(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
2023 年 1 月 26 日 (26.01.2023)

WIPO | PCT

(10) 国际公布号
WO 2023/000586 A1

(51) 国际专利分类号:
G06F 15/78 (2006.01) *G11C 16/34* (2006.01)

(21) 国际申请号: PCT/CN2021/137444

(22) 国际申请日: 2021 年 12 月 13 日 (13.12.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110823220.4 2021年7月21日 (21.07.2021) CN

(71) 申请人: 清 华 大 学 (TSINGHUA UNIVERSITY) [CN/CN]; 中 国 北 京 市 海 淀 区 清 华 园, Beijing 100084 (CN)。

(72) 发明人: 高滨(GAO, Bin); 中国北京市海淀区清华园, Beijing 100084 (CN)。 姚鹏(YAO, Peng); 中国北京市海淀区清华园, Beijing 100084 (CN)。 吴 华 强 (WU, Huaqiang); 中 国 北 京 市 海 淀 区 清 华 园, Beijing 100084 (CN)。 唐 建 石 (TANG, Jianshi); 中 国 北 京 市 海 淀 区 清 华 园, Beijing 100084 (CN)。 钱鹤(QIAN, He); 中国北京市海淀区清华园, Beijing 100084 (CN)。

(74) 代理人: 北 京 市 柳 沈 律 师 事 务 所 (LIU, SHEN & ASSOCIATES); 中国北京市海淀区彩和坊路10号1号楼10层, Beijing 100080 (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, 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,

(54) **Title:** STORAGE AND COMPUTATION INTEGRATED APPARATUS AND CALIBRATION METHOD THEREFOR

(54) 发明名称: 存算一体装置及其校准方法

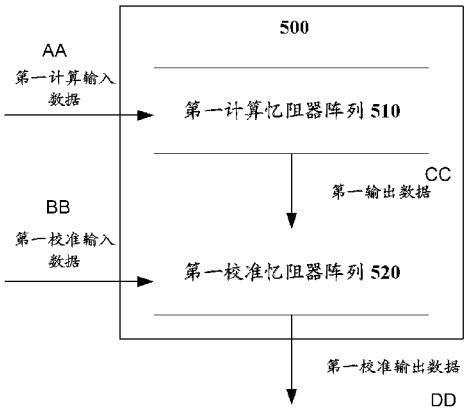


图 5

510 First computation memristor array
520 First calibration memristor array
AA First computation input data
BB First calibration input data
CC First output data
DD First calibration output data

(57) **Abstract:** A storage and computation integrated apparatus and a calibration method therefor. The storage and computation integrated apparatus comprises a first processing unit (500). The first processing unit (500) comprises: a first computation memristor array (510), which is configured to receive first computation input data and perform computation on the first computation input data, so as to obtain first output data; a first calibration memristor array (520), which is configured to receive first calibration input data and calibrate the first output data according to the first calibration input data, so as to obtain first calibration output data; and a first processing unit (500), which is configured to output the first calibration output data. The calibration method comprises: determining, by means of off-chip training, a first computation weight matrix which corresponds to a first computation memristor array (510), and writing the first computation weight matrix into the first computation memristor array (510); and on the basis of the first computation memristor array (510) where the first computation weight matrix has been written and the first computation weight matrix, performing on-chip training on a first calibration memristor array (520), so as to adjust a weight value of the first calibration memristor array (520).

WO 2023/000586 A1

[见续页]

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: 一种存算一体装置及其校准方法。存算一体装置包括第一处理单元（500），第一处理单元包括（500）：第一计算忆阻器阵列（510），被配置为接收第一计算输入数据，并对第一计算输入数据进行计算以得到第一输出数据；以及第一校准忆阻器阵列（520），被配置为接收第一校准输入数据，并根据第一校准输入数据对第一输出数据进行校准，以得到第一校准输出数据；第一处理单元（500）被配置为输出第一校准输出数据。校准方法包括：通过片外训练确定第一计算忆阻器阵列（510）对应的第一计算权重矩阵，将第一计算权重矩阵写入第一计算忆阻器阵列（510）；基于写入了第一计算权重矩阵的第一计算忆阻器阵列（510）和第一计算权重矩阵，对第一校准忆阻器阵列（520）进行片上训练，以调整第一校准忆阻器阵列（520）的权重值。

存算一体装置及其校准方法

本申请要求于 2021 年 7 月 21 日递交的中国专利申请第 202110823220.4 号的优先权，在此全文引用上述中国专利申请公开的内容以作为本申请的一部分。

技术领域

本公开的实施例涉及一种存算一体装置及其校准方法。

背景技术

基于忆阻器的存算一体装置可以在存储位置处直接进行原位乘累加等操作，实现计算和存储的器件级融合，打破传统硬件平台的算力、能效限制，是最具潜力的下一代硬件芯片技术之一。国内外企业、科研单位都投入大量人力物力，经过近十年的发展后，基于忆阻器的存算一体技术已逐渐从理论仿真阶段进入实际芯片、系统的原型演示阶段。

发明内容

本公开至少一实施例提供一种存算一体装置的校准方法，所述存算一体装置包括第一处理单元，所述第一处理单元包括第一计算忆阻器阵列和第一校准忆阻器阵列，所述第一计算忆阻器阵列被配置为接收第一计算输入数据，并对所述第一计算输入数据进行计算以得到第一输出数据；所述第一校准忆阻器阵列的被配置为接收第一校准输入数据，并根据所述第一校准输入数据对所述第一输出数据进行校准，以得到第一校准输出数据；所述第一处理单元被配置为输出所述第一校准输出数据，所述校准方法包括：通过片外训练确定所述第一计算忆阻器阵列对应的第一计算权重矩阵，将所述第一计算权重矩阵写入所述第一计算忆阻器阵列；基于写入了所述第一计算权重矩阵的所述第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值。

例如，在本公开至少一实施例提供的校准方法中，所述第一计算忆阻器阵

列包括M行*N列的忆阻器单元，所述第一校准忆阻器阵列包括K行*N列的忆阻器单元，所述第一校准忆阻器阵列的第i列与所述第一计算忆阻器阵列的第i列串联，所述第一校准忆阻器阵列的第i列的忆阻器单元被配置为根据所述第一校准输入数据对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第i列的第一校准输出数据，M、N、K、i均为正整数，且1≤i≤N，基于写入了所述第一计算权重阵列的所述第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值，包括：根据第一训练计算输入数据和所述第一计算权重矩阵，确定第一训练目标输出数据；向所述第一计算忆阻器阵列输入所述第一训练计算输入数据，并向所述第一校准忆阻器阵列输入第一训练校准输入数据，以得到第一训练校准输出数据；根据所述第一训练校准输出数据和所述第一训练目标输出数据的偏差，调整所述第一校准忆阻器阵列中对应于所述第一计算忆阻器阵列中需要校准的列的忆阻器单元的权重值。

例如，在本公开至少一实施例提供的校准方法中，所述第一校准忆阻器阵列包括固定偏差校准子阵列，所述第一训练校准输入数据包括固定训练校准输入数据，所述固定偏差校准子阵列包括1行*N列的忆阻器单元，向所述第一校准忆阻器阵列输入所述第一训练校准输入数据，包括：向所述固定偏差校准子阵列输入所述固定训练校准输入数据。

例如，在本公开至少一实施例提供的校准方法中，所述第一校准忆阻器阵列包括动态偏差校准子阵列，所述第一训练校准输入数据包括动态训练校准输入数据，所述动态偏差校准子阵列包括至少1行*N列的忆阻器单元，向所述第一校准忆阻器阵列输入所述第一训练校准输入数据，包括：根据所述第一训练计算输入数据，确定所述动态训练校准输入数据；以及，向所述动态偏差校准子阵列输入所述动态训练校准输入数据。

例如，在本公开至少一实施例提供的校准方法中，所述第一校准忆阻器阵列包括固定偏差校准子阵列和动态偏差校准子阵列，所述第一训练校准输入数据包括固定训练校准输入数据和动态训练校准输入数据，所述固定偏差校准子阵列包括1行*N列的忆阻器单元，所述动态偏差校准子阵列包括至少1行*N列的忆阻器单元，向所述第一校准忆阻器阵列输入所述第一训练校准输入数据，包括：向所述固定偏差校准子阵列输入所述固定训练校准输入数据；

根据所述第一训练计算输入数据，确定所述动态训练校准输入数据；以及，向所述动态偏差校准子阵列输入所述动态训练校准输入数据。

例如，在本公开至少一实施例提供的校准方法中，向所述固定偏差校准子阵列输入所述固定训练校准输入数据，包括：向所述固定偏差校准子阵列输入所述第一训练计算输入数据的最大值的5%-20%，以作为所述固定训练校准输入数据。

例如，在本公开至少一实施例提供的校准方法中，在基于写入了所述第一计算权重矩阵的所述第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值之前，所述校准方法还包括：确定所述第一计算忆阻器阵列的第i列的至少一个处于关键权重位置的忆阻器单元，并将所述动态偏差校准子阵列的第i列的至少一个忆阻器单元的输入设置为与所述第一计算忆阻器阵列的第 i 列的所述至少一个处于关键权重位置的忆阻器单元的输入相同，根据所述第一训练计算输入数据，确定所述动态训练校准输入数据，包括：将所述第一计算忆阻器阵列的第 i 列的所述至少一个处于关键权重位置的忆阻器单元的第一训练计算输入数据作为所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元的动态训练校准输入数据。

例如，在本公开至少一实施例提供的校准方法中，在基于写入了所述第一计算权重矩阵的所述第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值之前，所述校准方法还包括：确定所述第一计算忆阻器阵列的至少一行处于关键权重位置的忆阻器单元，并将所述动态偏差校准子阵列的至少一行忆阻器单元的输入设置为与所述第一计算忆阻器阵列的所述至少一行处于关键权重位置的忆阻器单元的输入相同，根据所述第一训练计算输入数据，确定所述动态训练校准输入数据，包括：将所述第一计算忆阻器阵列的所述至少一行处于关键权重位置的忆阻器单元的第一训练计算输入数据作为所述动态偏差校准子阵列的至少一行忆阻器单元的动态训练校准输入数据。

例如，在本公开至少一实施例提供的校准方法中，基于所述写入了所述第一计算权重矩阵的第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值，

还包括：在向所述第一计算忆阻器阵列输入所述第一训练计算输入数据，并向所述第一校准忆阻器阵列输入所述第一训练校准输入数据，以得到所述第一训练校准输出数据之前，向所述第一计算忆阻器阵列输入所述第一训练计算输入数据，以得到第一训练输出数据；根据所述第一训练输出数据和所述第一训练目标输出数据的偏差，确定所述第一计算忆阻器阵列中需要校准的列，并使所述第一校准忆阻器阵列中对应于所述第一计算忆阻器阵列中需要校准的列的忆阻器单元与所述第一计算忆阻器阵列中需要校准的列串联接通。

例如，在本公开至少一实施例提供的校准方法中，所述存算一体装置还包括第二处理单元，所述第二处理单元包括第二计算忆阻器阵列和第二校准忆阻器阵列，所述第二计算忆阻器阵列被配置为接收根据所述第一校准输出数据确定的第二计算输入数据，并对所述第二计算输入数据进行计算以得到第二输出数据；所述第二校准忆阻器阵列的被配置为接收第二校准输入数据，并根据所述第二校准输入数据对所述第二输出数据进行校准，以得到第二校准输出数据；所述第二处理单元被配置为输出所述第二校准输出数据，所述校准方法还包括：通过片外训练确定所述第二计算忆阻器阵列对应的第二计算权重矩阵，将所述第二计算权重矩阵写入所述第二计算忆阻器阵列；基于写入了所述第二计算权重矩阵的所述第二计算忆阻器阵列的计算输出和所述第二计算权重矩阵，对所述第二校准忆阻器阵列进行片上训练，以调整所述第二校准忆阻器阵列的权重值。

例如，在本公开至少一实施例提供的校准方法中，所述第二处理单元与所述第一处理单元耦接，在基于写入了所述第二计算权重矩阵的所述第二计算忆阻器阵列和所述第二计算权重矩阵，对所述第二校准忆阻器阵列进行片上训练，以调整所述第二校准忆阻器阵列的权重值的过程中，根据第一训练计算输入数据和所述第一计算权重矩阵，确定第一训练目标输出数据，并将所述第一训练目标输出数据作为所述第二计算忆阻器阵列的第二训练输入数据，或者，在训练好所述第一校准忆阻器阵列的情况下，向所述第一计算忆阻器阵列输入所述第一训练计算输入数据，向所述第一校准忆阻器阵列输入第一训练校准输入数据，以得到第一训练校准输出数据，并将所述第一训练校准输出数据作为所述第二计算忆阻器阵列的第二训练输入数据。

例如，在本公开至少一实施例提供的校准方法中，所述第一校准忆阻器阵

列中的每个忆阻器单元被配置为既能够实现正权重，又能够实现负权重。

本公开至少一实施例还提供一种存算一体装置，包括第一处理单元，所述第一处理单元包括第一计算忆阻器阵列和第一校准忆阻器阵列，所述第一计算忆阻器阵列被配置为接收第一计算输入数据，并对所述第一计算输入数据进行计算以得到第一输出数据；所述第一校准忆阻器阵列的被配置为接收第一校准输入数据，并根据所述第一校准输入数据对所述第一输出数据进行校准，以得到第一校准输出数据；所述第一处理单元被配置为输出所述第一校准输出数据。

例如，在本公开至少一实施例提供的存算一体装置中，所述第一计算忆阻器阵列包括 M 行*N 列的忆阻器单元，所述第一校准忆阻器阵列包括 K 行*N 列的忆阻器单元，所述第一校准忆阻器阵列的第 i 列与所述第一计算忆阻器阵列的第 i 列串联，所述第一校准忆阻器阵列的第 i 列的忆阻器单元被配置为根据所述第一校准输入数据对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据，M、N、K、i 均为正整数，且 $1 \leqslant i \leqslant N$。

