

量子现象、量子计算 及其潜在应用

周宗和
mileszhou@gmail.com

Table of Contents

1 量子现象和量子力学	3
1.1 量子现象和规律	3
1.2 量子力学	5
1.2.1 波函数	5
1.2.2 薛定谔 (SCHRÖDINGER) 方程	5
1.2.3 线性性和叠加性	7
1.2.4 时间逆转问题	7
1.2.5 测量	7
2 量子计算原理	8
2.1 量子位	8
2.1.1 量子位的复合	9
2.2 量子门	10
2.2.1 HADAMARD 门	10
2.2.2 受控“非”门	11
2.2.3 受控 U 门 (c- U 门)	12
2.2.4 移相门	12
2.3 量子电路	12
2.4 量子算法	12
2.4.1 DEUTSCH 算法	13
2.4.2 GROVER 算法	15
2.4.3 FOURIER 变换	18
2.4.4 大合数分解	20
3 量子计算机	22
4 量子计算的潜在应用	23
4.1 机器学习应用场景	23
4.2 大数据应用场景	24

5	结束语	24
---	-----------	----

量子计算被认为是人类征服自然的下一个可能的重大突破之一。量子计算有可能使很多现在难以完成的计算问题迎刃而解。在计算领域的多个公司都在花大力气研发，而且后续三到五年可能会有快速的发展，使量子位数从现在的 60 左右提高到 1000 到 2000 位，虽然还很难应付实际应用中提出的计算问题，但已经可以看出实现所谓的量子优势（quantum advantage）的可能性。

另一方面，现已有大量书籍、文章介绍量子计算，但由于量子计算涉及到很不直观的科学概念以及数学问题，致使很多人对量子计算仍感觉很难理解。本文试图以个人的视角谈一下这个问题。


量子计算机是怎么回事呢？概括地讲，就是构造由二状态的量子单元组成的量子系统，并根据问题的需要，利用量子系统的特性，对其进行符合某些数学规律的变换，使之变换成一个特定的目标量子状态。然后对其进行测量，从而得到我们所需要的结果。这里，二状态的量子系统指由多个二状态的量子单元组成的系统，各量子单元之间存在着确定的联系。至于什么是量子单元，这些单元之间有着什么联系，以及我们利用了量子系统的什么特性呢？这些问题我们在后文会做详细介绍。

1 量子现象和量子力学

1.1 量子现象和规律

量子现象是什么呢？量子现象是微观系统所表现出来的，不同于宏观系统所表现出来的现象。

我们都听到过测不准原理：测量一个微观粒子的位置¹，我们永远无法得到其准确的位置。那么，有的人会说了，即便是宏观系统，测量也有误差。但是，我们指出，微观系统的测量误差和宏观系统的测量误差有着本质的区别。例如，我们打靶，不能每枪都打十环，那是因为我们的举枪的角度、空气的流动、火药的作用力的大小和方向等等各种因素不可避免地存在着误差。随着控制能力的提高，理论上我们可以无限地缩小误差，例如到真空的地方去打枪，枪管更精确地固定，膛线精度的进一步提高，等等措施，总能使精度进一步地提高。但是微观系统的测量误差，却不能由这些因素的提高而无限地缩小。

我们来看一个实验（见  1. 电子衍射实验）：我们在一个金属板上开一个小孔。在金属板的一边放一把电子枪（阴极），另一边，平行于金属屏放一个屏幕，屏幕背后放一个阳极（施加一个正电压，图中没有画出来）。阴极释放的电子，在正电压的作用下，会奔向屏幕。如果屏幕接收到一个电子，就会在屏幕上留下一个小点。如果我们使阴极发射速率很慢，电子是可见的一个一个地被屏幕接收，屏幕上就会出现一个一个的小点。渐渐地，小点多了，会形成一个图形，我们称之为像。在整个实验过程中，我们准确地控制所有器材的位置和振动，实验放在一个真空的环境中，确保阳极的电场足够均匀，屏幕也可以是圆形的。甚至可以有严格对齐的两个屏幕，电子穿过第一个屏幕后还要飞一会儿才能到达第二个屏幕。这般等等，施加各种精度改进措

¹ 测不准原理不止针对位置。我们只以位置为例。

施。但我们却不能使像进一步缩小。更重要的是：我们看到的像是一组同心圆（环），而不是一个实心的圆。其实，这就是著名的电子衍射实验。

按照经典理论（电子直线传播），像应该是一个近似的实心圆，即便存在各种误差，根据概率论中心极限定理，其综合误差应该是正态分布的，而不是呈同心圆分布的！

这说明什么呢？这些电子到达屏幕的位置是不能准确控制的，但是它们以相同的方式制备，我们认为它们处于相同的量子状态。大量这种处于相同状态的电子，在完全相同的条件下到达屏幕，其到达的点分布出现准确的规律。也就是说，虽然我们不能准确控制或测量某一个特定的微观粒子的位置，但微观粒子的运动还是有确定的规律的。我们说：虽然对某个物理量的多次测量我们会得到不同的值，但是，这些微观系统是处于同一个量子状态的，也就是说，他们的量子状态是确定的！

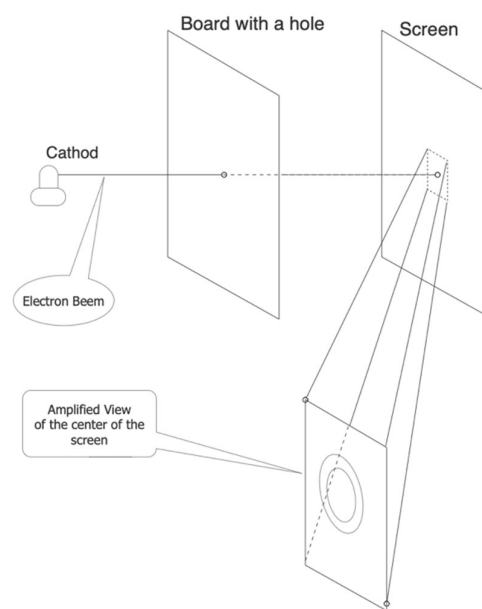


图1. 电子衍射实验

量子力学就是以量子系统的状态为基础，研究这些状态的规律的科学。

不同于经典力学，量子力学具有很不直观的特性，甚至让我们感到无法理解。无法理解的原因是我们总是以我们习惯的，直观的宏观世界的经验去解释微观现象。从伽利略到牛顿，再到爱因斯坦，我们研究的定量关系都是物理量之间的关系，也就是它们之间的微分方程。建立了微分方程，我们就有了解决问题的基础：它们的解就是物理量的动力学规律。但是，这一方法对于微观系统不适用，因为物理量的测量值存在着内在的不确定性！量子力学就是对这种微观系统建立的力学体系。不同于经典力学，量子力学并没有良好的哲学解释（后面还会提到出现的相关的难以理解的现象），究其原因，是我们总是以我们对宏观世界的经验去解释微观现象。例如，总相信一个粒子在某个时刻应该有个确定位置；如果测量不准确，一定是测量方法或仪器不准确所致。然而，量子力学告诉我们，“不准确”是物理现实，而不是测量仪器的问题。我们必须先承认“测不准”是基本规律，去解释其他现象，才能正确地理解它们。

我常和朋友们讲一个虚拟的“故事”：外国某地，经常发大水，造成生命财产损失。为了减小这种损害，当地的人找来很多技术人员，安装了各种仪器设备，花了好几年的时间，测量了水流的各种物理量，例如，比重、水压、流速、气泡因素、等等，不一而足，应有尽有，然后进行建模、仿真、等等研究。有一天，一个高中生放学回家路过这批技术人员的营地时，冲了进来，开口就问：“今天老师告诉我们，水的密度在摄氏 4°C 的时候出现一个峰值， $0-4$ 度之间出现冷胀热缩现象。我很想不通，为什么呢？你们应该是全世界最好的研究水的专家了，能不能帮我解释一下呢？”这些技术人员听完这问题，一下子全傻了。个个面面相觑，不知如何作答，...

我们知道，水在 0 到 4°C 之间出现冷胀热缩现象，是由于水分子的微观结构造成的，需要量子力学才能较好地机理上定性定量解答。不管这些技术人员研究了多少宏观水流的性质，却无法解释这个由水的微观结构特点产生的现象。

这个虚拟的故事告诉我们，量子力学之所以从哲学，或认识论上难以理解，是因为我们人类习惯于用宏观经验的思维模式描述或认识其他现象，包括微观现象。例如，说一个粒子没有位置，总感到不可思议。

虽然描述的对象是一个随机的现象，但量子力学为到现在为止的无数的实验，以极高的精度所验证。有不少文献说是人类至今为止最高精度的验证。不管这个说法是否片面，但量子力学确实是在极高的精度上得到了验证。

1.2 量子力学

这里所说的量子力学指狭义的量子力学，也就是由薛定谔方程所决定的力学体系。薛定谔方程对于很多不涉及到高速运动的，非高能的微观系统准确地成立。

刚才的电子衍射实验说明，电子存在衍射现象。我们知道，衍射是波动的特征。电子这样的典型地被认为是粒子的东西都有衍射，我们的先驱们大胆地假定，所有的微观粒子都有波动性。事实上，后来的实验证明，不仅电子呈现波动性，原子，甚至分子也呈现波动性。

1.2.1 波函数

不同于经典力学，量子力学不建立物理量之间的微分方程。相反，它用一个分布于空间的，依赖于时间的函数，叫做波函数来描述微观系统。而薛定谔方程就是这一波函数所满足的方程。波函数成为描述微观系统的基础。

波动通常由一个空间和时间的函数表示，我们称之为波函数，记为 $\Psi(x, t)$ ，它是一个坐标和时间的函数，其值通常取复数值。在量子力学中，波函数的物理意义是：其绝对值的平方是在该时刻该位置发现这个粒子的概率。也就是说，波函数描述的是一个“概率波”。这个不好理解，但我们说了，不要用宏观的经验去理解微观现象。基于波函数的量子力学与实验高度地吻合。相信我吧，不骗你。

