

## (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  
2020 年 11 月 19 日 (19.11.2020)



(10) 国际公布号  
**WO 2020/228140 A1**

- (51) 国际专利分类号:  
*G16C 60/00* (2019.01) *G06F 17/00* (2019.01)  
*G06F 7/00* (2006.01)
- (21) 国际申请号: PCT/CN2019/097972
- (22) 国际申请日: 2019 年 7 月 26 日 (26.07.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201910390012.2 2019 年 5 月 10 日 (10.05.2019) CN
- (71) 申请人: 北京大学深圳研究生院 (PEKING UNIVERSITY SHENZHEN GRADUATE SCHOOL) [CN/CN]; 中国广东省深圳市

南山区西丽深圳大学城北大园区, Guangdong 518055 (CN)。

(72) 发明人: 潘锋 (PAN, Feng); 中国广东省深圳市南山区西丽深圳大学城北大园区, Guangdong 518055 (CN)。 翁谋毅 (WENG, Mouyi); 中国广东省深圳市南山区西丽深圳大学城北大园区, Guangdong 518055 (CN)。 王志 (WANG, Zhi); 中国广东省深圳市南山区西丽深圳大学城北大园区, Guangdong 518055 (CN)。

(74) 代理人: 深圳鼎合诚知识产权代理有限公司 (DHC IP ATTORNEYS); 中国广东省深圳市福田区金田路与福华路交汇处现代商务大厦 1810, Guangdong 518048 (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,

(54) Title: METHOD FOR DESCRIBING CRYSTAL STRUCTURE OF MATERIAL AND APPLICATION THEREOF

(54) 发明名称: 一种描述材料晶体结构的方法及其应用

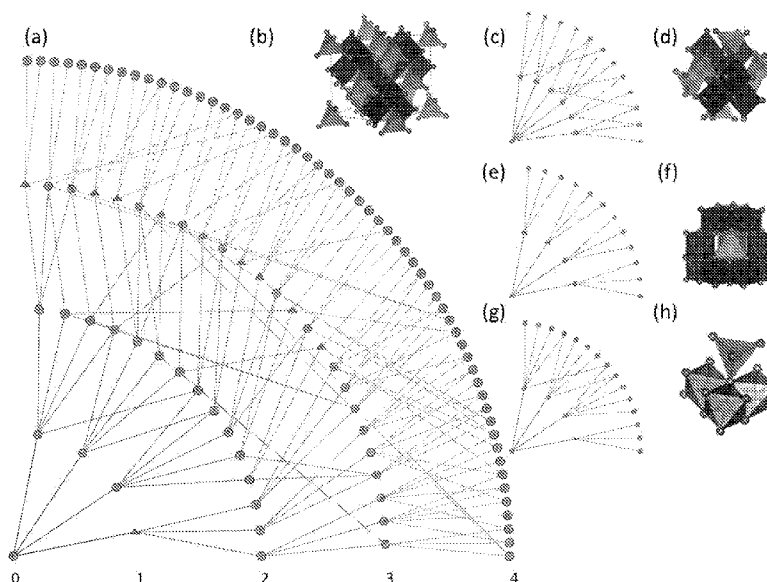


图 1

(57) Abstract: A method for describing the crystal structure of a material and an application thereof, comprising taking atoms and/or structural motifs in a crystal structure as points; taking connections between the atoms and/or structural motifs as edges; and using the points, the edges, and interrelationships between the points and the edges to describe connection relationships between the atoms and/or structural motifs in the crystal structure, thus achieving the description of the crystal structure. The creative use of points and edges to describe the crystal structure of the material may store a large amount of tedious point and line connection information by means of computer languages, which is convenient for computer operations. Conveniently and quickly constructing the atom and crystal structure

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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。

of the physical world into a computer language may directly utilize existing algorithms in the computer field to compare, analyze and calculate the properties of materials, may process crystal structures in large-scale batches, and lays a foundation for the construction of material databases and the computer analysis of material properties.

**(57) 摘要:** 一种描述材料晶体结构的方法及其应用, 包括将晶体结构中的原子和/或结构基元作为点; 把原子和/或结构基元之间的连接作为边; 采用点、边以及点与边的相互关系描述晶体结构中的原子和/或结构基元之间的连接关系, 从而实现晶体结构的描述。创造性的利用点和边的方式描述材料的晶体结构, 能通过计算机语言存储大量而又繁琐的点与线的连接信息, 方便计算机运算。便捷地将物理世界的原子、晶体结构构建到计算机语言中, 可直接利用计算机领域的现有算法对材料的性质进行比对、分析和运算; 可大规模批量处理晶体结构; 为材料数据库构建和材料性质计算机分析奠定了基础。

## 一种描述材料晶体结构的方法及其应用

### 技术领域

[0001] 本申请涉及材料晶体结构分析技术领域，特别是涉及一种描述材料晶体结构的方法及其应用。

### 背景技术

[0002] 材料晶体结构作为材料特征的重要组成部分，被广泛应用于生活中的各方各面。为了加速晶体结构的研发进程，计算机模拟晶体的方法正在逐渐发展并开始在新材料研发中发挥作用。在最近五年，随着材料基因组计划的发展，材料结构数据库和计算机算法应用于材料领域的趋势愈加明显。有越来越多的材料研发开始采用从计算机设计再到实验合成的过程。在这些过程中，材料数据库也发挥着重要作用。随着计算机方法逐渐使用到材料领域中，如何将材料的问题转换为计算机问题，就成为了目前材料问题研究的重点。

### 发明概述

### 技术问题

[0003] 传统的材料晶体结构描述使用了点群空间群和晶胞周期性的描述方法，与实验中X射线衍射数据可以直接关联，具有很好的被人理解和分析的能力和作用。但是传统的材料晶体结构描述方法不适用于大数据时代的晶体结构分析，这主要是由于在大数据时代，数据库中晶体结构数量常常会达到几十万乃至上百万的规模，对如此多的结构进行分析首先涉及到识别结构或者对重复结构进行去重。如果采用传统的点群空间群或者晶胞周期性的描述方法，一般是通过键长和键角去定义结构的差异，这样得到的差异是一个连续函数，如何设置差异阈值去识别不同结构是很困难的事情。如果设置全局阈值，得到的识别结果必然不合理，也会对于少数畸变结构产生识别错误；而如果针对特定结构去设置差异阈值，则需要人为的对上百万的晶体结构进行逐一分析和处理，也难以实现。因此，现有的传统的材料晶体结构描述方法难以实现通过计算机对材料晶体结构进行比对、分析和运算，不适用于大数据时代的晶体结构分析。

[0004] 因此，亟需研发一种能够便于计算机对材料晶体结构进行比对、分析和运算的新的材料晶体结构描述方法，以满足日益发展的材料晶体结构研发的需求。

## 问题的解决方案

### 技术解决方案

[0005] 本申请的目的是提供一种新的描述材料晶体结构的方法及其应用。

[0006] 本申请采用了以下技术方案：

[0007] 本申请的一方面公开了一种描述材料晶体结构的方法，包括：

[0008] 将晶体结构中的原子和/或结构基元作为点；

[0009] 把原子和/或结构基元之间的连接作为边；

[0010] 采用点、边以及点与边的相互关系描述晶体结构中的原子和/或结构基元之间的连接关系，从而实现材料晶体结构的描述。其中，结构基元是指由两个或多个原子组合在一起的基本单元。

[0011] 需要说明的是，本申请的材料晶体结构描述方法，将晶体结构中的原子和/或结构基元表述为点，将原子和/或结构基元的连接关系表述为边，通过对点和边的描述，实现晶体结构的描述，并通过分析点与边的相互关系，对晶体结构进行比对、分析和运算；这种描述方法能够很容易的采用计算机语言实现。因此，本申请的描述方法与传统的晶体结构描述方法相比：（1）本申请的描述方法，存储了大量而又繁琐的点与线的连接信息，这些内容虽然不适用于人直观的分析判断，但是能够方便计算机的运算，因此，本申请的描述方法特别适用于采用计算机进行大数据分析；（2）本申请的描述方法，能够便捷地将物理世界的原子、晶体结构构建到计算机语言中，可以利用计算机领域的现有的算法对材料的性质进行分析和运算，不仅准确、高效，而且大大减小了劳动力成本，对材料数据库的构建和应用具有重要意义。总的来说，传统的晶体结构描述方法，人类可以了解并分析现有的晶体结构，但是不利于计算机分析和运算；本申请的描述方法将晶体结构转变为可被计算机语言表述和识别的点和边的描述，计算机可以批量处理大量的晶体结构，并能够对这些结构进行比对、分析和运算。因此，本申请的描述方法可以方便的应用于基于计算机分析的材料数据库构建和材料性质分析等领域。

