

3

布莱叶盲文与 二进制码

塞缪尔·莫尔斯并不是第一个将书面语言的字母成功编码的人。实际上，在因自己的名字被用来命名某种编码方式而名声大噪的人中，莫尔斯先生也不是第一人。这项荣誉应该归一个法国盲人小伙子所有。尽管他比莫尔斯要晚出生 18 年，但是他很早就创建了自己的编码规范。他的生平鲜为人知，尽管关于他的故事流传至今的为数不多，但也足以构成一个引人注目的传奇。

路易斯·布莱叶 (Louis Braille) 于 1809 年生于法国的库普雷，这是一个距离巴黎市东只有 25 英里的小镇。布莱叶的父亲是一个马具匠人。在布莱叶 3 岁的时候——一个本不该在父亲的工作间玩耍的年纪——他不小心被一个尖锐的工具刺伤了一只眼睛。伤口感染也影响到了他的另外一只眼睛，最后使他的双目完全失明。按常理，布莱叶将在无知和贫困当中度过自己的一生（就像当时大部分盲人那样），但是小路易斯对知识的渴望和过人的智慧很快被



人们所发现。他最初同其他孩子一样在镇上的小学上学，后来在他 10 岁的时候，在小镇牧师和一名学校老师的帮助下，布莱叶被送往巴黎皇家盲人学校学习。

很显然，盲人教育中一个主要的障碍就是盲人无法阅读印刷的书籍。瓦伦丁·霍伊（1745-1822），巴黎皇家盲人学校的创始人，曾发明过一种在纸面上印下凸起文字的系统，这样盲人就可以通过触摸的方式来进行阅读了。但是这种文字系统使用起来很困难，而且使用这种方法来印刷的书籍也非常少。

霍伊先生视力健全，因此他被自身的感知模式所禁锢。对他来说，一个字母 A 就是一个 A，而且字母 A 在记录时也必须看起来（或者感觉起来）像一个 A（如果让他用手电筒交流的话，估计他会像我们最开始所做的那样，试图在空中比划出要表达的字母，但其实我们发现这法子根本不灵）。霍伊先生或许没有意识到，有一种完全不同于印刷字母的文字系统可能更加适合盲人阅读。

这种非常规编码方式的起源说出来可能有些出人意料。查尔斯·巴比尔（Charles Barbier），法国军队的一位军官，在 1819 年发明了一种他自称为“*écriture nocturne*”（也叫“夜间书写”）的文字系统。他在厚纸上使用凸起的点和划的组合来表示文字，这样当部队需要无声交流的时候，即使光线很暗，士兵们也可以通过这些符号互相传递信息。他们使用一个锥形的铁笔在厚纸的背面书写，这样纸的正面就会有相应的凸起。然后人们就可以使用手指触摸这些凸起的点和划来进行阅读了。

巴比尔文字系统的缺点是太过复杂了。该系统并非使用与字母表相对应的点划编码串来表示字母，而是用与读音相对应的编码串表示，因此有时仅仅是为了表示一个单词，就不得不使用很多的码字。如果只是传递简短的消息，这个系统用起来倒还不错；但是在表示长文本的时候，就明显力不从心，就更别说用来对整本书进行编码了。

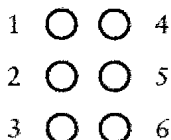
布莱叶在 12 岁的时候就很熟悉巴比尔的这种文字系统了。他特别喜欢使用凸起的点，不仅仅因为凸起的点通过手指就可以很容易实现阅读，还因为它们“写”起来也很简单。在教室里，一个盲人学生如果有了纸张和铁笔，他就可以做笔记，而且同时还能阅读记下来的文字。路易斯·布莱叶开始不辞辛劳地改进这个文字系统，3 年以后（那时他 15 岁）就创建成了自己的系统，而这个系统中的一些基本规范，直到今天仍在被人们所使用。在很长一段时间内，这种新的文字系统只被他们学校内部的人们所熟悉，但是渐渐地，布莱叶盲文传播到了世界的各个角落。1835 年，路易斯·布莱叶患上了肺结核。1852

年，就在布莱叶 43 岁生日刚过完不久，病魔无情地夺走了他的生命。

如今，在引导盲人进入文字殿堂的道路上，改进后的布莱叶盲文系统与磁带录音书进行着竞争，但是布莱叶盲文仍然是不可替代的工具，特别是对于又聋又盲的人来说，布莱叶盲文仍然是他们开启阅读大门的唯一钥匙。近些年来，布莱叶盲文被越来越多地应用到公共场所中，使得盲人朋友们可以越来越方便地使用电梯、自动取款机等设备。

