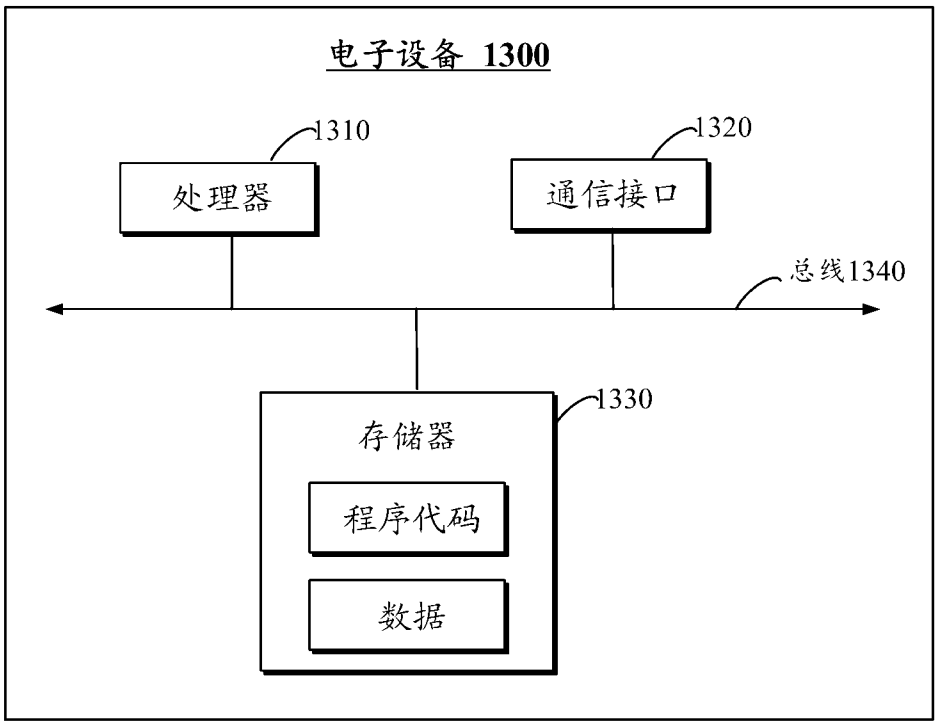


图 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/126090

A. CLASSIFICATION OF SUBJECT MATTER

G06F5/10(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)

IPC: 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; VEN; USTXT; WOTXT; EPTXT; IEEE; 百度, BAIDU; 读秀, DUXIU: 排序, 堆排序, 串联, 级联, Top-K, sort, heapsort, series, cascade

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 114547286 A (JINGDONG TECHNOLOGY HOLDING CO., LTD.) 27 May 2022 (2022-05-27) description, paragraphs [0043]-[0152]	1-22
A	CN 104317549 A (NO.771 INSTITUTE OF NO.9 RESEARCH INSTITUTE, CHINA AEROSPACE SCIENCE AND TECHNOLOGY CORP.) 28 January 2015 (2015-01-28) entire document	1-22
A	CN 111913955 A (YUSUR TECHNOLOGY CO., LTD.) 10 November 2020 (2020-11-10) entire document	1-22
A	CN 114968547 A (HUAWEI TECHNOLOGIES CO., LTD.) 30 August 2022 (2022-08-30) entire document	1-22
A	KR 101085967 B1 (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) 22 November 2011 (2011-11-22) entire document	1-22

☐ 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
"D" document cited by the applicant in the international application
"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	Date of mailing of the international search report
07 December 2023	**11 December 2023**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/126090

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 114547286 A	27 May 2022	None	
CN 104317549 A	28 January 2015	None	
CN 111913955 A	10 November 2020	None	
CN 114968547 A	30 August 2022	CN 115599541 A	13 January 2023
		WO 2022179023 A1	01 September 2022
KR 101085967 B1	22 November 2011	US 2012271838 A1	25 October 2012
		US 8296306 B1	23 October 2012

国际检索报告

国际申请号
PCT/CN2023/126090

A. 主题的分类

G06F5/10(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)

IPC：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;VEN;USTXT;WOTXT;EPTXT;IEEE;百度;读秀：排序，堆排序，串联，级联， Top-K，sort，heapsort，series，cascade

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
A	CN 114547286 A (京东科技控股股份有限公司) 2022年5月27日 (2022 - 05 - 27) 说明书第[0043]-[0152]段	1-22
A	CN 104317549 A (中国航天科技集团公司第九研究院第七七一研究所) 2015年1月28日 (2015 - 01 - 28) 全文	1-22
A	CN 111913955 A (中科驭数（北京）科技有限公司) 2020年11月10日 (2020 - 11 - 10) 全文	1-22
A	CN 114968547 A (华为技术有限公司) 2022年8月30日 (2022 - 08 - 30) 全文	1-22
A	KR 101085967 B1 (KOREA ADVANCED INST SCI & TECH) 2011年11月22日 (2011 - 11 - 22) 全文	1-22

☐ 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

“D” 申请人在国际申请中引证的文件

“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	国际检索报告邮寄日期
2023年12月7日	2023年12月11日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	受权官员 王丽娜 电话号码 (+86) 0512-88995937

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/126090

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	114547286	A	2022年5月27日	无			
CN	104317549	A	2015年1月28日	无			
CN	111913955	A	2020年11月10日	无			
CN	114968547	A	2022年8月30日	CN	115599541	A	2023年1月13日
				WO	2022179023	A1	2022年9月1日
KR	101085967	B1	2011年11月22日	US	2012271838	A1	2012年10月25日
				US	8296306	B1	2012年10月23日