



(12) 发明专利

(10) 授权公告号 CN 103025927 B

(45) 授权公告日 2015. 12. 16

(21) 申请号 201180021839. 8

代理人 王达佐 洪欣

(22) 申请日 2011. 04. 27

(51) Int. Cl.

(30) 优先权数据

C40B 20/02(2006. 01)

61/330, 130 2010. 04. 30 US

C40B 40/06(2006. 01)

13/092, 618 2011. 04. 22 US

(56) 对比文件

(85) PCT国际申请进入国家阶段日

WO 2009/059022 A1, 2009. 05. 07,

2012. 10. 30

US 6349144 B1, 2002. 02. 19,

(86) PCT国际申请的申请数据

US 2009/0250615 A1, 2009. 10. 08,

PCT/US2011/034178 2011. 04. 27

审查员 陈云华

(87) PCT国际申请的公布数据

W02011/137183 EN 2011. 11. 03

(73) 专利权人 考利达基因组股份有限公司

地址 美国加利福尼亚州

(72) 发明人 布赖恩·P·斯泰克

(74) 专利代理机构 北京英赛嘉华知识产权代理

有限责任公司 11204

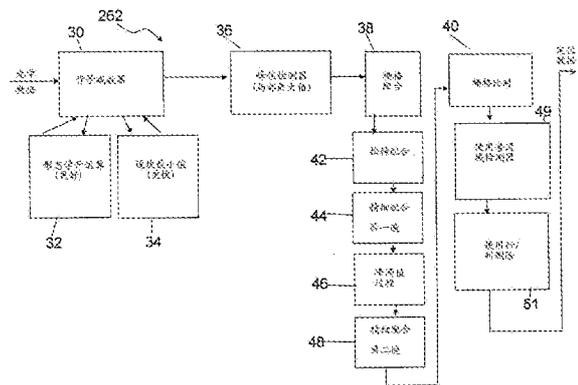
权利要求书2页 说明书15页 附图18页

(54) 发明名称

精确地比对与配准用于 DNA 测序的阵列的方法及系统

(57) 摘要

在基因组测序系统与方法中,提供工具以精确和准确地比对与配准用于光学分析的纳米球的平面阵列图像。通过使用最小化技术与 Moiré 求平均数校正亚期 x-y 偏移、比例与旋转中误差。在 Moiré 求平均数中,设置放大率从而使得成像元素的像素周期为位点周期的非整数倍数。通过提供预定义的假随机位点组实现精确的配准,其中防止纳米球粘附至基质从而使得阵列的位点可以用在匹配方案的模式中,作为用于绝对位置鉴定的配准标记。可以高度置信地提取与已知位置相关的信息,而同时使可包装在芯片上的信息量最大化。



1. 用于配准基质上生化位点的方法,其包括:  
制备用于成像的基质,以建立基质上生化位点的位置与配准;  
检测图像中基质上对应于生化位点的域点;  
比对二维栅格与基质上生物位点的域点以确定栅格偏移;以及  
配准所述栅格点与域中已知的位置,所述配准步骤包括将域点与基质上第一已知域点删除模式互相关以确定域中绝对位置。
2. 如权利要求 1 所述的方法,选择的删除模式为假随机的。
3. 如权利要求 1 所述的方法,所述配准步骤还包括将所述域点与第二已知删除模式互相关,鉴定基质上域点的位置。
4. 如权利要求 1 所述的方法,所述比对步骤包括迭代地测试旋转、比例、x- 偏移与 y- 偏移的不同组合,以寻找具有域点模式中每一栅格点与每一域点间最小二乘距离的测试组合。
5. 如权利要求 1 所述的方法,所述比对步骤包括:  
通过强度组收集域点;  
将组平均点亮度相对于至最近的栅格点的组平均距离绘图;以及  
在组中至最近的栅格点的组平均距离小于阈值的所有点上重复精细配合步骤。
6. 用于配准基质上生化位点的装置,其包括:  
具有离散光学可分辨的表面位点的固体平面基质,可修饰所述表面位点以支持生化实验;  
其中所述表面位点位于基质表面上点的二维周期阵列中;以及  
其中一部分的点根据至少一种已知删除模式从阵列删除。
7. 如权利要求 6 所述的装置,所述删除模式为假随机的。
8. 如权利要求 6 所述的装置,所述一部分为 1% 至 3%。
9. 如权利要求 6 所述的装置,所述生化实验包括与 DNA 纳米球的复合探针-锚定连接。
10. 用于配准基质上生化位点的方法,其包括:  
提供具有离散表面位点的固体平面基质,修饰表面位点以支持生化实验,所述位点位于基质表面上点的二维周期阵列中的点,并且一部分的点根据已知删除模式从阵列删除,否则保持;  
在所述位点处进行生化实验;  
获得所述生化实验的图像;  
将图像与删除模式互相关以确定点的绝对位置;以及  
用亚像素精确度观察图像以提取关于生化实验的信息。
11. 如权利要求 10 所述的方法,所述图像观察步骤包括根据点的精确位置加权来自邻近像素输出的贡献。
12. 用于配准基质上生化位点的系统,其包含:  
用于制备用于成像的基质的工具,以建立基质上生化位点的位置与配准;  
用于检测图像中基质上相应于生化位点的域点的工具;  
用于将域点与二维栅格的栅格点对比的工具;以及  
用于将所述栅格点与所述域中已知的位置配准的工具;所述配准工具包括适于执行以

下步骤的装置,所述步骤包括将栅格点与基质上域点的已知删除模式互相关以确定绝对位置。

## 精确地比对与配准用于 DNA 测序的阵列的方法及系统

[0001] 发明背景

[0002] 本发明涉及用于生化分析的成像,并且更具体地涉及用于对高通量基因组测序中所使用的高密度生化阵列进行成像的方法与系统。

[0003] 化学和 / 或生物种类的高通量分析为诊断学与治疗学领域中一重要的工具。生化阵列允许平行地进行多个生化实验。所述能力产生于在小容量中进行每一实验并且将实验紧密地打包在一起的技术的发展。基质上附着的化学和 / 或生物种类的阵列可被设计用于定义特异的靶序列,分析基因表达模式,鉴定特异的等位基因变体,确定 DNA 序列的拷贝数以及在全基因组基础上鉴定蛋白的结合位点(例如转录因子与其他的调控分子)。在具体的实例中,人基因组计划的出现需要开发 25 种改进的用于诸如 DNA(脱氧核糖核酸)与 RNA(核糖核酸)的核酸测序的方法。单倍体人基因组的全部 3,000,000,000 个碱基序列的测定提供了用于鉴定许多疾病的遗传基础的基础。然而,仍然要完成大量工作以鉴定与统计学上显著量的人基因组相关的遗传变异,并且在这一努力中,改进的用于分析的高通量方法可以很大地帮助。

[0004] 高通量分析方法常规地利用测定装置,被称为流式细胞,其包含用于分析的化学和 / 或生物种类的阵列。生物种类通常用多种荧光的颜色示踪,所述荧光的颜色可用成像系统读取。

[0005] 由于待观察、捕获以及分析的数据量庞大,基因组测序分析中关键因素为测定仪器的通量。通量对成本具有直接的影响。虽然与其他的技术比较,成像系统能够捕获大量的数据,但是此类系统的通量受限于照相机速度与每点的像素量。照相机速度受限于固有的物理限制,并且最小的每点像素量为一个。虽然理想的是将每点像素量减少至最小值,但是通常在实际的仪器中每点存在许多像素。

[0006] 来自于从与基质上附着位点有关的点所发射的光中按像素捕获的图像必须被比对并且配准以便可分析。涉及基质上的配准标记与参考线的常规配准技术需要基质上的空间,这减少了可用于分析的位点量并且因此减少每单元时间分析的体积。

[0007] 几种不同的 DNA 芯片的方法正在研发中。在一方法中,在芯片上建立 DNA 片段的组合阵列并且这些可用于通过杂交测序。在另一种方法中, DNA 随机地排列在表面以用于相同的目的。一研究组尝试使用 DNA 聚合酶的阵列以观察逐碱基的测序。仍然另一研究组使用通过组合的探针锚定连接所查询的自组装 DNA 纳米阵列。尽管这些方法各不相同,尤其是在它们的生化细节上,但是它们都取决于荧光成像技术以逐字“见到”阵列中个体的实验所产生的数据。

[0008] 通过设计生化反应从而使得不同的着色染料(例如,红色、绿色、蓝色或黄色)对应各个 DNA 碱基(A、C、G 或 T),荧光成像用于鉴定 DNA 碱基。然后研究人员可以用荧光显微镜观察 DNA 实验。所观察的颜色指示该特定步骤的 DNA 碱基。因此从 DNA 芯片提取数据取决于记录由芯片上几百万或甚至几十亿的生化实验所发射的荧光颜色。

[0009] 除非另有说明,本文所述的技术实践可使用有机化学、聚合物技术、分子生物学(包括重组技术)、细胞生物学、生物化学以及测序技术的常规的技术与描述,这些为本

领域技术人员的技术。此类常规的技术包括聚合物阵列合成、多聚核苷酸的杂交与连接以及利用标记检测杂交。参考本文的实施例,可以具体示例适当的技术。然而,当然也可以使用其他等同的常规程序。此类常规的技术与描述可以在标准的实验室指南中找到,诸如 Green 等,Eds. (1999), *Genome Analysis: A Laboratory Manual Series* (Vols. I-IV); Weiner, Gabriel, Stephens, Eds. (2007), *Genetic Variation: A Laboratory Manual*; Dieffenbach, Dveksler, Eds. (2003), *PCR Primer: A Laboratory Manual*; Sambrook 与 Sambrook (2003), *DNA Microarrays: A Molecular Cloning Manual*; Mount (2004), *Bioinformatics: Sequence and Genome Analysis*; Sambrook 与 Russell (2006), *Condensed Protocols from Molecular Cloning: A Laboratory Manual*; 以及 Sambrook 与 Russell (2002), *Molecular Cloning: A Laboratory Manual* (所有的都来自于冷泉港实验室出版社); Stryer, L. (1995) *Biochemistry* (第 4 版) W. H. Freeman, New York N. Y.; Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, London; Nelson 与 Cox (2000), *Lehninger, Principles of Biochemistry* 第 3 版, W. H. Freeman Pub., New York, N. Y.; 以及 Berg 等 (2002) *Biochemistry*, 第 5 版, W. H. Freeman Pub., New York, N. Y., 出于所有的目的,所有这些实验室指南通过引用全部并入本文。

[0010] 如本文所用的并且在所附的权利要求中,单数形式的“一 (a)”、“一个 (an)”以及“所述 (the)”包括复数指示物,除非上下文明确地另有指明。因此,例如提及“一 (a) 通道”指的是阵列基质上可利用的一个或多个通道,并且提及“所述 (the) 方法”包括本领域技术人员已知的等同的步骤与方法等等。

[0011] 除非另有定义,本文所用的所有技术的与科学的术语都具有如本发明所属于的领域中普通技术人员所通常理解的相同的意义。出于描述与公开装置、配方以及与目前描述的发明有关的可使用的方法的目的,本文提及的所有出版物通过引用并入本文。

[0012] 当提供了数值范围时,应理解的是在那一范围的最高限与最低限之间每一介于中间的数值,以及任一其他规定的或介于规定的范围内的数值都包括在本发明中。这些较小的范围的最高限与最低限可独立地包括于较小的范围内并且也包括在本发明中,受规定范围内任一特别排除的界限所影响。当规定范围包括一个或两个界限时,排除所包括的两个界限的任一的范围也包括在本发明中。

[0013] 在下文的描述中,列出了许多具体的细节以提供对本发明更加彻底的理解。然而,对本领域中技术人员明显的是阅读完本公开内容,没有这些具体细节之一或更多就可实践本发明。在其他的实例中,并没有描述熟知的特征与本领域技术人员熟知的那些程序以便避免模糊本发明。

[0014] 选择的定义

[0015] “扩增子”意指多聚核苷酸扩增反应的产物。即它为来自于一个或多个起始序列复制的多聚核苷酸的群体。各种各样的扩增反应可产生扩增子,其包括但不限于聚合酶链反应 (PCR)、直链聚合酶反应、基于核酸序列的扩增、循环依赖扩增以及相似的反应 (例如见美国专利第 4,683,195; 4,965,188; 4,683,202; 4,800,159; 5,210,015; 6,174,670; 5,399,491; 6,287,824 与 5,854,033 号以及美国专利申请公开号 2006/0024711)。

[0016] 本文的“附着位点”或“位点”指的是正常模式中基质上排列的官能化位置,生物活性结构可以结合在其上。实践中位点为活性阳性胺的亚微米区域,其经由硅烷化过程与

氧化物表面连接。

[0017] “循环依赖的复制”或“CDR”指的是循环模板的多重置换扩增,其利用与循环模板的同一链退火的一种或多种引物以产生只代表模板一条链的产物。在 CDR 中,不产生额外的引物结合位点并且产物的量只随着时间线性地增加。所用的引物可为随机序列(例如一种或多种随机六聚物)或可具有特异序列以选择用于所需产物的扩增。CDR 经常导致线性构建体的产生而不对终产物进行进一步的修饰,所述直链构建体具有串联循环模板链的多个拷贝,即模板链多个拷贝的线性单链串联体。

[0018] “循环依赖的扩增”或“CDA”指循环模板的多重置换扩增,其利用循与环模板的两条链退火引物以产生代表模板两条链的产物,从而造成多重杂交、引物延伸以及链置换事件的级联反应。这导致了引物结合位点数量的指数式增加,由此随着时间而产生的产物量的指数式增加。所用的引物可为随机序列(例如随机六聚体)或可具有特异序列以选择用于所需产物的扩增。CDA 导致串联体双链片段的形成。

[0019] 如本文所用的“域”为分析的二维亚单位,通常指的是通过照相机所捕获并且出于分析目的聚集在一起的数据。

[0020] 如本文所用的“栅格”指的是抽象的笛卡尔模式,其用于分析由像素构建的图像中信息的位置。出于本发明的目的,栅格具有 x 与 y 中恒定的周期性并且优选为平方的。在像素参考帧中常规地指定栅格的位置。

[0021] 如本文所用的“配体”指的是可共价地或非共价地,直接地或经由特异结合伴侣,与测定基质上分子连接的分子。本发明可以使用的配体的实例包括但不限于抗体、细胞膜受体、单克隆抗体以及与特异抗原决定簇(诸如在病毒、细胞或其他物质上)的抗血清反应、药物、多聚核苷酸、核酸、肽、辅酶因子、凝集素、糖类、多糖、细胞、细胞膜以及细胞器。