例如，在本公开至少一实施例提供的存算一体装置中，所述第一校准忆阻器阵列包括固定偏差校准子阵列，所述第一校准输入数据包括固定校准输入数据，所述固定偏差校准子阵列包括 1 行*N 列的忆阻器单元，所述固定偏差校准子阵列的第 i 列与所述第一计算忆阻器阵列的第 i 列串联，所述固定偏差校准子阵列的第 i 列的忆阻器单元被配置为根据所述固定校准输入数据对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据。

例如，在本公开至少一实施例提供的存算一体装置中，所述第一校准忆阻器阵列包括动态偏差校准子阵列，所述第一校准输入数据包括动态校准输入数据，所述动态偏差校准子阵列包括至少 1 行*N 列的忆阻器单元，所述动态偏差校准子阵列的第 i 列与所述第一计算忆阻器阵列的第 i 列串联，所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据是根据所述第一计算忆阻器阵列的第 i 列的 M 个忆阻器单元对应的第一计算输入数据确定的，所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元被配置为根据所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动

态校准输入数据，对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据。

例如，在本公开至少一实施例提供的存算一体装置中，所述第一校准忆阻器阵列包括固定偏差校准子阵列和动态偏差校准子阵列，所述第一校准输入数据包括固定校准输入数据和动态校准输入数据，所述固定偏差校准子阵列包括 1 行*N 列的忆阻器单元，所述动态偏差校准子阵列包括至少 1 行*N 列的忆阻器单元，所述固定偏差校准子阵列的第 i 列和所述动态偏差校准子阵列的第 i 列与所述第一计算忆阻器阵列的第 i 列串联，所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据是根据所述第一计算忆阻器阵列的第 i 列的 M 个忆阻器单元对应的第一计算输入数据确定的，所述固定偏差校准子阵列的第 i 列的忆阻器单元和所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元被配置为分别根据所述固定校准输入数据和所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据，对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据。

例如，在本公开至少一实施例提供的存算一体装置中，所述固定校准输入数据为所述第一计算输入数据的最大值的 5%-20%。

例如，在本公开至少一实施例提供的存算一体装置中，所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据与所述第一计算忆阻器阵列的第 i 列的至少一个处于关键权重位置的忆阻器单元对应的第一计算输入数据相同。

例如，在本公开至少一实施例提供的存算一体装置中，所述动态偏差校准子阵列的至少一行忆阻器单元对应的动态校准输入数据与所述第一计算忆阻器阵列的至少一行处于关键权重位置的忆阻器单元对应的第一计算输入数据相同。

附图说明

为了更清楚地说明本公开实施例的技术方案，下面将对实施例的附图作简单地介绍，显而易见地，下面描述中的附图仅仅涉及本公开的一些实施例，而非对本公开的限制。

图 1 图示了一种忆阻器交叉阵列结构的示意图;

图 2 图示了一种存算一体装置的处理单元的示意图;

图 3 图示了一种基于忆阻器的存算一体系统的示意图;

图 4A 图示了一种低层次、细粒度的元件级校准方法的示意图;

图 4B 图示了一种高层次、粗粒度的系统级校准方法的示意图;

图 4C 为本公开一实施例提供的一种中层次、中粒度的针对中间级组件的校准方法的示意图;

图 5 为本公开一实施例提供的一种存算一体装置的处理单元的示意图;

图 6 为本公开一实施例提供的一种存算一体装置的校准方法的流程图;

图 7 为本公开一实施例提供的第一处理单元的一个示例的示意图;

图 8 为本公开一实施例提供的动态确定动态训练校准输入数据的一个示例的示意图;

图 9A 为本公开至少一实施例提供的校准方法的一个示例的流程图;

图 9B 为本公开至少一实施例提供的校准方法的另一个示例的流程图;

图 10A 为本公开至少一实施例提供的一种独立并行校准方式的示意图;

图 10B 为本公开至少一实施例提供的一种逐级校准方式的示意图。

具体实施方式

为使本公开实施例的目的、技术方案和优点更加清楚，下面将结合本公开实施例的附图，对本公开实施例的技术方案进行清楚、完整地描述。显然，所描述的实施例是本公开的一部分实施例，而不是全部的实施例。基于所描述的本公开的实施例，本领域普通技术人员在无需创造性劳动的前提下所获得的所有其他实施例，都属于本公开保护的范围。

除非另外定义，本公开使用的技术术语或者科学术语应当为本公开所属领域内具有一般技能的人士所理解的通常意义。本公开中使用的“第一”、“第二”以及类似的词语并不表示任何顺序、数量或者重要性，而只是用来区分不同的组成部分。同样，“一个”、“一”或者“该”等类似词语也不表示数量限制，而是表示存在至少一个。“包括”或者“包含”等类似的词语意指出现该词前面的元件或者物件涵盖出现在该词后面列举的元件或者物件及其等同，而不排除其他元件或者物件。“连接”或者“相连”等类似的词语并非限定于物理的或者机

械的连接，而是可以包括电性的连接，不管是直接的还是间接的。“上”、“下”、“左”、“右”等仅用于表示相对位置关系，当被描述对象的绝对位置改变后，则该相对位置关系也可能相应地改变。

下面通过几个具体的实施例对本公开进行说明。为了保持本公开实施例的以下说明清楚且简明，可省略已知功能和已知部件的详细说明。当本公开实施例的任一部件在一个以上的附图中出现时，该部件在每个附图中由相同的参考标号表示。

忆阻器交叉阵列结构是忆阻器存算一体装置的模拟域计算核心。例如，图1图示了一种忆阻器交叉阵列结构100的示意图。如图1所示，忆阻器交叉阵列结构100可以包括以横纵交叉方式阵列排布的多个忆阻器，通过将输入数据排列为输入向量X（例如，包括图1所示出的x_1、x_2、……、x_n，并且输入向量可以为已编码幅值、宽度或脉冲个数的电压），将权重矩阵编码为忆阻器电导值G（例如，包括图1所示出的g_{11}、g_{21}、……、g_{1n}，以及图1中未示出的g_{m1}、g_{m1}、……、g_{mn}），利用高并行、低功耗的阵列读操作获得输出电流I（例如，包括图1所示出的I_1、I_2、……、I_m），可以实现深度学习中普遍的乘累加计算，进而对矩阵向量乘进行加速。例如，根据基尔霍夫定律，忆阻器交叉阵列结构的输出电流可以根据下述公式得出：$I = G \times V$。上述乘累加计算过程在模拟域利用物理定律实现，不同于布尔逻辑的数字电路实现方式，不需要频繁访存搬移权重数据，解决了经典计算系统的冯诺依曼瓶颈，可以实现高算力、高能效的智能计算任务。

忆阻器交叉阵列可以用于构建存算一体装置的基本处理单元（Processing Element，PE）。例如，图2图示了一种存算一体装置的处理单元200的示意图。如图2所示，处理单元200可以包括模拟域部分和数字域分，模拟域部分实现基于模拟信号的模拟计算，模拟域的输入、控制、输出都是数字信号；数字域部分控制、配合模拟域部分的功能，并和外部进行交互。例如，如图2所示，模拟域部分包括输入模块210、忆阻器阵列220、输出模块230和供电模块240。输入模块210是用来实现输入向量功能的相关模拟电路；忆阻器阵列220是忆阻器交叉阵列（例如图1所示的忆阻器交叉阵列结构100），可以被写入权重矩阵并进行乘累加计算；输出模块230是用来实现输出向量（例如图1所示的输出电流）量化的相关模拟电路；供电模块240是基本的模拟

电源电路。例如，数字域部分可以包括控制器、输入缓存、输出缓存、数字后处理模块和接口模块等（图 1 中未示出）。需要说明的是，图 1 所示的存处理单元 200 仅是示例性的，而不是对本公开的限制，处理单元可以根据实际情况进行模块的增减和变形。

图 3 图示了一种基于忆阻器的存算一体系统 300 的示意图。例如，存算一体系统 300 可以是存算一体芯片。如图 3 所示，存算一体系统 300 包括多个处理单元模组 310 和多个功能单元 320。如图 3 所示，每个处理单元模组 310 包括多个处理单元 311 和多个功能模块 312，例如，处理单元 311 可以是图 2 所示的处理单元 200，多个功能模块 312 由数字电路实现并可以包括以下中的一者或多者：互联模块、接口模块、控制模块等。例如，多个功能单元 320 由数字电路实现并可以包括以下中的一者或多者：互联单元、主控单元、片上存储单元、路由单元、接口单元、时钟单元、电源单元等。需要说明的是，图 3 所示的存算一体系统 300 仅是示例性的，而不是对本公开的限制，存算一体系统可以根据实际情况进行单元或模块的增减和变形。在基于忆阻器的存算一体系统中，所有数字域实现的信号处理和控制都非常稳定，但模拟域实现的模拟计算可能受到忆阻器天然的随机性、非理想特性、阵列寄生参数影响或者受到工艺、供电、温度等环境偏差和波动影响，从而在器件级（忆阻器）、电路级（模拟电路模块)、阵列级和系统级工作时都会偏离理想的功能和输出，这些误差、噪声等会造成计算结果精度损失和错误，使得硬件系统无法正确工作。例如，在图 3 所示的每个处理单元 311 是图 2 所示的处理单元 200 的情况下，每个处理单元 311 内的输入模块、忆阻器阵列、输出模块和供电模块都会存在误差和扰动，而图 3 所示的基于忆阻器的存算一体系统包括多个处理单元 311，误差和扰动会不断积累，从而严重影响系统正常工作，这时需要进行校准。

例如，图 4A-4C 图示了校准方法的示意图。传统的校准方法可以分为两类。第一类校准方法是低层次、细粒度的元件级校准，通过校准所有潜在的误差和波动来源保证每个部件可靠、准确度工作，进而保证包括多个部件的系统能够可靠、准确地工作。例如，采用多个忆阻器共同表征一个权重单元，利用平均效应抵消随机性的影响；设计全新的单元阵列结构，缓解阵列的寄生参数；等比例的增大电路设计尺寸，缓解工艺偏差；引入校准电路单元，补偿、

校准模拟电路的精度等。例如，图 4A 图示了一种低层次、细粒度的元件级校准方法的示意图，其中黑色圆圈表示校准位置，在图 4A 中，对处理单元的模拟域内的每个模块进行校准，从而保证处理单元可靠、准确地工作，进而保证包括多个处理单元的系统能够可靠、准确地工作。第二类校准方法是高层次、粗粒度的系统级校准，在第二类方案中不关注具体的误差和波动来源，而是从系统整体表现出发，强调系统功能层次的工作、输出正确。例如，算法和系统结合进行联合优化，对关键层权重或全部权重进行片上训练，采用集成学习的方法，采用更多的硬件资源实现同一算法等。例如，图 4B 图示了一种高层次、粗粒度的系统级校准方法的示意图，其中黑色圆圈表示校准位置，在图 4B 中，对基于忆阻器的存算一体系统整体通过算法优化和流程改进来实现校准。然而，第一类校准方法开销大，成本高，设计和实现复杂度高；第二类校准方法通常对器件特性要求较高，且校准可靠性难以保证，只能证明在有限的网络结构中起作用，无法推广到更广泛、复杂的网络模型上，此外，一些特定的第二类校准方法只能兼容器件级误差，若通过集成学习方法会成倍的增加硬件开销。

为实现一种高可靠、低成本的通用存算一体模拟计算误差校准，本公开提出一种中层次、中粒度的针对中间级组件的校准方法。在本公开所提出的中层次、中粒度的针对中间级组件的校准方法中，通过引入冗余校准单元对中间级组件的模拟域计算结果进行校准，从而保证系统能够可靠、准确地工作，并且由于仅调整引入的冗余校准单元而无需改变已有的忆阻器阵列所实现的权重矩阵，从而可以灵活根据校准精度的要求配置硬件开销，实现按需设计。例如，图 4C 为本公开一实施例提供的一种中层次、中粒度的针对中间级组件的校准方法的示意图，其中黑色圆圈表示校准位置，在图 4C 中，对每个处理单元的模拟域计算结果进行校准，由于各种底层的模拟元件的误差最终都会反映到该模拟计算结果上，所以校准该模拟域计算结果就相当于同时对各个底层的模拟元件进行校准，能够在处理单元级保证各个计算结果准确，从而保证系统正常工作。

下面结合附图对本公开所提出的中层次、中粒度的针对中间级组件的校准方法及其实施例和对应的示例进行详细说明。

图 5 为本公开一实施例提供的一种存算一体装置的处理单元的示意图。

图 6 为本公开一实施例提供的一种存算一体装置的校准方法的流程图。下面以图 5 所示的存算一体装置为例，对图 6 所示的校准方法进行详细说明。

例如，如图 5 所示，存算一体装置包括第一处理单元 500，第一处理单元 500 包括第一计算忆阻器阵列 510 和第一校准忆阻器阵列 520，第一计算忆阻器阵列 510 被配置为接收第一计算输入数据，并对第一计算输入数据进行计算以得到第一输出数据；第一校准忆阻器阵列 520 的被配置为接收第一校准输入数据，并根据第一校准输入数据对第一输出数据进行校准，以得到第一校准输出数据；第一处理单元 500 被配置为输出第一校准输出数据。

例如，第一计算忆阻器阵列 510 可以为图 2 所示的忆阻器阵列 220。第一计算输入数据可以为已编码的电压，第一计算忆阻器阵列 510 进行乘累加计算以得到第一输出数据，第一输出数据为电流值。

例如，第一校准忆阻器阵列 520 可以为忆阻器交叉阵列，第一校准输入数据可以为已编码的电压，第一校准忆阻器阵列 520 可以对第一输出数据的固定偏置偏差和动态偏差进行校准。

需要说明的是，为了清楚起见，图 5 仅示意性地示出了存算一体装置的第一处理单元 500，且仅示出了第一处理单元 500 的部分组件，实际上，存算一体装置可以包括与第一处理单元 500 类似的多个基本处理单元以及其他可能的组件（例如，图 3 所示的功能单元 320），并且第一处理单元 500 也可以包括其他可能的组件（例如，图 2 所示的各种模拟电路）。