波函数有什么好处呢，或者说有什么不同，或者优势呢？因为波函数包含有更多的信息（存在于比物理量更高维的空间中，这点对量子计算很重要，后面还会提到），波函数不可能为仪器所直接测量得到。但是，波函数可以确定物理量的数学期望和方差。这样，正好解释了测量的不准确性，以及不准确的程度。而这一程度也被实验所验证。

下面，我们来看看薛定谔方程长什么样子，波函数又长什么样子，以及算子如何作用于波函数。然后，我们来看看对于一个比较简单的二状态系统，就是只有两个状态的系统，其波函数又如何通过线性空间，其实是 Hilbert（希尔伯特）空间来表示。Hilbert 空间比较抽象，在有限维的情况下，它就是我们熟悉的抽象的 Euclid（欧几里得）空间。所谓抽象，是指它不是我们生活的物理空间，而是一个数学概念。Euclid 空间由向量组成，Euclid 空间的向量除了有长度和方向，还有内积（于是有夹角）。空间的维数就是独立的方向的个数。为了简化涉及的数学概念，我们后面尽量避免使用内积这个概念，看行不行。

1.2.2 薛定谔（Schrödinger）方程

先看看 Schrödinger 方程长什么样。为了简化问题，突出重点，我们考虑一种跟量子计算有关，而又最简单的情况。空间是三维的，但我们只考虑一维，即一个方向的“运动”。这不影响本质，避免了冗长和/或复杂的数学符号。

$$i\hbar \frac{d}{dt} \Psi(x, t) = \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right] \Psi(x, t)$$

这里， $\Psi(x, t)$ 就是所谓的波函数，我们的主角，闪亮登场，它是分布于空间 (x) 和时间 (t) 的函数。方括号内的部分称为系统的 Hamiltonian，即哈密顿算子，不妨记作 H ，是系统的总能量 Hamilton 函数的量子化，而 V 是所谓的势能场，代表我们问题的场景或环境，例如电磁场，束缚条件等。

注意，这里所说的量子化，不是像早期为解释黑体辐射的“紫外灾难”问题时所作的，人为地把能量强制为一份一份的假设，而是在代表系统总能量的 Hamiltonian 函数中用 $i\hbar \frac{\partial}{\partial x}$ 代替了动量 p 而得到了 Hamiltonian 算子。为什么这样代替呢？是 Schrödinger 猜出来的，后来发现这样很好。Hamilton 函数是分析力学的一个中心概念，是说，对每一个被研究的对象，我们建立一个 Hamilton 函数 H ，描述系统的总能量（动能+势能）。然后，把这个 H 代入一个方程，叫做 Hamilton 方程，就得到了运动方程（微分方程）。这个微分方程的解就是系统的行为。至于 Hamilton 方程长什么样，我们就不展开了。有兴趣的可以参考分析力学的相关著作。

波函数的物理意义是什么呢？或者说，它的物理解释是什么呢？按照哥本哈根学派的解释，波函数的绝对值的平方是在所论时间在空间某处发现这个粒子的概率。也就是说，其绝对值是概率分布。而且哥本哈根学派认为这是对量子系统的完整描述，而不存在未知因素（隐藏变量）的干扰，造成非确定的概率分布。后来 J. Bell（贝尔）假设存在隐藏变量，得到了 Bell 不等式，但 Bell 不等式被实验否定。于是，上述哥本哈根“解释”得到了学界的普遍认可。量子现象的这个概率解释是量子力学很难为人们理解的第一个原因。

好了，我们得到了描述微观系统的波函数，而不是这些系统的单元（如其中的一个粒子）的某些物理量，如位置、动量等，而这些物理量的测量结果可以由这个状态推出测到某些特定值的概率，以及（均）方差。

把括号里的 Hamiltonian 记为 H ，我们得到一个更清晰的 Schrodinger 方程：

$$i\hbar \frac{d}{dt} \Psi(x, t) = H \Psi(x, t)$$

H 刻画了一个量子系统。这里，我们故意忽略了边界条件和初始条件。边界条件和初始条件主要针对具体的问题。我们只做定性观察，暂时忽略这些细节。

这个方程是不是一下子提醒我们（代进去验证一下就知道了），如果我们知道系统某时刻，例如 t_0 ，的状态 $\Psi(x) = \Psi(x, t_0)$ ，那么系统随着时间的演变是： $\Psi(x, t) = e^{-\frac{it}{\hbar} H} \Psi(x)$ 。这里， $e^{-\frac{it}{\hbar} H}$ 理解为按照指数函数 e^x 的 Taylor 级数展开后把自变量 x 用算子 $-\frac{it}{\hbar} H$ 代替后得到的算子，其中，算子的乘积就是算子的重复作用。写出来就是（不要被这个复杂的式子吓到，后面我们不会用到，我只是在这里用它说点事儿）：

$$\begin{aligned} \Psi(x, t) &= e^{-\frac{it}{\hbar} H} \Psi(x) = \left[I + \left(-\frac{it}{\hbar} H \right) + \frac{1}{2!} \left(-\frac{it}{\hbar} H \right)^2 + \frac{1}{3!} \left(-\frac{it}{\hbar} H \right)^3 + \dots \right] \Psi(x) \\ &= \Psi(x) - \frac{it}{\hbar} H \Psi(x) - \frac{t^2}{2\hbar^2} H (H \Psi(x)) + \frac{it^3}{3! \hbar^3} H (H (H \Psi(x))) + \dots \end{aligned}$$

注意，这里， i 的高次方都被算了出来。这个方程就是一个量子状态（波函数）随着时间的演变规律。

我们看出几件事（你看，说事儿了）：

1. Schrödinger 方程是一个偏微分方程；
2. 有个 i ，所谓虚数单位，好像是个复数方程；
3. 左边是对时间的导数，而右边是对空间的两阶导数。对 $\Psi(x, t)$ 而言，是个齐次线性方程。

这些都很重要。我们来做些解释。

1. Schrödinger 方程就是波函数所需要满足的方程，是个偏微分方程。但别被它的形式上的简单迷惑，其内涵及其丰富。它描述了低能量微观系统的几乎所有的现象。它确实有一个缺点：它不符合 Einstein 的狭义相对论，而且微观粒子的自旋不是这个方程的解。这个缺点有后续的 Dirac 方程圆满解决，但对于我们的讨论并不必须，所以提到为止。
2. 含有 i ，说明函数会有虚数值。但是，这并不意味着波函数是复变函数。自变量仍然是空间和时间的坐标，而不是复数。

好，后面的说明比较重要，我们分节叙述。

1.2.3 线性性和叠加性

Schrödinger 方程是个齐次线性方程。从微分方程理论知道，齐次线性方程的解是一个线性空间。也就是说，如果方程有多个解，其线性组合也是该方程的解。而且对于一个多个输入的输出是各个输入造成的输出之和。这些都很重要，这个特性对量子计算至关重要：是量子叠加性的数学基础。波函数的自由度很大。方程的可能解“很多很多”，构成一个函数空间，远远多于我们所说的物理量的变化范围。在数学上，我们说 Schrödinger 方程的解空间是个高维的空间。这点对于量子计算也很重要：一个量子系统的状态可以比传统的二状态系统的 0、1 含有多得多的状态：如果 0 和 1 是两个可能的状态，那么，0 和 1 的线性组合也是可能的状态。特别是，因为波函数是个复值函数，这里的线性组合是复线性组合：如果 $\Psi_0(x)$ 和 $\Psi_1(x)$ 都是可能状态，例如，分别表示状态 0 和 1，则 $\Psi(x) = \alpha\Psi_0(x) + \beta\Psi_1(x)$ 也是。

1.2.4 时间逆转问题

和 e 指数函数一样， H 算子的 e 指数在 H 满足一定条件的情况下（有界性就足够了，物理的 Hamiltonian 都是满足的）， e 指数算子的行列式永远不会是 0，也就是说，它是可逆的。这个很重要：一个量子状态（波函数）随着时间的改变是可逆的，也就是说，量子状态所表示的量子信息不会随着时间的演变而丢失的！这句话听着似曾熟悉，就是这个意思。对于量子计算而言，它还有这一个重要的意义：对一个量子状态的变换（即量子计算步骤）都是可逆的。而传统的“与”门、“或”门都是不可逆的，所以其量子对应物一定是不同的。

1.2.5 测量

测量问题是微观系统最神奇而难以理解的问题之一。测量就是用宏观系统对微观系统进行观测。这里的关键字是宏观系统对微观系统。

稍微做一下数学展开，考虑方程 $H\Psi(x) = \lambda\Psi(x)$ ，或者 $(H - \lambda I)\Psi(x) = 0$ 。可以证明，存在一组独立的（即任何一个不是其余函数的线性组合）函数， H 的像都可以表示成这些函数的线性组合。或者说，这组函数称为 H 像空间的基（可能需要复习一下线性代数）。如果愿意，可以把上述的 H 看作为一个矩阵，可能会容易理解一点。为了简化叙述，避免过多的数学语言，我们假设这些特征值都是简单的，即对应于这个特征值的所有特征函数都只差一个常数。我们引用一些符号。设 $\Phi(x)$ 是 H 像空间中的一个函数， $\{\varphi_1, \varphi_2, \dots\}$ 是上述的基，则 $\Phi(x) = \alpha_1\varphi_1 + \alpha_2\varphi_2 + \dots$ 。因为一个波函数乘以一个常数代表同一个状态，不妨选择那些系数，使 $|\alpha_1|^2 + |\alpha_2|^2 + \dots = 1$ 。这个过程叫归一化。因为每个系数 α_i 的绝对值的平方代表这个状态出现的概率，归一化的物理意义就是，总概率为 1。