[0022] “微阵列”或“阵列”指的是具有表面的固相支持物,所述表面在本发明的实施方案中为必要地曲面或基本上平坦曲面,其运载含核酸的位点阵列从而使得阵列的每一位点包含许多的寡聚核苷酸或多聚核苷酸拷贝,位点为空间离散的。阵列的寡聚核苷酸或多聚核苷酸可共价地或非共价地与基质结合。例如在 Schena, Ed. (2000), *Microarrays: A Practical Approach* (IRL Press, Oxford) 中综述了常规的微阵列技术。

[0023] 本文所用的“核酸”或“寡聚核苷酸”意指核苷酸单体的聚合物。如本文所用的,术语也可指的是双链形式。装配核酸与寡聚核苷酸的单体能够特异地结合天然的多聚核苷酸,经由单体与单体相互作用的正常的模式形成二倍体或三倍体形式,所述正常模式诸如碱基对的沃森-克里克类型、碱基堆积、碱基对的 Hoogsteen 或反向的 Hoogsteen 类型,诸如此类。此类单体与它们的核苷酸间键可天然地形成或可为其类似物,例如天然形成的或非天然形成的类似物。非天然形成的类似物可包括肽核酸、锁核酸、硫代磷酸酯核苷酸间键、含有允许诸如荧光团或半抗原等标记粘附的连接基团的碱基。无论何时寡聚核苷酸或核酸的使用都需要酶促处理,诸如通过聚合酶延伸,通过连接酶连接等,普通技术人员应理解当此类类似物与酶促反应不相容时,在那些实例中寡聚核苷酸或核酸将不包含核苷酸间键、糖部分或在任何或一些位置上的碱基的某些类似物。核酸大小范围通常在几个单体单元-几十万或更多的单体单元,例如当它们通常被称为“寡聚核苷酸”时,核酸大小范围在 5 个-40 个单体单元。无论何时核酸或寡聚核苷酸都被字母序列(之上或之下的例子)所代表,诸如“ATGCCTG”,应理解的是核苷酸为自左到右 5' -3' 顺序并且“A”表示脱氧腺苷,“C”

表示脱氧胞苷,“G”表示脱氧鸟苷,以及“T”表示胸腺嘧啶核苷,“I”表示脱氧肌苷,“U”表示尿苷,除非另有指明或上下文中明确的。除非另有说明,术语与原子编号常规都将遵循下列 Strachan 与 Read, 人类分子遗传学 2(Human Molecular Genetics 2) (Wiley-Liss, New York, 1999) 中公开的那些。通常核酸包含通过磷酸二酯键连接的天然的核苷(例如对 DNA 而言脱氧腺苷、脱氧胞苷、脱氧鸟苷、脱氧胸腺嘧啶核苷或对 RNA 而言它们的核糖对应物);然而,他们也可包含非天然的核苷类似物,例如修饰的碱基、糖类或核苷酸间键。本领域中技术人员清楚的是当酶具有活性所需的特异的寡聚核苷酸或核酸基质时,例如单链的 DNA、RNA/DNA 二倍体等,然后寡聚核苷酸或核酸基质的适当组合的选择在技术人员的常识内,尤其来自于论述的指导,诸如 Sambrook 等, 分子克隆 (Molecular Cloning), 第二版(冷泉港实验室, 纽约, 1989) 等参考文献。如本文所用的,“靶向的核酸节段”指的是靶向的用于测序或重新测序的核酸。

[0024] “像素”为报告不可分割的位置的检测的光水平的照相机的不可分割的光敏感元件。单色的像素为单一的光检测元件。颜色过滤器可以用于检测像素上所接收光的光谱。

[0025] “引物”意指寡聚核苷酸,天然的或合成的,其一旦与多聚核苷酸模板形成二倍体就能够作为核酸合成的启动点使用并且从它的 3' 末端沿着模板延伸从而形成延伸的二倍体。延伸过程期间添加的核苷酸序列由模板多聚核苷酸的序列确定。通常通过 DNA 聚合酶延伸引物。引物通常具有 9 个 -40 个核苷酸,或在一些实施方案中,具有 14 个 -36 个核苷酸范围的长度。

[0026] 如本文所用的“探针”指的是寡聚核苷酸,天然的或合成的,其用于询问未知序列的核酸内互补的序列。特异探针与靶多聚核苷酸的杂交表示靶多聚核苷酸序列内与探针互补的特异序列。

[0027] 关于核酸的“测序”意指与核酸中核苷酸序列相关的信息的测定。此类信息可包括部分的和完整的核酸序列信息的鉴定或测定。可用不同程度的统计学的可信度或置信区间测定序列信息。在一方面,术语包括始于靶核酸中不同的核苷酸的核酸中多数连续的核苷酸的同一性与排序的测定。

[0028] 如本文所用的“点”指的是荧光分子所发射出光的位置。点不一定要位于附着位点的中心。

[0029] “基质”指的是具有刚性的或半刚性表面的材料或材料组。在现在的情况下,至少一个基质的表面基本上是平坦的,尽管在与本发明不相关的其他情况下,理想的是对不同的化合物而言,具有例如孔、凸起的区域、针、蚀刻的沟槽,诸如此类的物理上分隔合成区域。根据其他的实施方案,基质以珠子、树脂、凝胶、微球体或其他几何构型的形式呈现。在本发明中,基质的表面限于平面结构以促进分析。

[0030] 如本文所用的,术语“ $T_m$ ”作为“熔解温度”使用。熔解温度为双链的核酸分子群变成一半分离的单链时的温度。本领域中熟知用于计算核酸  $T_m$  的几个方程式。如标准的参考文献所指明的,可通过方程式  $T_m = 81.5 + 0.41 (\%G+C)$  计算  $T_m$  值的简单估算,当核酸在 1M NaCl 水溶液中时(例如见 Anderson 与 Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization(1985))。其他的参考文献(例如 Allawi, H. T. & Santa Lucia, J., Jr., Biochemistry 36, 10581-94(1997))包括可选择的计算方法,其计算  $T_m$  时考虑到了结构的与环境的以及序列的特征。

[0031] 常规的分析玻片由 1”x3” 硅芯片组成,其建立了官能化位点的阵列。位点为反应阳性的胺的亚微米区域,其经由硅烷化过程连接至氧化物表面。周边区域由中性的无反应的甲基基团组成。位点排列在分析玻片的狭窄方位的下面 4.5mm 宽的泳道中。目前利用胶水将 19mmx60mm 盖玻片粘合在芯片上。胶水形成最大为 4.5mmx19mm 的泳道。盖玻片与硅玻片间的间隔大约为 50  $\mu$ m。通过添加 50  $\mu$ m 玻璃球至胶水维持该 50  $\mu$ m 空间。

[0032] 19mm 宽的盖玻片基本上比最大 25mm 宽的硅玻片少,因为需要 5mm 用作入射口。入射口为在其上移液管吸取流体至泳道顶部的区域。毛细管力将泳道顶部的试剂移至盖玻片下面的间隙内。在玻片的底部,需要额外的 1mm 距离以清除过多的流体。

[0033] 在盖玻片正下方泳道的顶部和底部存在着 1mm-4mm 的阻进区。因为试剂蒸发,盖玻片比对精确性以及由于狭窄的入射口而造成的胶水侵蚀,需要该阻进区。考虑到所有这些允许偏差,可使用的分析玻片的宽度为常规玻片中总数可能 25mm 的约 12-15mm。

[0034] 在一已知的设计中,在分析玻片上构建 12 个 4.5mm 的泳道。这就产生了最大的可使用的宽度 54mm。然而 1mm/泳道专用于胶缝。这产生了最大的可使用的宽度只有 42mm。较少的泳道可为组合的,但是可以观察到这会使芯片不稳定,因为粘合层区域减小及比对指南的减少。

[0035] 考虑到这些尺寸大小,常规玻片全部可用的芯片区域百分比大约为  $(12.5\text{mm} \times 42\text{mm}) / (25\text{mm} \times 75\text{mm}) = 28\%$ 。人们需要的是提供增加可用面积的设计以及实现精确的比对的方法。

[0036] 下文为在其中应用本发明的研究中分析技术的描述。

[0037] 最近来自 Radoje Drmanac 带领的本发明的代理人的小组介绍了一种有前景的全基因组研究方法 (“Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays”, Radoje Drmanac 等, Science, 327, p 78-81, Jan 1, 2010 (其在美国法律下不是现有技术)。组合的探针锚定连接化学过程用于独立地测定来自于自组装的 DNA 纳米球的模式化纳米阵列的每一碱基。用约 1 个假变体 /100,000 碱基的精确度测序 3 个人类基因组。该平台的高精确度,低成本以及可伸缩性使完整的人类基因组测序成为可能,用于在大规模的遗传学研究中检测稀有的变体。

[0038] 在测量大约 25mmx75mm 大小的矩形的芯片上进行 Drmanac 研究中的生化实验。据报道每一芯片具有大约 10 亿个 DNA 纳米球以常规的矩形模式排列在其上。它对该阵列结构可视化是有用的。图 1 显示了此类生化阵列芯片 100 的概念图。因为大量的纳米球,芯片在概念上被分成域,例如域 105。典型的域大小可为 0.5mmx1.5mm,尽管精确的大小并不是关键的。可控规模的域使成像分析以可控的组块进行成为可能。在分步重复成像系统中域大小可对应于视野的系统域;在连续扫描系统中,域大小可为用于处理数据的便捷单元。

[0039] 参阅图 2,概念图为生化阵列芯片域 200 的显示。域包含点阵列(例如点 205、210、215),在其中进行 DNA 测序实验。尽管只用几百个点绘制了图 2 中的域,实际的域包含大约 10,000-1,000,000 个点。插图 220 显示了 6 个点,可以从这 6 个点中观察到任何四种颜色的荧光:蓝色 (“B”)、红色 (“R”)、黄色 (“Y”) 以及绿色 (“G”)。所用的实际颜色取决于所选择的荧光染料的选择并且可在染料发射光谱数据方面详细地说明。插图 220 中所示的 6 个点对应于自 6 个平行的 DNA 实验中读取的数据,每一个读取不同的光谱。在这种情况下,如插图 225 所示,荧光数据表示腺嘌呤 (“A”)、鸟嘌呤 (“G”)、胞嘧啶 (“C”) 以及

胸腺嘧啶 (“T”)。

[0040] 其目的是 DNA 芯片上的每个位点都包含一个它的序列待被测定的 DNA 链。图 2 中所示的插图 220 的阅读器对应测定 DNA 链序列的单一步骤。读取过程重复多次。

[0041] 重要的是与研究者正在看着的芯片上的点保持紧密的联系,否则通过记录荧光颜色获得的数据是无意义的。域点,即荧光染料分子发射光所处的位置在名义上位于常规的矩形模式中。实际的模式并不精确,因为 DNA 纳米球并不是总是精确地落在芯片上所定义的 DNA 附着位点的中心。用照相机观察域点,该照相机的图像传感器包含光传感像素的常规矩形阵列。

[0042] 人们需要的是使芯片上信息量最大的机制与方法,提供配准靶标并且提供成像系统的对照信息以便增加通量,从而改进测序熟练度。

[0043] 发明概述

[0044] 在基因组测序系统与方法中,提供工具以精确和准确地比对与配准用于光学分析的纳米球的平面阵列图像。通过使用最小化技术与 Moiré 求平均数校正亚周期 x-y 偏移、比例与旋转中误差。在 Moiré 求平均数中,设置放大率从而使得成像元素的像素周期为位点周期的非整数倍数。通过提供预定义的假随机位点组实现精确的配准,其中防止纳米球粘附至基质从而使得阵列的位点可以用在匹配模式的模式中,作为用于绝对位置鉴定的配准标记。

[0045] 在具体实施方案中,DNA 纳米球它们自己在阵列中除了保留位点以外的位点上自我组装,留下了假随机的位点或删除的微标记的二维模式,其相应的遮蔽(用于交互相关的数学模板)可以用于配准阵列至绝对位置。利用二维的空间误差的最小化,比例误差与旋转误差,以及还应用 Moiré 求平均数,可以在对来自图像分析域中所有 DNA 纳米球的阵列的光点的最佳匹配栅格中实现具有超过成像光学的光学分辨率的非常高程度的精确度的比对。因此,可以高度置信地提取与已知位置相关的信息,而同时使可包装在芯片上的信息量最大化。

[0046] 在具体实施方案中,删除模式包含超过 1% 并且少于 3% 的阵列。优选根据假随机模式选择删除点位置以辅助绝对位置配准。用常规的删除模式定位配准是可能的,只要明确的范围小于删除模式的周期性。

[0047] 其他的信息类型为行/列识别模式。每一域具有独特的删除模式,其鉴定芯片上的域位置,这允许芯片上域位置的独立确认。因此阵列中嵌入多重删除模式。选择每一彼此基本上正交的删除模式。

[0048] 在一具体实施方案中,精细比对技术使用基于雅可比矩阵 (Jacobian) 的距离最小化方法。

[0049] 本发明的主要目的为使基质上可用的官能化位点的量最大化,从而使得可以提取最大的信息量。本发明的其他目的为提供芯片设计,其允许附着位点上元件的快速分析,提供高密度的位点,使试剂使用最小化,使污染最小化,为机械强固的,为足够平坦用于成像的,为低成本的,并且易于制造。