在此情况下，例如，如图 6 所示，校准方法 600 包括步骤 S610 和步骤 S620。

步骤 S610：通过片外训练确定第一计算忆阻器阵列 510 对应的第一计算权重矩阵，将第一计算权重矩阵写入第一计算忆阻器阵列 510。

步骤 S620：基于写入了第一计算权重矩阵的第一计算忆阻器阵列 510 和第一计算权重矩阵，对第一校准忆阻器阵列 520 进行片上训练，以调整第一校准忆阻器阵列 520 的权重值。

对于步骤 S610，可以将第一计算权重矩阵编码为第一计算忆阻器阵列的电导值。

对于步骤 S620，写入了第一计算权重矩阵的第一计算忆阻器阵列 510 的计算输出可以视为实际的第一输出数据，利用第一计算权重矩阵可以确定理

想的目标输出数据，实际的第一输出数据与理想的目标输出数据之间存在偏差，对第一校准忆阻器阵列 520 进行片上训练，通过自适应地调整第一校准忆阻器阵列 520 的权重值可以减小该偏差直到该偏差达到合适的区间，从而实现校准的目的。

例如，上述校准方法 600 可以集成到系统的编译器等工具链中，实现全流程自动化。在这种情况下，引入的额外硬件开销仅仅只有第一校准忆阻器阵列 520 和简单的数字控制电路（图 5 中未示出）。

例如，片上训练可以采用线性回归方法，也可以采用其他算法，以调整第一校准忆阻器阵列 520 的权重值，使得实际的第一输出数据逼近理想的目标输出数据，以实现校准效果。

例如，调整第一校准忆阻器阵列 520 的权重值的顺序可以采用并行方式或串行方式，也可以采用其他可能的顺序方式，本公开的实施例对此不作限制。例如，调整第一校准忆阻器阵列 520 的权重值的操作可以是带读校验的精确定量更新，也可以是定性的符号更新，也可以采用能够保证权重按照需要确定的增大或减小的其他操作方式，本公开的实施例对此不作限制。

在本公开实施例中，对第一计算忆阻器阵列的第一输出数据进行校准，实际上是对该第一处理单元内的模拟域计算结果进行校准，该第一输出数据已经包含了第一处理单元内的第一计算忆阻器阵列以及其他可能的模拟电路（例如，图 2 所示的各种模拟电路）所带来的偏差，从而上述校准是处理单元级的校准，可以确保每个处理单元的模拟域计算结果准确，从而保证包括多个处理单元的系统正常工作，具有很好的可靠性和通用性。此外，由于不需要对底层的各个组件进行校准，引入的额外开销低，具有高效性和低成本的优势；由于仅对引入的第一校准忆阻器阵列的权重值进行校准，可以灵活根据校准精度的要求配置硬件开销，实现按需设计。因此，本公开所提出的校准方法解决了存算一体系统发展的关键瓶颈，是忆阻器存算一体技术走向实用化应用阶段的核心突破。

例如，在本公开至少一实施例提供的校准方法中，第一计算忆阻器阵列 510 包括 M 行*N 列的忆阻器单元，第一校准忆阻器阵列 520 包括 K 行*N 列的忆阻器单元，第一校准忆阻器阵列 520 的第 i 列与第一计算忆阻器阵列 510 的第 i 列串联，第一校准忆阻器阵列 520 的第 i 列的忆阻器单元被配置为根据

第一校准输入数据对第一计算忆阻器阵列 510 的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据，M、N、K、i 均为正整数，且 1≤i≤N。

需要说明的是，第一校准忆阻器阵列 520 的尺寸由校准精度的要求确定，期望模拟计算结果越精确、校准的范围越大，则需要越多的冗余权重，即需要越大尺寸的第一校准忆阻器阵列 520，在极端情况下，第一校准忆阻器阵列 520 的尺寸可以大于第一计算忆阻器阵列 510 的尺寸。

还需要说明的是，描述第一计算忆阻器阵列 510 和第一校准忆阻器阵列 520 的尺寸时所使用的“行”和“列”仅旨在区分忆阻器阵列的两个维度，而不是对忆阻器阵列的限制。

例如，在本公开至少一实施例提供的校准方法中，步骤 S620 包括步骤 S621 至步骤 S623。

步骤 S621：根据第一训练计算输入数据和第一计算权重矩阵，确定第一训练目标输出数据。

步骤 S622：向第一计算忆阻器阵列 510 输入第一训练计算输入数据，并向第一校准忆阻器阵列 520 输入第一训练校准输入数据，以得到第一训练校准输出数据。

步骤 S623：根据第一训练校准输出数据和第一训练目标输出数据的偏差，调整第一校准忆阻器阵列中对应于第一计算忆阻器阵列中需要校准的列的忆阻器单元的权重值。

对于步骤 S621，向第一计算忆阻器阵列 510 输入第一训练输入数据，第一计算忆阻器阵列 510 可以根据第一计算权重矩阵计算得到第一训练目标输出数据。

例如，第一训练输入数据可以随机生成，或者根据编译器等工具链预先确定的训练集的输入分布采样生成，或者直接为的编译器等工具链预先确定的训练集的一部分，本公开的实施例对此不作限制。

对于步骤 S622，在第一计算忆阻器阵列 510 输入第一训练计算输入数据，第一校准忆阻器阵列 520 输入第一训练校准输入数据的情况下，第一计算忆阻器阵列 510 和第一校准忆阻器阵列 520 两者的整体可以计算得到第一训练校准输出数据。

对于步骤 S623，根据第一训练校准输出数据和第一训练目标输出数据的偏差，仅调整第一校准忆阻器阵列 520 中对应于第一计算忆阻器阵列 510 中需要校准的列的忆阻器单元的权重值，而不调整第一计算忆阻器阵列 510 中需要校准的列的忆阻器单元的权重值，即可实现对偏差的校准。

例如，第一校准忆阻器阵列 510 中的每个忆阻器单元被配置为既能够实现正权重，又能够实现负权重。

需要说明的是，上述实施例示意性地说明了第一校准忆阻器阵列的片上训练的一次迭代过程，在实际情况中，通常需要进行多次迭代过程以使得校准结果满足需求，由于多次迭代过程中的每一次迭代过程均与步骤 S621 至 S623 相同，故在此不再赘述。

还需要说明的是，为了防止迭代过程次数过多，可以设定最大迭代过程次数以节省资源。

例如，图 7 为本公开一实施例提供的第一处理单元的一个示例的示意图。如图 7 所示，第一处理单元 700 包括第一计算忆阻器阵列 710 和第一校准忆阻器阵列 720。在图 7 中，第一校准忆阻器阵列 720 包括固定偏差校准子阵列 720a 和动态偏差校准子阵列 720b。固定偏差校准子阵列 720a 被配置为校准固定偏置偏差，动态偏差校准子阵列 720b 被配置为校准动态偏置偏差。需要说明的是，在本公开的实施例中，第一校准忆阻器阵列 720 可以包括固定偏差校准子阵列 720a 和动态偏差校准子阵列 720b 中任一者或两者，并且在第一校准忆阻器阵列 720 包括固定偏差校准子阵列 720a 和动态偏差校准子阵列 720b 两者的情况下，可以使用固定偏差校准子阵列 720a 和动态偏差校准子阵列 720b 中的任一者或两者。例如，可以通过开关（图 7 中未示出）控制方式来选择固定偏差校准子阵列 720a 和动态偏差校准子阵列 720b 中的任一者或两者，以根据需要灵活控制第一校准忆阻器阵列。

还需要说明的是，图 7 中的忆阻器采用 1R（即一个电阻元件）类型仅是示例性的而不是对本公开的限制，忆阻器可以根据实际情况采用不同的类型。例如，忆阻器还可以采用 1T1R（即一个电阻元件加一个晶体管）类型。

还需要说明的是，图 7 中的第一校准忆阻器阵列 720 的尺寸仅是示例性的，而不是对本公开的限制，第一校准忆阻器阵列 720 的尺寸可以根据所面向的特定的智能计算场景和任务及其精度要求而确定。并且，在图 7 中，固

定偏差校准子阵列 720a 采用阵列拓展模式，动态偏差校准子阵列 720b 采用独立的多列模式，也是示例性的而不是对本公开的限制，固定偏差校准子阵列 720a 和动态偏差校准子阵列 720b 的模式可以更具具体情况而定。

例如，在本公开的实施例的一个示例中，第一校准忆阻器阵列包括固定偏差校准子阵列（例如，图 7 所示的固定偏差校准子阵列 720a），第一训练校准输入数据包括固定训练校准输入数据，固定偏差校准子阵列包括 1 行*N 列的忆阻器单元。在这种情况下，步骤 S622 中的向第一校准忆阻器阵列输入第一训练校准输入数据，包括：向固定偏差校准子阵列输入固定训练校准输入数据。

例如，在本公开实施例的另一个示例中，第一校准忆阻器阵列包括动态偏差校准子阵列（例如，图 7 所示的动态偏差校准子阵列 720b），第一训练校准输入数据包括动态训练校准输入数据，动态偏差校准子阵列的包括至少 1 行*N 列的忆阻器单元。在这种情况下，步骤 S622 中的向第一校准忆阻器阵列输入第一训练校准输入数据，包括：根据第一训练计算输入数据，确定动态训练校准输入数据；以及，向动态偏差校准子阵列输入动态训练校准输入数据。

例如，在本公开实施例的又一个示例中，第一校准忆阻器阵列包括固定偏差校准子阵列和动态偏差校准子阵列（图 7 所示的固定偏差校准子阵列 720a 和动态偏差校准子阵列 720b），第一训练校准输入数据包括固定训练校准输入数据和动态训练校准输入数据，固定偏差校准子阵列包括 1 行*N 列的忆阻器单元，动态偏差校准子阵列包括至少 1 行*N 列的忆阻器单元。在这种情况下，步骤 S622 中的向第一校准忆阻器阵列输入第一训练校准输入数据，包括：向固定偏差校准子阵列输入固定训练校准输入数据；根据第一训练计算输入数据，确定动态训练校准输入数据；以及，向动态偏差校准子阵列输入动态训练校准输入数据。

在本公开的实施例中，固定训练校准输入数据对于固定偏差校准子阵列的 N 列中的每一列都是相同的。

例如，固定训练校准输入数据是固定偏置的电压脉冲。

例如，在第一计算忆阻器阵列各行的第一训练计算输入数据是 0 或 1 的情况下，固定训练校准输入数据可以固定设置为 1 且不随第一训练计算输入数据的变化而变化。

例如，考虑到实际的第一输出数据与理想的目标输出数据之间的偏差往往相对较小，固定训练校准输入数据可以设置为第一训练计算输入数据的最大值的 5%-20%。在这种情况下，向固定偏差校准子阵列输入固定训练校准输入数据，包括: 向固定偏差校准子阵列输入第一训练计算输入数据的最大值的 5%-20%，以作为固定训练校准输入数据。

在本公开的实施例中，动态训练校准输入数据是动态确定的，且对于动态偏差校准子阵列的 N 列中的每一列可能是不同的。

例如，在本公开实施例的一个示例中，在基于写入了第一计算权重矩阵的第一计算忆阻器阵列和第一计算权重矩阵，对第一校准忆阻器阵列进行片上训练，以调整第一校准忆阻器阵列的权重值之前，校准方法还包括步骤 S630a: 确定第一计算忆阻器阵列的第 i 列的至少一个处于关键权重位置的忆阻器单元，并将动态偏差校准子阵列的第 i 列的至少一个忆阻器单元的输入设置为与第一计算忆阻器阵列的第 i 列的至少一个处于关键权重位置的忆阻器单元的输入相同。在这种情况下，根据第一训练计算输入数据，确定动态训练校准输入数据，包括: 将第一计算忆阻器阵列的第 i 列的至少一个处于关键权重位置的忆阻器单元的第一训练计算输入数据作为动态偏差校准子阵列的第 i 列的至少一个忆阻器单元的动态训练校准输入数据。

例如，可以根据实际情况确定第一计算忆阻器阵列中的各列的关键权重位置的比例，并按照所确定的比例，在第一计算忆阻器阵列中定义每列的关键权重位置，比如可以按照波动程度以及波动对结果影响程度对每列的各个权重进行排序，按照所确定的比例选择影响程度靠前的忆阻器的相应位置作为该列的关键权重位置，第一计算忆阻器阵列的各列的关键权重位置可能不尽相同; 同时，第一计算忆阻器阵列的各列的关键权重位置的个数和动态偏差校准子阵列的各列接入的权重单元数相同，如果第一计算忆阻器阵列的第 i 列和动态偏差校准子阵列的第 i 列一一对应，第一计算忆阻器阵列的第 i 列的关键权重的输入则和动态偏差校准子阵列的第 i 列的各行输入共享，从而确定动态偏差校准子阵列的第 i 列的动态输入，即动态偏差校准子阵列的第 i 列共享第一计算忆阻器阵列的第 i 列的关键权重位置的第一训练计算输入数据。

例如，图 8 为本公开一实施例提供的动态确定动态训练校准输入数据的一个示例的示意图。在图 8 所示的示例中，确定第一计算忆阻器阵列 810 的

各列的关键权重位置数目为 2，对应地，动态偏差校准子阵列 820b 的尺寸为 2 行。例如，在图 8 中，确定第一计算忆阻器阵列 810 的第 1 列的第 1 行和第 j 行的忆阻器单元为处于关键权重位置的忆阻器单元，并将动态偏差校准子阵列 820b 的第 1 列的第 1 行和第 2 行的输入分别设置为与第一计算忆阻器阵列 810 的第 1 列的第 1 行和第 j 行的忆阻器单元的输入相同；确定第一计算忆阻器阵列 810 的第 2 列的第 1 行和第 2 行的忆阻器单元为处于关键权重位置的忆阻器单元，并将动态偏差校准子阵列 820b 的第 2 列的第 1 行和第 2 行的输入分别设置为与第一计算忆阻器阵列 810 的第 2 列的第 1 行和第 2 行的忆阻器单元的输入相同；确定第一计算忆阻器阵列 810 的第 3 列的第 2 行和第 j 行的忆阻器单元为处于关键权重位置的忆阻器单元，并将动态偏差校准子阵列 820b 的第 3 列的第 1 行和第 2 行的输入分别设置为与第一计算忆阻器阵列 810 的第 3 列的第 2 行和第 j 行的忆阻器单元的输入相同，j 为大于等于 2 的正整数。这样，就可以确定动态偏差校准子阵列 820b 的各列的动态输入，动态偏差校准子阵列 820b 的各列的输入共享第一计算忆阻器阵列 810 对应列的关键权重位置的输入。