- [0012] 优选的，本申请的材料晶体结构描述方法中，采用点、边以及点与边的相互关系描述晶体结构中的原子和/或结构基元之间的连接关系，具体包括采用原子半径重叠方法判断连接关系，或者采用寻找最近邻原子方法判断连接关系。
- [0013] 优选的，采用原子半径重叠方法判断连接关系，具体包括以下步骤，
- [0014] 对于晶体结构中元素的配对成键判断提供一个键长参数作为距离的参考阈值，其中，参考阈值为两个原子的半径之和；
- [0015] 对给定晶体结构进行扩胞操作，扩胞操作包括以给定的晶体结构为原始晶胞，以原始晶胞为基础单元，并将原始晶胞作为中心，在三维空间中连续复制若干次后堆积成超胞；
- [0016] 对超胞所包含的每个原子进行查找，每次以选定原子作为起点，依次计算选定原子和其他所有原子的距离，若该距离小于或等于参考阈值，则视作两个原子之间有连接关系；查找完成所有的原子，即构建获得超胞代表的晶体结构的原子之间的连接关系；
- [0017] 结构基元之间的连接关系通过结构基元之间是否有共用原子判断，如果有共用原子则两个结构基元之间有连接关系，如果没有共用原子则通过结构基元中的原子与其它结构基元中的原子之间的连接关系进行判断。
- [0018] 需要说明的是，参考阈值，例如对于碳-碳而言，其键长为1.89002埃，该值是两个碳原子的原子半径之和，该值即判断两个原子是否为碳-碳连接关系的参考阈值。另外，对于结构基元，可以理解，一个结构基元是由两个或多个原子组成，如果两个结构基元共用原子，则可以直接判断两个结构基元有连接；如果两个结构基元没有共用原子，则结构基元的连接判断方法实际上就是结构基元中的原子与其它结构基元中的原子之间的连接关系的判断方法，即采用参考阈值判断两个结构基元中的原子是否有连接，以此判断结构基元是否有连接。
- [0019] 优选的，采用寻找最近邻原子方法判断连接关系，其特点在于实现并不需要给出一系列两个原子之间的参考阈值，而只通过晶体结构中的晶胞常数和原子位置信息便可判断原子的连接关系，具体包括以下步骤，
- [0020] 对给定晶体结构进行扩胞操作，扩胞操作包括以给定的晶体结构为原始晶胞，以原始晶胞为基础单元，并将原始晶胞作为中心，在三维空间中连续复制若干

次后堆积成超胞；

[0021] 对超胞中原始晶胞的每个原子进行查找，每次选定原子之后，构建该选定原子的魏格纳-塞兹原胞（Wigner-Seitz cell，缩写WS cell），如果选定原子与某一原子连线的垂直平分面在最小的魏格纳-塞兹原胞表面，则选定原子与该原子有连接，查找完成所有的原子，即构建获得超胞代表的晶体结构的原子之间的连接关系；

[0022] 结构基元之间的连接关系通过结构基元之间是否有共用原子判断，如果有共用原子则两个结构基元之间有连接关系，如果没有共用原子则通过结构基元中的原子与其它结构基元中的原子之间的连接关系进行判断。

[0023] 需要说明的是，其中，构建该选定原子的魏格纳-塞兹原胞的具体方法可以参考现有的魏格纳-塞兹原胞构建，例如，本申请的一种实现方式中，具体包括，每次选定原子之后，从该选定原子出发依次和其他所有原子连接一条线段，并找出空间中该线段的垂直平分面，假设超胞内共计有 $n$ 个原子，即可获得 $n-1$ 个垂直平分面；从获得的 $n-1$ 个垂直平分面内，任意连续不重复的取出三个面，利用行列式判断三面是否存在共点并利用相关几何知识求解三个平面的交点坐标，如果存在平行情况则跳过，重新任选三个面；从三个平面的交点开始依次计算该交点和超胞内的所有原子的距离，如果该交点和选定原子之间的距离是最小的，且连续有三个相等的最小值，并且没有重复交点情况，则视作该交点有效，记录下获得该交点的三个平面，并由三个平面回溯找到对应的超胞内的原子，记录这三个平面对应的原子的元素种类和位置；当完成所有交点的遍历判断，即完成针对选定原子的魏格纳-塞兹原胞的建立；最后，根据记录的结果，对于晶体结构中的给定原子，判断出的其最近邻原子，认为与给定原子均有连接关系，即构建获得晶体结构的原子之间的连接关系。

[0024] 优选的，本申请的材料晶体结构的描述方法还包括从每个原子或结构基元为中心出发向外扩展若干层，将向外扩张的晶体结构转变为一个有限大小的图，具体包括，构建图步骤和向外扩展点的层数判断步骤；

[0025] 构建图的步骤包括，用点和边描述晶体结构中的连接关系，再将这些点和边构建为图；

- [0026] 向外扩展点的层数判断步骤包括，根据研究目标，设定构建的图的扩充层数，将层数以内的点和边，构建成一个有限大小的图。
- [0027] 需要说明的是，向外扩展的层数，一方面，取决于晶胞的大小，可以理解，晶胞越大向外扩展的层数越多，反之则反；另一方面，也取决于所设定的研究目标，例如，如果需要区分多相物质之间的不同相，则扩展层数就必须扩展到至少能够有效区分不同相的层数，并且随着层数的继续增加区分效果不会增强。
- [0028] 本申请的另一面公开了本申请的材料晶体结构描述方法在材料晶体结构中独立原子或独立结构基元研究、材料晶体结构比对、材料晶体结构差异分析、二维材料研发、材料数据库构建或者基于机器学习的材料性质预测、分析或新材料设计中的应用。
- [0029] 可以理解，本申请的材料晶体结构描述方法，可以使用计算机对材料晶体结构进行大批量的比对、分析和运算，因此，本申请的材料晶体结构描述方法特别适用于材料晶体结构中独立原子或独立结构基元研究、材料数据库构建、材料晶体结构比对、材料晶体结构差异性分析、二维材料构建，以及基于机器学习的材料性质预测和分析、基于机器学习的新材料设计等。
- [0030] 本申请的再一面公开了一种发现材料晶体结构中的独立原子或独立结构基元的方法，包括采用本申请的材料晶体结构描述方法对材料的晶体结构进行描述，以晶体结构中的每一个原子或结构基元作为起始点，从起始点出发，分别向外构建有限层数的图，然后分析这些图是否完全相同；所有不完全相同的图对应的起始点即为独立原子或独立结构基元。
- [0031] 需要说明的是，基于本申请的发现材料晶体结构中的独立原子或独立结构基元的方法，本申请进一步的还可以利用发现的独立原子或独立结构基元进行晶体结构比对，例如，如果两个晶体结构中的独立原子或独立结构基元的种类比例和类型安全相同，则这两个晶体结构完全相同。
- [0032] 另外，如果不进行独立原子或独立结构基元的判断，也可以直接根据构建的图的同构性判断两个晶体结构是否相同。本申请的同构性是指按照特定的运算规则比较两个图是否完全相同，例如图4，图中左边的图G和右边的图H，通过肉眼观察，两个结构为不同的形状；但是如果使用最右侧的运算规则，两个结构

的连接方式完全相同，两个结构为同构。

[0033] 本申请的再一面公开了一种材料晶体结构差异性分析的方法，其包括采用本申请的材料晶体结构描述方法对材料的晶体结构进行描述，将晶体结构构建为图，并利用所构建的图中点和边的差异性描述材料晶体结构的差异性。对于两个不同的晶体结构转变后的两个图，为了得到两个图的差异性，我们采用图编辑距离（Graph edit distance，简称GEM）方法。具体的，从其中一个图若变为另一个图，中间需要经过一系列操作，在这里定义所有的操作共6种，分别是针对点和边的插入、删除和替换操作，对每一种操作规定操作代价。然后利用A\*启发式搜索算法，遍历第一个图的每个点，每次考虑上述的6种操作并且尽可能的优先使用操作代价低的操作，每次延伸某种操作后更新A\*启发函数的评估代价项，当完成遍历后，即获得了从其中一个图变为另一个图的一系列顺序操作及总的操作代价和，这里代价和即为图的差异性。

[0034] 需要补充说明的是，任意给定两个图，两个图之间的相互转化经过的一系列操作是不唯一的，但上述的方法可以在给定操作代价的前提下，获得所有可能的一系列操作中使操作代价和最小的一种，保证了获得图的差异性的唯一性和完备性。