在这一章中我们将解析布莱叶盲文，来看看它是如何工作的。我们并不是要真的学习布莱叶盲文，而且也无须刻意记住什么关于它的内容。我们仅仅希望从中窥探到编码的一些本质。

在布莱叶盲文中，每个在书写文字中用到的符号——具体来说就是字母、数字和标点符号——都被编码成为 2×3 的点码单元中的一个或者多个凸起的点。这个点码单元包含的点通常使用 1 到 6 的数字来编号。



在现代应用中，使用特殊的打印机或轧花机可以将布莱叶盲文印到纸张上。

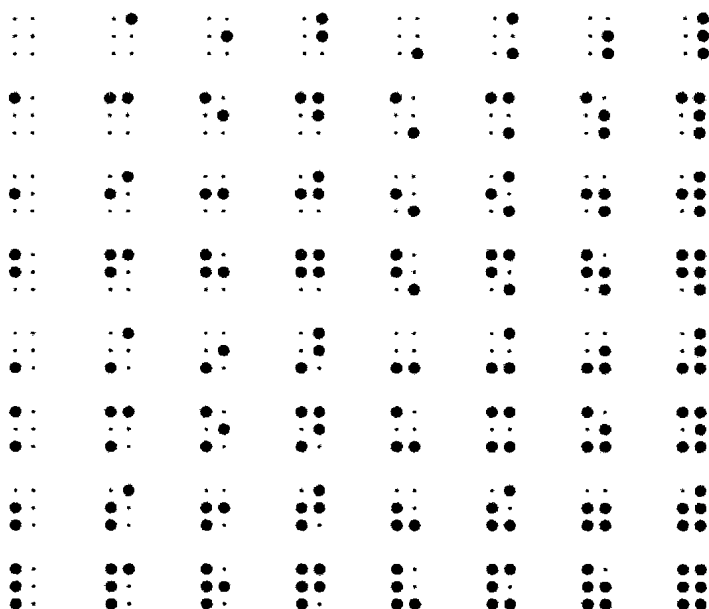
即使只是用布莱叶盲文在本书中印上几页，造价也实在是有点儿高，所以我在书中使用一种符号来表示布莱叶盲文。在这种表示方法中，点码单元中所有的 6 个点都会被表示出来。大点表示这是一个凸起的点，小点则表示其对应的位置是平的。例如，下面的布莱叶盲文：



其中，第 1，3，5 点是凸起的，而第 2，4，6 点的位置是平的。

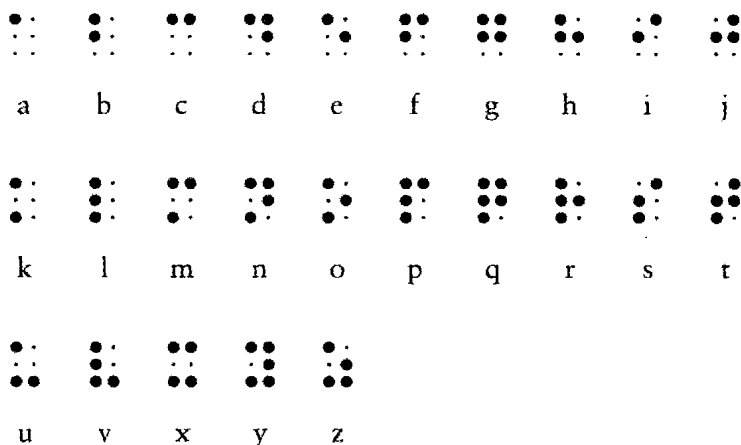
令我们感兴趣的是，这些点码都是二进制的。一个单独的点不是平的就是凸起的。也就是说我们可以把在莫尔斯编码中学到的知识应用到布莱叶盲文的分析中！我们现在已经知道，每组有 6 个点，并且每个点有平和凸两种状态，因此 6 个可平可凸的点的组合数就是 $2 \times 2 \times 2 \times 2 \times 2 \times 2$ ，即 2^6 ，也就是 64。