[0050] 参考下列与附图有关的详细描述将更好地理解本发明。

[0051] 附图简要说明

[0052] 图 1 为生化测定芯片的概念图(现有技术)。

- [0053] 图 2 为生化测定芯片的域的概念图（现有技术）。
- [0054] 图 3 为示例根据本发明用于成像与在阵列芯片测定中生化处理的双重方法的概念图。
- [0055] 图 4 为根据本发明一测定系统的实施方案的透视图。
- [0056] 图 5A 为用于测定芯片的代表性顶侧载体的侧视图。
- [0057] 图 5B 为图 5A 的测定芯片与载体的顶视图。
- [0058] 图 6 为根据本发明比对引擎的原理框图。
- [0059] 图 7 为显示用于比对测定的实域与相关的分析栅格的概念图。
- [0060] 图 8A、8B 以及 8C 为示例旋转、比例以及偏移的图。
- [0061] 图 9 为为了说明 Moiré 求平均数的需要而显示一维的域 - 点 - 像素比对的图。
- [0062] 图 10 为用于示例 Moiré 求平均数优势的图。
- [0063] 图 11 为图 6 比对系统的基本流程图。
- [0064] 图 12 为用于示例域点成像的图。
- [0065] 图 13A 与图 13B 为显示比对之前及之后域与栅格模式的图。
- [0066] 图 14 为用于重复粗适应程序步骤的伪代码清单。
- [0067] 图 15 为显示根据本发明涉及分析的、很好的栅格适应程序矢量的图。
- [0068] 图 16 为显示根据本发明峰阈值辨别的结果的图。
- [0069] 图 17A 为显示栅格比对但是没有用域配准的图。
- [0070] 图 17B 为显示栅格比对并且用域配准的图。
- [0071] 图 18 为根据本发明具有处于假随机位置的删除的芯片域的图。
- [0072] 图 19 为显示栅格偏移数据的图。
- [0073] 图 20 为用于示例亚像素栅格比对的图。
- [0074] 发明的详细描述
- [0075] 图 3 为用于本文所述阵列芯片测定的成像与生化处理步骤的概念图。图 3 中，插图 300-320 与图 2 中插图 220 相似，因为它们显示了芯片上一些点的观察结果。插图 300、310 以及 320 显示了比对与成像步骤，在其中读取荧光数据；插图 305 与 315 显示了在成像步骤之间所进行的生化处理步骤。通过循环地成像与生化处理步骤测定芯片上某点的 DNA 序列。例如插图 300 中一个蓝色的荧光点“B”为圆圈的。该相同点在插图 310 中发红色荧光（圆圈的“R”）并在插图 320 中发绿色荧光（圆圈的“G”）。因此某点上成像与生化处理的 3 个循环显示了荧光序列“BRG”（蓝色、红色、绿色），其出于说明的目的示例于图 3 中，对应于 DNA 序列 AGT（腺嘌呤、鸟嘌呤、胸腺嘧啶）。
- [0076] 实际上图 3 中每一“比对与成像”步骤可涉及对应于所用的四种染料颜色的四种图像的获得。在超高密度系统中，将光学与检测设备调谐用于待被检测的每一波长。
- [0077] 当利用复合探针 - 锚定连接在自组装的 DNA 纳米球阵列上进行测序时，生化处理步骤（例如 305, 315）花费的时间显著长于比对与成像步骤（例如 300, 310, 320）。生化处理期间从成像装置上移除芯片，从而使得其他的芯片可以在那段时间期间成像。
- [0078] 图 4 为代表性测定系统 10 的示例，显示了具有在其上 DNA 样品待被光学分析的官能化硅芯片 12 的流式玻片，其安装于载体 14，该载体为真空夹盘 16 所持，该真空夹盘受花岗岩桌子 20 上所放的二维的转化台 18 所支撑。放置物镜 22 以观察硅芯片 12 的表面并且

向成像系统 24 提供图像以向比对引擎 26 供给信息 (图 6)。

[0079] 成像系统 24 包含四通道的荧光检测系统 (不可见的), 其中每一通道测量一种类型的碱基的荧光, 该碱基为位于芯片 12 上每一附着位点的遗传物质链的碱基。光出现点或区域, 具有来源于位点某处的光峰。在目前的情况下, 分别用 Fam, Cy3, 德克萨斯红以及 Cy5 染料进行 T, G, C, & A (酪氨酸、鸟嘌呤、胞嘧啶以及腺嘌呤) 存在的测量。将针对每一颜色的过滤器 28 移到位并且针对每个域的全部四种图像捕捉每一颜色的图像。通过使用多个成像系统与多个过滤器将步骤组合。

[0080] 参阅图 5A 与图 5B, 示意图显示了其上安装有芯片 12 的载体 14 的实施方案。在侧视图中 (图 5A), 示例了一个通道 13, 样品在表面 15 上, 流体经由输入口 17 可进入并且从输出口 19 排水。在顶视图中 (图 5B), 见到的是存在着许多与通道 13 相似的平行通道, 其每一个都具有基本上大于芯片与载体之间间隔的宽度。可使用其他的芯片 / 载体设计, 只要存在沿着通道 13 观察位点荧光的方法。

[0081] 参阅图 6, 通过功能元件示例合适的引擎 26, 该功能元件用于进行 2D 比对与配准过程。在下文中还进一步地解释了涉及的方法, 例如关于图 11。分析引擎 26 有效地进行三种主要的操作: 首先, 捕获阵列中基本模式下的图像; 第二, 比对该模式, 以及第三配准图像。作为该过程的准备, 并且作为输入来自于芯片 12 上物质的模式化阵列的光学数据, 首先通过用于去除噪音的背景遮蔽表征模式化阵列。通过背景减数器 30 从分析栅格中删除最暗的位点, 通过称为形态学开运算 (元件 32) 或逐块最小值 (元件 34) 的方法。形态学开运算方法被认为是更精确的但是这两种方法中较慢的, 并且它可以单独地使用或者与其他的方法联合以实现背景减除功能的输出。

[0082] 下一步, 二维的峰值检测器 36 对从芯片 12 成像的小块像素进行峰值检测 (见图 12), 通常在 5x5 或 3x3 的阵列中进行以光学地鉴定荧光的局部最大值以收集可用的数据。小块可能重叠, 因为物镜横穿并沿着芯片 12 的 X 轴与 Y 轴步进。

[0083] 此后激活栅格比对亚系统 38。理解成为栅格比对亚系统 38 操作基础的因素是有帮助的。第一步操作, 建立栅格模式以配合芯片 12 的图像, 这受制于四个主要的自由度: 比例、旋转以及 X 与 Y 偏移, 如下文所解释的。在光学像散性或歪曲像差存在的情况下, 可包括更多的自由度。像散性导致栅格的 X 轴与 Y 轴间不同比例的系数。径向放大率变异为几个可能的歪曲像差的其中之一。每一额外的变异度将增加其他轴线中一定量的不确定性。每一额外的自由度也耗费分析时间, 所以通常只使用四个自由度。在栅格比对方法中通常存在着四个嵌套的过程。首先为松转配合 42, 其中进行强力的参数取样以迅速大体地使栅格与点配合。第二为第一遍精细配合 44, 其中分析优化松转配合 42 中所获得的配合。此后, 数据接受峰阈值过程 46 以使额外的用于处理的候选点具有资格, 如下文所解释的。此后为第二遍精细配合分析优化 48 以微调栅格比对 38。

[0084] 通过栅格配准亚系统 40 进行第二次操作, 直到最后配准栅格模式, 从而使得配准是绝对的。在该操作中, 找到图像中栅格模式的位置。目标是鉴定已成像哪一种栅格模式。通过观察位点占据完成该操作。如下文所解释的, 存在基于删除模式进行的两种类型的位点占据测试: 在所有域 49 中使用普通的栅格删除, 和使用行 / 列删除 51 以鉴定特定栅格的行与列。

[0085] 因此, 参与上述的 2D 比对与配准引擎的四个过程为: (1) 背景减除, (2) 峰值检测,

(3) 栅格比对, 以及 (4) 栅格配准。下文中对它们进行了更为详细的解释。

#### [0086] 背景减除

[0087] 背景减除为抑制杂散光的必需过程。荧光系统中的杂散光可具有梯度。这些梯度可为流式玻片上不完整性与污点的结果。因此, 背景减除总是必需的。

[0088] 存在许多背景减除的不同方法。图像存储为强度并且不允许为负数。背景减除通常由以下组成: 找到区域内最低强度的像素并且在它的强度上校正。最简单的技术为块背景减除, 其中在小的亚区域内找到最低像素的强度。提供稍微更好的结果的方法为找到区域内最低像素的形态学背景方法, 其中区域定义为扫描整个图像的窗口。它比块减除慢, 因为它需要多得多的计算时间。可使用其他的背景减除方法。

#### [0089] 峰值检测

[0090] 通过鉴定在  $3 \times 3$  或  $5 \times 5$  区域或由  $(2n+1) \times (2n+1)$  所定义的更大的区域上的最大值的像素找到峰值, 其中  $n=1, 2, 3, 4, \dots$ 。所选过滤器的大小为栅格固有间距与考虑中点的大小的函数。出于本发明的目的, 间距  $\leq 4$  个像素 / 点并且点大小占据 1-10 个像素, 因此优选  $3 \times 3$  与  $5 \times 5$  栅格。

#### [0091] 栅格比对

[0092] 使用雅可比变换 (Jacobian transformation) 的特征比对栅格模式。通过使抽象的点栅格 405 适合点的实际域 400 示例该过程。

[0093] 参阅图 7, 为辅助追踪域点, 使用抽象的栅格 405。栅格代表理想阵列的坐标轴, 其使域点尽可能紧密地适合域 400。就它是完全平方的意义而言, 栅格是理想的并且具有恒定的周期性。

[0094] 确切地知道栅格位于像素参考坐标系何处是有用的, 因为在图像传感器处可观察到的仅有的量为像素强度。图 8A、8B 以及 8C 分别示例了栅格旋转误差、栅格比例误差以及栅格偏移误差的概念。找到栅格位置的过程为尽可能多地减小这些误差的过程。当如图 8A 中域与栅格相对彼此旋转时, 旋转误差发生。当如图 8B 中栅格具有不同于域的周期性的周期性时, 比例误差发生。当如图 8C 所示例如在 X 或 Y 方向平面内置换栅格时, 偏移误差发生。一种类型的偏移误差为小于栅格点间距离的一半的偏移。另一种类型的偏移误差, 称为配准误差, 为整数栅格周期的偏移。

[0095] 人们将栅格点与域点配合, 然而仅经由离散的照相机像素观察, 因为在无限高的分辨率下域点为不可观察的。谨慎选择域点周期与照相机像素周期之间比率, 从而避免整数关系, 提高了栅格比对操作的精确度。首先考虑当域周期为像素周期的整数倍数时使域点成像的一维的实例, 如图 9 所示。图 9 中, 包括点 600、605 等的一行域点具有  $\Delta f$  周期或点之间的间隔, 包括像素 620、625、630 等的一行照相机像素具有  $\Delta p$  周期。在图 9 的实例中,  $\Delta f = n \Delta p$ , 其中  $n$  为整数。考虑用像素行观察点行。在图像“ A ”中比对像素中心的点。在图像“ B ”中比对像素左上角的点。然而, 不可能只使用像素区分图像“ A ”与“ B ”。当光落在其上时不能确定像素。因此在图 9 的实例中, 像素以及因此栅格可以偏离多达  $0.707 \Delta p$  并且很难检测到这种误差。

[0096] 一个解决方案为选择不是像素周期整数倍数的域 (与栅格) 周期, 如图 10 的域点成像的一维的概念性实例所示。图 10 中, 包括点 700、705 等的一行域点具有  $\Delta f$  周期或点之间的间隔, 包括像素 720、725、730 等的一行照相机像素具有  $\Delta p$  周期。在图 10 的实例

中,  $8\Lambda_f = 9\Lambda_p$  (相当于  $\Lambda_f = 1.125\Lambda_p$ )。考虑用像素行观察点行, 如在 740 所示, 其中像素被“1”直至“9”标记。域点 1 与 9 及像素 1 与 9 显示只在每第 9 个像素处存在精确的比对。在介于中间的像素中点与像素的相对位置席卷彼此。框 750 为像素“1”直至“9”彼此叠加的放大视图。点均匀地分布在放大叠加的像素。点与像素周期的不同导致在相同的步骤中点取样像素的长度。叠加 750 中所有点位置的平均数引起对最适合域位置 (即栅格位置) 的估计, 系数减小误差  $\frac{1}{\sqrt{N}}$ , 其中 N 为重复间的像素数; 在本实例中 N=9。

[0097] 显然地, 域点没有无限小的程度, 但是实际上光按空间分布发射。象形文字 760 显示了从点光源到具有标准差  $\sigma$  所表征的高斯程度的光源的概括。在那种情况下, 可获得的匹配误差为与  $\frac{\sigma}{\sqrt{N}}$  成比例的。

[0098] 选择不是像素周期整数倍数的域 (与栅格) 周期允许配合亚像素分辨率的栅格。图 9 与 10 只示例了在域阵列与像素阵列间建立有意的 Moiré 模式的一维的实例。在实际的系统中, 将相同的原理应用于二维。然而, 二维的 Moiré 模式的好处大于正交一维的 Moiré 模式的简单连续应用。

[0099] 在叠加的像素 750 中, 取样的域点位置自左向右横过像素在一条直线上。以垂直的方向应用 Moiré 原理, 但是进行两个单独的 (例如 X 与 Y) 验配步骤, 将引起点取样像素如所示的在一条直线上并且沿着垂直线。然而, 如下文所见的, 二维的栅格对齐涉及在一步骤中使栅格位置与域点间平方距离 (例如  $\Delta x^2 + \Delta y^2$ ) 最小而不是分别使 X 与 Y 误差最小。当平方距离为最小化数量时, 整个叠加的像素区域被取样。

[0100] 在图 10 的实例中, 域点与像素每九个像素排成一行, 因此定义 Moiré 边缘或拍合期。一个域中所见到的边缘数量通常选为大概 10 个 -20 个。边缘的平直度可用于检测域失真。

[0101] 用于栅格在二维中尽可能精确地配合域的程序的结果为常规的矩形的栅格, 其为适于观察在照相机成像传感器中像素的坐标系中的域。图 11 显示了在根据本发明的系统中使用的栅格比对和配准程序并且对应于图 6 的引擎。图 11 中, 步骤 805, 810, 815, 820, 825, 830 以及 835 通常以所示的顺序进行, 然而, 步骤的顺序并不是程序的要求。而且, 并不是所有步骤在所有的情况下都需要, 因此可以省略一些步骤。例如, 不是所有情况都需要峰阈值区别 825 与精细栅格配合 830, 例如当不要求高精度度时。

[0102] 开始于通过照相机中的成像传感器所记录的域的图像, 步骤 805 为从图像减去背景信号水平。步骤 810 为检测域点, 即找到图像中出现的通过荧光染料所发射的信号在哪里。步骤 815 为使用迭代程序进行松转栅格配合域点。步骤 820 为使用分析程序进行精细栅格配合域点。步骤 825 中域点通过峰阈值区别重整资格。步骤 830 为第二遍精细栅格配合域点, 使用在重整资格的域点上进行的分析程序。步骤 835 为在适当的偏移处配准栅格, 换句话说, 校正任何模栅格周期误差。

[0103] 图 12 显示了域点成像的概念图。在图 12 中, 盒 905 中显示了几个域点。点具有不同的亮度, 一些点完全没有发射出任何荧光, 而另一些为光亮的。在 DNA 测序系统中所获得的一典型的域中, 只有四分之一或更少的域点为可见的, 因为一次只有一种荧光颜色成像。因此, 域可为稀疏的。