[0035] 本申请的再一面公开了一种二维材料研发的方法，其包括采用本申请的材料晶体结构描述方法对材料的晶体结构进行描述，将晶体结构构建为图，并利用所构建的图中的任意一点是否能通过边的连接在二维平面内扩展，而这些二维平面之间没有边的连接，判断晶体结构能否形成二维材料；如果是则晶体结构能形成二维材料，否则不能形成二维材料。

[0036] 本申请的一种实现方式中，具体的，对给定的晶体结构进行扩胞操作获得超胞后，将超胞转化为多张图，每张图仅由原始晶胞的某个原子组成，然后分别对这多张图进行图的连通性操作，根据深度优先算法，由某个原子出发，遍历与其相连的原子，利用递归算法对每个遍历到的原子继续利用深度优先去继续遍历所有可以连接的原子，直到无法继续搜索，并返回已经搜索到的原子的信息，储存到连通分量中；通过对每张图进行图连通搜索，获得基于每张图的不同连通分量；对找到的连通分量进行判断，利用点线面的几何知识，排除一维

线性材料和小块体材料的情况；最后针对得到的二维材料，根据二维材料的扩展方向和原始晶胞晶格常数的关系，构造产生新的二维材料的结构，即完成二维材料构建。

[0037] 本申请的再一面公开了一种材料数据库，该材料数据库中包含了本申请的材料晶体结构描述方法描述的材料晶体结构、本申请的发现材料晶体结构中的独立原子或独立结构基元的方法获得的独立原子或独立结构基元信息、本申请的材料晶体结构差异性分析的方法获得的材料晶体结构差异性分析信息，以及本申请的二维材料研发的方法获得的二维材料研发信息中的至少一种。

[0038] 本申请的再一面公开了一种基于机器学习的智能设计新材料的方法，其包括采用本申请的材料晶体结构描述方法对材料的晶体结构进行描述，使用本申请的数据库和机器学习方法找到所描述的晶体结构的特征与材料性能的相关性；根据所找到的相关性以及所需要的目标性能，得到具有目标性能的晶体结构特征，根据得到的晶体结构特征设计具有所需要的目标性能的新材料。

[0039] 可以理解，在相同的思路下，同样可以通过机器学习方法，实现材料性质的预测和分析，例如，将描述的晶体结构的特征作为特征输入进行机器学习，然后对材料的分类和带隙信息等关键性质进行回归预测，从而实现材料性质的预测和分析。

[0040] 需要说明的是，本申请的发现材料晶体结构中的独立原子或独立结构基元的方法、材料晶体结构差异性分析的方法、二维材料研发的方法、材料数据库，以及基于机器学习的智能设计新材料的方法，都是基于本申请的材料晶体结构描述方法而提出的，因此，其同样具有本申请材料晶体结构描述方法的优点，即能够方便的利用计算机对材料晶体结构进行大规模的批量比对、分析和运算。

## 发明的有益效果

### 有益效果

[0041] 本申请的有益效果在于：

[0042] 本申请的材料晶体结构描述方法，创造性的利用点和边的方式描述材料的晶体结构，能够通过计算机语言存储大量而又繁琐的点与线的连接信息，方便计算机的运算。本申请的方法可以便捷地将物理世界的原子、晶体结构构建到计算

机语言中，因此，可直接利用计算机领域的现有的算法对材料的性质进行比对、分析和运算；可大规模的批量处理晶体结构；为材料数据库的构建以及材料性质的计算机分析奠定了基础。

## 对附图的简要说明

### 附图说明

- [0043] 图1是本申请实施例中关于将结构转变为计算机图运算的示意图；其中，（a）图表示一个尖晶石 $\text{Co}_3\text{O}_4$ 结构在经过计算机图变换之后的形状；（b）图展示了尖晶石 $\text{Co}_3\text{O}_4$ 结构原始的晶体结构；（c）、（e）和（g）三个图分别展示了在不同的原子环境下，从一个原子出发，周围的原子环境排布情况；（d）、（f）和（h）三个图依序展示了（c）、（e）和（g）三个图转化为计算机图后的形状；
- [0044] 图2是本申请实施例中使用“魏格纳-塞兹原胞”方法获取原子连接方式的示意图；以其中灰色点为中心，黑色的点为与其相邻有链接的点；
- [0045] 图3是本申请实施例中关于图连接层数的示意图；其中分别展示了 $\text{NaCl}$ 结构从其中一个原子出发，周围一层到五层的图连接方式示意图；图中第一排为计算机图变换之后的形状，第二排位相应的原始晶体结构；
- [0046] 图4是本申请实施例中关于图同构的示意图；
- [0047] 图5是本申请实施例中使用图编辑距离分析两种晶体结构差异性的示意图；
- [0048] 图6是本申请实施例中根据图5的两种晶体结构，利用图编辑距离得到的图编辑路径的示意图；
- [0049] 图7是本申请实施例中根据构建的图研发二维材料的示意图。

## 发明实施例

### 本发明的实施方式

- [0050] 图论作为计算机科学中常见的问题之一，在计算机领域已经有了相对较多的研究。本申请创造性的将图论算法应用到材料研究中，即利用图论算法描述材料的晶体结构，具体的，本申请创造性的将晶体结构中的原子和/或结构基元作为点；把原子和/或结构基元之间的连接作为边；采用点、边以及点与边的相互关系描述晶体结构中的原子和/或结构基元之间的连接关系，从而实现材料晶体结

构的描述；这种描述方法可以方便的将晶体结构的描述转变为计算机语言。利用本申请的方法，材料结构将可以被转变为计算机可以处理的图论问题，从而可以被应用于材料数据库构建、材料高通量计算筛选等多个领域。

[0051] 相比于之前结晶学中的描述晶体结构的方法，使用本申请的方法描述晶体结构，打破了传统的晶体结构描述模式，这种新的描述方法相比较于传统的晶体结构描述方法有如下的特点：（1）传统的晶体结构描述方法使用了点群空间群和晶胞周期性的描述方法，与实验中X射线衍射数据可以直接关联，具有很好的被人理解和分析的能力；但是传统的方法不适用于大数据时代的晶体结构分析，人也无法对于上百万的晶体结构进行逐一分析；本申请的描述晶体结构的方法，存储了大量而又繁琐的点与线的连接信息，这些内容不适用于人的分析判断，但是方便计算机的运算。（2）使用这种运算方法，可以便捷地将物理世界的原子、晶体结构到计算机语言中的图；此后，可以利用计算机领域的现有的算法对材料的性质进行分析和运算。

[0052] 总结起来，使用传统的晶体结构描述方法，人类可以了解并分析现有的晶体结构；而使用本申请的将晶体结构转变为点和边的描述方法，计算机可以批量处理大量的晶体结构，并对这些结构进行比对、分析、运算，本申请的方法将会被应用于材料数据库的构建，材料性质的计算机分析等领域。

[0053] 下面通过具体实施例和附图对本申请作进一步详细说明。以下实施例仅对本申请进行进一步说明，不应理解为对本申请的限制。

[0054] 实施例

[0055] 本例描述材料晶体结构的方法，将晶体结构中的原子和/或结构基元作为点；把原子和/或结构基元之间的连接作为边；采用点、边以及点与边的相互关系描述晶体结构中的原子和/或结构基元之间的连接关系，从而实现材料晶体结构的描述。进一步的，根据点和边描述的晶体结构，将这些点和边构建为图，即从每个原子或结构基元为中心出发向外扩展若干层，将向外扩张的晶体结构转变为一个有限大小的图。本例的方法描述的晶体结构，以及所构建的图，都可以方便的通过计算机语言实现，从而可以实现通过计算机对晶体结构进行比对、分析和运算。

[0056] 其中，从某一原子出发，寻找与其相邻有连接关系的原子的算法，这一步骤，本例分别采用了两种不同的方法，即原子半径重叠方法和寻找最近邻原子方法。

[0057] 原子半径重叠方法，也就是“距离判断”方法，是基于由原子半径之和获得的键长参数连接原子的连接关系的构建方法，具体包括，对于晶体材料中元素的配对成键判断提供一个距离的参考阈值，这个参考阈值实际上就是两个原子的半径之和，例如C-C为1.89002埃等；然后，对给定晶体结构中首先进行扩胞操作，即以给定的晶体结构为原始晶胞，以原始晶胞为基础单元，并将原始晶胞作为中心，在三维空间中连续复制若干次后堆积成超胞；对超胞所包含的每个原子进行遍历，每次从选定的原子出发，依次计算选定原子和其他所有原子的距离，根据两个原子的对应元素，如果返回的距离值小于或者等于上述中提到的该元素配对的距离参考阈值，则视作两个原子之间有成键关系；遍历完成所有的原子之后，就完成了晶体材料的原子之间的连接关系的判断。