在本公开的实施中，动态偏差校准子阵列也可以采用阵列拓展模式。例如，在动态偏差校准子阵列采用阵列拓展模式时，可以以行为单位地确定第一校准忆阻器阵列的关键权重位置。

例如，在本公开实施例的另一个示例中，在基于写入了第一计算权重矩阵的第一计算忆阻器阵列和第一计算权重矩阵，对第一校准忆阻器阵列进行片上训练，以调整第一校准忆阻器阵列的权重值之前，校准方法还包括步骤 S630b: 确定第一计算忆阻器阵列的至少一行处于关键权重位置的忆阻器单元，并将动态偏差校准子阵列的至少一行忆阻器单元的输入设置为与第一计算忆阻器阵列的至少一行处于关键权重位置的忆阻器单元的输入相同。在这种情况下，根据第一训练计算输入数据，确定动态训练校准输入数据，包括：将第一计算忆阻器阵列的至少一行处于关键权重位置的忆阻器单元的第一训练计算输入数据作为动态偏差校准子阵列的至少一行忆阻器单元的动态训练校准输入数据。

在本公开的实施中，可以对第一计算忆阻器阵列中的全部列进行校准，也可以确定第一计算忆阻器阵列中需要校准的列，仅对第一计算忆阻器阵列中

需要校准的列进行校准以节省校准的计算量。

例如，在本公开的实施例中，基于写入了第一计算权重矩阵的第一计算忆阻器阵列和第一计算权重矩阵，对第一校准忆阻器阵列进行片上训练，以调整第一校准忆阻器阵列的权重值（步骤 S620），还包括步骤 S624：在向第一计算忆阻器阵列输入第一训练计算输入数据，并向第一校准忆阻器阵列输入第一训练校准输入数据，以得到第一训练校准输出数据（步骤 S622）之前，向第一计算忆阻器阵列输入第一训练计算输入数据，以得到第一训练输出数据；根据第一训练输出数据和第一训练目标输出数据的偏差，确定第一计算忆阻器阵列中需要校准的列，并使第一校准忆阻器阵列中对应于第一计算忆阻器阵列中需要校准的列的忆阻器单元与第一计算忆阻器阵列中需要校准的列串联接通。

例如，图 9A 为本公开至少一实施例提供的校准方法的一个示例的流程图，校准方法 900a 包括步骤 S910a 至步骤 S950a。如图 9A 所示，在步骤 S910a 处，确定第一校准忆阻器阵列的结构和尺寸，确定第一校准忆阻器阵列的结构指代确定第一校准忆阻器阵列包括或者选择固态偏差校准子阵列和动态偏差校准子阵列中的一者或两者，确定第一校准忆阻器阵列的尺寸指代根据校正精度要求确定第一校准忆阻器阵列的行列数；在步骤 S920a 处，映射第一权重计算矩阵，例如，由编译器等工具链将第一权重计算矩阵部署、映射到每个处理单元的第一计算忆阻器阵列；在步骤 S930a 处，确定第一计算忆阻器阵列的各列的关键权重位置，例如，根据所确定的比例选择、定位出各列相应的关键权重位置；在步骤 S940a 处，确定第一训练校准输入数据，例如，固定训练校准输入数据为第一训练计算输入数据的最大值的 5%-20%，动态偏差标准子阵列的各列的不同行的输入和第一计算忆阻器阵列对应列的关键权重位置的输入实时共享；在步骤 S950a 处，对第一校准忆阻器阵列的全部列进行片上训练以实现校准，例如，利用线性回归的方法，片上训练、调整固定偏差校准子阵列和动态偏差校准子阵列的权重值，直到实际的第一输出数据与理想的目标输出数据之间的偏差达到合适的区间或片上训练达到最大迭代过程次数。在本示例中，未预先确定第一计算忆阻器阵列中需要校准的列，对第一计算忆阻器阵列中全部列进行校准。

例如，图 9B 为本公开至少一实施例提供的校准方法的另一个示例的流程

图，校准方法 900b 包括步骤 S910b 至步骤 S960b。如图 9B 所示，在步骤 S910b 处，确定第一校准忆阻器阵列的结构和尺寸，确定第一校准忆阻器阵列的结构指代确定第一校准忆阻器阵列包括或者选择固态偏差校准子阵列和动态偏差校准子阵列中的一者或两者，确定第一校准忆阻器阵列的尺寸指代根据校正精度要求确定第一校准忆阻器阵列的行列数；在步骤 S920b 处，映射第一权重计算矩阵，例如，由编译器等工具链将第一权重计算矩阵部署、映射到每个处理单元的第一计算忆阻器阵列；在步骤 S930b 处，确定第一计算忆阻器阵列的各列的关键权重位置，例如，根据所确定的比例选择、定位出各列相应的关键权重位置；在步骤 S940b 处，确定第一训练校准输入数据，例如，固定训练校准输入数据为第一训练计算输入数据的最大值的 5%-20%，动态偏差标准子阵列的各列的不同行的输入和第一计算忆阻器阵列对应列的关键权重位置的输入实时共享；在步骤 S950b 处，确定并接通第一计算忆阻器阵列中需要校准的列，例如，如步骤 S624 所述的确定并接通第一计算忆阻器阵列中需要校准的列；在步骤 S960b 处，对第一校准忆阻器阵列中需要校准的列进行片上训练以实现校准，例如，利用线性回归的方法，片上训练、调整固定偏差校准子阵列和动态偏差校准子阵列的权重值，直到实际的第一输出数据与理想的目标输出数据之间的偏差达到合适的区间或片上训练达到最大迭代过程次数。在本示例中，预先确定第一计算忆阻器阵列中需要校准的列，仅对第一计算忆阻器阵列中需要校准的列进行校准。

在本公开的实施例中，存算一体装置包括多个处理单元。例如，存算一体装置还包括第二处理单元，第二处理单元包括第二计算忆阻器阵列和第二校准忆阻器阵列，第二计算忆阻器阵列被配置为接收根据第一校准输出数据确定的第二计算输入数据，并对第二计算输入数据进行计算以得到第二输出数据；第二校准忆阻器阵列的被配置为接收第二校准输入数据，并根据第二校准输入数据对第二输出数据进行校准，以得到第二校准输出数据；第二处理单元被配置为输出第二校准输出数据；校准方法还包括：通过片外训练确定第二计算忆阻器阵列对应的第二计算权重矩阵，将第二计算权重矩阵写入第二计算忆阻器阵列；基于写入了第二计算权重矩阵的第二计算忆阻器阵列的计算输出和第二计算权重矩阵，对第二校准忆阻器阵列进行片上训练，以调整第二校准忆阻器阵列的权重值。

需要说明的是，对于存算一体装置所包括的多个处理单元，可以采用多种校准顺序，本公开的实施例对此不作限制。例如，多个处理单元可以独立并行地进行校准。例如，多个处理单元也可以逐级地进行校准。

例如，在本公开的实施例中，第二处理单元与第一处理单元耦接，在基于写入了第二计算权重矩阵的第二计算忆阻器阵列和第二计算权重矩阵，对第二校准忆阻器阵列进行片上训练，以调整第二校准忆阻器阵列的权重值的过程中，根据第一训练计算输入数据和第一计算权重矩阵，确定第一训练目标输出数据，并将第一训练目标输出数据作为第二计算忆阻器阵列的第二训练输入数据，或者，在训练好第一校准忆阻器阵列的情况下，向第一计算忆阻器阵列输入第一训练计算输入数据，向第一校准忆阻器阵列输入第一训练校准输入数据，以得到第一训练校准输出数据，并将第一训练校准输出数据作为第二计算忆阻器阵列的第二训练输入数据。

例如，图 10A 为本公开至少一实施例提供的一种独立并行校准方式的示意图。如图 10A 所示，在编译阶段划分、部署各个处理单元计算权重矩阵时，就确定了各个处理单元的理想的计算输入数据，使用各个处理单元的理想的计算输入数据分别构建各个处理单元的训练计算输入数据，可以独立并行地对各个处理单元进行校准。

例如，图 10B 为本公开至少一实施例提供的一种逐级校准方式的示意图。如图 10B 所示，按照深度学习时权重所属的层来对应划分多个处理单元中的各个处理单元的级别，第一级处理单元先进行校准，对应的输入是使用第一级处理单元的理想的计算输入数据所构建的训练计算输入数据；校准的第一级处理单元在使用理想的计算输入数据所构建的训练计算输入数据下的输出被作为下一级处理单元的输入，并对下一级处理单元进行校准，依此类推。

在图 10A 所示的示例中，各级处理单元的校准输入基于编译阶段确认的理想的计算输入数据，然而，在图 10B 所示的示例中，各级处理单元的校准输入（除了第一级以外）基于校准后的上一级处理单元的实际输出（可能和编译阶段的理想的计算输入数据存在偏差）。相比而言，图 10B 所示的方案更复杂，但误差适应性更好，系统校准精度更高。

综上，本公开提出了一种针对模拟域计算结果的中层次、中粒度的校准方法，并设计了配套的架构和操作流程。根据以上所述的各种实施例及其示例，

本公开所提出的校准方法可以实现处单元级的校准，确保每个处理单元的模拟域计算结果准确，从而保证包括多个处理单元的系统正常工作，具有很好的可靠性和通用性。此外，由于不需要对底层的各个组件进行校准，引入的额外开销低，具有高效性和低成本的优势；由于仅对引入的第一校准忆阻器阵列的权重值进行校准，可以灵活根据校准精度的要求配置硬件开销，实现按需设计。因此，本公开所提出的校准方法解决了存算一体系统发展的关键瓶颈，是忆阻器存算一体技术走向实用化应用阶段的核心突破。

本公开至少一实施例提供的一种存算一体装置，包括第一处理单元，第一处理单元包括第一计算忆阻器阵列和第一校准忆阻器阵列，第一计算忆阻器阵列被配置为接收第一计算输入数据，并对第一计算输入数据进行计算以得到第一输出数据；第一校准忆阻器阵列的被配置为接收第一校准输入数据，并根据第一校准输入数据对第一输出数据进行校准，以得到第一校准输出数据；第一处理单元被配置为输出第一校准输出数据。例如，第一处理单元可以参考第一处理单元 500 的相关描述，在此不再赘述。

例如，在本公开至少一实施例提供的存算一体装置中，第一计算忆阻器阵列包括 M 行*N 列的忆阻器单元，第一校准忆阻器阵列包括 K 行*N 列的忆阻器单元，第一校准忆阻器阵列的第 i 列与第一计算忆阻器阵列的第 i 列串联，第一校准忆阻器阵列的第 i 列的忆阻器单元被配置为根据第一校准输入数据对第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据，M、N、K、i 均为正整数，且 $1 \leqslant i \leqslant N$。例如，第一处理单元可以参考第一处理单元 500 的相关描述，在此不再赘述。

例如，在本公开至少一实施例提供的存算一体装置中，第一校准忆阻器阵列包括固定偏差校准子阵列，第一校准输入数据包括固定校准输入数据，固定偏差校准子阵列包括 1 行*N 列的忆阻器单元，固定偏差校准子阵列的第 i 列与第一计算忆阻器阵列的第 i 列串联，固定偏差校准子阵列的第 i 列的忆阻器单元被配置为根据固定校准输入数据对第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据。例如，固定偏差校准子阵列可以参考固定偏差校准子阵列 720a 的相关描述，在此不再赘述。

例如，在本公开至少一实施例提供的存算一体装置中，第一校准忆阻器阵列包括动态偏差校准子阵列，第一校准输入数据包括动态校准输入数据，动态

偏差校准子阵列的包括至少 1 行*N 列的忆阻器单元，动态偏差校准子阵列的第 i 列与第一计算忆阻器阵列的第 i 列串联，动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据是根据第一计算忆阻器阵列的第 i 列的 M 个忆阻器单元对应的第一计算输入数据确定的，动态偏差校准子阵列的第 i 列的至少一个忆阻器单元被配置为根据动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据，对第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据。例如，动态偏差校准子阵列可以参考动态偏差校准子阵列 720b 的相关描述，在此不再赘述。

例如，在本公开至少一实施例提供的存算一体装置中，第一校准忆阻器阵列包括固定偏差校准子阵列和动态偏差校准子阵列，第一校准输入数据包括固定校准输入数据和动态校准输入数据，固定偏差校准子阵列包括 1 行*N 列的忆阻器单元，动态偏差校准子阵列的包括至少 1 行*N 列的忆阻器单元，固定偏差校准子阵列的第 i 列和动态偏差校准子阵列的第 i 列与第一计算忆阻器阵列的第 i 列串联，动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据是根据第一计算忆阻器阵列的第 i 列的 M 个忆阻器单元对应的第一计算输入数据确定的，固定偏差校准子阵列的第 i 列的忆阻器单元和动态偏差校准子阵列的第 i 列的至少一个忆阻器单元被配置为分别根据固定校准输入数据和动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据，对第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据。例如，固定偏差校准子阵列可以参考固定偏差校准子阵列 720a 的相关描述，动态偏差校准子阵列可以参考动态偏差校准子阵列 720b 的相关描述，在此不再赘述。

例如，在本公开至少一实施例提供的存算一体装置中，固定校准输入数据为第一计算输入数据的最大值的 5%-20%。

例如，在本公开至少一实施例提供的存算一体装置中，动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据与第一计算忆阻器阵列的第 i 列的至少一个处于关键权重位置的忆阻器单元对应的第一计算输入数据相同。例如，动态偏差校准子阵列可以参考动态偏差校准子阵列 820b 的相关描述，动态校准输入数据的配置可以参考步骤 S630a 的相关描述，在

此不再赘述。

例如，在本公开至少一实施例提供的存算一体装置中，动态偏差校准子阵列的至少一行忆阻器单元对应的动态校准输入数据与第一计算忆阻器阵列的至少一行处于关键权重位置的忆阻器单元对应的第一计算输入数据相同。例如，动态校准输入数据的配置可以参考步骤 S630b 的相关描述，在此不再赘述。