[0104] 插图 910 显示了通过成像传感器中像素所记录的一些域点。可在超过一个像素中

检测到来自于单一的域点的光。还有一些背景噪音 - 像素,其记录大于零亮度,当没有光从最近的域点发射出来时。图 12 中概述的程序的第一步为背景减除,如上文所提及的。用于背景减除的一个方法为取走在图像中通过最暗的像素所记录的亮度并且减去来自于图像中的每个像素的亮度。例如在插图 910 中,像素 915 为最暗的并且可从插图中的每个像素减去它的亮度。通常在邻近的像素块上进行背景减除,其代表域中点的子集,即不是一次在域的图像中所有像素上。

[0105] 一旦将背景水平从图像中减除,就检测域点。用于检查像素子集的一种技术,即 5x5 阵列,并且指定最亮的像素的位置为域点的位置。另一种用于检测域点的方法为使高斯分布配合像素亮度水平并且使用最接近最佳配合分布的峰的像素的位置。插图 910 中,白色的三角形表示像素,诸如像素 920,其已被鉴定为域点的位置。

[0106] 一旦鉴定了域点,就可开始栅格配合域点的过程。作为起始点,可确定与图像传感器中的像素比对的栅格。栅格配合涉及调整栅格以将它尽可能紧密地与域点对比。

[0107] 调整栅格的四个参数,旋转、比例(即放大率)以及侧面的(x与y)偏移以找到最佳的配合。研究者可用允许像散性、失真扭曲以及其他类型的误差更精确配合的更多参数设计栅格。这些更高阶的误差是由于诸如透镜、芯片基质或转化台瑕疵的因素。这些误差不如旋转、比例以及偏移便以在成像系统中动态地抵消。

[0108] 图 13A 与图 13B 为在二维的比对之前(“A”)以及之后(“B”)域 1000 与栅格 1005 的概念图。比对后,域点与栅格点的位置基本上匹配,除了偏移误差取模栅格周期之外(通过等于栅格周期的距离常规的栅格转化留下了不带电荷的栅格,没有边缘位置的知识或相当于栅格的起源。下文描述了用于找到起源或“配准”栅格的方法)。

[0109] 栅格比对涉及迭代的松转配合程序与分析的精细配合程序。迭代的松转配合程序的结果为其中每个栅格点比任何其他域点更接近正确的域点。图 13 中,域 1000 与栅格 1005 在“A”中对偏从而使得栅格点有时候比正确的那些域点更接近错误的域点。在“B”中每个栅格点最接近,为正确的域点,按模计算栅格周期。

[0110] 图 14 为迭代的松转栅格配合程序中步骤的伪代码列表。盒 1100 中步骤为四倍地嵌套圈,其评估参数值的四维网格上的功能。不同的参数为:栅格旋转(“ $\theta$ ”)、栅格放大率或比例(“S”)、栅格 X 转化或偏移(“X”)并且栅格 Y 转化或偏移(“Y”)。在每一网格点上评估,功能, F, 为每一栅格点与最近的域点间平方距离的总和:  $F = \sum_i \| \mathbf{r}_{p_i} - \mathbf{r}_{d_i} \|^2$

[0111] 其中矢量  $\mathbf{r}_{p_i}$  为域点的位置,矢量  $\mathbf{r}_{d_i}$  为最近的栅格点的位置或栅格中参考并且总和 F 为遍及域点。

[0112] 对测试的每一(  $\theta$ , S, X, Y ) 组合而言记录 F 值。产生最小值的 F 的组合为使用该迭代程序的最佳可用配合。有限步骤大小  $\Delta \theta$ ,  $\Delta S$ ,  $\Delta X$  及  $\Delta Y$  影响进行松转配合所要求的时间长度,因为它要花费更多的时间以通过参数值的细网格。松转配合程序不可能产生最佳的配合,因为最佳的(  $\theta$ , S, X, Y ) 组合几乎肯定位于网格点间。选择步骤范围  $[\theta_{\min}, \theta_{\max}]$ ,  $[S_{\min}, S_{\max}]$ ,  $[X_{\min}, X_{\max}]$  以及  $[Y_{\min}, Y_{\max}]$  从而使得其中每一栅格点最接近于它的正确的域点的解决方案包括在参数网眼中。

[0113] 总之,当处理一组域点时,研究者并没有提前知道要寻找多少荧光信号(例如来自 DNA 纳米球)。来自只是几个百分点至 50% 或更多的域点,可照亮任何地方。其他的像

素,其似乎代表域点,简单地可为噪音。因此,在松转配合程序中,只使用亮的域点。通常检测到的只有最亮的 15% 域点包括在松转配合中,尽管特别的阈值为工程选择的问题。

[0114] 一旦进行迭代的或松转的配合,精细栅格配合程序使配合最佳化。概念上地,精细栅格配合相当于用橡皮筋将每一栅格点(然而定义的)连接最近的域点,然后允许栅格调整其自身。图 15 显示了精细栅格配合程序中涉及的矢量。图 15 中  $r = (x_i, y_i)$  为域点,而  $r_{grid} = (p_{mi}, p_{ni})$  为平方栅格的栅格点。此处  $p$  为栅格点(可选择地被称为“斜度”、“周期”、“比例”或“放大率”)并且  $m_i$  与  $n_i$  为整数索引,其选择最接近于域点的栅格点。矢量  $\vec{r}_0 = (x_0, y_0)$  为栅格置换矢量并且矢量  $\vec{r}' = (x', y')$  为域点间的误差并且矢量栅格总和以及栅格置换。误差与域有关并且通过以下置换:

[0115]  $x'_i = x_i \cos \theta - y_i \sin \theta + x_0 - m_i p,$

[0116]  $y'_i = x_i \sin \theta + y_i \cos \theta + y_0 - n_i p,$

[0117] 其中  $\theta$  代表栅格的旋转,对垂直于  $x$ - $y$  平面的轴周围的域而言。精细配合程序的目标为找到  $x_0, y_0, \theta$  及  $p$  的最佳值,其使栅格与域间的误差最小。为找到这些值,平方误差的总和写为:

[0118] 
$$\chi^2 = \sum_{i=1}^N ((x'_i)^2 + (y'_i)^2)$$

[0119] 其中  $N$  为域点的数量。下一步,  $\chi^2$  的偏微分,对每一四个独立的变量  $x_0, y_0, \theta$  及  $p$  都设定等于零,其产生了下列的方程组(称为雅可比变换):

[0120] 
$$\begin{bmatrix} -X_i \\ -Y_i \\ M_i X_i + N_i Y_i \\ 0 \end{bmatrix} = \begin{bmatrix} -Y_i & -M_i & N & 0 \\ X_i & -N_i & 0 & N \\ M_i Y_i - N_i X_i & M_i^2 - N_i^2 & -M_i & -N_i \\ X_i^2 + Y_i^2 & M_i Y_i - N_i X_i & -Y_i & X_i \end{bmatrix} \begin{bmatrix} \theta \\ p \\ x_0 \\ y_0 \end{bmatrix}$$

[0121] 
$$\begin{bmatrix} -X_i \\ -Y_i \\ M_i X_i + N_i Y_i \\ 0 \end{bmatrix} = \begin{bmatrix} -Y_i & -M_i & N & 0 \\ X_i & -N_i & 0 & N \\ M_i Y_i - N_i X_i & M_i^2 - N_i^2 & -M_i & -N_i \\ X_i^2 + Y_i^2 & M_i Y_i - N_i X_i & -Y_i & X_i \end{bmatrix} \begin{bmatrix} \theta \\ p \\ x_0 \\ y_0 \end{bmatrix}$$

[0122] 此处,小角度近似值  $\cos \theta = 1$  与  $\sin \theta = \theta$  已用于使方程式线性化,并且根据常规代表总和:以及

[0123]

$$A_i = \sum_{j=1}^N a_j$$

[0124] 
$$A_i = \sum_{j=1}^N a_j$$

[0125] 
$$A_i B_j = \sum_{k=1}^N a_k b_k$$

[0126] 矩阵方程式的解决方案根据需要产生  $x_0, y_0, \theta$  以及  $p$ 。

[0127] 一旦获得第一遍精细栅格配合,通过峰阈值区别以及随后的附加的精细配合可获得配合中进一步的改进。在松转配合与第一遍精细配合中,只考虑了一小部分的域点。通常只有检测到的最亮的 15% 域点包括在第一遍精细配合中,但是实际上百分比为工程选择

的问题。在第一遍中只使用最亮的点的理由为许多较暗的点可能是噪音。松转与精细配合后,然而,研究者可重新评价哪一个点经由峰阈值区别包括在配合中。

[0128] 为建立用于包括配合中的域点的可用的亮度阈值,所有域点,并不仅仅是在最初的松转与精细配合中所用的那些,被分成组。例如组为 50 个最亮的点,接下来 50 个最亮的点,或第一个几千与接下来的几千等(数目只需要足够大从而使得对所有组中的点而言均方距离误差具有统计学意义)。下一步在均方距离误差对亮度的图上标绘组。图 16 为针对典型的域峰阈值区别数据的图 1305。图 16 中,组中具有低均方距离误差的域点比具有高均方距离误差的那些亮。图具有相当急剧的弯曲,其中对较暗点的组而言均方距离误差迅速增大。在靠近该点处画上阈值,诸如图 16 中所画的阈值 1310。具有小于阈值的均方距离误差的组中域点的数量通常显著地多于松转与第一遍精细配合中所用的 15% 的域点数量。然后大量的域点用于重复所述的与图 15 有关的精细配合程序。图 16 中,例如,所有比阈值 1310 下最暗的亮点的域点可包括在第二遍精细配合中,阈值 1310 下最暗的点即虚线 1315 右边的那些。

[0129] 在这一点上,比对栅格而且它可以为荧光点的域。栅格可能以 X 与 Y 方向通过栅格周期的整数仍然偏移域。图 17A 显示了栅格的概念图,将其与域比对,但是并不在正确的偏移处配准。图 17A 中,比对域 1400 与栅格 1405,但是在“A”处没有配准。图 17B 中,在“B”处正确地配准域 1400 与栅格 1405。然而迄今为止在图 17A 与图 17B 的状况“A”与“B”间不能区别所述的比对与配合程序。因此需要额外的配准程序。

[0130] 使用域点的假随机删除模式高效地完成根据本发明配准。图 18 显示了根据本发明具有假随机删除模式的二维的域 1500 的概念图。应理解的是阵列可包括支持诸如与 DNA 纳米球的探针-锚定连接的生化实验的数十亿位点,并且基质的区域除了表面位点之外并不支持生化实验。此类区域为化学处理的以减少生物分子的非特异性结合。域中的一些点,诸如点 1505 与 1510 为缺少的,即删除的。荧光标记的生物分子不可能在那些点处出现,因此知道哪个点缺少可用于域与栅格的精确配准。

[0131] 使用半导体光刻技术以及电子束直写技术与微触印刷技术使用于自组装 DNA 纳米球阵列的芯片模式化。例如带正电的胺基团可模式化以结合域阵列中带负电的 DNA 纳米球。通过去除用于模式化胺基团的光刻掩模的特征完成域阵列中点的删除。如果特定的域点位置没有胺基团, DNA 纳米球就不可能粘附。

[0132] 通过使用相互关联技术使假随机删除模式用于配准以匹配已知的“模板”模式或来自于域阵列的假随机模式。如果用域正确地配准模板,将见到很少的光线,因为用删除的点将洞排成了一行。如果通过一些数量的域周期从域置换模板,碰巧落在洞下面的来自于点的光线将穿过模板。穿过模板的光线数量大致相当于所有置换,除了当模板与域配准时之外。

[0133] 图 19 显示了图 1605 中栅格配准数据的实例。图显示了偏移 0-81 个域周期的亮度。这些 81 个光栅模式偏移为二维的 X 与 Y 偏移的 9x9 域。对除了栅格周期之外每一偏移而言,数字 40,出现在假随机模式中的域点的亮度为大致 425 个任意亮度单位。在偏移 40 处,亮度接近于零。因为 81 偏移周期的光栅模式,偏移 40 指的是 X=4, Y=4。这是配准栅格与域所要求的偏移。

[0134] 假随机删除模式可用于鉴定域以及配准栅格点与域点。参阅图 18 的举例说明,用

第一个假随机删除模式提供每一域,该假随机删除模式对所有域都是常见的并且用于配准的。另外,每一域也可具有两个进一步的假随机删除模式:一个是鉴定芯片上特异域的行而另一个是鉴定芯片上特异域的列,如与图 6 的系统有关的所示。因此,每一观察到的域模式互相关抗普通的用于配准假随机删除模式并且抗许多其他的假随机删除模式以确定域位于芯片上何处。理想的是在假随机模式中删除大约 3% 域点;然而,至少一组包含只有 1% 删除的模式产生了可接受的结果。大于 3% 的删除模式只提供了精确度中临界的增加并且消耗了可用的空间,而少于 1% 的删除模式挑战配准中潜在地不可接受的误差。

[0135] 本文所述的比对与配准技术允许在高密度、生化阵列实验中所观察到的域点的快速准确的鉴定。这些技术允许在具有较少的像素/域点的笛卡尔坐标阵列上进行实验,否则是可能的。用亚像素精确度确定域点的位置,如配合理想的栅格。因此将每一域点的亮度计算为通过接近域点的像素记录的像素强度的加权平均数。图 20 示例了如何使用亚像素栅格比对的实例。

[0136] 图 20 中,显示了像素的代表性 9x9 块(像素 715, 1720, 1725, 1730, 1735, 1740, 1745, 1750 以及 1755)。通过上文所述的栅格比对与配准技术测定域点 1705 的位置。圆圈 1710 为域点 1705 所发射出来的光线的横向程度的概念表征。研究者可将圆圈 1710 认为是高斯分布的 1-西格玛半径,例如,如果已知域点 1705 的位置只有一个像素精确度,那么研究者可估算点的亮度如通过像素 1735 所报道的亮度那样或研究者可加权像素 1735 50% 的亮度并且每一像素 1720, 1730, 1740 以及 1750 12.5% 的亮度。考虑到高精度度,域点位置的亚像素估算,然而,根据本发明更精确的加权是可能的。例如,在图 20 的实例中,像素 1730, 1735, 1745 以及 1750 的加权分别可为 20%, 50%, 10% 以及 20%, 代表每一像素与圆圈 1710 之间的重叠。

[0137] 基因组研究与其他的高密度生化阵列的使用要求先进的成像方法,诸如本文所述的那些以获得商业上可行的数据采集率。高密度有序的芯片与先进的比对及配准技术为系统的关键要素,使大规模的人类基因组研究强有力。