量子力学告诉我们，如果对系统进行测量，相当于我们把一个微观系统放入一个具有一个特定算子决定的系统中，这一系统只允许被测对象处于这一算子的特征向量张成的空间中。一经测量，微观系统就进入某一个特征向量的状态中，而对应的 λ 就是观察值（或跟观察值有关的值）。我们称这一现象为状态崩塌（collapse）。而崩塌到某一个状态的概率就是上述对应的那个 $|\alpha_i|^2$ 。归一化保证了总概率为 1，就是说，怎么滴，也能测出一个值来。

如果某些状态的 $|\alpha_i|^2$ 很大，甚至于为 1，则观测得到这个对应值的可能性就极大。这个对量子计算非常重要。量子计算算法设计的一个重要目标就是使所要的结果被測到的概率变得很大，最好是 1。

2 量子计算原理

量子现象可以从多个方面发展应用。本文只介绍其中的一个方向：模仿数字计算机的模式，从构造量子位、量子门、量子电路来实现量子算法。量子现象还有几种其他的利用方案被提出，例如，利用量子纠缠进行量子密钥交换；利用量子模拟对各种分子（包括蛋白质）进行仿真等等，不一而足。

前面说过，量子计算就是构造一个二状态的量子系统，并根据问题的需要，利用量子系统的特性，对其进行符合某些数学规律的变换，使之变换成一个特定的目标量子状态。然后对其进行测量，从而得到我们所需要的结果。二状态量子系统的基本单元就是二状态量子单元，我们叫量子位。实施数学变换的装置叫量子门。由量子位和量子门组成的系统叫量子电路。量子电路进行的演算的方案叫量子算法。本章介绍量子计算的算法原理，而不涉及到量子计算的物理原理，正如我们进行经典算法分析时，无需涉及组成计算机的电路的晶体管或 MOS 器件的物理原理一样。

2.1 量子位

量子位基于一个二状态的量子单元。正如前面已经说到的，我们不涉及单元的物理机理，而从计算的角度看看量子位。

一个经典的位（或叫 bit）有两个状态，通常叫做 0 和 1。一个量子位也有两个状态，也可以叫做 0 和 1。不同的是，正如我们在量子现象一节中讲过的，量子位的两个状态是可以叠加的，两个独立的状态，也就是说，0 和 1 的线性组合也都是可能的状态，所以，一个量子位所存储的信息比一个经典位多得多。

存储更多信息还不是量子位优越性的全部。更重要的是，所说的线性组合是复数的线性组合，用数学语言描述，就是，一个量子位的状态构成一个复线性空间。我们知道，复数除了有大小信息（叫绝对值）外，还有相位信息。当只有一个量子位时，相位信息并不重要，但当两个或多个量子位相互作用时，不同相位的量子位就会发生干涉；可以增强、也可以互相抵消。这对于量子计算至关重要。

大多数的量子计算的著作中，量子位的状态用量子力学中常用的 Dirac 记号表示，即状态 0 和 1 分别写成 $|0\rangle$ 和 $|1\rangle$ 。究其物理本质，它们是二状态单元的两个状态所对应的波函数。它们的叠加，就是波函数的叠加。

为了行文的简洁，我们参照向量的表示，状态 $|0\rangle$ 用黑体字 **0** 表示，状态 $|1\rangle$ 用黑体字 **1** 表示。

如果一个量子位处于一个 **0** 和 **1** 的叠加态，则用它们的线性组合表示： $\alpha\mathbf{0} + \beta\mathbf{1}$ ，这里， α 和 β 都是复数，而且要求 α 和 β 满足 $|\alpha|^2 + |\beta|^2 = 1$ 。这个要求叫做归一化。当把两个状态叠加时，每一个都必须归一化。也为了行文方便，我们有时也只在必要时才做归一化，叫做怠滞归一化。

注意，-1 也是复数。当 $\alpha = 1$, $\beta = -1$ 时，上述状态为 $\mathbf{0}-\mathbf{1}$ ，它与 $\mathbf{0}+\mathbf{1}$ 的和为 $2\cdot\mathbf{0}$ 。那个常数因子 2 不重要，归一化为 1。注意，这里就是怠滞归一化的一个例子。其实， $\alpha = 1$, $\beta = \pm 1$ 不符合归一化要求。如果不进行滞后归一化，则 $\alpha = \frac{1}{\sqrt{2}}$, $\beta = \pm \frac{1}{\sqrt{2}}$ 。于是 $\frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}}\mathbf{0} + \frac{1}{\sqrt{2}}\mathbf{1}) + \frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}}\mathbf{0} - \frac{1}{\sqrt{2}}\mathbf{1}) = \frac{1}{2}\mathbf{0} + \frac{1}{2}\mathbf{1} + \frac{1}{2}\mathbf{0} - \frac{1}{2}\mathbf{1} = \mathbf{0}$ 。结果是一样的。

2.1.1 量子位的复合

在实际的量子计算中单个量子位不能做什么有用的工作。通常要把很多量子位合在一起，形成复合量子位。复合量子位的状态用各个量子位的状态组合表示。例如，第一个量子位为 x ，第二个为 y ，第三个为 z ，则复合量子位的联合状态可以写为 xyz 。这里的顺序是很重要的，因为为了行文的简洁，没有下标标注，只用位置表示符号和量子位的对应关系。

两个量子位的复合有 4 个线性独立的基础状态：**00**、**01**、**10**、**11**，也分别叫做 **0**、**1**、**2**、**3**。值得注意的是，它们是线性独立的状态，并成为 4 维 Hilbert 空间的基，也叫计算基。跟单个量子位的情形一样，这四个线性独立状态可以线性叠加：

$$\psi = a_{00}\mathbf{00} + a_{01}\mathbf{01} + a_{10}\mathbf{10} + a_{11}\mathbf{11}$$

所以，两个量子位所表示的量子状态空间是一个 4 维的 Hilbert 空间。

很多情况下，复合状态中的各个状态是独立的，也有时候是不独立的。以两个位组成的复合状态为例，状态 $xy = \mathbf{00}+\mathbf{01}+\mathbf{10}+\mathbf{11}$ 中，两个位是独立的，可以表示为

$$\begin{aligned} xy &= \mathbf{00} + \mathbf{01} + \mathbf{10} + \mathbf{11} \\ &= (\mathbf{0} + \mathbf{1})(\mathbf{0} + \mathbf{1}) \end{aligned}$$

当一个复合系统的状态表示为上式的叠加态时，这些系数 a_{00} 、 a_{01} 、 a_{10} 和 a_{11} 除了表示各个基础状态的相位和大小关系外，它们的绝对值平方还表示着这两个量子位的联合概率分布。从这个分布可以看出两个量子位在联合概率分布意义下是否独立的。

注意，这里的加法和乘法符合通常的分配律和交换律，也就是说，最后的那行的和的乘积可以像独立的代数量一样地处理，结果得到第一行的结果。这种情况下，我

们称这两个位组成的状态是两个量子位的张量积。在这种情况下，对两个量子位中的如何一个的测量不影响另一个量子位。例如，对第一个量子位测量，不管得到 0 或 1，再对第二个量子位进行测量，仍然是 50% 为 0，50% 为 1。跟没对第一个量子位进行测量是概率分布一样。所以，我们说这两个量子位是独立的。

但是，不是所有的复合状态都能表示成为这样的张量积的。例如状态 $00+11$ 就不能表示为两个单个量子位的张量积。读者可以自行验证，这个状态无法表示为张量积。事实上，它们确实不是独立的：例如，第一个量子位测量得到 1，联合状态就进入 11 。对第二个量子位的测量比得到 1。所以要么测到 00 ，要么测到 11 ，测到一个 0，一个 1 的概率为 0。其实，两个量子位的四个系数决定了两个量子位的联合概率分布，它们是对两个量子位的联合概率分布的完整描述。

对于多个量子位复合的情况，可以类似地推广。 n 个量子位的复合可以得到 2^n 维 Hilbert 空间。所有的 n 个 0 和 1 的任意组合都是一个基础状态。这 2^n 个基础状态构成 n 位复合系统的计算基。

2.2 量子门

量子门对量子位的状态进行变换。究其物理本质，是将量子位处于一个量子环境（不是测量环境）中，例如电场、磁场或者其他物理环境中，让其发生时间演变，变成一个新的状态。因为是时间演变，我们在量子现象中说过，时间演变都是可逆的。所以，量子门都是可逆的。这和经典的门很不一样。经典的门中，“非”门是可逆的，但是“与”门和“或”门却是不可逆的。但是，所有的量子门都是可逆的。

这里介绍的量子门远不是全部。我们挑了几个，试图通过它们让我们能体会到量子计算的奥秘。

2.2.1 Hadamard 门

Hadamard 门如图 1 所示。其逻辑为，输入 x 为 0 时，输出为 $0+1$ ；输入为 1 时，输出为 $0-1$ 。在输入为 0 和 1 的叠加态时，输出为这两个状态（即 $0+1$ 和 $0-1$ ）的叠加。

于是，当输入为 $0+1$ 时，输出为 $(0+1)+(0-1)=2\cdot 0$ 。因为我们不在乎状态前的常数因子 2，可以认为输出为 0。而在输入为 $(0-1)$ 时，输出为 $(0+1)-(0-1)=2\cdot 1$ 。我们也认为输出为 1。



图 2. Hadamard 门

现在我们注意到，如果把两个 Hadamard 串联起来，我们会发现，0 被转换成 $0+1$ ，再被转换回 0；而 1 被转换为 $0-1$ 后又被转换回 1。也就是说，Hadamard 门是它自己的逆。

在目标位为 $0-1$ ，控制位为非叠加态 x 时，Hadamard 门的输出是：

$$Hx = \frac{1}{\sqrt{2}}(0 + (-1)^x 1) = \frac{1}{\sqrt{2}} \sum_{z=0,1} (-1)^{xz} z$$