需要说明的是，为表示清楚、简洁，本公开实施例并没有给出该存算一体装置的全部组成单元，也未给出该存算一体装置所包括的处理单元的全部组成单元。为实现存算一体装置及其所包括的处理单元的必要功能，本领域技术人员可以根据具体需要提供、设置其他未示出的组成单元，本公开的实施例对此不作限制。

关于不同实施例中的存算一体装置及其所包括的处理单元的技术效果可以参考本公开的实施例中提供的存算一体装置的校准方法的技术效果，在此不再赘述。

有以下几点需要说明:

（1）本公开实施例附图只涉及到与本公开实施例涉及到的结构，其他结构可参考通常设计。

（2）在不冲突的情况下，本公开的实施例及实施例中的特征可以相互组合以得到新的实施例。

以上所述仅是本公开的示范性实施方式，而非用于限制本公开的保护范围，本公开的保护范围由所附的权利要求确定。

权利要求书

1、一种存算一体装置的校准方法，其中，

所述存算一体装置包括第一处理单元，所述第一处理单元包括第一计算忆阻器阵列和第一校准忆阻器阵列，所述第一计算忆阻器阵列被配置为接收第一计算输入数据，并对所述第一计算输入数据进行计算以得到第一输出数据；所述第一校准忆阻器阵列的被配置为接收第一校准输入数据，并根据所述第一校准输入数据对所述第一输出数据进行校准，以得到第一校准输出数据；所述第一处理单元被配置为输出所述第一校准输出数据，

所述校准方法包括：

通过片外训练确定所述第一计算忆阻器阵列对应的第一计算权重矩阵，将所述第一计算权重矩阵写入所述第一计算忆阻器阵列；

基于写入了所述第一计算权重矩阵的所述第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值。

2、根据权利要求 1 所述的校准方法，其中，

所述第一计算忆阻器阵列包括 M 行*N 列的忆阻器单元，所述第一校准忆阻器阵列包括 K 行*N 列的忆阻器单元，所述第一校准忆阻器阵列的第 i 列与所述第一计算忆阻器阵列的第 i 列串联，所述第一校准忆阻器阵列的第 i 列的忆阻器单元被配置为根据所述第一校准输入数据对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据，其中，M、N、K、i 均为正整数，且 $1 \leq i \leq N$，

基于写入了所述第一计算权重阵列的所述第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值，包括：

根据第一训练计算输入数据和所述第一计算权重矩阵，确定第一训练目标输出数据；

向所述第一计算忆阻器阵列输入所述第一训练计算输入数据，并向所述第一校准忆阻器阵列输入第一训练校准输入数据，以得到第一训练校准输出数据；

根据所述第一训练校准输出数据和所述第一训练目标输出数据的偏差，调整所述第一校准忆阻器阵列中对应于所述第一计算忆阻器阵列中需要校准的列的忆阻器单元的权重值。

3、根据权利要求 2 所述的校准方法，其中，

所述第一校准忆阻器阵列包括固定偏差校准子阵列，所述第一训练校准输入数据包括固定训练校准输入数据，所述固定偏差校准子阵列包括 1 行*N 列的忆阻器单元，

向所述第一校准忆阻器阵列输入所述第一训练校准输入数据，包括：向所述固定偏差校准子阵列输入所述固定训练校准输入数据。

4、根据权利要求 2 所述的校准方法，其中，

所述第一校准忆阻器阵列包括动态偏差校准子阵列，所述第一训练校准输入数据包括动态训练校准输入数据，所述动态偏差校准子阵列包括至少 1 行*N 列的忆阻器单元，

向所述第一校准忆阻器阵列输入所述第一训练校准输入数据，包括：根据所述第一训练计算输入数据，确定所述动态训练校准输入数据；以及，向所述动态偏差校准子阵列输入所述动态训练校准输入数据。

5、根据权利要求 2-4 中任一项所述的校准方法，其中，

所述第一校准忆阻器阵列包括固定偏差校准子阵列和动态偏差校准子阵列，所述第一训练校准输入数据包括固定训练校准输入数据和动态训练校准输入数据，所述固定偏差校准子阵列包括 1 行*N 列的忆阻器单元，所述动态偏差校准子阵列包括至少 1 行*N 列的忆阻器单元，

向所述第一校准忆阻器阵列输入所述第一训练校准输入数据，包括：

向所述固定偏差校准子阵列输入所述固定训练校准输入数据；

根据所述第一训练计算输入数据，确定所述动态训练校准输入数据；以及，

向所述动态偏差校准子阵列输入所述动态训练校准输入数据。

6、根据权利要求 3 或 5 所述的校准方法，其中，向所述固定偏差校准子阵列输入所述固定训练校准输入数据，包括：向所述固定偏差校准子阵列输入所述第一训练计算输入数据的最大值的 5%-20%，以作为所述固定训练校准输入数据。

7、根据权利要求 4 或 5 所述的校准方法，其中，

在基于写入了所述第一计算权重矩阵的所述第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值之前，所述校准方法还包括:

确定所述第一计算忆阻器阵列的第 i 列的至少一个处于关键权重位置的忆阻器单元，并将所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元的输入设置为与所述第一计算忆阻器阵列的第 i 列的所述至少一个处于关键权重位置的忆阻器单元的输入相同，

其中，根据所述第一训练计算输入数据，确定所述动态训练校准输入数据，包括:

将所述第一计算忆阻器阵列的第 i 列的所述至少一个处于关键权重位置的忆阻器单元的第一训练计算输入数据作为所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元的动态训练校准输入数据。

8、根据权利要求 4 或 5 所述的校准方法，其中，

在基于写入了所述第一计算权重矩阵的所述第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值之前，所述校准方法还包括:

确定所述第一计算忆阻器阵列的至少一行处于关键权重位置的忆阻器单元，并将所述动态偏差校准子阵列的至少一行忆阻器单元的输入设置为与所述第一计算忆阻器阵列的所述至少一行处于关键权重位置的忆阻器单元的输入相同，

其中，根据所述第一训练计算输入数据，确定所述动态训练校准输入数据，包括:

将所述第一计算忆阻器阵列的所述至少一行处于关键权重位置的忆阻器单元的第一训练计算输入数据作为所述动态偏差校准子阵列的至少一行忆阻器单元的动态训练校准输入数据。

9、根据权利要求 2-5 中任一项所述的校准方法，其中，

基于所述写入了所述第一计算权重矩阵的第一计算忆阻器阵列和所述第一计算权重矩阵，对所述第一校准忆阻器阵列进行片上训练，以调整所述第一校准忆阻器阵列的权重值，还包括:

在向所述第一计算忆阻器阵列输入所述第一训练计算输入数据，并向所述第一校准忆阻器阵列输入所述第一训练校准输入数据，以得到所述第一训练校准输出数据之前，

向所述第一计算忆阻器阵列输入所述第一训练计算输入数据，以得到第一训练输出数据；

根据所述第一训练输出数据和所述第一训练目标输出数据的偏差，确定所述第一计算忆阻器阵列中需要校准的列，并使所述第一校准忆阻器阵列中对应于所述第一计算忆阻器阵列中需要校准的列的忆阻器单元与所述第一计算忆阻器阵列中需要校准的列串联接通。

10、根据权利要求 1-9 中任一项所述的校准方法，其中，

所述存算一体装置还包括第二处理单元，所述第二处理单元包括第二计算忆阻器阵列和第二校准忆阻器阵列，所述第二计算忆阻器阵列被配置为接收根据所述第一校准输出数据确定的第二计算输入数据，并对所述第二计算输入数据进行计算以得到第二输出数据；所述第二校准忆阻器阵列的被配置为接收第二校准输入数据，并根据所述第二校准输入数据对所述第二输出数据进行校准，以得到第二校准输出数据；所述第二处理单元被配置为输出所述第二校准输出数据，

所述校准方法还包括：

通过片外训练确定所述第二计算忆阻器阵列对应的第二计算权重矩阵，将所述第二计算权重矩阵写入所述第二计算忆阻器阵列；

基于写入了所述第二计算权重矩阵的所述第二计算忆阻器阵列和所述第二计算权重矩阵，对所述第二校准忆阻器阵列进行片上训练，以调整所述第二校准忆阻器阵列的权重值。

11、根据权利要求 10 所述的校准方法，其中，所述第二处理单元与所述第一处理单元耦接，

在基于写入了所述第二计算权重矩阵的所述第二计算忆阻器阵列和所述第二计算权重矩阵，对所述第二校准忆阻器阵列进行片上训练，以调整所述第二校准忆阻器阵列的权重值的过程中，

根据第一训练计算输入数据和所述第一计算权重矩阵，确定第一训练目标输出数据，并将所述第一训练目标输出数据作为所述第二计算忆阻器阵列

的第二训练输入数据，或者，

在训练好所述第一校准忆阻器阵列的情况下，向所述第一计算忆阻器阵列输入所述第一训练计算输入数据，向所述第一校准忆阻器阵列输入第一训练校准输入数据，以得到第一训练校准输出数据，并将所述第一训练校准输出数据作为所述第二计算忆阻器阵列的第二训练输入数据。

12、根据权利要求 1-11 中任一项所述的校准方法，其中，所述第一校准忆阻器阵列中的每个忆阻器单元被配置为既能够实现正权重，又能够实现负权重。

13、一种存算一体装置，包括第一处理单元，其中，所述第一处理单元包括第一计算忆阻器阵列和第一校准忆阻器阵列，

所述第一计算忆阻器阵列被配置为接收第一计算输入数据，并对所述第一计算输入数据进行计算以得到第一输出数据;

所述第一校准忆阻器阵列的被配置为接收第一校准输入数据，并根据所述第一校准输入数据对所述第一输出数据进行校准，以得到第一校准输出数据;

所述第一处理单元被配置为输出所述第一校准输出数据。

14、根据权利要求 13 所述的装置，其中，所述第一计算忆阻器阵列包括 M 行*N 列的忆阻器单元，所述第一校准忆阻器阵列包括 K 行*N 列的忆阻器单元，所述第一校准忆阻器阵列的第 i 列与所述第一计算忆阻器阵列的第 i 列串联，

所述第一校准忆阻器阵列的第 i 列的忆阻器单元被配置为根据所述第一校准输入数据对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据，

其中，M、N、K、i 均为正整数，且 $1 \leqslant i \leqslant N$。

15、根据权利要求 14 所述的装置，其中，所述第一校准忆阻器阵列包括固定偏差校准子阵列，所述第一校准输入数据包括固定校准输入数据，所述固定偏差校准子阵列包括 1 行*N 列的忆阻器单元，所述固定偏差校准子阵列的第 i 列与所述第一计算忆阻器阵列的第 i 列串联，

所述固定偏差校准子阵列的第 i 列的忆阻器单元被配置为根据所述固定校准输入数据对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，

以得到第 i 列的第一校准输出数据。

16、根据权利要求 14 所述的装置，其中，所述第一校准忆阻器阵列包括动态偏差校准子阵列，所述第一校准输入数据包括动态校准输入数据，所述动态偏差校准子阵列包括至少 1 行*N 列的忆阻器单元，所述动态偏差校准子阵列的第 i 列与所述第一计算忆阻器阵列的第 i 列串联，

所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据是根据所述第一计算忆阻器阵列的第 i 列的 M 个忆阻器单元对应的第一计算输入数据确定的，

所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元被配置为根据所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据，对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据。

17、根据权利要求 14-16 所述的装置，其中，

所述第一校准忆阻器阵列包括固定偏差校准子阵列和动态偏差校准子阵列，所述第一校准输入数据包括固定校准输入数据和动态校准输入数据，所述固定偏差校准子阵列包括 1 行*N 列的忆阻器单元，所述动态偏差校准子阵列包括至少 1 行*N 列的忆阻器单元，所述固定偏差校准子阵列的第 i 列和所述动态偏差校准子阵列的第 i 列与所述第一计算忆阻器阵列的第 i 列串联，

所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据是根据所述第一计算忆阻器阵列的第 i 列的 M 个忆阻器单元对应的第一计算输入数据确定的，

所述固定偏差校准子阵列的第 i 列的忆阻器单元和所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元被配置为分别根据所述固定校准输入数据和所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据，对所述第一计算忆阻器阵列的第 i 列的第一输出数据进行校准，以得到第 i 列的第一校准输出数据。

18、根据权利要求 15 或 17 所述的装置，其中，所述固定校准输入数据为所述第一计算输入数据的最大值的 5%-20%。

19、根据权利要求 16 或 17 所述的装置，其中，所述动态偏差校准子阵列的第 i 列的至少一个忆阻器单元对应的动态校准输入数据与所述第一计算

忆阻器阵列的第 i 列的至少一个处于关键权重位置的忆阻器单元对应的第一计算输入数据相同。

20、根据权利要求 16 或 17 所述的装置，其中，所述动态偏差校准子阵列的至少一行忆阻器单元对应的动态校准输入数据与所述第一计算忆阻器阵列的至少一行处于关键权重位置的忆阻器单元对应的第一计算输入数据相同。

100

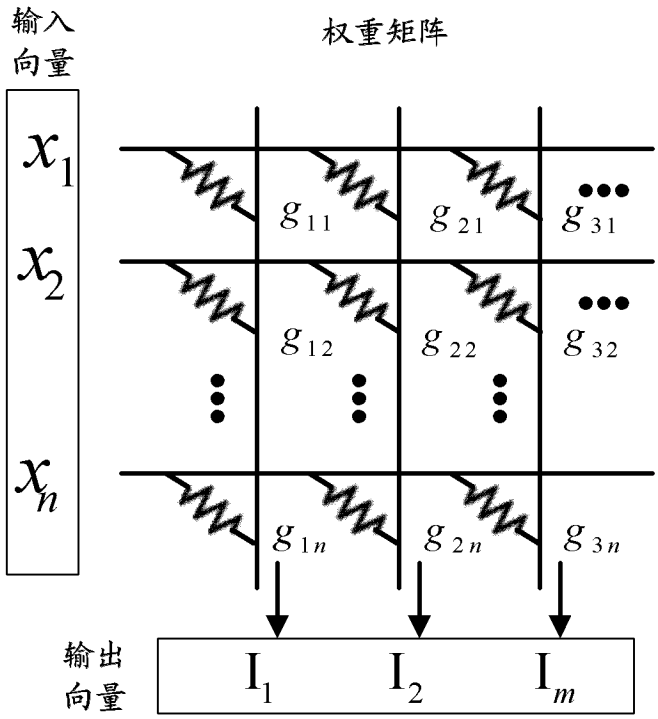

输入
向量
权重矩阵
x_1
x_2
x_n
g_{11}
g_{21}
g_{31}
g_{12}
g_{22}
g_{32}
g_{1n}
g_{2n}
g_{3n}
输出
向量
I_1
I_2
I_m