[0138] 另外,本发明的设备与机械在许多用于未知的序列的核酸的生化询问方法中 useful。例如,本发明的分析玻片可与基于杂交的方法使用,诸如美国专利 6,864,052;6,309,824;以及 6,401,267 与美国出版的专利申请 2005/0191656 中公开的;通过合成方法测序,诸如美国专利 6,210,891;6,828,100;6,833,246;6,911,345;Ronaghi 等(1998), *Science*, 281:363-365;以及 Li 等, *Proc. Natl. Acad. Sci.*, 100:414-419(2003);以及基于连接的方法,例如 W01999019341, W02005082098, W02006073504 以及 Shendure 等(2005), *Science*, 309:1728-1739。达到相关法律所允许的程度,出于所有的目的将这些出版物的内容通过引用并入本文。

[0139] 在特定的方面,在具有多个生化测序反应的高通量分析中使用多个分析玻片。例如可并排地排列流式细胞类型分析玻片或前面在测序反应系统中一个在另一个的前面。多个流式细胞任选地包括核酸或引物粘附于流式细胞的基质,随机地或以预定的方式,从而使得贯穿反应过程可以监控多个流式细胞中的每一核酸。核酸或引物可粘附于表面从而使得至少一部分的核酸或引物为单独地光学上可分辨的。

[0140] 在一实施方案的优选方面,供用于本发明的系统中的流式玻片包含基质,将未知序列的核酸固定于其上。在实施方案的某些方面,夹紧工具能够一起夹紧许多的流式细胞。

通常,1 个 -12 个或 16 个左右的流式细胞通过单一的夹紧工具同时地夹紧。以基本上水平的或基本上垂直的方式在夹紧方法中安排流式玻片,尽管介于这两个位置之间的任何位置都是可能的。

[0141] 本说明书在目前描述的技术的实例方面提供了方法学、系统和 / 或结构及其用途的完整描述。尽管上文已描述了具有一定程度特殊性的该技术的各方面,或参考一个或多个个体的方面,本领域中技术人员可做许多公开方面的修改而没有背离其技术的精神或范围。因为可以做许多方面而没有背离目前描述的技术的精神与范围,适当的范围存在于下文所附的权利要求中。因此考虑其他的方面。任何顺序今夕而且,应理解的是可以任何顺序进行任何操作,除非另有明确的声明或特定的顺序通过权利要求语言为固有必需的。其目的是包含在以上描述与显示于附图的所有物质应解释为只是特定方面示例的并且不限于所示的实施方案。可进行细节或结构的变化而没有背离如下列的权利要求中所定义的本发明技术的基本要素。依据法条 35U. S. C. § 112 ¶ 6,在任一相应的实用性的申请的权利要求中,除非使用术语“方法”,其中列举的特征或元素都不应解释为方法加上功能局限性。