[0058] 寻找最近邻原子方法，本例具体采用的是魏格纳-塞兹原胞方法，即基于魏格纳-塞兹原胞方法连接原子的原子之间的连接关系的构建方法。具体的，对给定晶体结构进行扩胞操作，扩胞操作包括以给定的晶体结构为原始晶胞，以原始晶胞为基础单元，并将原始晶胞作为中心，在三维空间中连续复制若干次后堆积成超胞；对超胞中原始晶胞的每个原子进行查找，每次选定原子之后，构建该选定原子的魏格纳-塞兹原胞，如果选定原子与某一原子连线的垂直平分面在最小的魏格纳-塞兹原胞表面，则选定原子与该原子有连接，查找完成所有的原子，即构建获得超胞代表的晶体结构的原子之间的连接关系。

[0059] 其中，魏格纳-塞兹原胞，是来源于数学上“沃罗诺伊晶胞”的概念，该方法提供直观上创建了一种最小体积的原胞，且该原胞跨域了整个布拉维晶格而不留下任何空隙和孔洞，在使用上以此作为判断原子是否是最近邻的方法。具体来说，在给定的晶体结构构建的超胞体系内，对原始晶胞的每个原子进行遍历，其中，原始晶胞处在整个超胞体系内的中心位置，每次选定原子之后，从该原子出发依次和其他所有原子连接一条线段，并找出空间中该线段的垂直平分面，假设超胞内共计有 $n$ 个原子，此时获得了 $n-1$ 个垂直平分面。下面需要判断，该平

面是否是“魏格纳塞斯原胞”构建所需要的平面。具体的，从找到的 $n-1$ 个垂直平分面内，任意不重复的取出三个面，利用行列式判断三面是否存在共点并利用相关几何知识求解三个平面的交点坐标，如果存在平行情况则跳过，每找到一个交点坐标需要对该点进行判断，判断方法是从找到的交点开始依次计算交点和所有超胞内的原子的距离，如果该交点和初始遍历的原子之间的距离是最小的，且连续有三个相等的最小值，并且没有重复交点情况，则视作该交点有效，记录下获得该交点的三个平面，并由三个平面回溯找到对应的超胞内原子，记录这三个原子的元素种类和位置。

[0060] 需要说明的是，每次判断一个交点有效之后，实际上记录了三个最近邻的原子的信息，当完成交点的遍历判断后也就完成了针对该初始原子的魏格纳塞斯原胞的建立。最后，根据记录的结果，对于晶体结构中的给定原子，判断出的其最近邻原子，认为与给定原子均有连接关系，即构建获得晶体结构的原子之间的连接关系。

[0061] 本例以尖晶石 $\text{Co}_3\text{O}_4$ 结构为例，“魏格纳-塞兹原胞”构建的计算机图变换形状如图1所示；图中，（a）图表示一个尖晶石 $\text{Co}_3\text{O}_4$ 结构在经过计算机图变换之后的形状，（b）图展示了尖晶石 $\text{Co}_3\text{O}_4$ 结构原始的晶体结构，（c）、（e）和（g）三个图分别展示了在不同的原子环境下，从一个原子出发，周围的原子环境排布情况，（d）、（f）和（h）三个图依序展示了（c）、（e）和（g）三个图转化为计算机图后的形状。

[0062] 在构建好了原子之间的连接关系之后，将晶体结构转变为有限大、有限数量的计算机语言的图的方法包括构建图步骤和向外扩展点的层数判断步骤。这两个步骤相应的解决的问题即如何构建图、如何判断最大找到第几层原子。下面分别进行详细说明。

[0063] 构建图的方法，即构建图步骤，该步骤主要是用点和边描述晶体结构中的连接关系，再将这些点和边构建为图。该方法同样是基于图论中图存储的“邻接矩阵”方法。给定某个晶体材料结构，同样经过扩胞操作后，假定超胞内现有 $n$ 个原子，所以应当构建 $n \times n$ 的二维矩阵，矩阵中只有0和1两种元素，且第 $i$ 行，第 $j$ 列的元素代表晶胞内第 $i$ 个原子和第 $j$ 个原子的连接情况，0表示未连接，1表示连接。

基于上文的“采用原子半径重叠方法判断连接关系，或者采用寻找最近邻原子方法判断连接关系”的具体步骤，对每个原子进行遍历，每次计算选定的原子和其他所有的原子距离，若小于或等于之前的距离参考阈值，则矩阵的对应位置为1，否则为0。显然，由于连接的相对性，获得的二维矩阵应是沿对角线的对称阵。藉由此，即完成构建图的方法。

[0064] 向外扩展点的层数判断步骤包括，根据研究目标，设定构建的图的扩充层数，将层数以内的点和边，构建成一个有限大小的图。由前文所述，对于给定的晶体结构，由某一原子出发，由于晶体材料的周期性，理论上都可以按照无限层去寻找原子继而构成图，而且随着图的层数的增多，图的规模会越大，表示的信息也越来越详细。如图2所示，使用上，本例限制去寻找原子的层数，具体方法是预先设定好图的扩充层数，例如3层、4层、5层、6层和7层，然后分别按照对应层数在材料数据库内进行测试，特别是测试某些结构相近的结构，例如，SiC是典型的多相物质，不同相之间的差异较小，测试的内容是检测当图扩充层数是多少是可以有效的识别不同结构构建的图，当超过某一个层数后，继续增加层数对于检测的效果没有提升。通过这种方法，获得合适的最多找到几层原子的方法。可以理解，扩展层数一方面是取决于晶胞本身的大小，另一方面也取决于构建图的用途，即需要采用构建图进行怎样的研究。

[0065] 根据以上方案即可实现本例的材料晶体结构描述方法。在本例的新的描述方法的基础上，本例进一步提供了一系列基于本例的材料晶体结构的应用实例。具体的，这些应用包括将本例的材料晶体结构描述方法应用于材料晶体结构中独立原子或独立结构基元研究、材料晶体结构比对、材料晶体结构差异性分析、二维材料研发、材料数据库构建，以及基于机器学习的材料性质预测、分析、新材料设计等。详细的应用方式介绍如下：

[0066] 第一种应用，材料晶体结构中独立原子或独立结构基元研究，本例具体提供了一种发现材料晶体结构中的独立原子或独立结构基元的方法，包括采用本例的晶体结构描述方法对材料的晶体结构进行描述，以晶体结构中的每一个原子或结构基元作为起始点，从起始点出发，分别向外构建有限层数的图，然后分析这些图是否完全相同；所有不完全相同的图对应的起始点即为独立原子或独立

结构基元。通过发现材料晶体结构中的独立原子或独立结构基元，实际上也可以实现材料晶体结构比对，例如，如果两个晶体结构中的独立原子或独立结构基元的种类比例和类型安全相同，则这两个晶体结构完全相同。

[0067] 基于相同的思路，本例同时提供了一种材料晶体结构比对的方法，即使用图的同构性判断结构的是否相同，其包括采用本例的材料晶体结构描述方法对材料的晶体结构进行描述，并利用所获得的构建图的同构性质判断材料晶体结构是否相同。图的同构是图论中的一个概念，指的是两个图的完全等价的关系；以NaCl结构从其中一个原子出发，周围一层到五层的图连接方式为例进行说明，如图3所示。判断图的同构关系核心是判断图之间的顶点的映射关系，若存在情况，即对于其中一个图的任何一个顶点，在另外一个图中都有且仅有一个不重复的顶点与之构成映射关系。在判断同构时，使用了基于深度优先（Depth-First-Search，缩写DPS）算法配合剪枝去尽可能减少计算的时间复杂度。具体做法是，遍历其中一个图的每个顶点，每次遍历时，顺序的依次选择第二个图的每个顶点，去判断两个顶点是否存在映射关系。如果存在映射关系，则保留找到的顶点，从左边的图中增加一个顶点与之前找到的顶点放到一起，再从第二个图去遍历是否存在映射顶点。该算法的核心是“回溯”以及“剪枝”，具体来说，每次当对于第一个图的当前顶点，遍历右边的图完成后依然无法找到存在映射关系的顶点，则回溯到上一个寻找情况，两边找到的顶点数-1，从上次第二个图找到的映射顶点开始继续寻找映射顶点，若遍历后依然不存在，则继续回溯，直到找到了映射顶点。如果回溯到第一层，依然没有映射顶点，则算法返回不同构并输出。如果结果为同构，则可以判断这两种结构是相同的。图3分别展示了NaCl结构从其中一个原子出发，周围一层到五层的图连接方式示意图；图中第一排为计算机图变换之后的形状，第二排位相应的原始晶体结构；连接层数越多，图就越复杂，包含的信息也就越多。