图 1

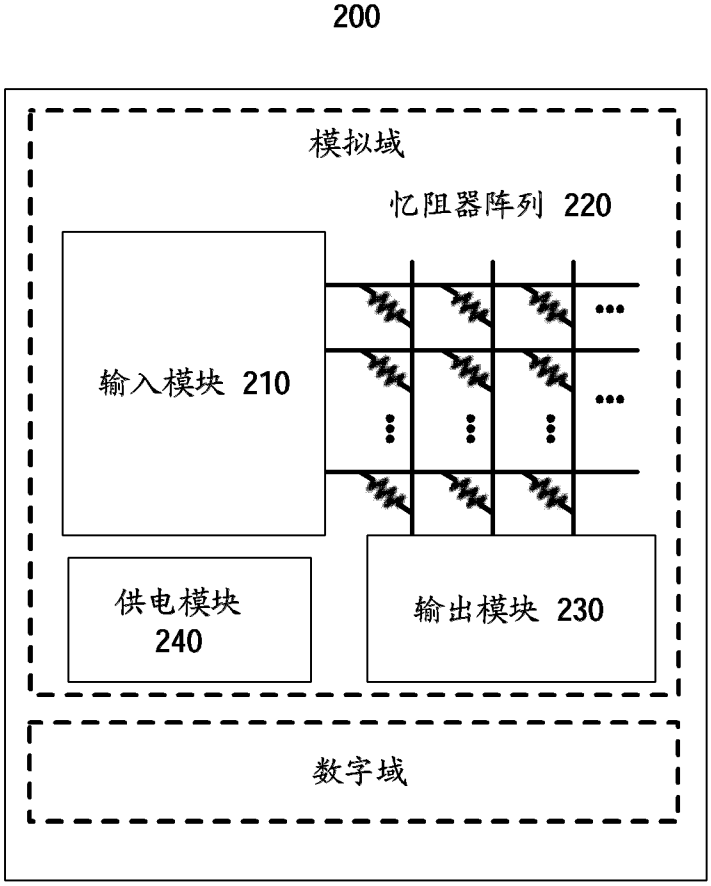


图 2

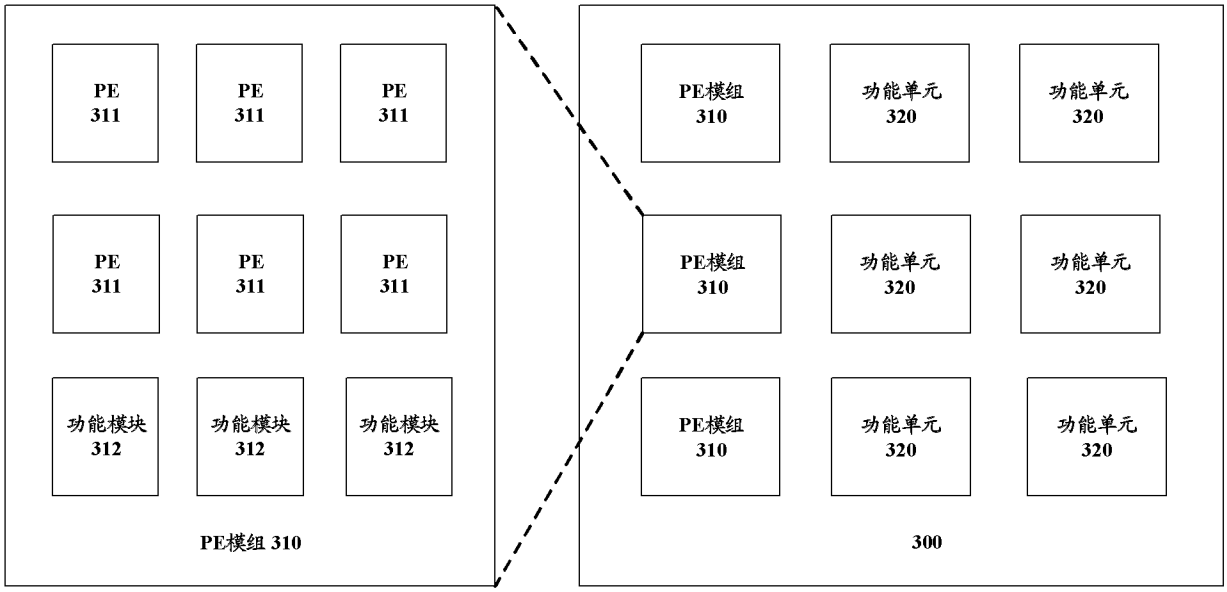

PE
311
PE
311
PE
311
PE
311
PE
311
PE
311
功能模块
312
功能模块
312
功能模块
312
PE模组 310
PE模组
310
功能单元
320
功能单元
320
PE模组
310
功能单元
320
功能单元
320
PE模组
310
功能单元
320
功能单元
320
300

图 3

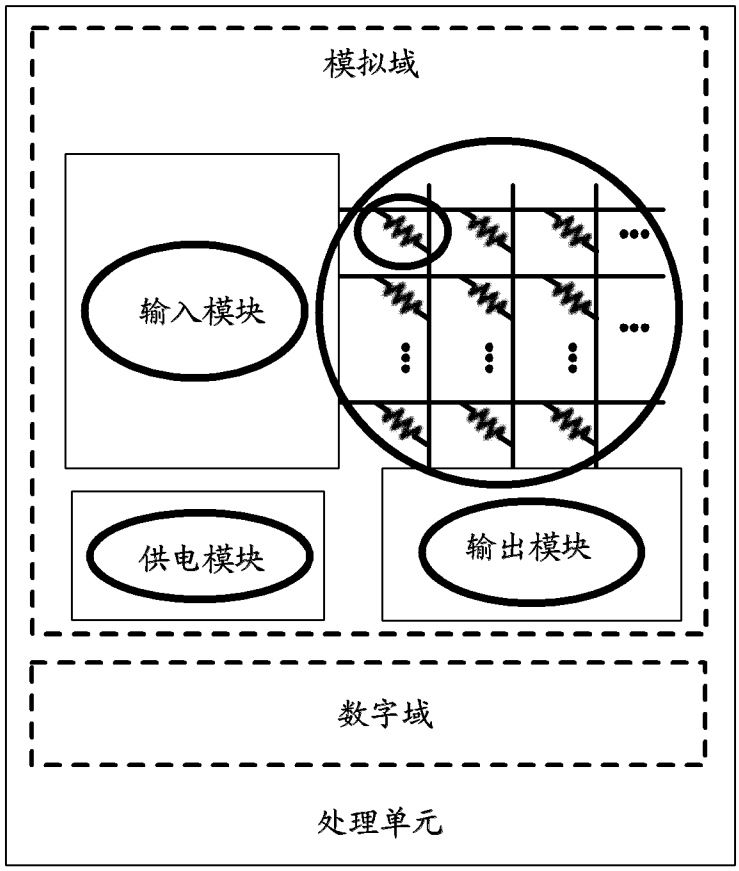


图 4A

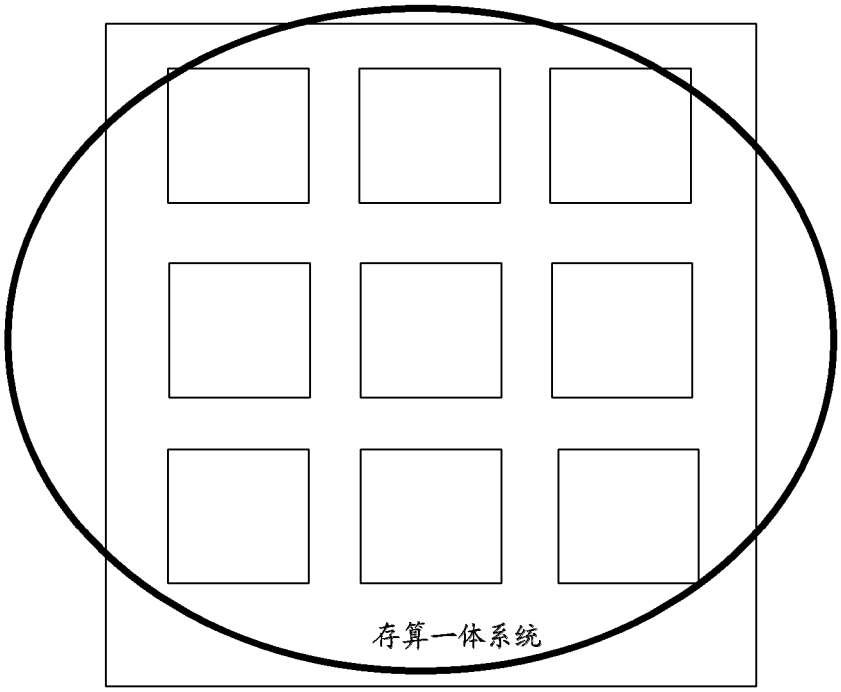


图 4B

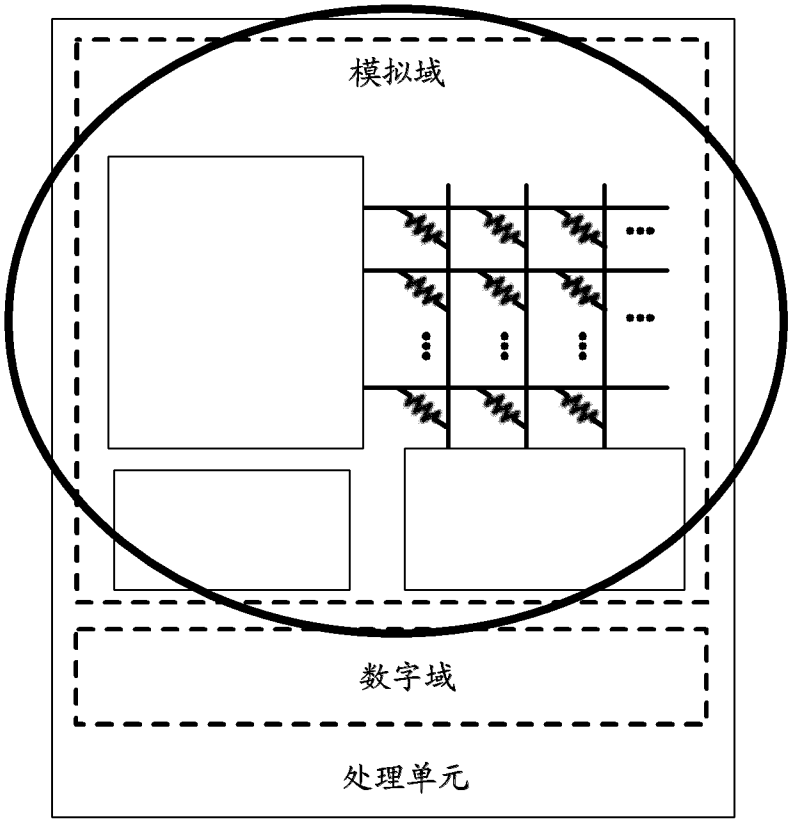


图 4C

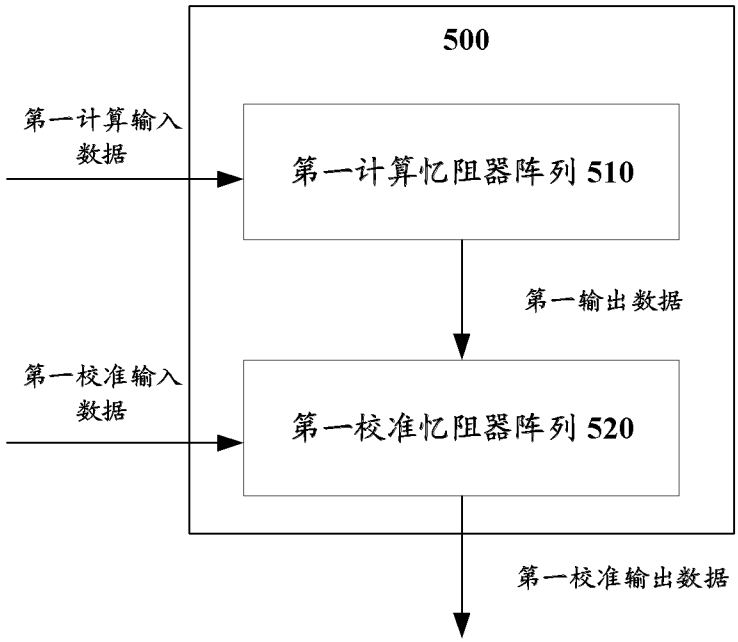

500
第一计算输入数据
第一计算忆阻器阵列 510
第一输出数据
第一校准输入数据
第一校准忆阻器阵列 520
第一校准输出数据

图 5

600

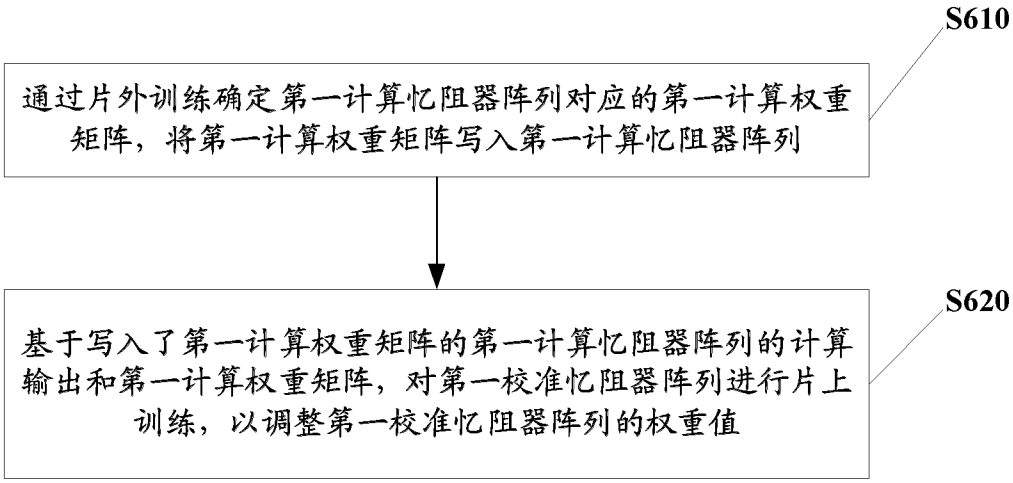

S610
通过片外训练确定第一计算忆阻器阵列对应的第一计算权重矩阵，将第一计算权重矩阵写入第一计算忆阻器阵列
S620
基于写入了第一计算权重矩阵的第一计算忆阻器阵列的计算输出和第一计算权重矩阵，对第一校准忆阻器阵列进行片上训练，以调整第一校准忆阻器阵列的权重值