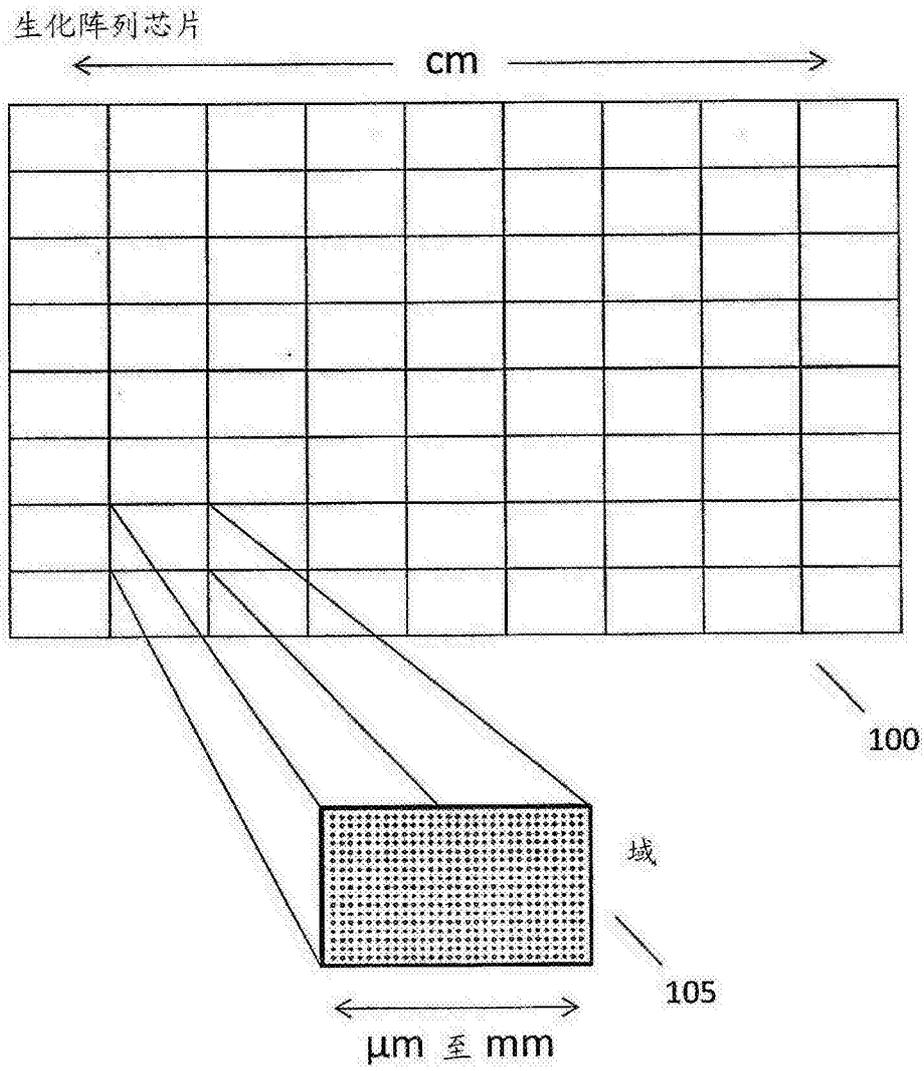


图 1

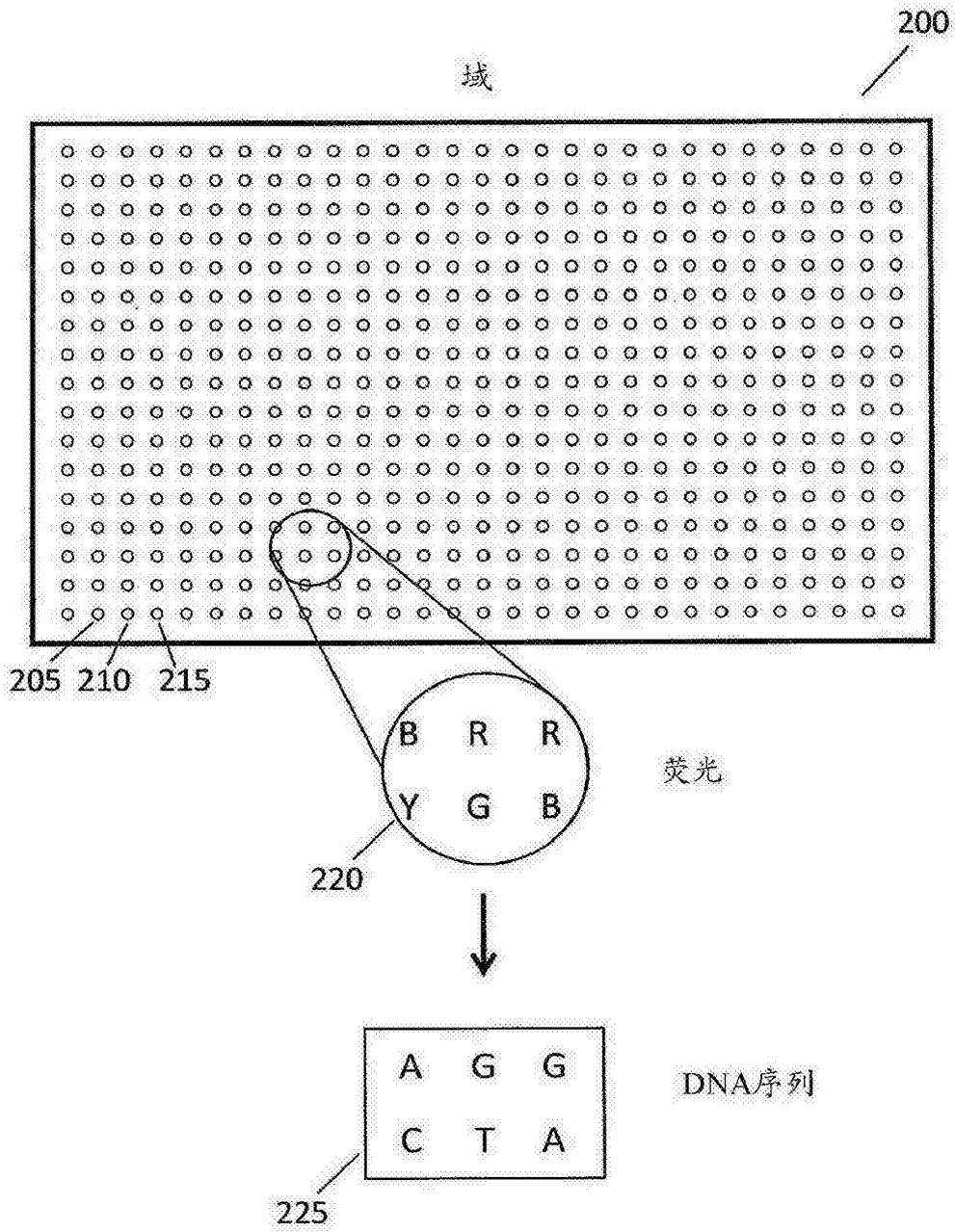


图 2

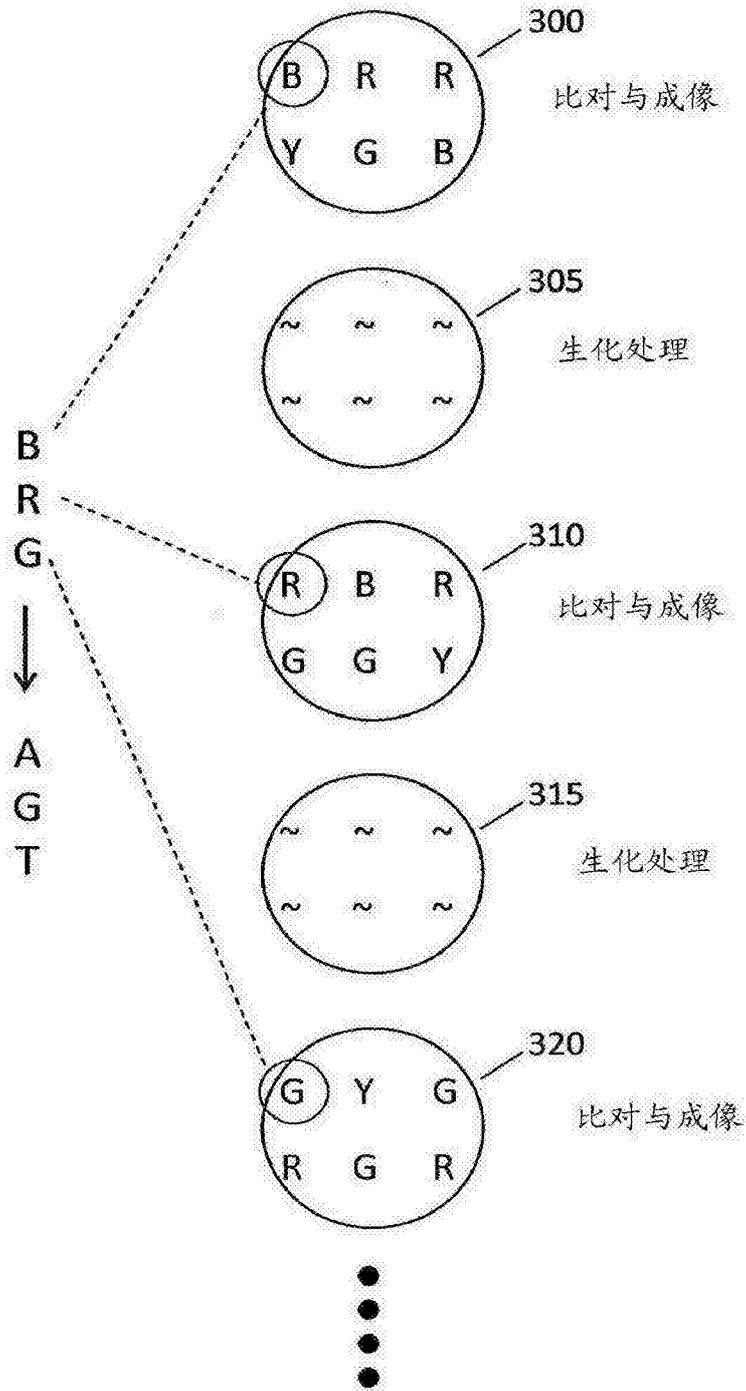


图 3

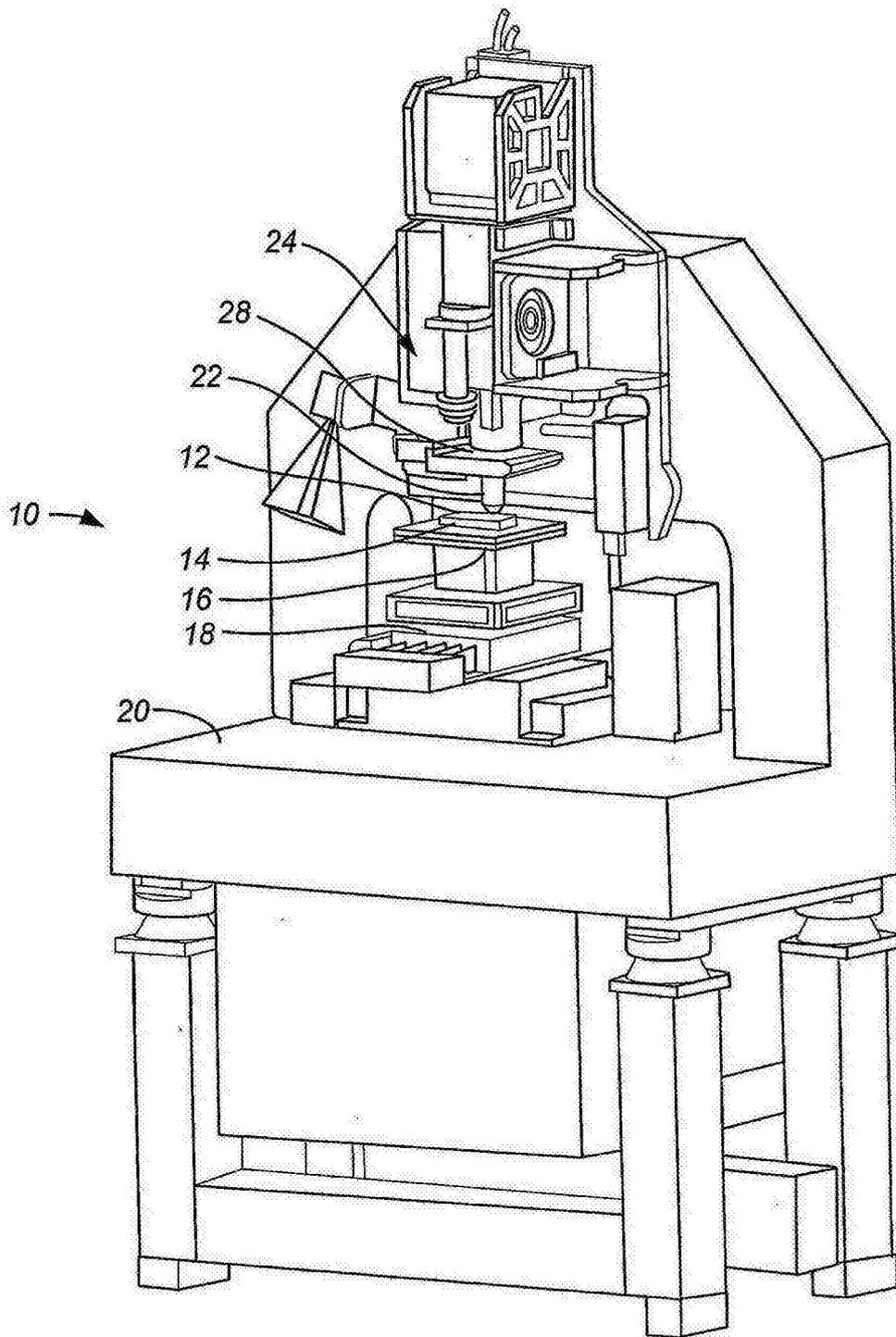


图 4

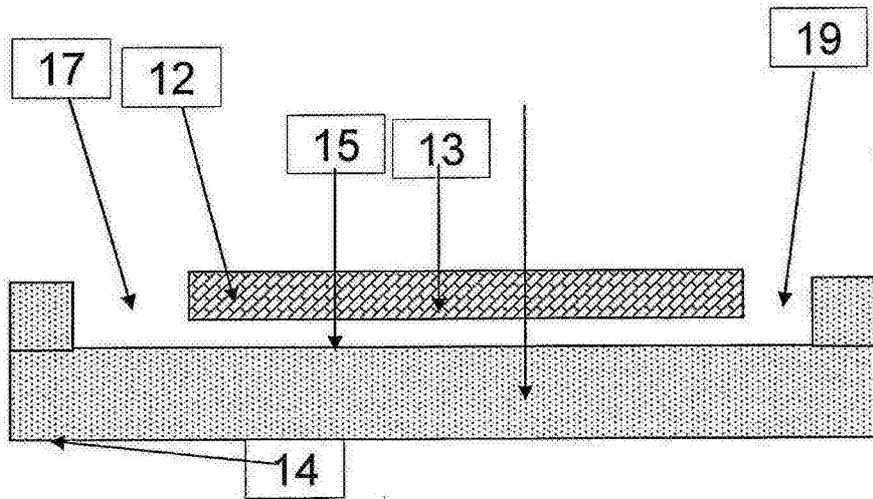


图 5A

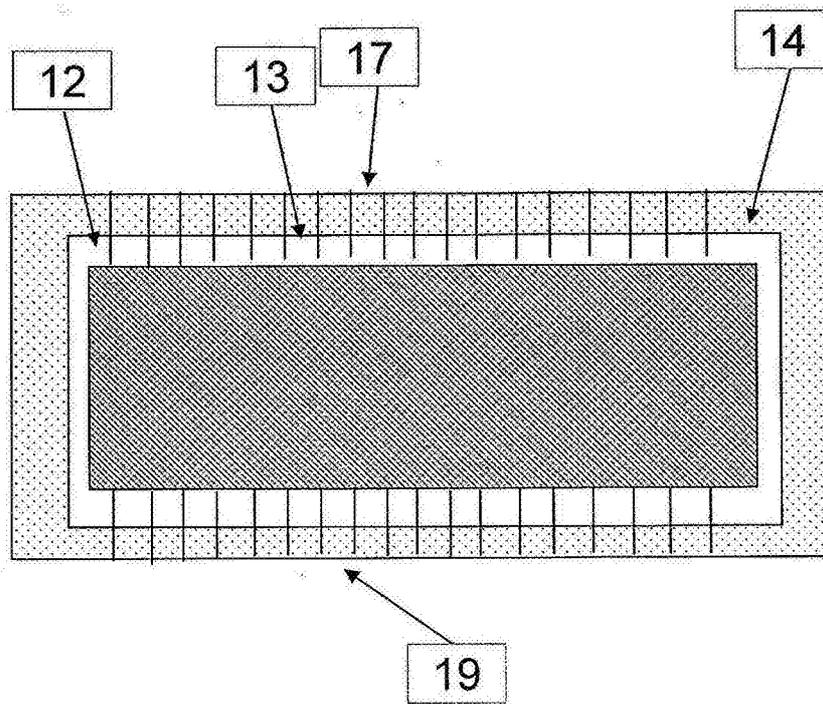


图 5B

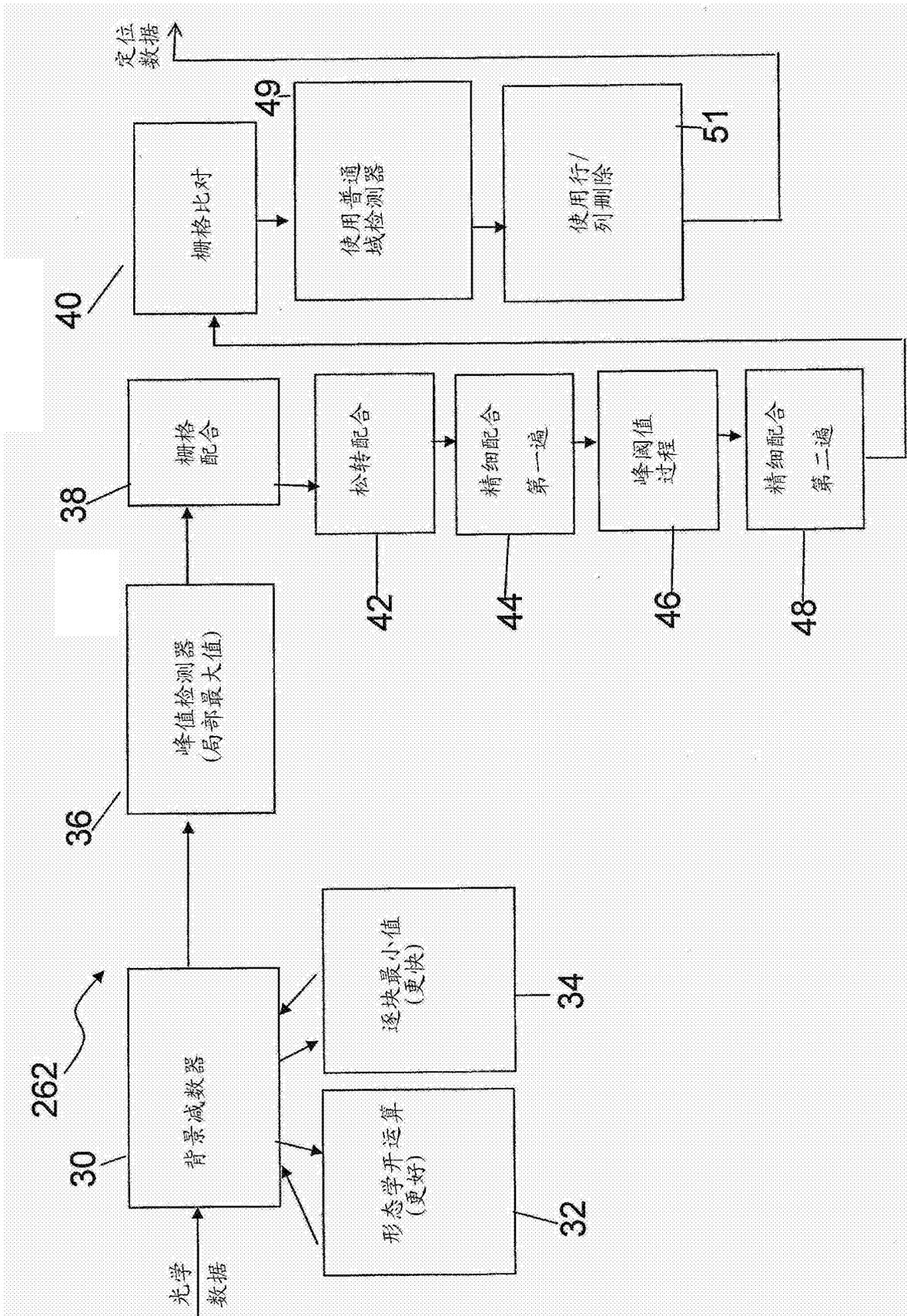
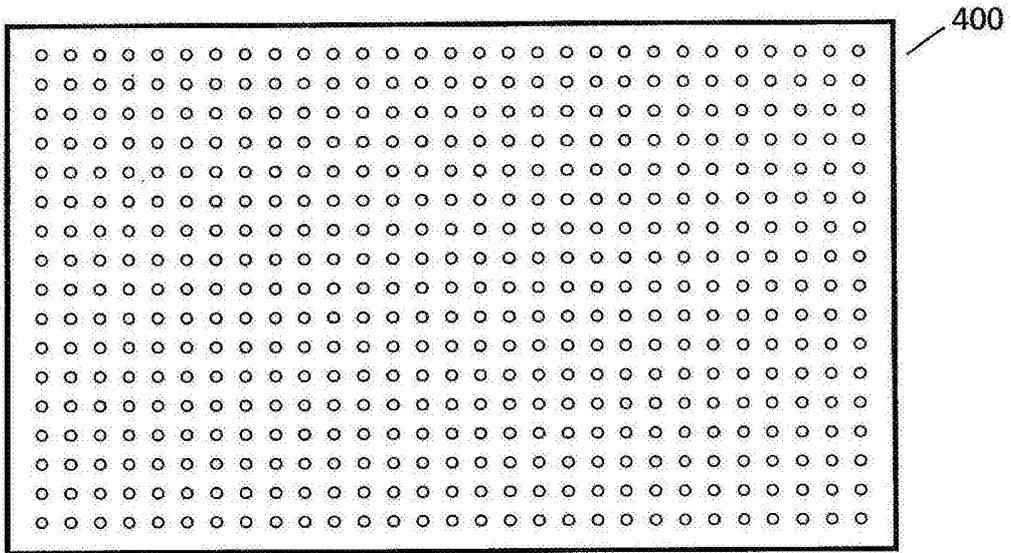
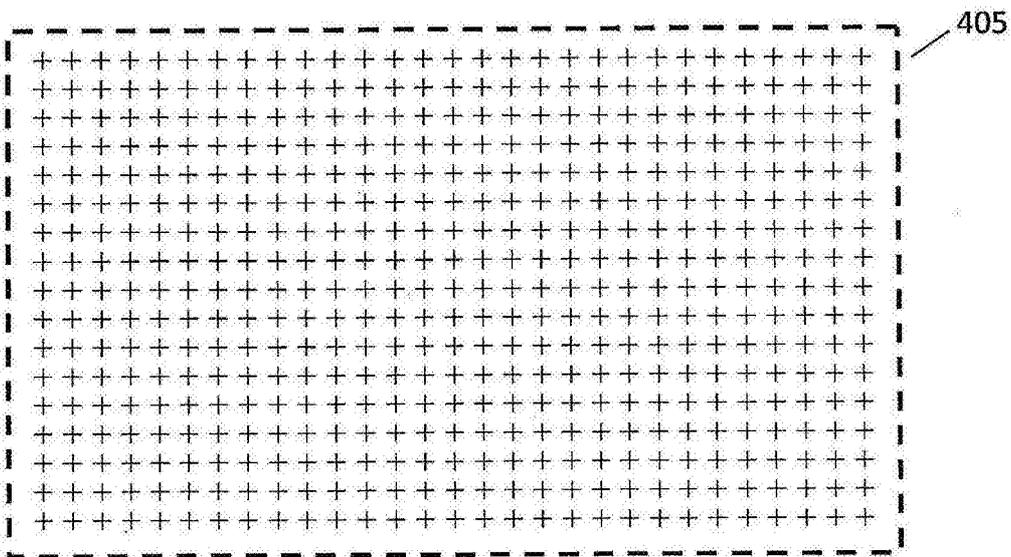


