

(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



(10) 国际公布号
WO 2024/183577 A1

(43) 国际公布日
2024 年 9 月 12 日 (12.09.2024)

(51) 国际专利分类号:
G06F 7/483 (2006.01)

(21) 国际申请号: PCT/CN2024/078993

(22) 国际申请日: 2024 年 2 月 28 日 (28.02.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
2023 10238789.3 2023 年 3 月 3 日 (03.03.2023) 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: 伍玮翔 (WU, Wei Hsiang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。罗元勇 (LUO, Yuanyong); 中国广东省深圳市龙岗区坂田华为总部办公楼,

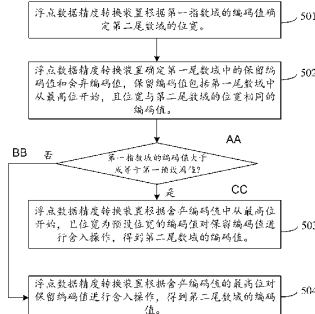
Guangdong 518129 (CN)。陈敏琪 (CHEN, Minqi); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。张忠星 (ZHANG, Zhongxing); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 北京中博世达专利商标代理有限公司 (BEIJING ZBSD PATENT & TRADEMARK AGENT LTD.); 中国北京市海淀区交大东路 31 号 11 号楼 8 层, Beijing 100044 (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: FLOATING-POINT DATA PRECISION CONVERSION METHOD AND APPARATUS

(54) 发明名称: 浮点数据精度转换方法和装置



501 A floating-point data precision conversion apparatus determines the bit width of a second mantissa domain according to an encoded value of a first exponential domain.
502 The floating-point data precision conversion apparatus determines a reserved encoded value in a first mantissa domain, the reserved encoded value comprising an encoded value starting from the highest bit in the first mantissa domain and having the same bit width as the bit width of the second mantissa domain.
503 The floating-point data precision conversion apparatus performs a rounding operation on the reserved encoded value according to an encoded value starting from the highest bit in the discarded encoded value and having a bit width of a preset bit width to obtain an encoded value of the second mantissa domain.
504 The floating-point data precision conversion apparatus performs the rounding operation on the reserved encoded value according to the highest bit of the discarded encoded value to obtain the encoded value of the second mantissa domain.
AA Is the encoded value of the first exponential domain greater than or equal to a first preset threshold?
BB No
CC Yes

图 5

(57) Abstract: Embodiments of the present application relate to the technical field of chips, and provided thereby are a floating-point data precision conversion method and apparatus, which reduce conversion errors when converting high-precision data to low-precision data. A specific solution is: according to an encoded value of a first exponential domain, determining the bit width of a second mantissa domain; determining a reserved encoded value and a discarded encoded value in a first mantissa domain, the reserved encoded value comprising an encoded value starting from the highest bit in the first mantissa domain and having the same bit width as the bit width of the second mantissa domain; if the encoded value of the first exponential domain is greater than or equal to a first preset threshold, performing a rounding operation on the reserved encoded value according to an encoded value starting from the highest bit in the discarded encoded value and having a bit width of a preset bit width to obtain an encoded value of the second mantissa domain; and if the encoded value of the first exponential domain is less than the first preset threshold, performing the rounding operation on the reserved encoded value according to the highest bit of the discarded encoded value to obtain the encoded value of the second mantissa domain. The embodiments of the present application are used for a floating-point data precision conversion process.

[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: 本申请实施例提供一种浮点数据精度转换方法和装置, 涉及芯片技术领域, 降低了高精度数据向低精度数据转换时的转换误差。具体方案为: 根据第一指数域的编码值确定第二尾数域的位宽; 确定第一尾数域中的保留编码值和舍弃编码值, 保留编码值包括第一尾数域中从最高位开始, 且位宽与第二尾数域的位宽相同的编码值; 若第一指数域的编码值大于或等于第一预设阈值, 根据舍弃编码值中从最高位开始, 且位宽为预设位宽的编码值对保留编码值进行舍入操作, 得到第二尾数域的编码值; 若第一指数域的编码值小于第一预设阈值, 根据舍弃编码值的最高位对保留编码值进行舍入操作, 得到第二尾数域的编码值。本申请实施例用于浮点数据精度转换的过程。