图 6

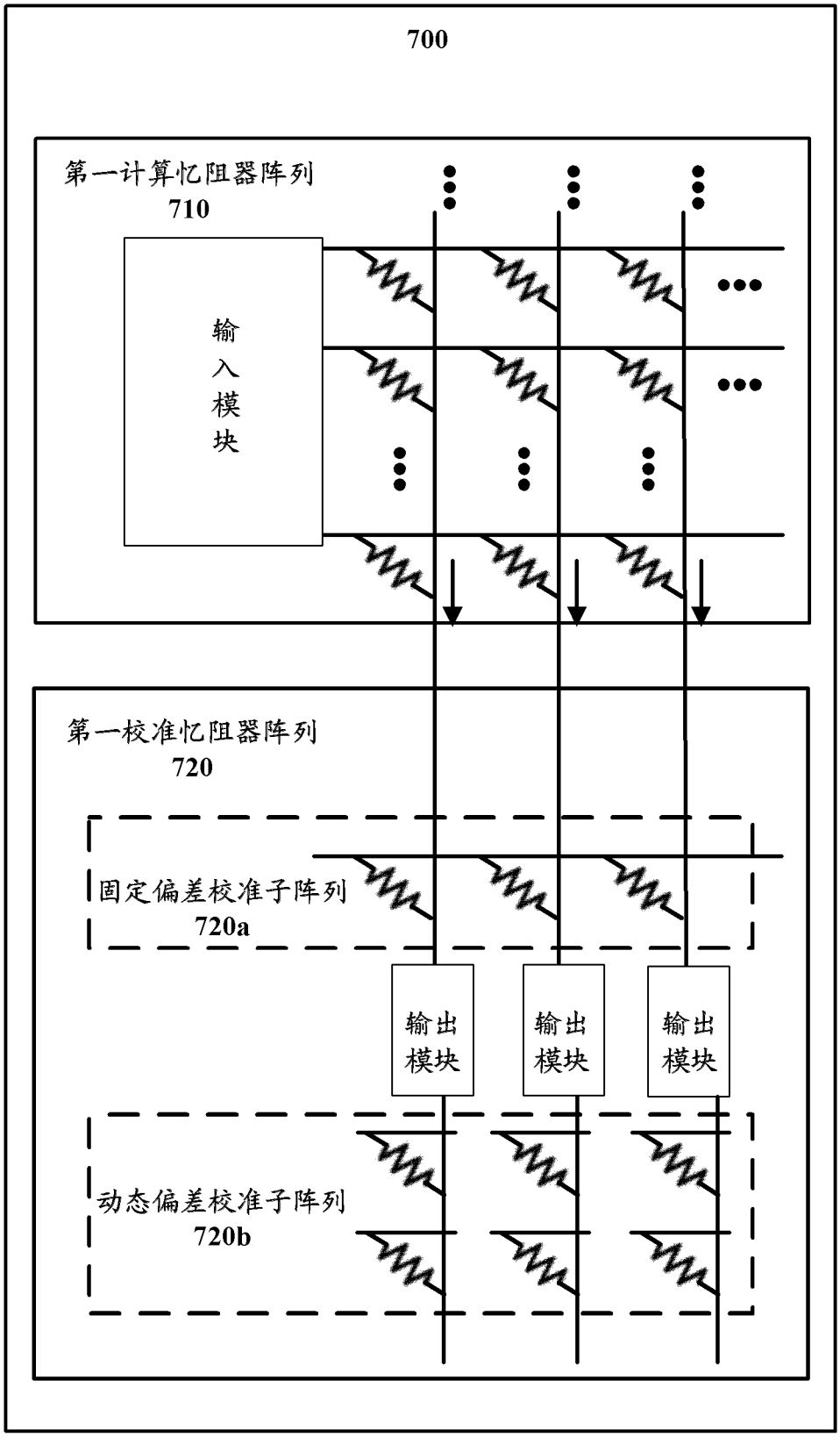

700
第一计算忆阻器阵列
710
输入模块
第一校准忆阻器阵列
720
固定偏差校准子阵列
720a
输出模块
输出模块
输出模块
动态偏差校准子阵列
720b

图 7

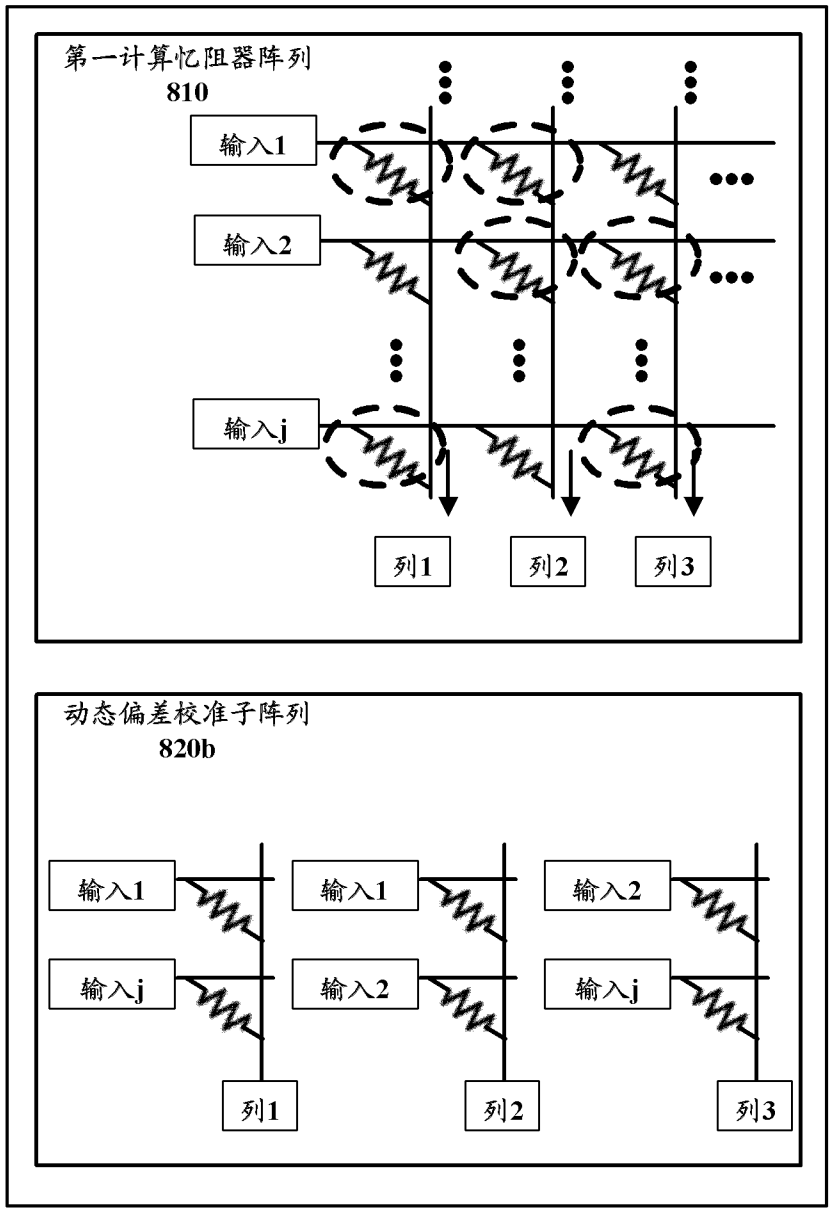

第一计算忆阻器阵列
810
输入1
输入2
输入j
列1
列2
列3
动态偏差校准子阵列
820b
输入1
输入j
输入1
输入2
输入2
输入j
列1
列2
列3

图 8

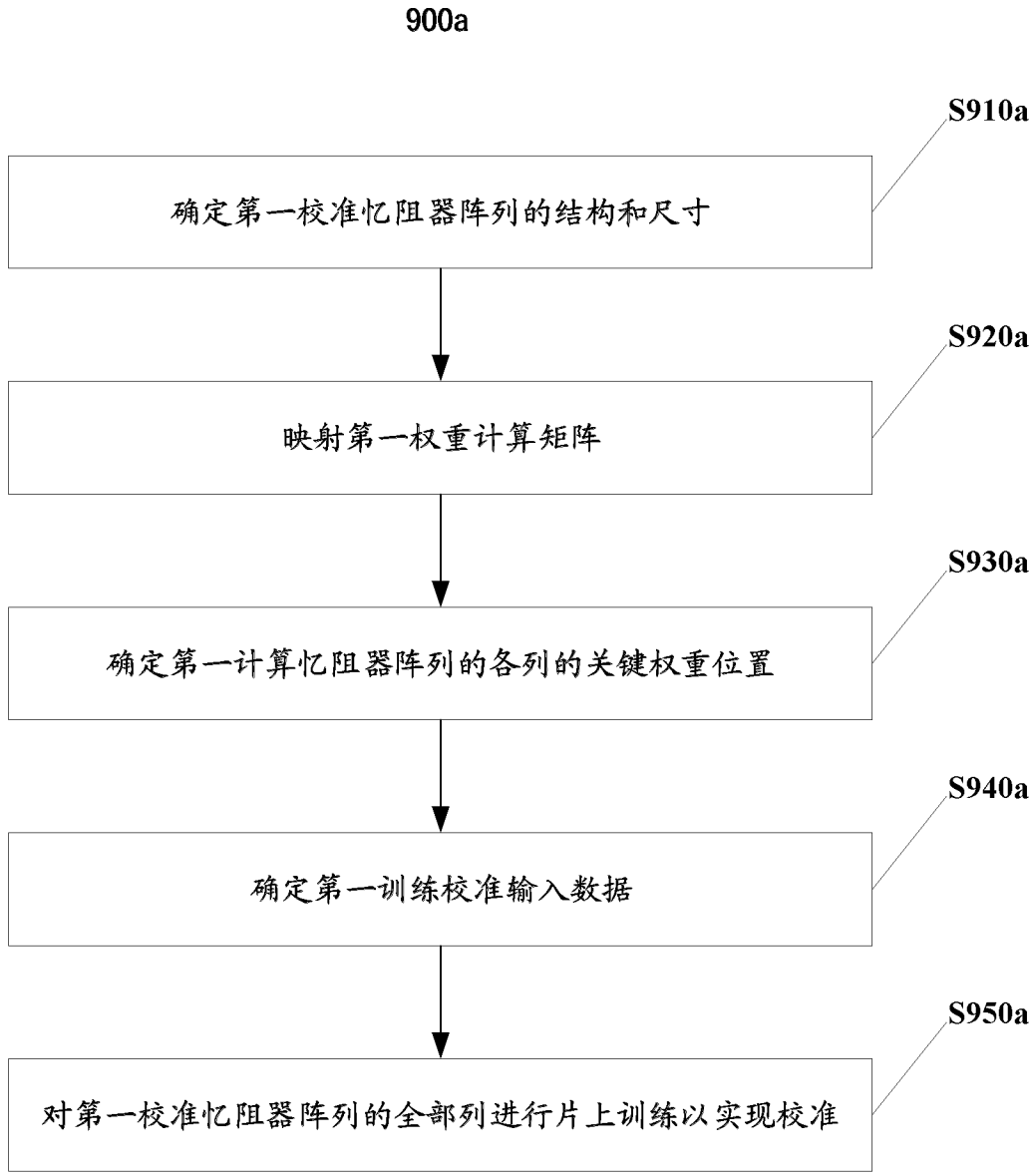

900a
S910a
确定第一校准忆阻器阵列的结构和尺寸
S920a
映射第一权重计算矩阵
S930a
确定第一计算忆阻器阵列的各列的关键权重位置
S940a
确定第一训练校准输入数据
S950a
对第一校准忆阻器阵列的全部列进行片上训练以实现校准