上式右边和中间是一个量的两个不同表达而已，这个式子容易验证：右边 $z=0$ 时（即状态为 0 的那项），指数 $xz=0$ ，系数为 $(-1)^{xz} = (-1)^0 = 1$ ；对于 $z=1$ 那项，指

数为 x ，系数为 $(-1)^{xz} = (-1)^x$ 。两项都和中间那个式子的两项分别相等。这个关系等会儿要用，但不需要背。正如前面所说的，这里的公式都是来说事儿的。

Hadamard 门在量子计算中很重要：它可以把一个基础状态 0 转换为叠加态 $0+1$ 。也可以用它把基础状态 1 转换为反相的叠加态 $0-1$ 。也可以把这些叠加态转换为基础状态 0 或 1 。

总结一下：

1. Hadamard 门在基础状态 0 或 1 和叠加态 $0+1$ 或 $0-1$ 之间互相转换；
2. Hadamard 门所代表的运算的逆运算门就是 Hadamard 门。

2.2.2 受控“非”门

受控“非”门有两个量子位：一个目标位，一个控制位。从信息流向看，有输入端和输出端，如图 3. 受控非门 (CNOT 门) 所示。当控制位为 0 时，输出不变（即 = 输入）；当控制位为 1 时，输出跟输入相反，即 $0 \rightarrow 1$ ，而 $1 \rightarrow 0$ 。

受控“非”门的符号如图 2 所示。

仔细想一想，这个就是“异或”门，而且，输出对两个输入确实是对称的！

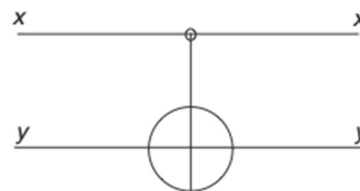


图 3. 受控非门 (CNOT 门)

当控制位是叠加态时会发生什么呢？这正是量子计算的魔力所在：控制作用也叠加！这在量子计算中很重要。后面会看到，很多量子算法都是利用控制位上的叠加实现的。

但是有趣的事情还远未停止。下面看看目标位对控制位的反射作用。

这里要用一点数学了。设目标位为状态 $0-1$ ，控制位为 $0+1$ ，联合状态为 $xy=(0+1)(0-1)$ 输出端是什么呢？输出端为输入分别为 0 和 1 时输出的叠加。对输入 0 ，输出端联合状态为 $(xy)_0=0(0-1)$ ；对输入 1 ，输出端联合状态为 $(xy)_1=1(1-0)=1(-1)(0-1)=(-1)1(0-1)$ 。合起来，得到输出端的状态为： $(xy)_0+(xy)_1=0(0-1)-1(0-1)=(0-1)(0-1)$ 。

好，神奇的事情发生了：本来要用 x 去控制 y 的，结果 x 倒落得个自己从 $(0+1)$ 变成 $(0-1)$ 了。 x 对 y 的作用被反射到输入端 x 。而且对控制位为 0 和 1 的反射作用不一样。叠加的两个状态中的一个（即对应于 1 那个）被反了个向。有兴趣的读者可以把 x 和 y 反过来，看看能得到什么（提示：结果是一样的）。

这种反射作用并不需要从目标位对控制位的物理信息的传输，而是因为控制位和目标位作为复合量子位系统的联合分布由于数据的干涉作用发生了关联。

总结一下，

1. CNOT 门在计算基非叠加态输入时，其作用类似于“异或”门；
2. 输入和输出都可以是叠加态；
3. 在某些叠加情况下输出会反射到输入端；
4. 在叠加态输入时，对参与叠加到两个状态的反射是不一样的；形成干涉。

2.2.3 受控 U 门 (c- U 门)

考虑一个整数函数 $f(x)$: 一个一位输入一位输出的函数。理论上可以用一个量子电路来模拟这个函数。

现在假设有一个量子电路 U_f 模拟一个函数 $f(x)$, 并把它的输出作为我们上述 CNOT 的控制位, 如图 4. 受控 U 门

图 2 所示。当函数的输出值为 0 时, 输出 y 不变, 否则, 输出 y 与输入相反。

同样地, 如果函数输出为 0 和 1 的叠加态时, 输出也要反射到控制位 x 。在目标位输入为 $y = 0-1$ 时, 这个结果可以比较有趣地写成: $(-1)^{f(x)}x(0-1)$ 。读者可以自行验证一下。

事情渐渐地好玩了。继续看下去。

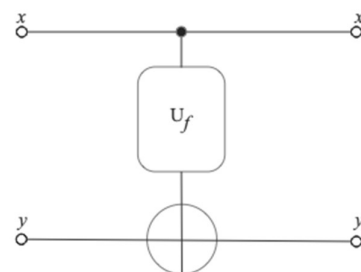


图 4. 受控 U 门

2.2.4 移相门

上述的受控非门在目标位为 (0-1) 时, 在控制位为 0 时, 两个为都不变; 在控制位为 1 是, 将其反相。其实就是旋转了 180° , 也就是一个 π 。在某些情况下, 我们需要一个单量子位门, 其作用就是保持在输入为 0 时, 保持不变, 而在输入为 1 时, 将其旋转一个角度 $2\pi/2^k$ (以弧度计), 称为 k -移相门, 用 R_k 表示, 其作用为:

$$R_k(x) = \begin{cases} 0, & x = 0 \\ e^{\frac{2\pi i}{2^k}} 1, & x = 1 \end{cases}$$

注意, 移相门通常作用于叠加态。因为作用于简单状态没有意义: 单独一个位, 1 和 $e^{\frac{2\pi i}{2^k}} 1$ 其实是一样的状态。但是对于叠加态, 就表示 0 和 1 的不同相位了。

2.3 量子电路

量子电路我们就不再多啰嗦了。就是由量子位、量子门组成的, 先后一步一步进行的链路。这个可以从后面的量子算法例子中体会到。

2.4 量子算法

量子电路的一步一步的演算的方案就是量子算法。

每一本量子计算的书籍都告诉我们, 量子计算具有高度的并行性。怎么实现的呢?

我们设想一个量子电路, 其输入端的每一位都是 0 和 1 的叠加态 (这个可以用前面说过的 Hadamard 门用 0 作为输入而得到), 把这些叠加态输入到一个模拟某个函数 f 的量子电路 U_f , 在输出端我们得到了自变量所有叠加状态下这个函数的值的叠加态。然后怎么办呢? 这时如果我们在输出端测量, 我们得到一个输入下的一个输出, 但无法控制哪个特定输入时的函数值被读出。而且, 一旦读取, 输出就崩塌到了已经读出的状态, 而继续测量得不到新的信息。所以这个办法不能解决什么问题。

为了能让量子计算做些有用的工作, 我们需要把输出端做些重组, 使之形成干涉, 让某些输出值得到加强, 而另一些输出值因相位不一致互相抵消。于是, 在输出端的测量就能得到加强的结果了。所谓量子算法, 就是根据我们所要进行的计算, 安排这

些叠加、干涉，从而得到所要的结果。也就是说，利用量子现象来模拟我们的数学问题，然后测量其结果。从这个意义来看，量子计算倒有几分过去模拟计算的意思。

2.4.1 Deutsch 算法

Deutsch 算法不怎么实用，但是对于理解通过干涉得到结果的量子算法很有启发。我们来看一下。

考虑一个二进制函数 $y=f(x)$ ：输入 x 是 n 位二进制数，输出 y 为 0 或 1，的一个函数。除此之外，我们还知道这个函数的取值要么全是 0 或全是 1（但不知道会是那个），要么 0 和 1 各一半。除此以外，没有别的规律了。请仔细阅读这里的文字，这里说的有点像数学语言，不像人话，不喜欢数学的读者请忍耐一下，需要很字面地理解它。

为了知道这个函数到底是两种情况中的哪一种，在最坏的情况下，我们需要调用这个函数多少次呢？就算不用算完全部的 2^n 个组合，也要算一半加一次：也就是说，算了一半的自变量，如果发现全是一样的，我们还需要再算另一个还没算过的自变量。如果结果还一样，则可以断定全是一样的，否则，就是 0 和 1 各一半，所以需要 $2^{n-1}+1$ 次函数调用。如果 n 是 1000，那么，就是 $2^{999}+1$ ，很长，大约 300 位十进制数那么大，那是算到天荒地老也算不完的！

那么，用量子计算机要调用函数多少次呢？一次！看看怎么做到的。

算法框图如图 5 所示。输入 x 为 n 位的复合。这个全 0 的状态可以写成 $x = 0^{\otimes n}$ 。 $0^{\otimes n}$ 经过一个 Hadamard 门，成为 0 和 1 的所有的叠加态，代表了所有的可能的输入。这是量子算法的第一个优越性：可以把很多，甚至全部输入叠加在一起，让一个函数一下子算出全部的结果！这里我们就一下子算出了全部的结果，在输出端成为结果的叠加态。我们前面说过，如果这时去测量输出，没什么用，只能得到这些输出之中的一个，而且还不能事先确定是哪一个。我们来看看，后面我们怎么用相位的干涉来得到有用的结果的。

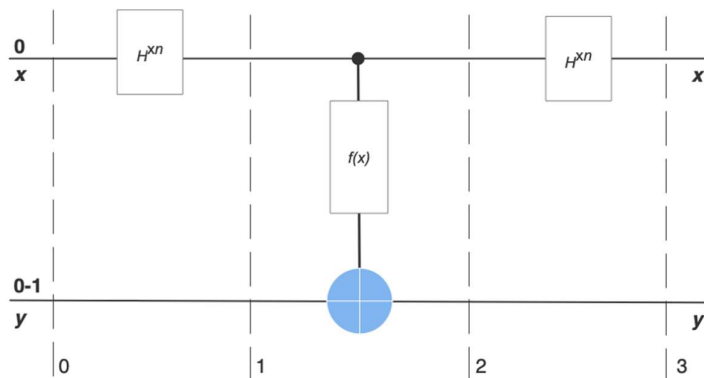


图 5. Deutsch 算法

算出结果后，把结果送去和下面的“异或”门进行“异或”计算。注意到“异或”门的目标位输入为 $y = 0 - 1$ 。根据前面介绍过的异或门在目标位为 $0 - 1$ 时，控制位为 0 时，输出不变，还是 $0 - 1$ ；而控制位为 1 时，输出会反射到输入，使输入反相，即乘以 -1 。也就是说，对于函数输出为 1 的那些输入 x ，其状态会反相；而对于函数目标位为 0 的那些 x 状态， x 不反相。