[0068] 另外，关于图同构的情况，如图4所示，图中左边的图G和右边的图H，通过肉眼观察，两个结构为不同的形状；但是如果使用最右侧的运算规则，两个结构的连接方式完全相同，两个结构为同构。

[0069] 需要说明的是，由于适应于材料领域问题，因此存在根据材料相关知识剪枝的

方法。例如，预先进行顶点数和边数的判断；每次判断映射情况时，如果元素不相同，则自动跳过后面的判断等。

[0070] 第二种应用，一种材料晶体结构差异性分析的方法，即使用图编辑距离（Graph edit distance，简称GEM）判断图的差异性；本例具体包括，采用本例的材料晶体结构描述方法对材料的晶体结构进行描述，将晶体结构构建为图，并利用所构建的图中点和边的差异性描述材料晶体结构的差异性。图同构可以回答结构是否相同的问题，但是无法回答结构之间差异大小的问题。本例采用图编辑距离去度量两个图之间的差异大小，具体的做法是，针对两个不同构的图，总可以找到一定的操作，点或者边的插入删除等，对其中一个图进行修改使得修改后的图和第二个图同构。具体做法是，定义所有可以进行的操作共6种，分别是对点的替换、插入和删除以及对于边的替换、插入和删除。需要说明的是，实际上修改的操作集合有很多，本例计算最短的编辑路径完成图修改。另外，针对两个图的差异度，或者称为“距离”，定义为编辑路径操作的代价函数之和。首先通过定义6种操作的不同代价，利用A\*启发式搜索算法，遍历第一个图的每个点，每次考虑上述的6种操作并且尽可能的优先使用操作代价低的操作，每次延伸某种操作后更新A\*启发函数的评估代价项，当完成遍历后，即获得了从其中一个图变为另一个图的一系列顺序操作及总的操作代价和，这里代价和即为图的差异性，即为两图的“距离”或差异度。

[0071] 具体的，如图5所示，图5展示了两种简单的材料晶体结构，即结构1和结构2。利用图编辑距离，首先，将晶体结构转为图的表示，如图6的g1和g2所示。接着，考虑点的插入、删除和替换操作，并且定义插入和删除的操作代价为1，替换的代价为0.5，利用A\*启发式搜索算法，获得了由g1到g2的操作序列，分别是删除一个点，替换三个点以及删除一条边，得到操作代价和为3.5，即为两个图的差异度，这个差异度同样也表征了两个晶体结构的差异性。

[0072] 第三种应用，一种二维材料研发的方法，即使用图的连通性寻找二维材料。具体包括，采用本例的材料晶体结构描述方法对材料的晶体结构进行描述，将晶体结构构建为图，并利用所构建的图中的任意一点是否能通过边的连接在二维平面内扩展，而这些二维平面之间没有边的连接，判断晶体结构能否形成二维

材料；如果是则晶体结构能形成二维材料，否则不能形成二维材料。二维材料在微电子、半导体行业应用广泛，被视为未来最具潜力的材料领域。我们通过开发图的连通性方法，来实现从体材料切割出二维材料，并应用到晶体材料库进行高通量操作，筛选创造出有潜力的二维材料结构。具体做法是：针对给定的晶体材料结构，首先进行扩胞操作，然后再将扩胞后的超胞，转化为多张图，每张图仅由每个晶胞的某个原子组成，然后分别对这多张图进行图的连通性操作，根据DFS深度优先算法，由某个原子出发，去遍历与其相连的原子，利用递归算法对每个遍历到的原子继续利用DFS深度优先去继续遍历所有可以连接的原子，直到无法继续搜索，并返回已经搜索到的原子的信息，储存到连通分量中。通过对每张图进行图连通搜索，得到基于每张图的不同连通分量。然后对找到的连通分量进行判断，利用点线面的几何知识，排除一维线性材料和小块体材料的情况。最后针对得到的二维材料，根据二维材料的扩展方向和原晶胞晶格常数的关系，构造产生新的二维材料的结构。

[0073] 需要说明的是，由于该连通算法的要求，为了尽可能的排除由于原始体材料晶格常数的少数畸变情况带来的二维材料的延伸方向难以确定的问题，以及尽可能的在更多的方向上去筛选二维材料存在的可能性，需要建立尽可能大的超胞来包括更多的方向。事实上，对于常见的高通量筛选情况，超胞的扩充层数有合适的数值，依据此可以较好的完成图连通的二维材料切割问题。

[0074] 如图7所示，针对某个给定的晶体结构材料，在将结构转变为图并且完成了图的连通判断完成后，直观上将原始的晶体结构划分为了两个不连通的部分，如图7的左边部分，但每个部分内部彼此之间连通。接着，针对连通的部分判断其在二维平面可以进行扩展，而在垂直方向无法扩展，确定其为二维材料。最后生成新的二维材料结构，如图7的右边部分即展示了新的二维材料。

[0075] 第四种应用，一种基于机器学习的智能设计新材料的方法，其包括采用本申请的材料晶体结构描述方法对材料的晶体结构进行描述，使用本申请的数据库和机器学习方法找到所描述的晶体结构的特征与材料性能的相关性；根据所找到的相关性以及所需要的目标性能，得到具有目标性能的晶体结构特征，根据得到的晶体结构特征设计具有所需要的目标性能的新材料。随着机器学习的日益

火爆，近年间利用机器学习或深度学习来对材料某些性质进行回归预测变得越来越普遍，其中，某些性质通常是传统第一性原理难以计算或计算消耗资源过大的材料性质。然而，由于机器学习的效果直接取决于训练数据的质量和规模，特别是深度学习需要的训练集规模更大，而晶体材料由于不同取晶胞方法，导致难以将材料的结构信息转化为数值输入，或者即使输入和结构有关的一些传统晶体结构学的参数，也会面临着特征向量长度不统一和丢失过多信息的问题。

- [0076] 在本例的材料晶体结构描述方法的基础上，本例开发了将图有关信息作为特征输入，进行机器学习来对材料进行分类和带隙等关键性质的回归预测的方法。具体做法是，对每一个样本，除了提供宏观物理量之外，主要包括电负性、元素序数的统计分析、原子最外层电子数的统计分析等，再引入有关图论的特征，主要为图的拉普拉斯变换后的矩阵的特征值。为了保证每个样本的特征长度相同，对得到的特征值进行归一化。需要说明的是，为了保证特征长度相等而进行的特征归一化会在一定程度上损失输入信息，但是借鉴于晶体材料数据库的样本广泛以及深度学习自我尝试特征组合的特点，所带来的预测的误差实际上是可以忽略不计的。
- [0077] 需要补充的是，拉普拉斯转换前的图实际上是对该结构中任意两种元素的存在点的连接情况产生的图，经过拉普拉斯变换可以计算得到矩阵特征值。基于此，选择组合机器学习算法中的随机森林，设置决策树个数为200，限制最大树扩展深度为12层，并利用网格化寻参来进行参数优化。
- [0078] 本例描述材料晶体结构的方法，采用点和边对晶体结构进行描述，不仅可以准确详细的记载大量的原子以及原子之间的连接信息，而且采用本例的描述方法，可以直接采用计算机对材料晶体结构进行比对、分析、运算，能够大规模的对晶体结构进行批量处理，从而满足材料数据库的使用需求。并且，本例的晶体结构描述方法，借助于计算机的比对、分析和运算，还能够很好的应用于晶体结构中独立原子或独立结构基元的发现、材料晶体结构比对、材料晶体结构差异性分析、二维材料研发和基于机器学习的材料性质预测、分析或新材料的智能设计等领域，为材料晶体结构的研究和开发提供了一种简单、有效的途径

。

[0079] 以上内容是结合具体的实施方式对本申请所作的进一步详细说明，不能认定本申请的具体实施只局限于这些说明。对于本申请所属技术领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干简单推演或替换。