图 9A

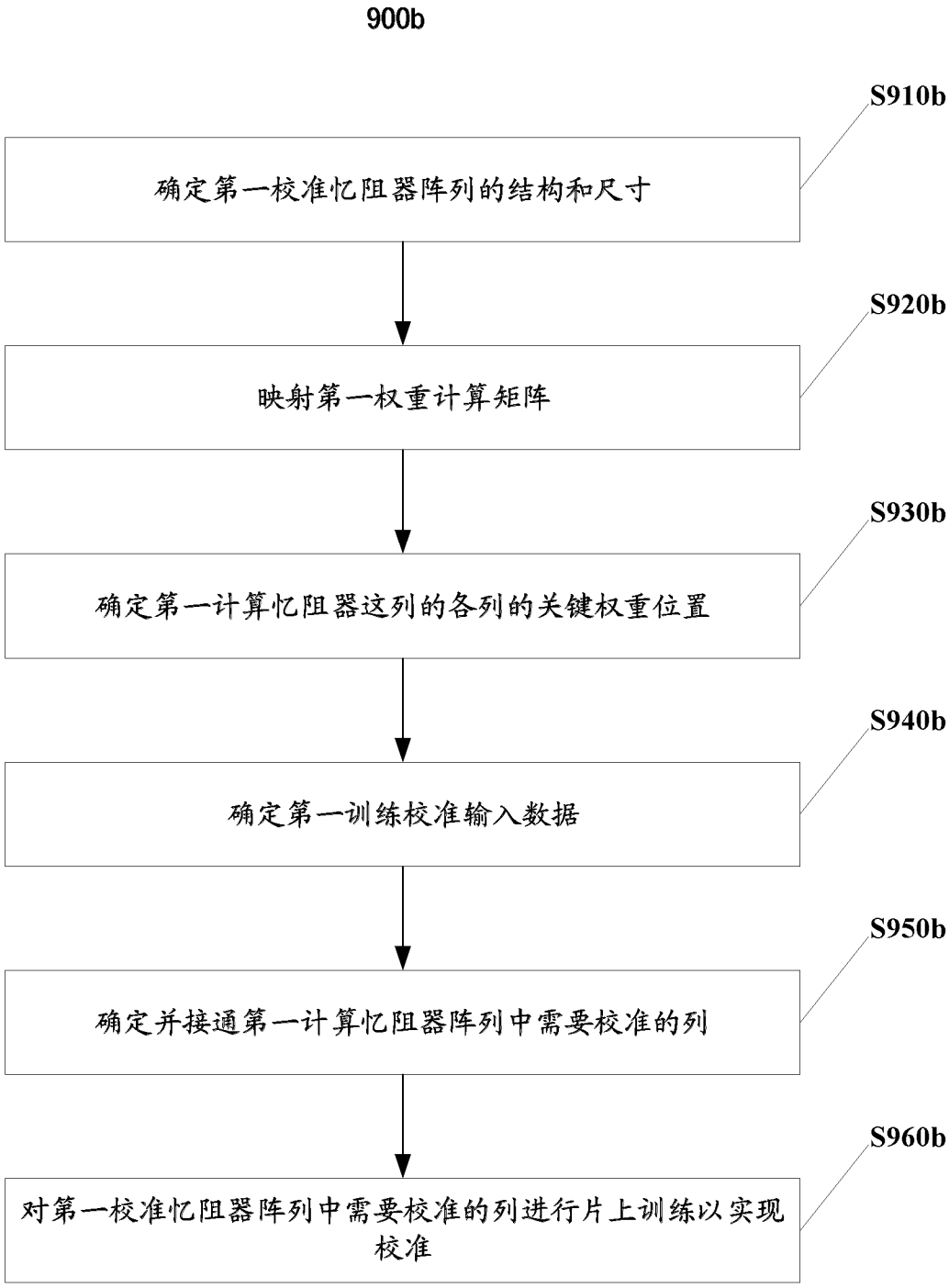

900b
S910b
确定第一校准忆阻器阵列的结构和尺寸
S920b
映射第一权重计算矩阵
S930b
确定第一计算忆阻器这列的各列的关键权重位置
S940b
确定第一训练校准输入数据
S950b
确定并接通第一计算忆阻器阵列中需要校准的列
S960b
对第一校准忆阻器阵列中需要校准的列进行片上训练以实现校准

图 9B

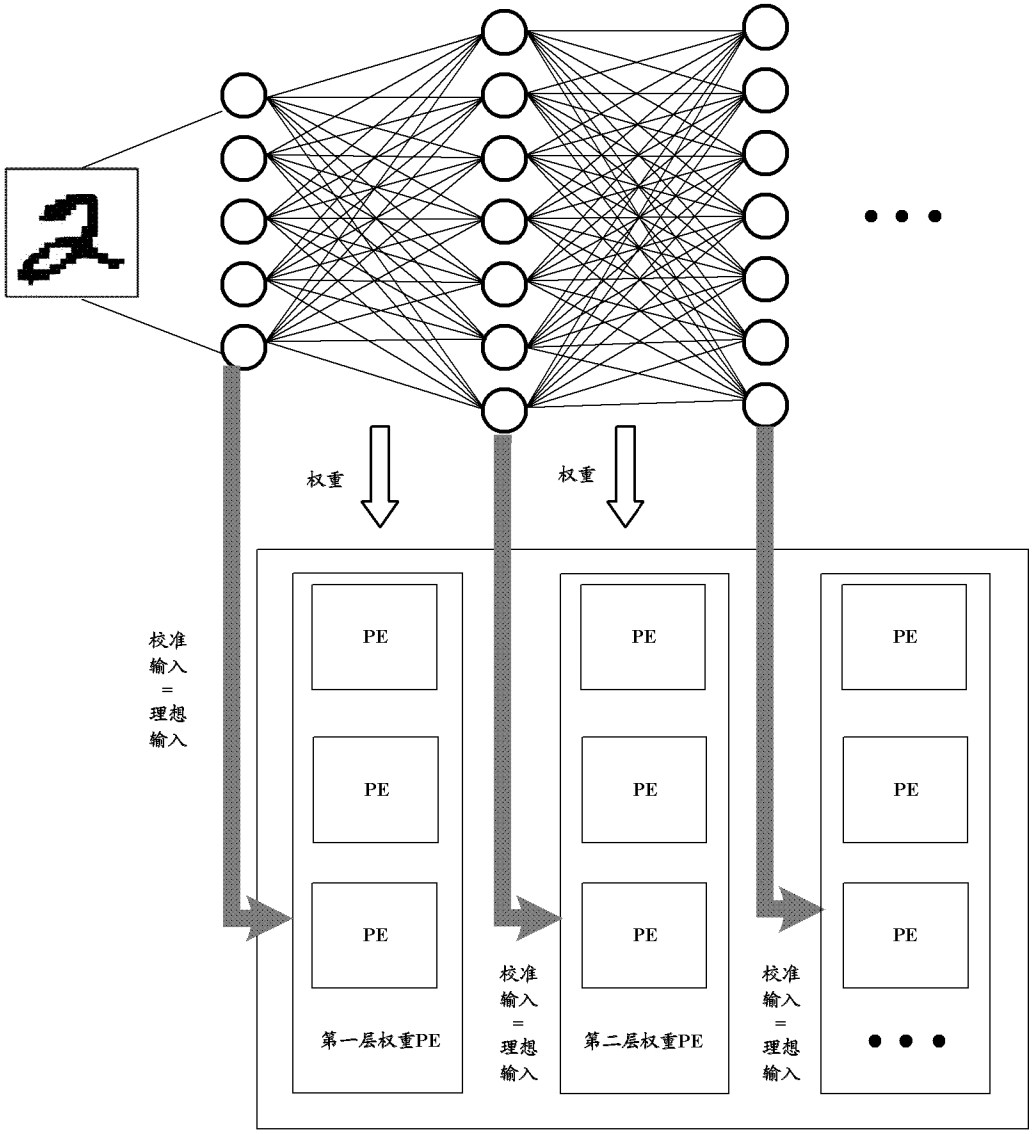

权重
权重
校准输入 = 理想输入
PE
PE
PE
PE
PE
PE
PE
PE
PE
第一层权重PE
第二层权重PE
校准输入 = 理想输入
校准输入 = 理想输入

图 10A

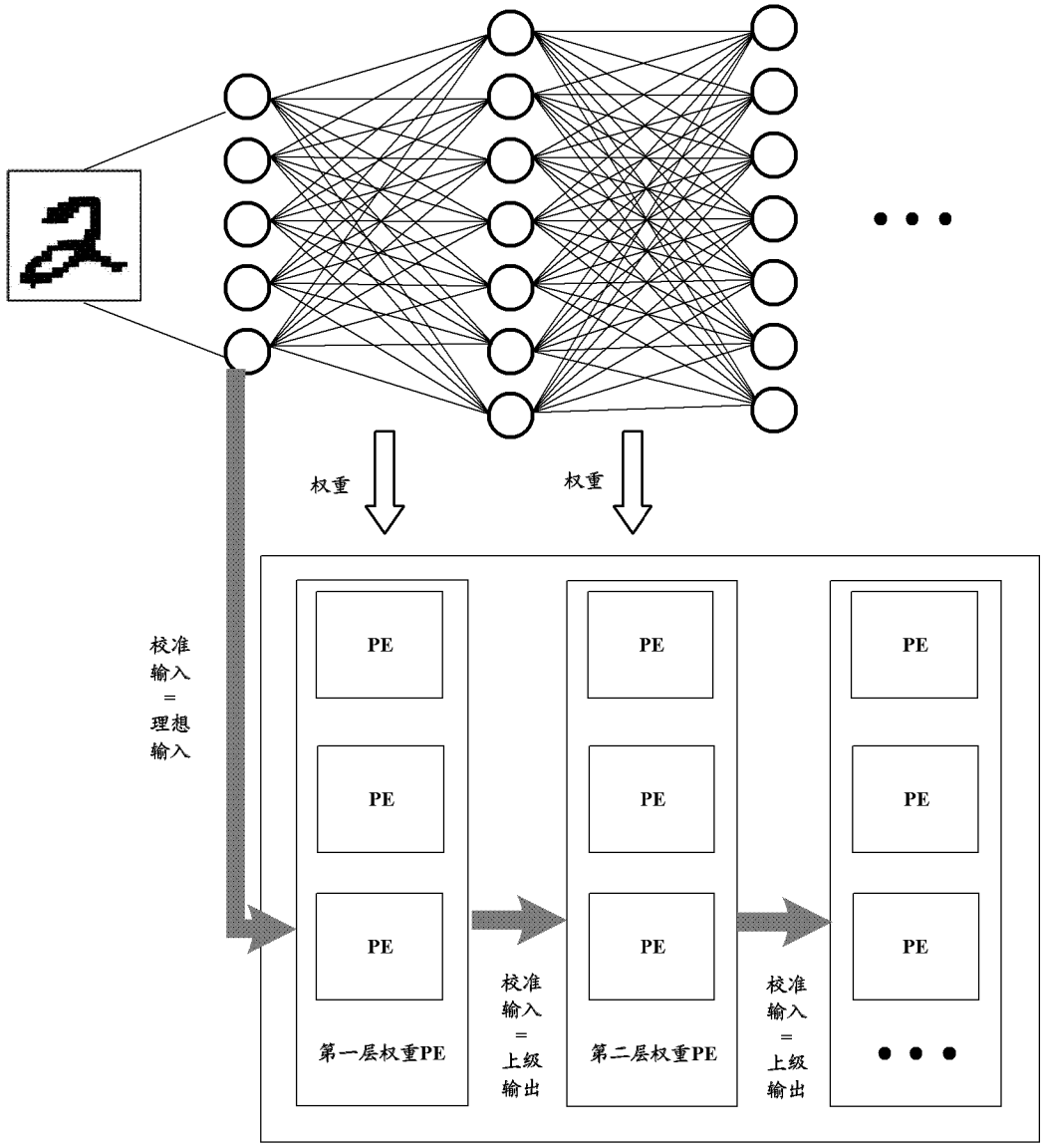

权重
权重
校准输入=理想输入
PE
PE
PE
PE
PE
PE
PE
PE
PE
第一层权重PE
第二层权重PE
校准输入=上级输出
校准输入=上级输出

图 10B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2021/137444

A. CLASSIFICATION OF SUBJECT MATTER

G06F 15/78(2006.01)i; G11C 16/34(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; G11C

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)

CNKI, CNAPT, WPI, EPODOC, IEEE: 忆阻器, 存算, 校准, 阵列, 权重, memristor, memory, calibration, array, weight

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113553293 A (TSINGHUA UNIVERSITY) 26 October 2021 (2021-10-26) description, paragraphs [0040]-[0109], and figures 1-10	1-20
A	CN 112836814 A (TSINGHUA UNIVERSITY) 25 May 2021 (2021-05-25) description, paragraphs [0041]-[0059], and figures 1-5	1-20
A	CN 113052307 A (SHANGHAI JIAO TONG UNIVERSITY) 29 June 2021 (2021-06-29) entire document	1-20
A	CN 108470746 A (PEKING UNIVERSITY) 31 August 2018 (2018-08-31) entire document	1-20
A	US 2021201136 A1 (HEWLETT PACKARD ENTERPRISE DEVELOPMENT LP) 01 July 2021 (2021-07-01) entire document	1-20

☐ 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	Date of mailing of the international search report
31 March 2022	**19 April 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/CN2021/137444

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 113553293 A	26 October 2021	None	
CN 112836814 A	25 May 2021	None	
CN 113052307 A	29 June 2021	None	
CN 108470746 A	31 August 2018	None	
US 2021201136 A1	01 July 2021	WO 2019212488 A1	07 November 2019
		CN 112041928	04 December 2020

国际检索报告

国际申请号
PCT/CN2021/137444

A. 主题的分类

G06F 15/78(2006.01)i; G11C 16/34(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; G11C

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）)

CNKI, CNAPT, WPI, EPODOC, IEEE: 忆阻器, 存算, 校准, 阵列, 权重, memristor, memory, calibration, array, weight

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 113553293 A (清华大学) 2021年10月26日 (2021-10-26) 说明书第[0040]-[0109]段，附图1-10	1-20
A	CN 112836814 A (清华大学) 2021年5月25日 (2021-05-25) 说明书第[0041]-[0059]段，附图1-5	1-20
A	CN 113052307 A (上海交通大学) 2021年6月29日 (2021-06-29) 全文	1-20
A	CN 108470746 A (北京大学) 2018年8月31日 (2018-08-31) 全文	1-20
A	US 2021201136 A1 (HEWLETT PACKARD ENTERPRISE DEVELOPMENT LP) 2021年7月1日 (2021-07-01) 全文	1-20

☐ 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	国际检索报告邮寄日期
2022年3月31日	2022年4月19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 白桦 电话号码 86-(10)-53961425

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2021/137444

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 113553293 A	2021年10月26日	无	
CN 112836814 A	2021年5月25日	无	
CN 113052307 A	2021年6月29日	无	
CN 108470746 A	2018年8月31日	无	
US 2021201136 A1	2021年7月1日	WO 2019212488 A1	2019年11月7日
		CN 112041928	2020年12月4日