

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022 年 7 月 7 日 (07.07.2022)



(10) 国际公布号
WO 2022/143432 A1

- (51) 国际专利分类号:
G06F 17/16 (2006.01) **G06F 9/30** (2006.01)
- (21) 国际申请号: PCT/CN2021/141000
- (22) 国际申请日: 2021 年 12 月 23 日 (23.12.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202011617575.X 2020 年 12 月 30 日 (30.12.2020) CN
202110181498.6 2021 年 2 月 8 日 (08.02.2021) 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)。
- (72) 发明人: 林腾毅 (LIN, Tengyi); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (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, 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))。

(54) Title: MATRIX CALCULATION APPARATUS, METHOD, SYSTEM, CIRCUIT, CHIP, AND DEVICE

(54) 发明名称: 一种矩阵计算装置、方法、系统、电路、芯片及设备

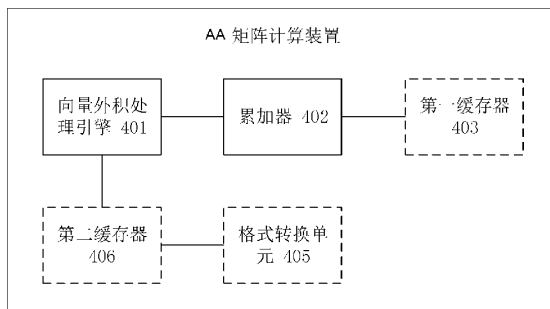


图 4A

401 Vector outer product processing engine
402 Accumulator
403 First buffer
405 Format conversion unit
406 Second buffer
AA Matrix calculation apparatus

(57) Abstract: A matrix calculation apparatus, a method, a system, a circuit, a chip and a device, being able to directly calculate a matrix in a compressed format, thereby improving the calculation efficiency of the matrix in a compressed format. The matrix calculation apparatus comprises: a vector outer product processing engine and an accumulator. The matrix calculation apparatus calculates a first matrix and a second matrix in a compressed format on the basis of a vector outer product. During the calculation process, the vector outer product processing engine performs vector outer product calculation on a first column vector with row coordinates retained and a second column vector with column coordinates retained, and then the accumulator accumulates, on the basis of the index of position coordinates, third element values with the same position coordinates, so as to obtain a result matrix obtained by calculating the two matrices in a compressed format. Compared with traditional methods in which a matrix in a compressed format needs to be decompressed first and then matrix calculation is performed on the decompressed matrix, the matrix calculation apparatus provided in the embodiments of the present application can effectively improve the calculation efficiency of a matrix in a compressed format.

[见续页]

(57) 摘要：一种矩阵计算装置、方法、系统、电路、芯片及设备，能够直接对压缩格式的矩阵进行计算，从而提升压缩格式的矩阵的计算效能。矩阵计算装置包括：向量外积处理引擎和累加器，矩阵计算装置基于向量外积对压缩格式的第一矩阵和第二矩阵进行计算，在计算的过程中，向量外积处理引擎对保留了行坐标的第一列向量和保留了列坐标的第二列向量进行向量外积计算，然后累加器基于位置坐标的索引，将位置坐标相同的第三元素值进行累加，从而得到两个压缩格式的矩阵进行计算的结果矩阵，相对于传统的需要将压缩格式的矩阵进行先解压缩，然后对解压缩后的矩阵进行矩阵计算的方法，本申请实施例中提供的矩阵计算装置能够有效提升压缩格式的矩阵的计算效能。