为了比较好地说明，我们把整个过程分为四个阶段，如图中的四根竖线所示，这些位置的状态分别为 ψ_0, ψ_1, ψ_2 和 ψ_3 。

先看 ψ_0 ，这个比较简单，显然：

$$\psi_0 = 0^{\otimes n}(0 - 1)$$

再看 ψ_1 。这时，函数的值还没有算，所以， x 是全部0和1的叠加态。于是， $\psi_1 = \sum_{0 \leq x < 2^n} x(0-1)$ 。

然后 ψ_2 。我们知道，对于某个 x ，如果 $f(x) = 0$ ，输入和输出都不变，而如果 $f(x) = 1$ ，则输入得到来自输出的反射-1。把两种情况合起来就是，乘以 $(-1)^{f(x)}$ 。于是，

$$\begin{aligned}\psi_2 &= \sum_{0 \leq x < 2^n} (-1)^{f(x)} x(0-1) \\ &= \left(\sum_{0 \leq x < 2^n} (-1)^{f(x)} x \right) (0-1)\end{aligned}$$

式中，第二个量子位为0-1，后面不再用到它，所以不必理会。第一组 n 个量子位（上面大括号中的那个部分）是一个全叠加态（即 $0 \leq x < 2^n$ ），但相位不一定都一样。好像有点意思了。对于不同的函数值，控制位的有些叠加成分的相位相反。结果好像要呼之欲出了。但事情还没完。如果这时直接对 ψ_2 进行测量，这个相反的相位是测不出来的，因为1和-1的绝对值是一样的。

最后是 ψ_3 。这里需要一些数学演算。先回顾一下，一个Hadamard门在控制位为一个非叠加状态 x ，而目标位为0-1时，其输出是： $Hx = \frac{1}{\sqrt{2}}(0 + (-1)^x 1) = \frac{1}{\sqrt{2}} \sum_{z=0,1} (-1)^{xz} z$ 。看看 $H^{\otimes n}$ 对于 n 位复合位（记为 y ）的作用：

$$\begin{aligned}H^{\otimes n} x &= Hy_1 Hy_2 \dots Hy_n \\ &= (0 + (-1)^{y_1} 1)(0 + (-1)^{y_2} 1) \dots (0 + (-1)^{y_n} 1) \\ &= \sum_{z_1, z_2, \dots, z_n} (-1)^{x_1 z_1 + x_2 z_2 + \dots + x_n z_n} z_1 z_2 \dots z_n \\ &= \sum_{z_1, z_2, \dots, z_n} (-1)^{xz} z\end{aligned}$$

注意，这里，-1上面的幂 xz 是其各个对应分量的相乘。因为结果只用于-1的幂，采用二进制算术，相乘就是“与”，相加就是“异或”，是不是进位都一样的，反正我们只需要最后一位。

对于 ψ_2 作为输入，这里的 $y = \sum_{0 \leq x < 2^n} (-1)^{f(x)} x$ ，于是 ψ_3 的第一个量子位组输出为

$$\begin{aligned}H^{\otimes n} y &= Hy_1 Hy_2 \dots Hy_n \\ &= \sum_{x_1, x_2, \dots, x_n} \sum_{z_1, z_2, \dots, z_n} (-1)^{f(x)} (-1)^{x_1 z_1 + x_2 z_2 + \dots + x_n z_n} z_1 z_2 \dots z_n \\ &= \sum_{0 \leq z < 2^n} \sum_{0 \leq x < 2^n} (-1)^{f(x) + xz} z\end{aligned}$$

好，现在我们来分析 $z = 00\dots 0$ 的那项。因为指数中的 $z = 0$ ，这项的系数是 $\sum_{0 \leq x < 2^n} (-1)^{f(x)}$ 。如果 $f(x)$ 是平衡的（回忆一下，就是说，对一半的 x ， $f(x) = 0$ ，对另一半的 x ， $f(x) = 1$ ），那么，这个和=0，因为正负项数相等而抵消，系数为0，也就是说，我们几乎一定不会测到全0；否则，如果 $f(x)$ 是常数的，则所有的项都一样，这个和为 $\pm 2^n$ 。经归一化后，得到这个状态的概率为1。也就是说，测量前 n 个量子位几乎一定得到全0。

最后的结论：如果函数 $f(x)$ 是常数，Deutsch算法的结果为全0，否则，不是全0。

2.4.1.1 Deutsch 算法的人类语言解释

结论有了，但看得还是糊里糊涂。我们用人话来说明一下。

首先，我们用一个 Hadamard 门把全 0 的状态转换成一个全叠加态，它包含了自变量可以取值的所有组合。然后，把这个全叠加状态送到函数计算器，并把结果用受控异或门把结果反射到输入。这时，输入还是很乱，是个高度叠加的状态。最后，把这个高度叠加的状态在用一个 Hadamard 门把这些结果进行干涉，要么正负反相抵消（对平衡函数），要么全部叠加（对常数函数）。于是，我们要么得到全 0，要么不是。所以，在这个算法中，我们用到了量子状态的叠加和干涉这两个重要特性。

现在有点悟出来了吧！Deutsch 算法是最早提出的量子算法（之一？），但也是太简单了一点，不容易看出其具体应用场景。在这里提及这个算法，主要是试图演示量子计算是如何通过叠加、并行、干涉来进行“计算”的。

2.4.2 Grover 算法

Grover 算法是个搜索算法。当被搜索数据为无结构数据时（例如，未排序的数据），需要多少次尝试才能找到所要的项目呢？如果运气不好，最坏的情况下要读完所有的数据才能找出所要的数据。那么，用量子计算能不能更快地完成呢？Grover 算法就试图解决这一问题。

为了明确起见，我们说所有数据成为 N 个数据条目。 N 通常认为是一个 2 的整数幂。即 $N = 2^n$ 。如果没那么多条，就在尾部加一些空的数据。我们还假设，可以用一个函数 $f_a(k)$ 描述数据：其中， a 为待搜索内容，如果第 k 条数据为 a ，则 $f_a(k) = 1$ ，否则 $f_a(k) = 0$ 。因为整个讨论不涉及多个内容项的搜索，所以，此后都忽略了下标 a 。

这个假设很重要：如果数据真的没有任何结构，或者我们对其结构一无所知，或者无法构造一个这样的函数，Grover 算法就无效。所以，Grover 算法不一定适用于某些场合的搜索，因为在那些应用中，我们可能无法构造这么个函数。这里先打个预防针。

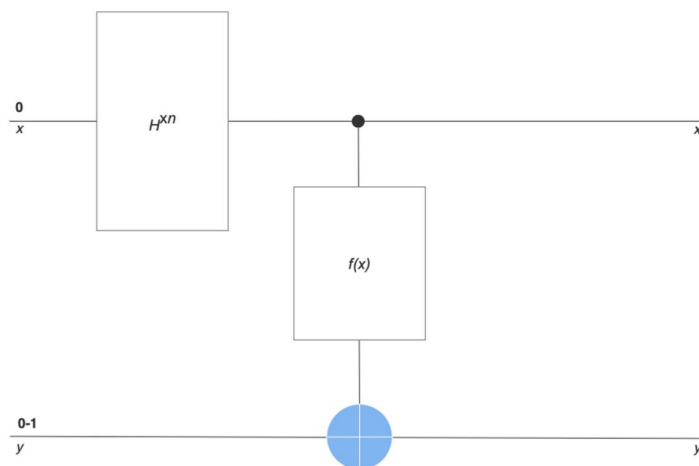


图6. Grover 第一草案

假设我们能用一个量子电路构造那个函数，如果在这个函数的输入端加上一个 0 到 $N-1$ 的所有整数的叠加态，则一次该函数的调用就可以算出全部的条目的判别函数的值。但这个叠加态还不能被利用，因为对其测量只能得到一个下标以及这个下标的数据是否为所要搜索的 a 。为了利用量子计算加速。首先，试着把这个函数作用于一个受控异或门的控制位，而在其目标位上施加 $0-1$ 。如图 6. Grover 第一草案所示。这时，输出端的 x 是一个 0 到 $N-1$ 的所有整数的叠加态，但是，对于 $f(k) = 1$ 的那些 k ，其相位是反相的。也就是说，其状态为

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} (-1)^{f(x)} x$$

这个反相的信息仍然无法直接利用，但是找到了一个能对是否命中的那些 x 区别对待的方法，从而一定能设法放大这种区别，以达到最后被测量到的办法。

下面要用一点数学推演，没耐心不要紧，我会把结论用人话总结一下的。

我们定义一个集合 S ，函数对于 S 内的数，其值为 1： $S = \{x: f(x) = 1\}$ ，以及余集 $S' = \{x: f(x) = 0\}$ 。并设 S 中有 M 个元素（即数据中有 M 个 a ）。

图6. Grover 第一草案中， $H^{\otimes n}$ 作用于 $\mathbf{0} = \langle 00 \dots 0 \rangle$ 后，得到

$$\begin{aligned} \psi &= \frac{1}{\sqrt{N}} \sum_{0 \leq x < N} x \\ &= \frac{1}{\sqrt{N}} \sum_{x \in S} x + \frac{1}{\sqrt{N}} \sum_{x \in S'} x \\ &= \frac{\sqrt{M}}{\sqrt{N}} \frac{1}{\sqrt{M}} \sum_{x \in S} x + \frac{\sqrt{N-M}}{\sqrt{N}} \frac{1}{\sqrt{N-M}} \sum_{x \in S'} x \\ &= \frac{\sqrt{M}}{\sqrt{N}} \psi_S + \frac{\sqrt{N-M}}{\sqrt{N}} \psi_{S'} \end{aligned}$$