图 6



域



栅格

图 7

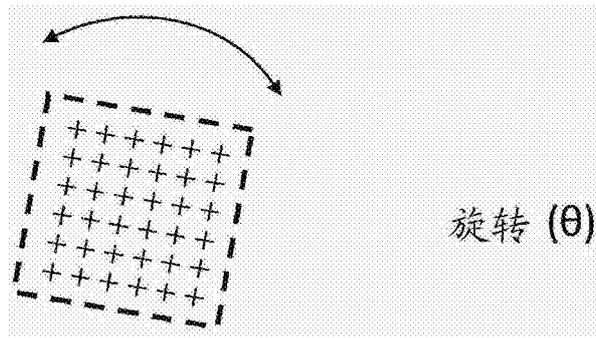


图 8A

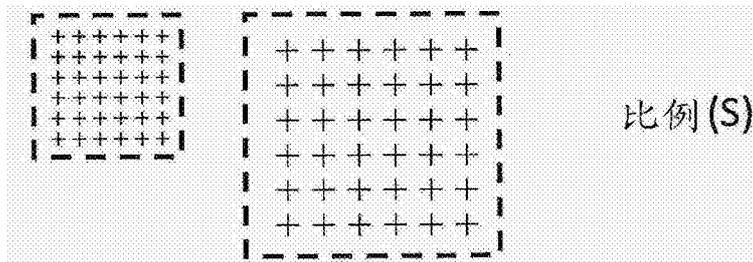


图 8B

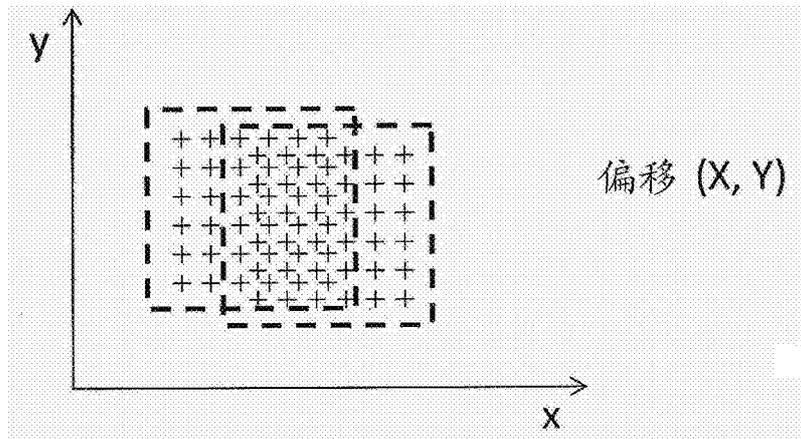


图 8C

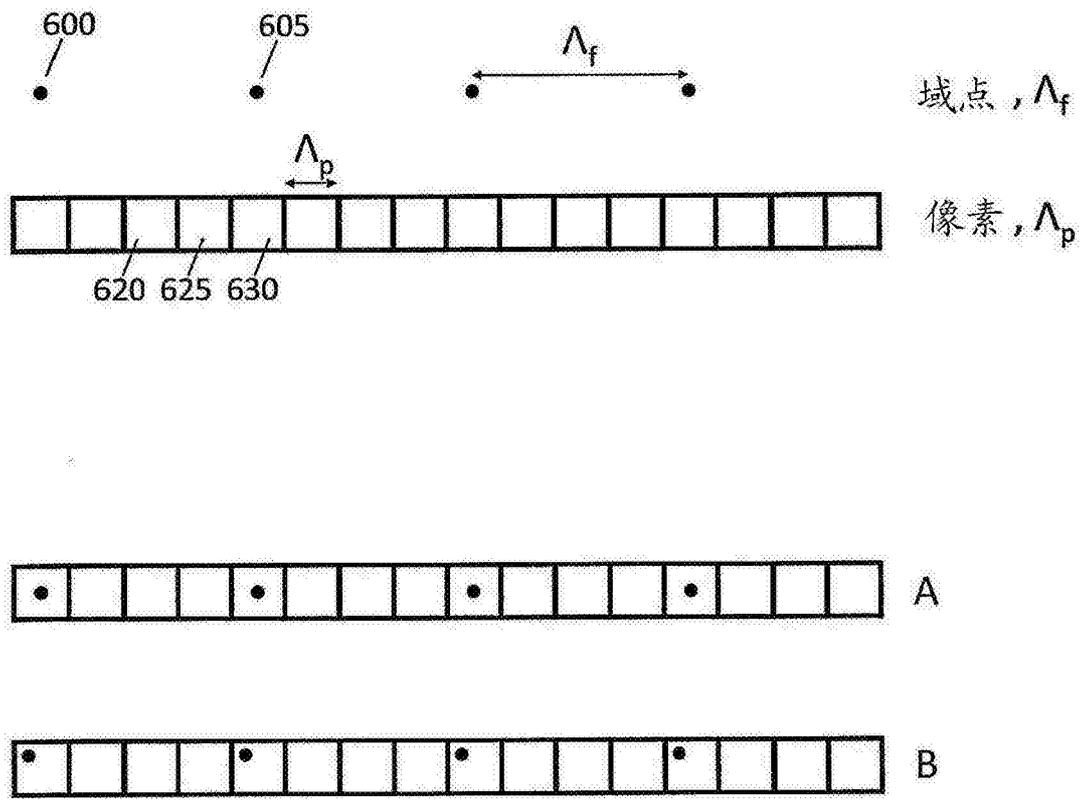


图 9

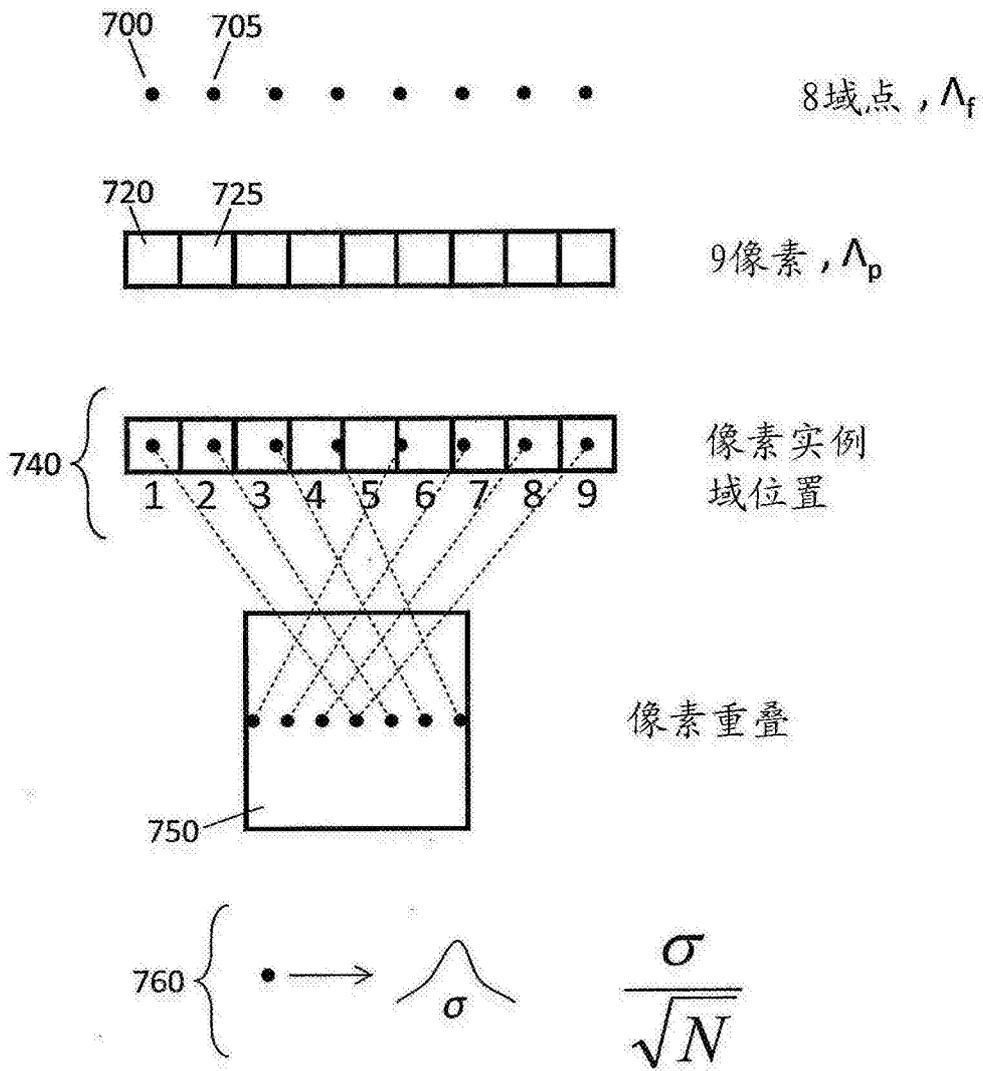


图 10

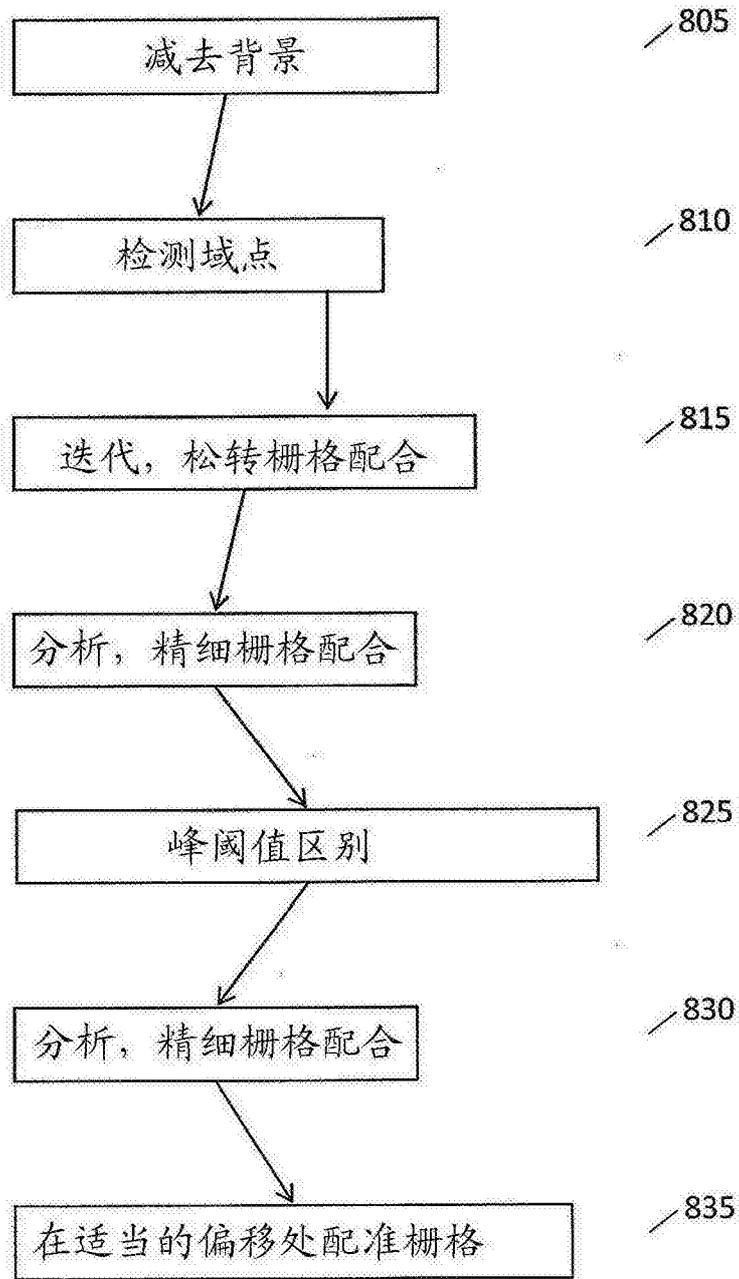


图 11

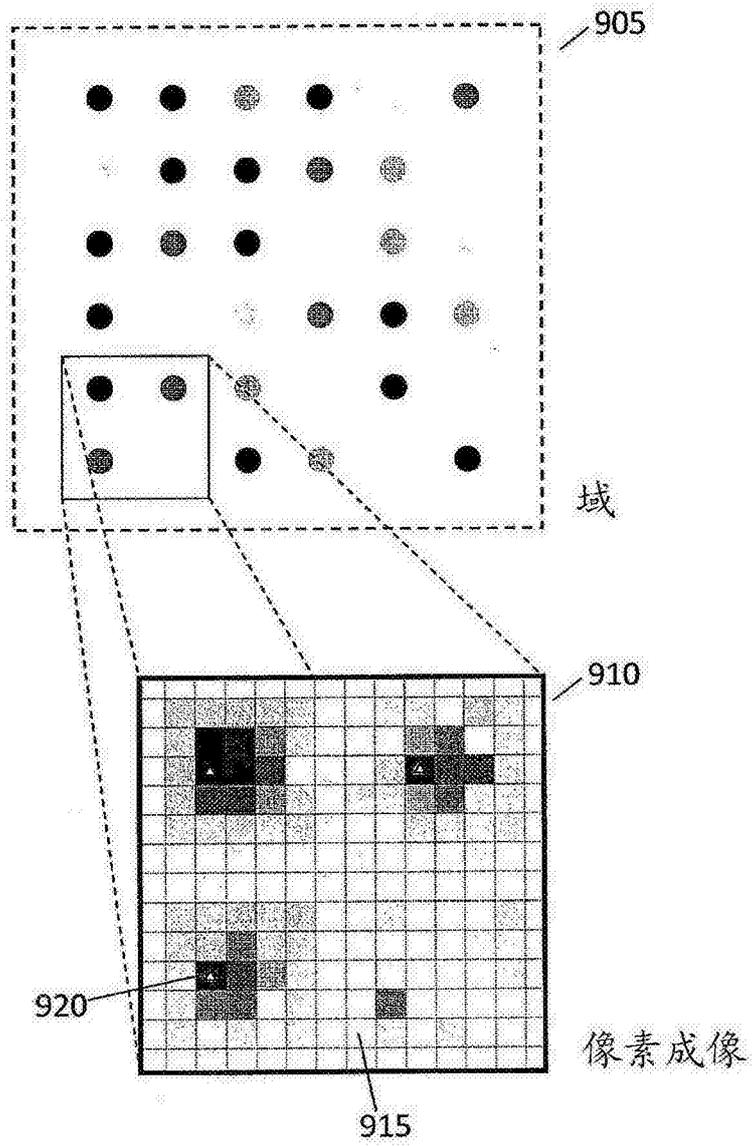


图 12

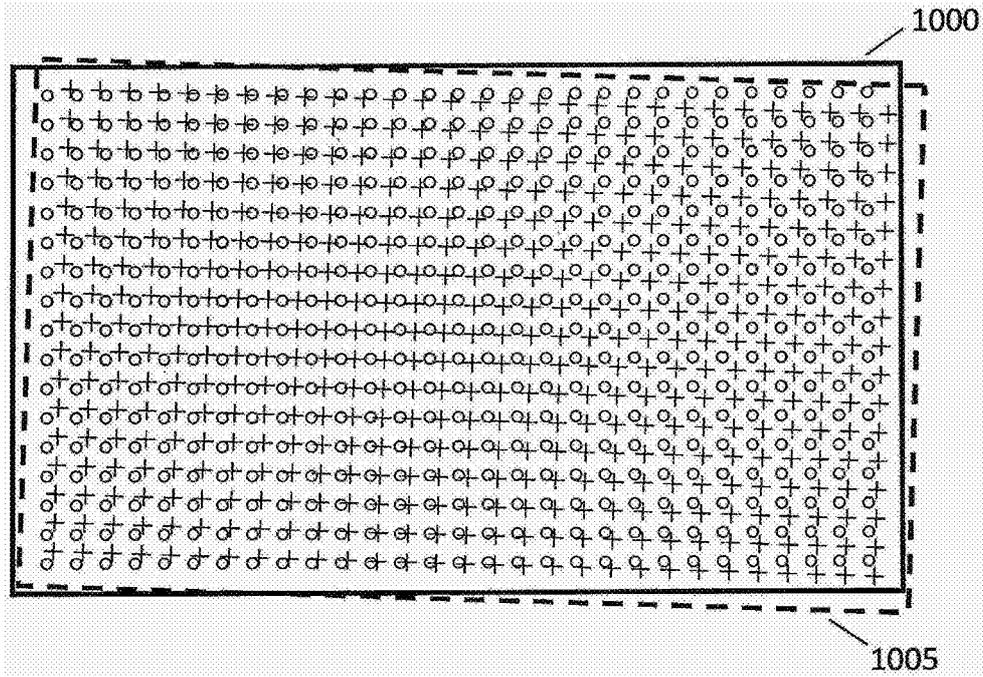


图 13A

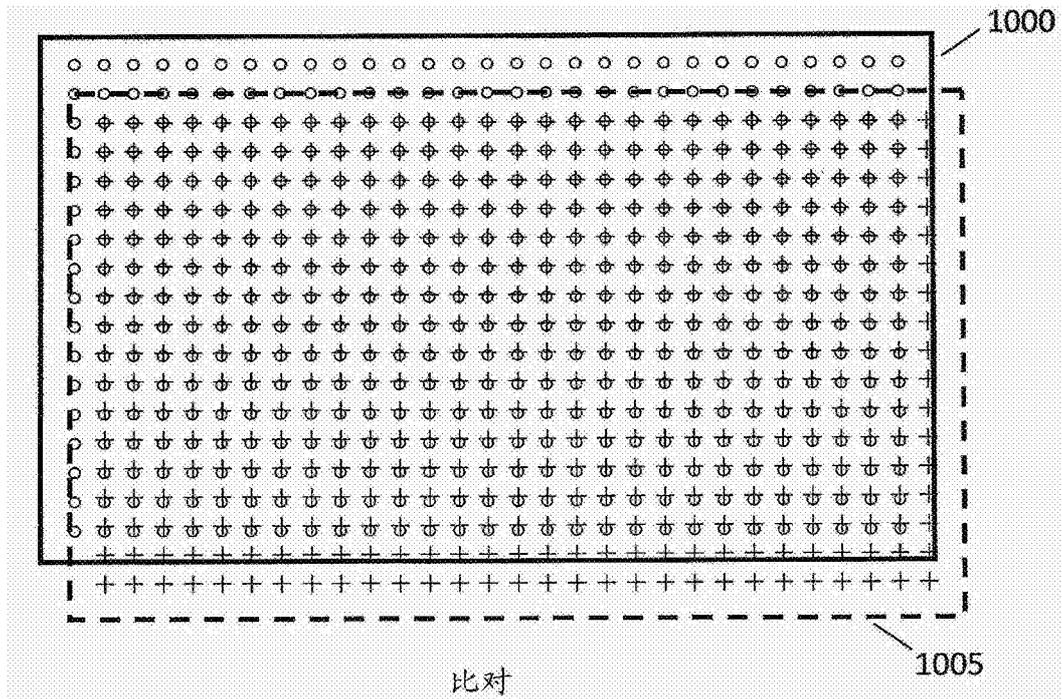


图 13B

## 迭代松转栅格配合

```

for ( $\theta = \theta_{\min}$  ;  $\theta \leq \theta_{\max}$  ;  $\theta += \Delta\theta$ ){