## 权利要求书

- [权利要求 1] 一种描述材料晶体结构的方法，其特征在于：  
将晶体结构中的原子和/或结构基元作为点；  
把原子和/或结构基元之间的连接作为边；  
采用点、边以及点与边的相互关系描述晶体结构中的原子和/或结构基元之间的连接关系，从而实现材料晶体结构的描述。
- [权利要求 2] 根据权利要求1所述的方法，其特征在于：所述采用点、边以及点与边的相互关系描述晶体结构中的原子和/或结构基元之间的连接关系，具体包括采用原子半径重叠方法判断连接关系，或者采用寻找最近邻原子方法判断连接关系。
- [权利要求 3] 根据权利要求2所述的方法，其特征在于：所述采用原子半径重叠方法判断连接关系，具体包括以下步骤，  
对于晶体结构中元素的配对成键判断提供一个键长参数作为距离的参考阈值，所述参考阈值为两个原子的半径之和；  
对给定晶体结构进行扩胞操作，所述扩胞操作包括以给定的晶体结构为原始晶胞，以原始晶胞为基础单元，并将原始晶胞作为中心，在三维空间中连续复制若干次后堆积成超胞；  
对超胞所包含的每个原子进行查找，每次以选定原子作为起点，依次计算选定原子和其他所有原子的距离，若该距离小于或等于参考阈值，则视作两个原子之间有连接关系；查找完成所有的原子，即构建获得超胞代表的晶体结构的原子之间的连接关系；  
结构基元之间的连接关系通过结构基元之间是否有共用原子判断，如果有共用原子则两个结构基元之间有连接关系，如果没有共用原子则通过结构基元中的原子与其它结构基元中的原子之间的连接关系进行判断。
- [权利要求 4] 根据权利要求2所述的方法，其特征在于：所述采用寻找最近邻原子方法判断连接关系，具体包括以下步骤，  
对给定晶体结构进行扩胞操作，所述扩胞操作包括以给定的晶体结构

为原始晶胞，以原始晶胞为基础单元，并将原始晶胞作为中心，在三维空间中连续复制若干次后堆积成超胞；

对超胞中原始晶胞的每个原子进行查找，每次选定原子之后，构建该选定原子的魏格纳-塞兹原胞，如果选定原子与某一原子连线的垂直平分面在最小的魏格纳-塞兹原胞表面，则选定原子与该原子有连接，查找完成所有的原子，即构建获得超胞代表的晶体结构的原子之间的连接关系；

结构基元之间的连接关系通过结构基元之间是否有共用原子判断，如果有共用原子则两个结构基元之间有连接关系，如果没有共用原子则通过结构基元中的原子与其它结构基元中的原子之间的连接关系进行判断。

[权利要求 5]

根据权利要求1-4任一项所述的方法，其特征在于：还包括从每个原子或结构基元为中心出发向外扩展若干层，将向外扩张的晶体结构转变为一个有限大小的图，具体包括，构建图步骤和向外扩展点的层数判断步骤；

所述构建图的步骤包括，用点和边描述晶体结构中的连接关系，再将这些点和边构建为图；

所述向外扩展点的层数判断步骤包括，根据研究目标，设定构建的图的扩充层数，将层数以内的点和边，构建成一个有限大小的图。

[权利要求 6]

一种发现材料晶体结构中的独立原子或独立结构基元的方法，其特征在于：包括采用权利要求1-5任一项所述的方法对材料的晶体结构进行描述，以晶体结构中的每一个原子或结构基元作为起始点，从起始点出发，分别向外构建有限层数的图，然后分析这些图是否完全相同；所有不完全相同的图对应的起始点即为独立原子或独立结构基元。

[权利要求 7]

一种材料晶体结构差异性分析的方法，其特征在于：包括采用权利要求1-5任一项所述的方法对材料的晶体结构进行描述，将晶体结构构建为图，并利用所构建的图中点和边的差异性描述材料晶体结构的差异性。

- [权利要求 8] 一种二维材料研发的方法，其特征在于：包括采用权利要求1-5任一项所述的方法对材料的晶体结构进行描述，将晶体结构构建为图，并利用所构建的图中的任意一点是否能够通过边的连接在二维平面内扩展，而这些二维平面之间没有边的连接，判断晶体结构能否形成二维材料；如果是则晶体结构能形成二维材料，否则不能形成二维材料。
- [权利要求 9] 一种材料数据库，其特征在于：所述材料数据库中包含权利要求1-5任一项所述的方法描述的材料晶体结构、权利要求6所述的方法获得的独立原子或独立结构基元信息、权利要求7所述的方法获得的材料晶体结构差异性分析信息，以及权利要求8所述的方法获得的二维材料研发信息中的至少一种。
- [权利要求 10] 一种基于机器学习的智能设计新材料的方法，其特征在于：包括采用权利要求1-5任一项所述的方法对材料的晶体结构进行描述，使用权利要求9的数据库和机器学习方法找到所描述的晶体结构的特征与材料性能的相关性；根据所述相关性以及所需要的目标性能，得到具有目标性能的晶体结构特征，根据得到的晶体结构特征设计具有所需要的目标性能的新材料。