其中， $\psi_S = \frac{1}{\sqrt{M}} \sum_{x \in S} x$ 和 $\psi_{S'} = \frac{1}{\sqrt{N-M}} \sum_{x \in S'} x$ 都是单位向量，分别表示 S 和 S' 中的向量的归一化平均和。因为 ψ_S 和 $\psi_{S'}$ 完全由不同的（ S 中的和 S' 中的）基向量组成，所以，它们正交（垂直），并且张成一个两维子空间（平面）。

ψ_S 前面的系数 $= \frac{\sqrt{M}}{\sqrt{N}}$ 一般很小，就是说，盲猜命中率很低。而 $\frac{\sqrt{N-M}}{\sqrt{N}}$ 很接近于 1。两个系数的平方和为 1。这不就是 \sin 和 \cos 的平方和吗？好的，不妨设 $\sin \theta = \frac{\sqrt{M}}{\sqrt{N}}$ ($0 < \theta < \frac{\pi}{2}$)，那么 $\cos \theta = \frac{\sqrt{N-M}}{\sqrt{N}}$ 。这样，

$$\psi = \cos \theta \psi_{S'} + \sin \theta \psi_S$$

再令

$$\psi' = -\sin \theta \psi_{S'} + \cos \theta \psi_S$$

这下更清楚了： $\{\psi, \psi'\}$ 就是 $\{\psi_{S'}, \psi_S\}$ （沿逆时针）旋转了一个小角度 θ ，所以， $\{\psi, \psi'\}$ 组成的平面也在 $\{\psi_{S'}, \psi_S\}$ 平面上。上述旋转的逆变换是：

$$\psi_{S'} = \cos \theta \psi - \sin \theta \psi'$$

$$\psi_S = \sin \theta \psi + \cos \theta \psi'$$

现在的目的就是要对 ψ 进行转换，得到，例如， ψ_1 ，它在 $\{\psi_{S'}, \psi_S\}$ 平面上的表示中 ψ_S 前面的系数（现在是 $\sin \theta = \frac{\sqrt{M}}{\sqrt{N}}$ ）增大到接近 1，类似于 $\psi_1 = 1\psi_S + 0\psi_{S'}$ ，于是能一测一个准，得到 ψ_S 。

为了增强 ψ_S 的信号，我们构造一个量子算子 U_ψ^\dagger （其物理实现就不考虑了），其作用于状态 ϕ 的作用如下：

$$U_{\psi}^{\perp} \phi = \begin{cases} \phi, & \phi = \psi \\ -\phi, & \phi \in \psi^{\perp} \end{cases}$$

这里， ψ^{\perp} 是 ψ 的正交子空间，正像与一个向量垂直的平面上的所有向量那样。

并令 $Q = U_{\psi}^{\perp} U_f$ 。 Q 作用于 $\{\phi_{S'}, \psi_S\}$ 平面上的一个向量 $\phi = \cos \phi \psi_{S'} + \sin \phi \psi_S$ 的作用为：

$$\begin{aligned} Q\phi &= U_{\psi}^{\perp} U_f (\cos \phi \psi_{S'} + \sin \phi \psi_S) \\ &= U_{\psi}^{\perp} (\cos \phi \psi_{S'} - \sin \phi \psi_S) \\ &= U_{\psi}^{\perp} (\cos \phi (\cos \theta \psi - \sin \theta \psi') - \sin \phi (\sin \theta \psi + \cos \theta \psi')) \\ &= U_{\psi}^{\perp} (\cos(\phi + \theta) \psi - \sin(\phi + \theta) \psi') \\ &= \cos(\phi + \theta) \psi + \sin(\phi + \theta) \psi' \\ &= \cos(\phi + \theta) (\cos \theta \psi_{S'} + \sin \theta \psi_S) + \sin(\phi + \theta) (-\sin \theta \psi_{S'} + \cos \theta \psi_S) \\ &= \cos(\phi + 2\theta) \psi_{S'} + \sin(\phi + 2\theta) \psi_S \end{aligned}$$

也就是说，一个 $\{\phi_{S'}, \psi_S\}$ 平面上的向量 ϕ ，经过 Q 的作用，被逆时针转动了 2θ 。

现在，我们从 $\psi = \cos \theta \psi_{S'} + \sin \theta \psi_S$ 开始，经过多次（例如 k 次） Q 的作用，我们就得到

$$\begin{aligned} \psi_k &= Q^k \psi \\ &= \cos((2k+1)\theta) \psi_{S'} + \sin((2k+1)\theta) \psi_S \end{aligned}$$

我们希望 ψ_S 前面的系数接近1，也就是说， $\sin((2k+1)\theta) \approx 1$ 。也就是说， $(2k+1)\theta \approx \pi/2$ 。这样，我们就可以计算出最佳的次数。

回忆一下， $\sin \theta = \frac{\sqrt{M}}{\sqrt{N}}$ ，由于 θ 很小， $\theta \approx \sin \theta = \frac{\sqrt{M}}{\sqrt{N}}$ ，所以， $(2k+1) \frac{\sqrt{M}}{\sqrt{N}} \approx (2k+1)\theta \approx \pi/2$ ，得到：

$$k \approx \frac{\left(\frac{\sqrt{N}}{\sqrt{M}} \frac{\pi}{2} - 1\right)}{2} \approx \frac{\pi \sqrt{N}}{4 \sqrt{M}}$$

如果 $M=1$ （即只有一个答案，这是常见的情形），则

$$k \approx \frac{\pi}{4} \sqrt{N}$$

可以看到，我们需要 $O(\sqrt{N})$ 次的运算可以得到结果，比直接的野蛮试验法得到了平方根级的加速。

注意，不能转太多。太多了，超过了 $\pi/2$ ，效果就又下降了，甚至于消失。所以要控制次数，也就是说，要对可能有多少个命中值有个靠谱的估计。更为深刻的分析可以用量子计算对这个命中率进行估计，但由于涉及较多的数学演算，我们就不再深入了。值得一提的是：量子算法的设计其实就是这样的分析和综合，以得出可以解决实际问题的量子电路。

最后用人话总结一下：

Grover 算法利用一个量子电路所实现的判别函数，利用叠加态平行计算该函数值。并且，利用命中点对输入的反相和旋转，不断加强命中点的信号，直到加强到出现的概率接近1。然后对其进行测量，而得到命中点。这一算法可以实现搜索算法的

平方根级的加速，但对于被搜索空间必须能定义一个量子电路，实现一个判别函数。这个条件对于很多应用不一定满足的。特别是：如果需要调用一个数据库扫描，则这个算法是不现实的，因为数据库扫描无法比野蛮搜索更快，更何况要扫 \sqrt{N} 次呢！

2.4.2.1 比特币加速

说到搜索，比特币的 Hash 是一个搜索。我们知道，影响矿机性能的最主要因素是 SHA256 算法的速度。我们来看看，Grover 算法是否可以加速矿机的 Hash 算法。

我们需要一个判别函数 f ，当 Hash 小于一个特定的阈值时返回 1，否则返回 0。这样的量子电路原则上是可以实现的（因为 SHA256 基本上是个多层门电路，是个不深的电路）。

考虑 SHA256，Hash 的值为 256 位，但要求 Hash 小于某一个随着网络计算能力而变的阈值 T 。所有小于这个阈值的 Hash 值都可以接受。因为 Hash 的分布比较均匀，近似地，可以认为，搜索空间 $N = 2^{256}$, $M = T$ 。如果 $T = 2^{210}$ （我没有细查目前 T 为多少，给个大致估计，256 位二进制数 46 个二进制 0 开头），则

$$k \approx \frac{\pi \sqrt{N}}{4 \sqrt{M}} = \frac{\pi \sqrt{2^{256}}}{4 \sqrt{2^{210}}} \approx 2^{23}$$

而经典的计算，要让 Hash 落入小于 T 的范围，则其概率为 $p = T/N = 2^{210}/2^{256} = 2^{-46}$ 。所以，平均需要 $1/p = 2^{46}$ 次试算。确实是个平方根级的加速。就算一次量子计算耗时相当于 1000 次经典计算，也能加速 $2^{46-23-10} = 2^{13}$ ：一万亿倍！

但是，量子计算机真的获利吗？如果量子计算机成为成熟技术，那么就会出现水涨船高的局面。根据比特币协议，上述阈值会随之改变，以保持每 10 分钟产生一次比特币（个数也会随着时间的推移而减少）的速度。

那么，会不会计算太快，最后击垮比特币体系呢？极限情况下，阈值变成 1（即只有 256 个 0 才是合格的 hash），则 Grover 算法也需要 $2^{128} \approx 10^{43}$ ，仍然是一个工程上无法设想的次数。所以，Grover 算法不会击垮比特币。只会让阈值 T 变得较小。究其原因，是因为 Grover 算法不是一个指数级加速算法。平方根级加速是个多项式级的加速，无法改变问题的本质：一个具有指数级难度的问题不会被一个多项式级的加速所击垮！

那么，如果全世界只有你有一台量子计算机，而且成本合理，你会不会因为比特币极大地获利呢？会，但我劝你做点更有意义的事儿！例如，药物的搜索，举一个例子。这对你个人和对社会都更有意义！

2.4.3 Fourier 变换

下面看一个至少听起来还比较有用的算法：量子傅立叶变换（QFT）。大家都听过 FFT（快速数字傅立叶变换），QFT 要解决同样的问题，但速度比 FFT 快得多。这不是快多少倍的问题，而是一个指数级的加速。