for ( $S = S_{\min}$  ;  $S \leq S_{\max}$  ;  $S += \Delta S$ ){
for ( $X = X_{\min}$  ;  $X \leq X_{\max}$  ;  $X += \Delta X$ ){
for ( $Y = Y_{\min}$  ;  $Y \leq Y_{\max}$  ;  $Y += \Delta Y$ ){

calculate  $F(\theta, S, X, Y)$ 

}
}
}
}

find minimum  $F$ 

```

1100

图 14

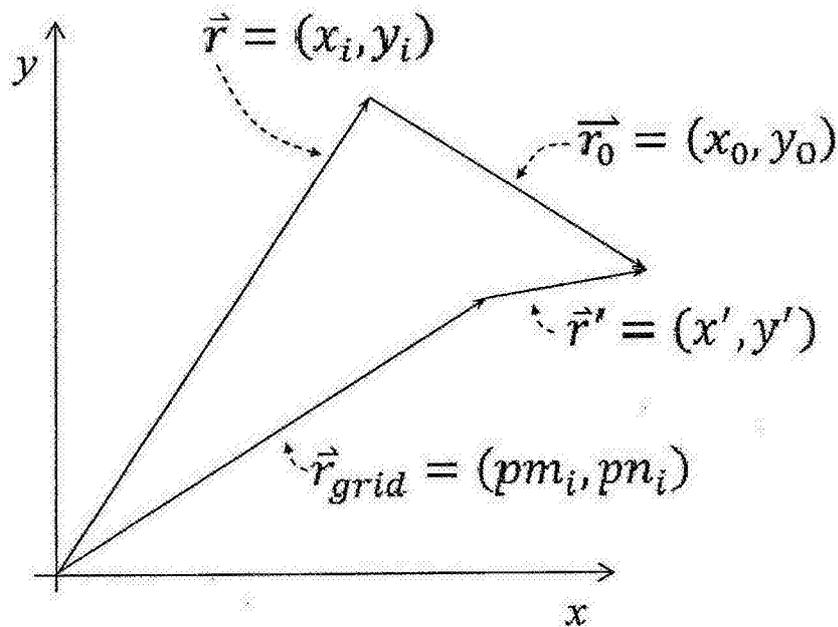
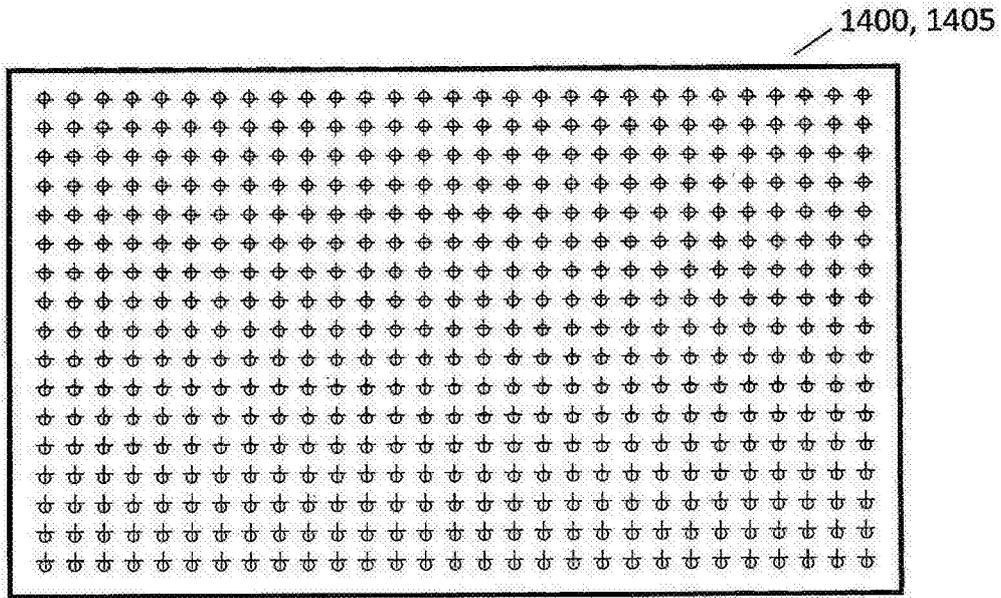


图 15





配准

图 17B

具有假随机删除的域

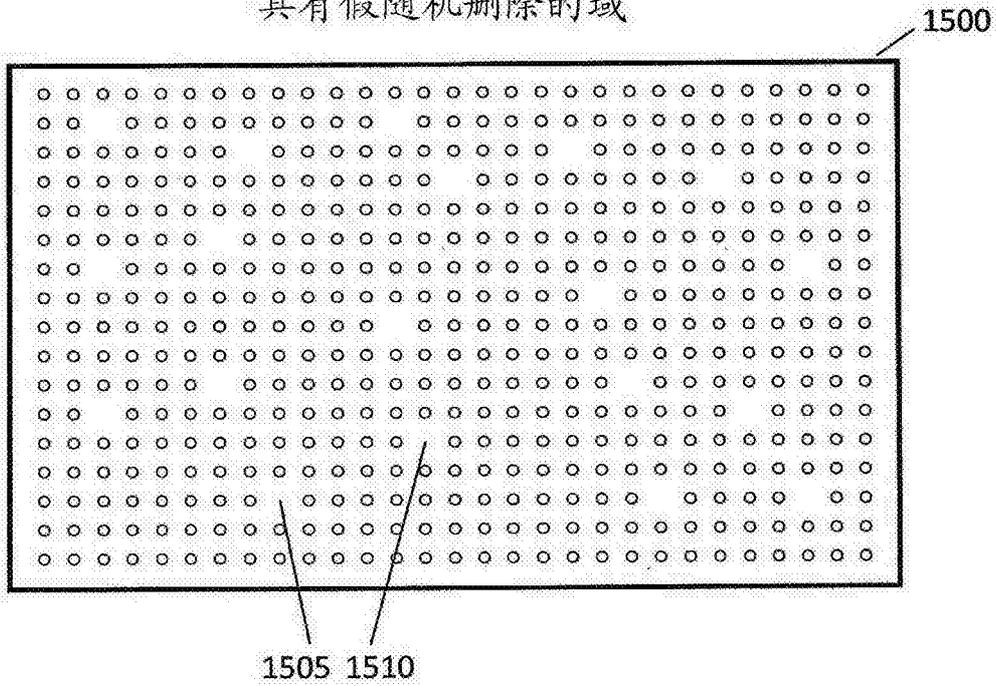


图 18

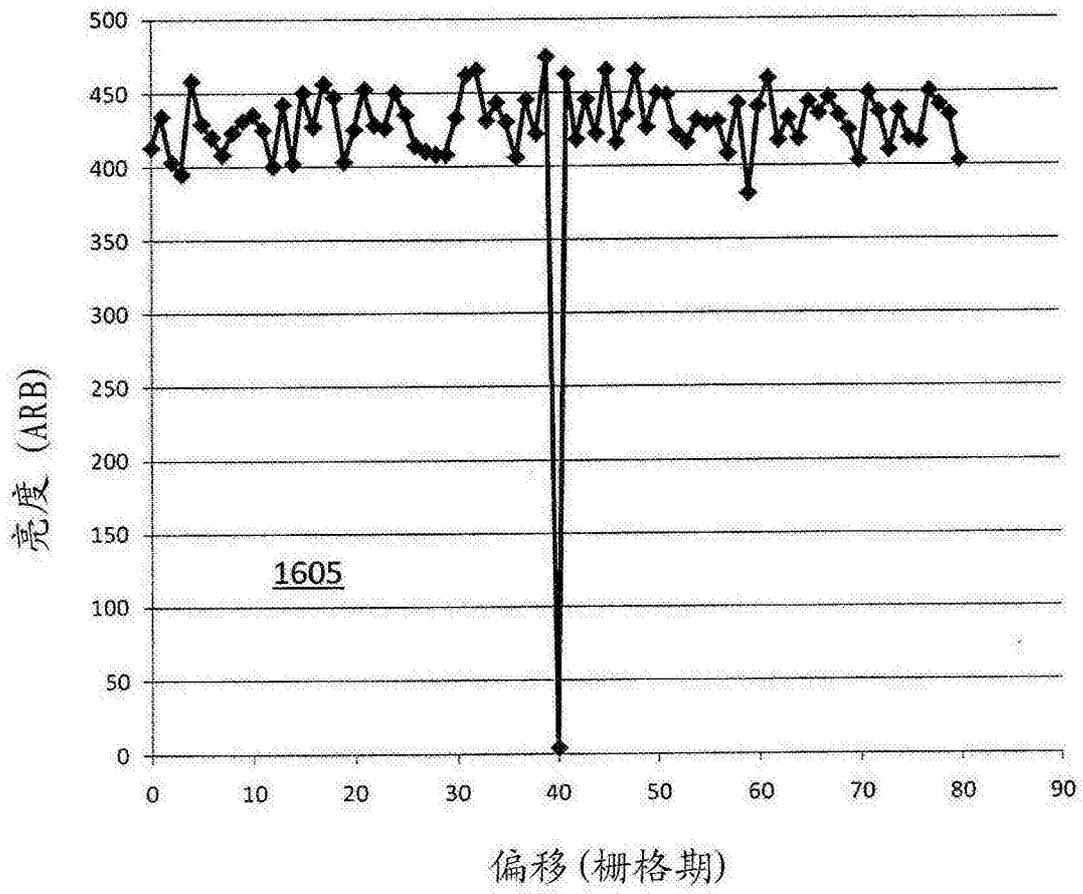


图 19

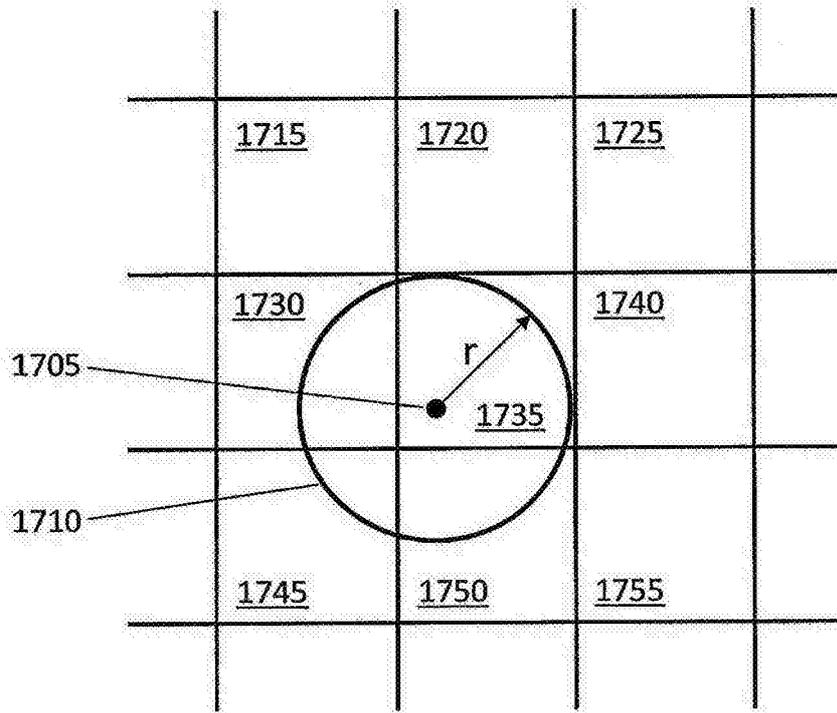


图 20