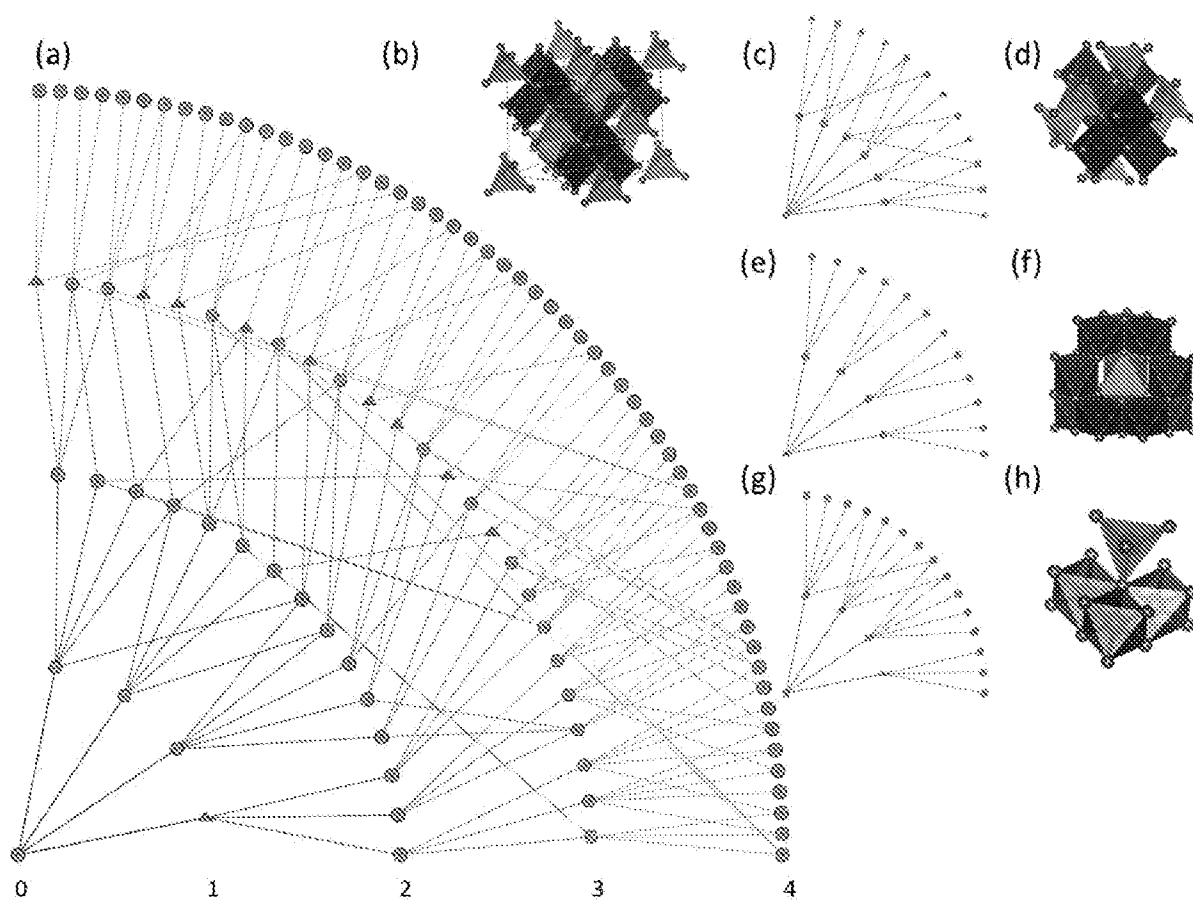


图 1

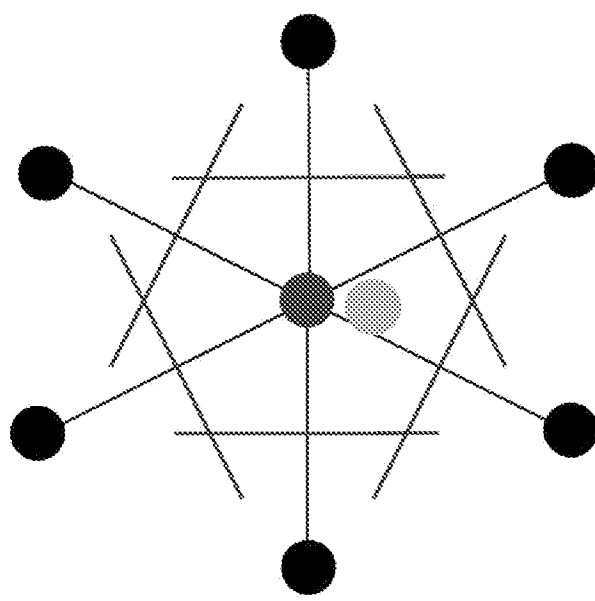


图 2

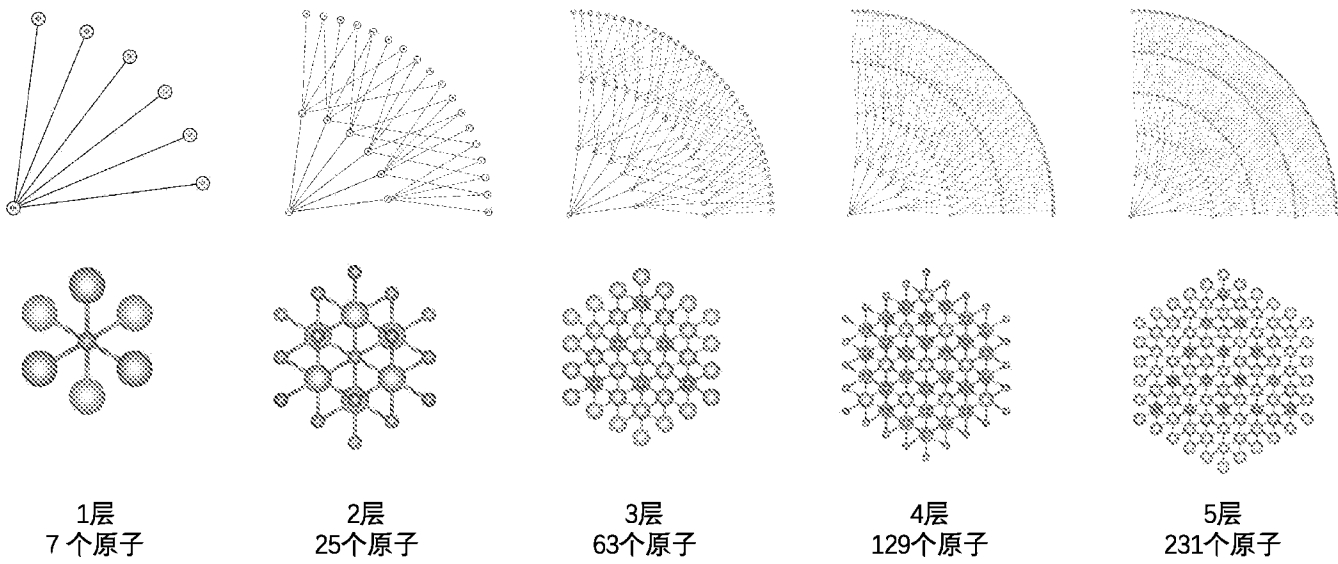


图 3

| 图G | 图H | 从图G到图H的同构映射 $\sigma$                                                                                                                                 |
|----|----|------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |    | $\sigma(a) = 1$<br>$\sigma(b) = 6$<br>$\sigma(c) = 8$<br>$\sigma(d) = 3$<br>$\sigma(g) = 5$<br>$\sigma(h) = 2$<br>$\sigma(i) = 4$<br>$\sigma(j) = 7$ |

图 4

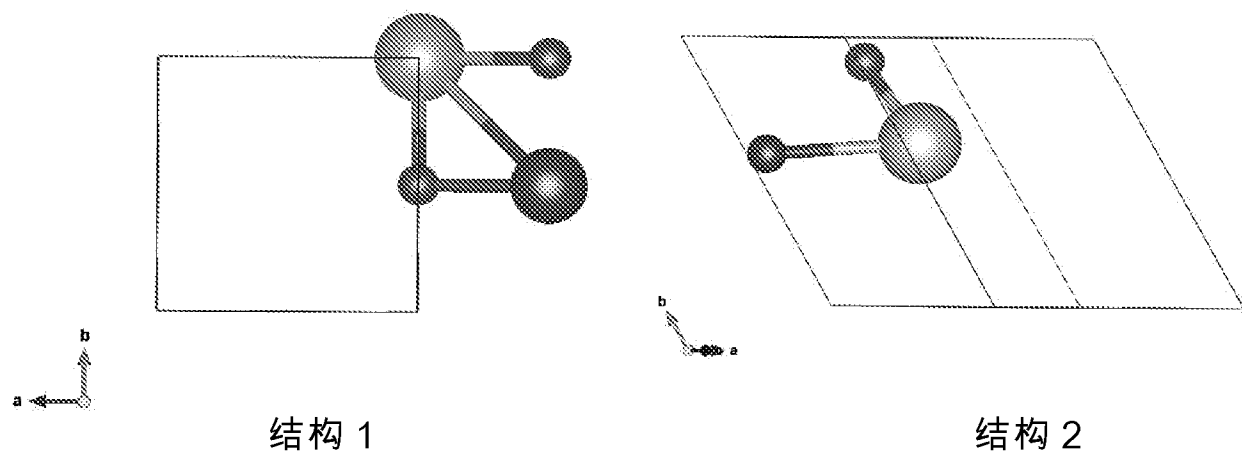


图 5

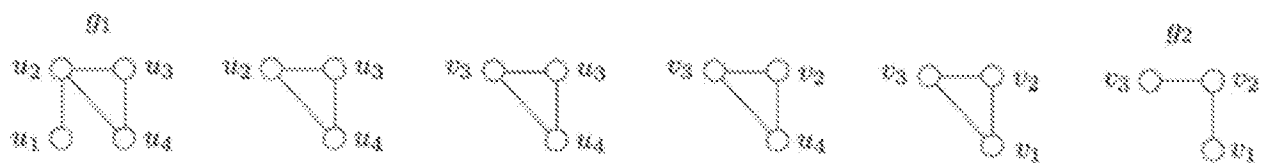


图 6

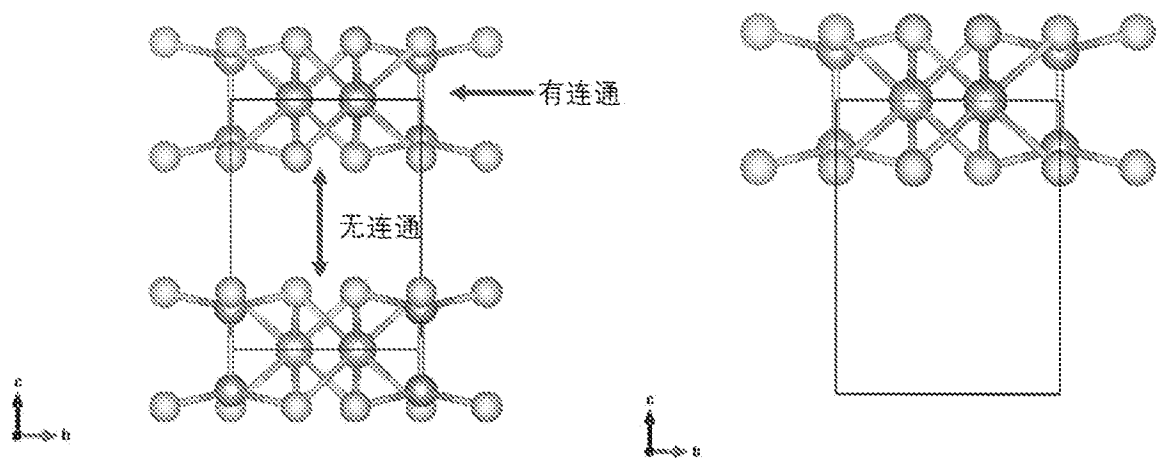


图 7

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/097972

**A. CLASSIFICATION OF SUBJECT MATTER**

G16C 60/00(2019.01)i; G06F 7/00(2006.01)i; G06F 17/00(2019.01)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)

G16C; 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)

CNKI, CNABS, CNTXT, DWPI, SIPOABS, ISI WEB OF KNOWLEDGE: 晶体结构, 描述, 检索, 计算机, 电脑, 数据库, 边, 点, 原子, crystal+, structure, describ+, retriev+, computer, data, base, point?, edge?, atom

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                           | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | 霍卫峰等 (HUO, Weifeng et al.). "一种在无机晶体结构中检索特定子结构的计算机方法 (A Computational Method for Specified Substructure Search in Inorganic Crystal Structures)" 物理化学学报 ( <i>Acta Physico-Chimica Sinica</i> ), Vol. 28, No. (3), 31 March 2012 (2012-03-31), ISSN: 1000-6818, pages 536-540 | 1-10                  |
| A         | CN 105719005 A (QU, Hui) 29 June 2016 (2016-06-29)<br>entire document                                                                                                                                                                                                        | 1-10                  |
| A         | CN 107273670 A (COMPUTER NETWORK INFORMATION CENTER, CHINESE ACADEMY OF SCIENCES) 20 October 2017 (2017-10-20)<br>entire document                                                                                                                                            | 1-10                  |
| A         | CN 107357971 A (COMPUTER NETWORK INFORMATION CENTER, CHINESE ACADEMY OF SCIENCES) 17 November 2017 (2017-11-17)<br>entire document                                                                                                                                           | 1-10                  |
| A         | CN 108427861 A (INSTITUTE OF MATERIALS, CHINA ACADEMY OF ENGINEERING PHYSICS) 21 August 2018 (2018-08-21)<br>entire document                                                                                                                                                 | 1-10                  |



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

15 January 2020

Date of mailing of the international search report

22 January 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration  
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing  
100088  
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

**INTERNATIONAL SEARCH REPORT**

International application No.