数字傅立叶变换可以表示为：

$$x \mapsto \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n-1} e^{2\pi i \frac{x}{2^n} y}$$

看看它的指数部分，不计那个虚数单位 i ，是 $\frac{2\pi x}{2^n} y$ ， y 为整数。可以看出， $\frac{x}{2^n}$ 相当于频率，乘以 2π 就是圆频率。 $\frac{x}{2^n}$ 是个基础频率。 y 是整数，表示基础频率的整数倍。然后把这些整数倍加起来，得到傅立叶变换。前面的 $\frac{1}{\sqrt{2^n}}$ 是个归一化系数。如果参考数字信号处理的书籍，这个系数不一样，可以不管这个。在量子计算中，所有状态都需要归一化，前面那个系数只为归一化。

注意，这里的变换只是对一个输入矢量 (x ，视为一个 n 位的矢量) 的傅立叶变换。有意义的变换要对大量这样的矢量 (视为值) 进行变换，然后把结果加起来，才得到这些输入信号的频谱。也就是说，傅立叶变换可以找出一系列输入值的频谱。

我们知道， e 的纯虚指数函数，其指数被乘以一个倍数，相当于在复平面旋转了一个角度，再把它加起来。那么，当输入值具有周期性时，转了一圈又回来了，某些信号会被叠加，而非周期的信号会被互相抵消。当输入信号是严格周期时，信号被明显的加强。

现在，我们把上述的数字傅立叶变换用量子状态来表示：

$$x \mapsto \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n-1} e^{2\pi i \frac{1}{2^n} xy} y$$

把 x 用二进制小数表示： $x = (x_1 x_2 \dots x_n) = 2^{n-1} x_1 + 2^{n-2} x_2 + \dots + 2^0 x_n$ ，则

$$(x_1 \otimes x_2 \otimes \dots \otimes x_n) \mapsto \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n-1} e^{2\pi i (\frac{x_1}{2^1} + \frac{x_2}{2^2} + \dots + \frac{x_n}{2^n}) y} y$$

这里， x 、 y 是整数。黑体字的 x 、 y 表示 x 、 y 的二进制表示所对应的由量子位组成的 n 位状态。

从这个表达式不容易看出怎么用量子门实现这个算法。用跟前面 Deutsch 算法类似的方法 (经过比较冗长的推导，并注意到，对于整数 k ， $e^{2\pi i k} = 1$ ，也就是说， $1/2^n$ 被 y 乘了以后，大于 1 的部分是可以忽略的)，把右边的和式用 n 个量子位的张量积表示，可以得到：

$$(x_1 \otimes x_2 \otimes \dots \otimes x_n) \mapsto \left(\mathbf{0} + e^{2\pi i (\frac{x_1}{2^1} + \frac{x_2}{2^2} + \dots + \frac{x_n}{2^n})} \mathbf{1} \right) \otimes \left(\mathbf{0} + e^{2\pi i (\frac{x_2}{2^1} + \dots + \frac{x_n}{2^{n-1}})} \mathbf{1} \right) \otimes \dots \otimes \left(\mathbf{0} + e^{2\pi i (\frac{x_n}{2^1})} \mathbf{1} \right)$$

右边的运算可以“方便”地实现，见图 7. 量子 Fourier 变换算法框图。

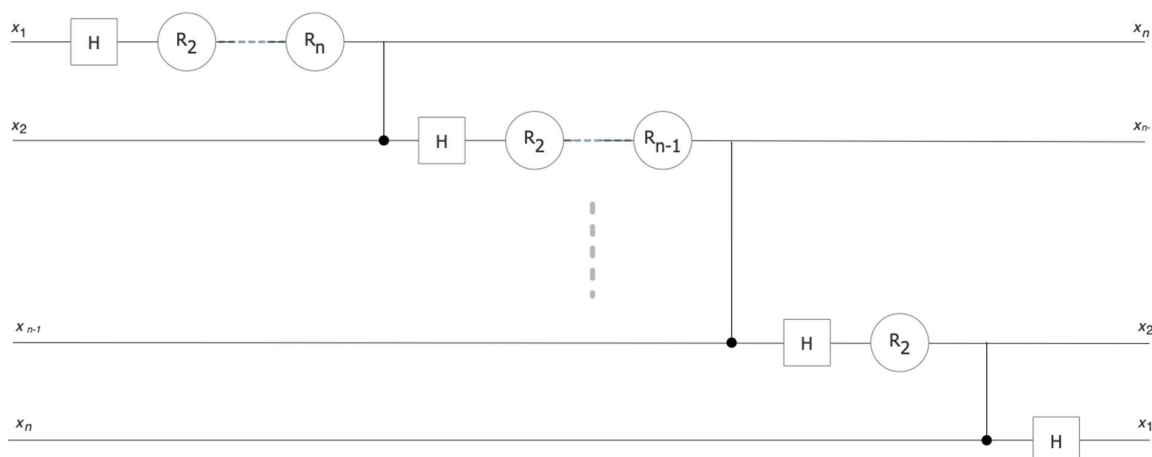


图 7. 量子 Fourier 变换算法框图

有没有点奇怪： R_1 怎么没有？不奇怪，Hadamard 门 H 就是 R_1 。不是吗？ H 将输入 1 反了相，就是旋转了 π ，也就是 $\frac{2\pi}{2^1}$ ，所以， H 就是一个 R_1 。

有一个细节说明一下：输出端的量子位的标注和输入端是相反的：从 x_n, x_{n-1}, \dots ，到 x_1 ，这跟上面的公式一致：输出端 x_n 只和输入端 x_1 有关，输出端 x_{n-1} 和输入端 x_1 和 x_2 有关， \dots ，而输出端 x_1 和所有的输入端 x_1, x_2, \dots, x_n 都有关。

傅立叶变换可以把一个数用 1 和 0 的不同的相位差进行编码。当把很多杂乱无章的数据进行编码时，它们的相位各不相同。如果把这些不同相位的编码加起来，则不同的相位互相抵消，结果比较平均。但是，如果输入的数据是周期性的，那么情况就不同了：如果周期是 r ，则过了 r 个，一样的数据又来了。如果数据里面含有很多个 r 个数据，那么，这种叠加就可以很强，产生峰值。也就是说，以周期性数据作为输入，而且数据含有很多个长度为 r 的周期时，傅立叶变换的输出在频率为 $\frac{2\pi k}{2^n} = \frac{2\pi}{r}$ ， k 为某个整数，处出现峰值。通过这个关系，可以得到 r 。

2.4.4 大合数分解

可能完成大合数分解是量子计算机的第一个引起人们巨大兴趣的问题之一。我们知道，我们现在普遍使用的 RSA 算法就是建立在大合数分解的困难性之上。如果大合数分解被解决，则 RSA 将会被一举攻破。除此以外，基于离散对数的椭圆曲线也可以被上述的量子傅立叶（QFT）算法破解（本文就不再深入了）。

用 QFT 破解大合数分解，是基于一个叫做 Shor 算法（Shor's Algorithm）。所以我们先介绍一下 Shor 算法。

2.4.4.1 Shor 算法

分解大合数是个数论问题，其说明需要一点点最基本的数论知识。

大合数指一个大正整数，它由若干个（主要考虑两个）大素数相乘得到。素数是指不能表示为比它小的两个正整数之乘积的正整数。例如， 5 就是素数，而 10 就不是，因为 $10 = 2 \times 5$ ，而 2 和 5 都比 10 小。虽然听起来很简单，但是对于很大的合数来说，

要找到其素因子分解，一般而言，极其困难，至今没有找到亚指数的通用算法。亚指数算法指比指数级复杂度低的算法。

下面我们来介绍一个用找阶的方法来求素因子分解的方法。

考虑由两个大素数 p 、 q 相乘得到的合数 $N=pq$ 。我们知道 N ，但不知道 p 或 q （这个很容易做到：他们随机找两个大素数 p 和 q ，相乘得到 N 。他们只告诉我们 N ，但不告诉我们 p 和 q ）。我们任取一个跟 N 互素的数，例如 a 。所谓 a 和 N 互素，指 a 和 N 没有大于 1 的公因子，也就是说，没有一个数（1 不算）同时整除 a 和 N 。这个 a 很容易找，其实，随机找一个，几乎不可能和 N 有公因子的（否则，我们已经找到 N 的因子了）。我们把 a 自乘若干次，得到 a^k ，因为 a 和 N 互素，自乘不管多少次， a^k 和 N 还是互素（ a 中没有 N 的因子， a^k 中也没有）。考虑 a^k 除以 N 的余数 r ，记为

$$a \equiv r \pmod{N}$$

在具体计算时，我们认为 r 是余数，于是 $0 \leq r < N$ 。在数学中，上述式子也表示 a 和 r 关于模 N 同余，即除以 N 的余数相同。同余之间的计算很有意思，只要参与计算的数与 N 互素，通常的加、减、乘法和整除计算都是成立的。（为避免引进太多概念，我们不涉及非整除的情况）。为便于叙述，我们记 $r = \text{mod}(a, N)$ 为 a 除以 N 的余数。注意，余数 r 为 0，即 $a \equiv 0 \pmod{n}$ 则表示 N 能整除 a ，也记为 $N | a$ 。

注意，上述的幂 $\{a^k, k=1, 2, \dots\}$ 的余数都小于 N （余数的定义），于是，总有一些 k 和 $k' > k$ ， $a^{k'} \equiv a^k \pmod{n}$ 。于是，两边同整除 a^k ，得到 $a^{k'-k} \equiv 1 \pmod{n}$ 。我们称最小的这样的 $k' - k$ 为 a 对于 N 的阶（order），记为 $\text{order}_N(a)$ 。