**PCT/CN2019/097972****C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages        | Relevant to claim No. |
|-----------|-------------------------------------------------------------------------------------------|-----------------------|
| A         | US 7174343 B2 (ORACLE INT. CORP. et al.) 06 February 2007 (2007-02-06)<br>entire document | 1-10                  |
| A         | US 2008014572 A1 (SEARETE LLC) 17 January 2008 (2008-01-17)<br>entire document            | 1-10                  |

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2019/097972**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) |            |    | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|------------|----|--------------------------------------|
| CN                                        | 105719005  | A  | 29 June 2016                         | None                    |            |    |                                      |
| CN                                        | 107273670  | A  | 20 October 2017                      | None                    |            |    |                                      |
| CN                                        | 107357971  | A  | 17 November 2017                     | None                    |            |    |                                      |
| CN                                        | 108427861  | A  | 21 August 2018                       | None                    |            |    |                                      |
| US                                        | 7174343    | B2 | 06 February 2007                     | US                      | 2003212692 | A1 | 13 November 2003                     |
| US                                        | 2008014572 | A1 | 17 January 2008                      | US                      | 2009055138 | A1 | 26 February 2009                     |
|                                           |            |    |                                      | US                      | 2009094003 | A1 | 09 April 2009                        |
|                                           |            |    |                                      | US                      | 2008015787 | A1 | 17 January 2008                      |
|                                           |            |    |                                      | US                      | 2009082974 | A1 | 26 March 2009                        |

## 国际检索报告

国际申请号

PCT/CN2019/097972

| <b>A. 主题的分类</b><br>G16C 60/00(2019.01)i; G06F 7/00(2006.01)i; G06F 17/00(2019.01)i<br>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>B. 检索领域</b><br>检索的最低限度文献(标明分类系统和分类号)<br>G16C; G06F<br>包含在检索领域中的除最低限度文献以外的检索文献<br>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))<br>CNKI, CNABS, CNTXT, DWPI, SIPOABS, ISI WEB OF KNOWLEDGE: 晶体结构, 描述, 检索, 计算机, 电脑, 数据库, 边, 点, 原子, crystal+, structure, describ+, retriev+, computer, data, base, point?, edge?, atom                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                         |                                       |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| <b>C. 相关文件</b> <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>霍卫峰等. “一种在无机晶体结构中检索特定子结构的计算机方法”<br/>物理化学学报, 第28卷, 第3期, 2012年 3月 31日 (2012 - 03 - 31),<br/>ISSN: 1000-6818,<br/>第536页-第540页</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 105719005 A (渠汇) 2016年 6月 29日 (2016 - 06 - 29)<br/>全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 107273670 A (中国科学院计算机网络信息中心) 2017年 10月 20日 (2017 - 10 - 20)<br/>全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 107357971 A (中国科学院计算机网络信息中心) 2017年 11月 17日 (2017 - 11 - 17)<br/>全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 108427861 A (中国工程物理研究院材料研究所) 2018年 8月 21日 (2018 - 08 - 21)<br/>全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>US 7174343 B2 (ORACLE INT. CORP. et al.) 2007年 2月 6日 (2007 - 02 - 06)<br/>全文</td> <td>1-10</td> </tr> </tbody> </table> |                                                                                                                         |                                       | 类 型* | 引用文件, 必要时, 指明相关段落 | 相关的权利要求 | X | 霍卫峰等. “一种在无机晶体结构中检索特定子结构的计算机方法”<br>物理化学学报, 第28卷, 第3期, 2012年 3月 31日 (2012 - 03 - 31),<br>ISSN: 1000-6818,<br>第536页-第540页 | 1-10 | A | CN 105719005 A (渠汇) 2016年 6月 29日 (2016 - 06 - 29)<br>全文 | 1-10 | A | CN 107273670 A (中国科学院计算机网络信息中心) 2017年 10月 20日 (2017 - 10 - 20)<br>全文 | 1-10 | A | CN 107357971 A (中国科学院计算机网络信息中心) 2017年 11月 17日 (2017 - 11 - 17)<br>全文 | 1-10 | A | CN 108427861 A (中国工程物理研究院材料研究所) 2018年 8月 21日 (2018 - 08 - 21)<br>全文 | 1-10 | A | US 7174343 B2 (ORACLE INT. CORP. et al.) 2007年 2月 6日 (2007 - 02 - 06)<br>全文 | 1-10 |
| 类 型*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 引用文件, 必要时, 指明相关段落                                                                                                       | 相关的权利要求                               |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| X                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 霍卫峰等. “一种在无机晶体结构中检索特定子结构的计算机方法”<br>物理化学学报, 第28卷, 第3期, 2012年 3月 31日 (2012 - 03 - 31),<br>ISSN: 1000-6818,<br>第536页-第540页 | 1-10                                  |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | CN 105719005 A (渠汇) 2016年 6月 29日 (2016 - 06 - 29)<br>全文                                                                 | 1-10                                  |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | CN 107273670 A (中国科学院计算机网络信息中心) 2017年 10月 20日 (2017 - 10 - 20)<br>全文                                                    | 1-10                                  |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | CN 107357971 A (中国科学院计算机网络信息中心) 2017年 11月 17日 (2017 - 11 - 17)<br>全文                                                    | 1-10                                  |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | CN 108427861 A (中国工程物理研究院材料研究所) 2018年 8月 21日 (2018 - 08 - 21)<br>全文                                                     | 1-10                                  |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | US 7174343 B2 (ORACLE INT. CORP. et al.) 2007年 2月 6日 (2007 - 02 - 06)<br>全文                                             | 1-10                                  |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| <input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                         |                                       |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| * 引用文件的具体类型:<br>“A” 认为不特别相关的表示了现有技术一般状态的文件<br>“E” 在国际申请日的当天或之后公布的在先申请或专利<br>“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)<br>“O” 涉及口头公开、使用、展览或其他方式公开的文件<br>“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<br>“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<br>“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<br>“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<br>“&” 同族专利的文件                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                         |                                       |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| 国际检索实际完成的日期<br>2020年 1月 15日                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                         | 国际检索报告邮寄日期<br>2020年 1月 22日            |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |
| ISA/CN的名称和邮寄地址<br>中国国家知识产权局(ISA/CN)<br>中国北京市海淀区蓟门桥西土城路6号 100088<br>传真号 (86-10)62019451                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                         | 受权官员<br>宋国英<br>电话号码 86-(010)-53962902 |      |                   |         |   |                                                                                                                         |      |   |                                                         |      |   |                                                                      |      |   |                                                                      |      |   |                                                                     |      |   |                                                                             |      |

### 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     | US 2008014572 A1 (SEARETE LLC.) 2008年 1月 17日 (2008 - 01 - 17)<br>全文 | 1-10    |
| <hr/> |                                                                     |         |

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2019/097972

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 |            |    | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|------------|----|----------------|
| CN          | 105719005  | A  | 2016年 6月 29日   | 无    |            |    |                |
| CN          | 107273670  | A  | 2017年 10月 20日  | 无    |            |    |                |
| CN          | 107357971  | A  | 2017年 11月 17日  | 无    |            |    |                |
| CN          | 108427861  | A  | 2018年 8月 21日   | 无    |            |    |                |
| US          | 7174343    | B2 | 2007年 2月 6日    | US   | 2003212692 | A1 | 2003年 11月 13日  |
| US          | 2008014572 | A1 | 2008年 1月 17日   | US   | 2009055138 | A1 | 2009年 2月 26日   |
|             |            |    |                | US   | 2009094003 | A1 | 2009年 4月 9日    |
|             |            |    |                | US   | 2008015787 | A1 | 2008年 1月 17日   |
|             |            |    |                | US   | 2009082974 | A1 | 2009年 3月 26日   |