现在，假如我们对上述的 a 和 N 能找到这样一个阶 r ，而且，运气好的话，这个 r 是偶数（概率很大），也就是说， $r = 2r'$ ，这样，则 $a^r = a^{2r'} = (a^{r'})^2 \equiv 1 \pmod{n}$ 。记 $b \equiv \text{mod}(a^{r'}, n)$ ，于是

$$(b+1)(b-1) = b^2 - 1 \equiv 0 \pmod{N}$$

即 $N | (b+1)(b-1)$ 。因为 $b \equiv a^{r'} \pmod{n}$ ，其幂 $r' = r/2 < r$ 。根据 r 的定义， $b \equiv a^{r'} \not\equiv 1 \pmod{n}$ ，（否则，阶就比 r 小了）。所以， $b-1 \not\equiv 0 \pmod{N}$ ，意即 N 不能整除 $b-1$ 。如果运气不错， $b+1 \not\equiv 0 \pmod{N}$ 也成立，那么，我们有以下结论： N 不能整除 $b+1$ 和 $b-1$ ，但能整除其乘积： $N | (b+1)(b-1) = b^2 - 1$ 。这说明， $b+1$ 和 $b-1$ 都含有 N 的一个小于 N 因子。用欧几里得（Euclid）辗转相除法可以很容易地找出这些公因子。既然是公因子，当然也是 N 的因子。

如果 r 不是偶数，或者后来得到的 $b+1$ 正好能被 N 整除怎么办呢？这种可能性不大。如果这样，再选一个 a ，重新做一次，直到找到满足条件的 a 。可以证明（不再卷入这些细节了），这个概率小于 $1/2$ 。也就是说，多试几次就能以很大的概率成功。

就这样， N 的因子被找出来了。但是，我们还缺一个细节：怎么才能得到那个阶 r 呢？用经典计算机计算阶很困难，要计算所有的 a 的幂，并且还要从这些幂中找出周期。虽然可以有比这些步骤简化的算法，但是至今没有亚指数级的算法。然而，量子计算机能！

2.4.4.2 用量子计算找阶

找阶其实就是找频率。序列 $\{a^k\}$ 以阶为周期反复出现。只要把这样的序列进行傅立叶变换，其输出就会在这个周期 r 对应的频率 $2\pi/r$ 处出现峰值。

但是，如果用一个经典计算机在做这事儿，要计算至少很多个这样的 a^k 。 r 又是数量级为 N 的巨大的数，根本就不现实。这时量子计算 comes to rescue（来救我们了）。构造一个 $\text{mod}(a^k, N)$ 的量子计算电路，把对所有的 k 的结果利用量子计算机的平行计算能力一口气计算出来，并把这些结果统统叠加起来，一起去输入到一个 QFT 量子电路，则在其输出端就得到一个概率分布，在输入数据的周期 r 这个地方有个峰值。如果这时对 QFT 的输出进行测量，我们就有可观的概率得到 r 。

得到一个 r 的候选者后，可以用经典计算机进行试算， r 是不是偶数，如果是，则算出上述的 $b-1$ 和 $b+1$ 。如果符合条件，就能得到 N 的分解。否则，再试一次或若干次，可以以很大的概率成功。

不过，有一事需要提醒。如果 r 是一个周期， $2r, 3r, \dots$ 也是周期（想得通吗？当然啦）。比如说， 2π 是 \sin 函数的周期，其实 $4\pi, 6\pi$ 等等也是。所以，我们在 QFT 输出端的测量只能得到 r 的一个整数倍，而不是 r 。所以，我们还需要一些数学的技巧来找到这个最小的周期。我们就不再纠结到这些细节了。如果有兴趣，只能请各位“读原著”了。

我们这里理解到量子计算机可以用来求解大合数分解的核心是，1、高度的平行计算：一口气算出全部的指数余数，2、并把其结果放进一个叠加态；3、把这个叠加态又送进一个量子傅立叶变换中进行干涉，找出其周期；4、用经典计算机对周期进行进一步加工；如果成功，万事大吉，否则，有必要的話：5、重试若干次，以极大的概率得到最后的结果。

3 量子计算机

本来还应该介绍一下量子计算机的。量子计算机就是在严格控制条件下准确实现上述量子现象的装置，构造成千上亿个量子位，并让他们通过量子门、量子电路准确地叠加、干涉、纠缠²。目前（就个人所知，至写到这里时），人类构造了由 53 个量子位构成的专用量子电路，进行了量子计算模拟实验。但这个实验所实施的算法很难 practically 想象任何（has little practical）价值。要想让量子计算机真的解决实用问题，距今还非常遥远。

为了使上述量子现象精确地、稳定地存在并被准确测量，需要满足一系列的要求。简单概括一下：

1. 需要极低的温度，说个概念：用绝对温度表示，目前大约需要低到 20 毫度（就是一度的千分之二十）以内。
2. 由于量子现象只在微观系统中存在，而微观系统很脆弱，很容易受到干扰而失去应有的数学关系（decoherence）。
3. 还是这个问题，要解决一个实际的问题，往往需要至少成千，乃至上百万个量子单元的共同参与，且它们都必须处于稳定的纠缠之中。也是由于上述的原因，目前，这种纠缠的持续时间很短，很脆弱。

² 文章中没有具体介绍纠缠在量子计算中的作用。但事实上，还是出现的。为了聚焦重点，文章中忽略了。

在可见的将来能否成功，还真的需要人们用大量的人力物力，在毫度的温度下对其进行豪赌！具体内容，个人理解也颇为粗浅，就不在这里介绍了。

4 量子计算的潜在应用

目前阶段，量子计算机还处于 speculation（期盼？）的阶段，硬件能力还远远不能支撑任何有意义的应用的需要。这里聊一下个人对量子计算的潜在应用的一下期盼。

本文中，我们主要介绍了两种量子算法：量子傅立叶变换（QFT）和量子搜索算法（Grover 算法）。

QFT 可以广泛地应用于各类信号处理的领域。特别地，可能被应用于具有那些数据具有某些内在数学规律，但这种特征不容易被描述。其实，在 QFT 找阶的算法中， $\{a^k\}$ 就是这样的序列。把它们打印出来，看不出任何规律，用统计算法分析，很像均匀分布的随机数序列。但是，它却来自一个简单的函数（连续的乘法求余）。这是量子计算最容易发挥作用的场合：用量子门构造一个量子电路，去模拟那个规律，然后用 QFT 将问题转入频域分析。

非常类似地，Grover 算法也可能被广泛地应用于优化、新结构的搜索（前面提到的药物搜索就是一例）。其基本方法就是用量子电路构造一个判别函数，当满足搜索条件时，返回结果 1，否则，返回结果 0。将这个电路插入 Grover 算法的框架中，就可能找出符合判别函数的一个结果，be it a satisfying structure or an optimized outcome.

这里，对领域的数据有两个方面的要求：

1. 这些数据具有某种内在的数学规律，虽然这种规律除了穷举计算不易利用。上述的 QFT 找阶就是这样的情况。这种情况下，数据量可以是非常巨大，例如 2^n ，而 n 取决于技术实现的规模，例如 100，或 2000（对于大合数分解就要有这么大）。前面说到的药物都搜索也是这样的情况。
2. 这些数据来自社会（或实验），我们不知道其内在的数学规律。这是更有吸引力的应用（可能有争议，个人意见）。但是，在这种情况下，我们无法构造 Grover 算法中的那个判别函数，也难以构造类似于 QFT 找阶算法中的序列生成器。
3. 此外，一种量子随机存储器模型（QRAM）也很有吸引力。类似于经典计算机的内存，量子随机存储器存储大量数据，但以叠加态的方式存储于量子电路（主要有量子位组成的电路）中，其特点是，QRAM 可以用叠加态的地址访问数据。存储器返回与相关地址对应的子数据集的叠加态。后面就可以把这些数据用于后续的平行计算，大大加快运算速度。

4.1 机器学习应用场景

在上述 2 的这种情况下，量子搜索也还是有可能实现的：我们要把所涉及的数据都放入判别函数中（启动时装载），利用量子计算平行地计算所有候选空间中的对象的判别函数。但这类方案只能在较小的数据集上应用，例如，机器学习。

机器学习，就是在模型的参数空间中的搜索，只要能构造学习效果的判别函数，搜索就能实现。利用量子计算进行的机器学习的优势在于，可以对搜索空间进行平行

搜索，达到平方根级的优化。同时，如果涉及到策略的生成，同样地，可以用序列生成器同时生成所有的策略，进行平行优化。

4.2 大数据应用场景

这里所涉及的情况是，海量数据完全来自社会或实验，没有一个构造函数可以生成这些数据。在这种情况下，Grover 算法是失败的。因为 Grover 算法要多次地调用判别函数，每次判别函数的调用都涉及到扫描数据。事实上，比穷举搜索还慢！但是，对于 QFT（以及其他类似数学变换）的应用，还是有可能的。因为 QFT 中，被分析数据只被使用一次，有可能在初始时刻装载。

5 结束语

本文以个人对于量子计算的粗浅理解，介绍了量子现象和量子计算的主要概念、机制和潜在应用。由于篇幅和资源所限，加之个人学识浅陋，挂一漏万在所难免。只希望对各位理解这一领域有所帮助。

本文没有涉及量子计算机的硬件构造和机理。这还是一个处于婴儿阶段的工业，涉及很多物理机理的利用，尚未形成主流的方向。本人也只是道听途说，不便在这里班门弄斧。

由于量子计算究其本质，是利用量子规律模拟数学关系，所以文章难免涉及到一些数学概念和方法。本文的目的全无在教会读者那些数学概念和方法，而是借用那些概念和方法说明问题，就是前面说的，说事儿。所以，文章中已经尽可能压缩了对数学概念的引用，且保持所用的概念和符号都停留在所涉及学科的最肤浅的层面上。特别值得一提的是，跟绝大多数量子力学和量子计算相关的书籍不同，本文没有采用 Dirac 符号。原因是我已经在文章中避免使用了内积这个 Dirac 符号体现优势的概念，所以，就没有必要使用 Dirac 符号了。Dirac 符号的松散形式有时也给初学者带来麻烦。