因此，布莱叶盲文系统能够表示 64 个不同的码字。下图就是所有可能的 64 种码字。



假如发现布莱叶盲文中用到的码字数目少于 64，就会有人问了，为什么 64 个可能的码字中有一些被遗弃不用呢？假如我们发现布莱叶盲文中用到的码字数目超过 64，问题就更严重了，我们会怀疑自己的智商，甚至怀疑基本的数学原理——到底 2 加 2 是不是等于 4。

还是让我们来开始解析布莱叶盲文吧，首先看看基本的小写字母表。



例如，词组“you and me”用布莱叶盲文表示成如下组合：

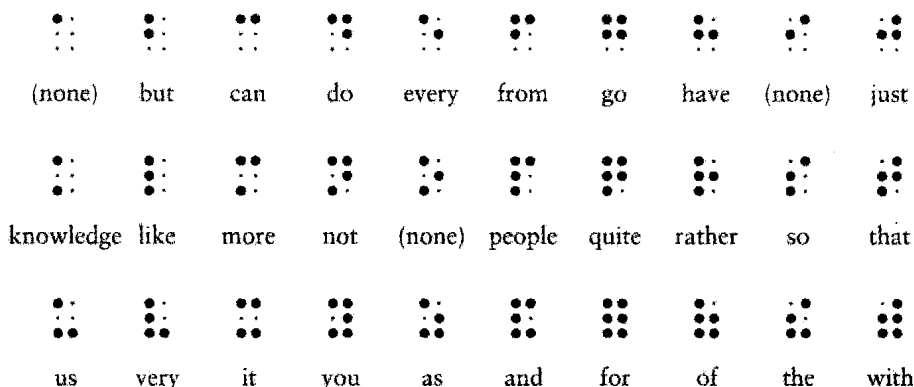


注意，一个单词中每个字母所对应的点码单元之间都用一小块空白分开；而单词之间则用一个大的空格（实际上是没有凸点的点码单元）所分隔开来。

这就是路易斯·布莱叶制订的布莱叶盲文的基本规范，至少它对于拉丁字母很适用。布莱叶还制订了表示重音符号字母的编码，这在法语中要经常用到。注意这里没有“w”这个字母的编码，因为在传统的法语中不会用到它（别担心，这个字母最后会出现的，下文将作讲解）。到此为止，64个码字中只使用了25个。

经过仔细的检查，你会发现，从我们列举的那个三排布莱叶盲文的例子（小写字母表）中，可以总结出一个规律。第一排（字母a到j）只用到了点码单元中最上面的四个点——第1、2、4和5点。第二排在复用了第一排的编码的基础上，把第3点改为凸点。第三排也沿用了同样的规律，只是将第3和第6点改为凸点。

自从路易斯·布莱叶发明布莱叶盲文以来，其应用已经扩展到各个领域。目前在英文出版物中最常用的盲文系统被称为二级布莱叶盲文（Grade 2 Braille）。二级布莱叶盲文使用了很多缩写，以便于保存树型结构和提高阅读速度。例如，如果字母的码字单独出现，它们就表示一个普通的单词。以下三排图样（包含“完整的”第三排）为我们展示了这些单词的码字。

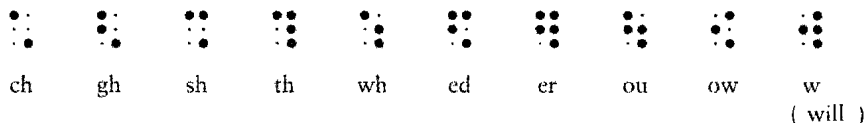


因此，短语“you and me”使用二级布莱叶盲文就可以表示为：



到目前为止，我们已经描述了 31 个码字——单词间的大空格（即没有凸点的点码单元）以及总共 3 排每排 10 个的字母和单词码字。与理论上可以达到的最多码字数——64 相比，我们还差得远呢。正如从下面的分析中我们将看到的那样，在二级布莱叶盲文里，没有任何的码字会被浪费。

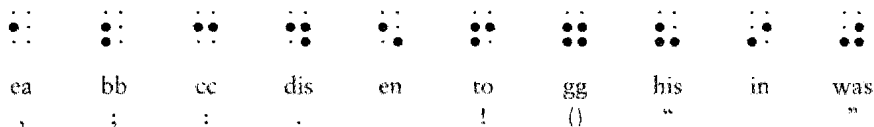
首先，我们可以使用字母 a 到 j 的码字加 6 号凸点的组合。这些新码字通常用来表示单词中字母串的缩写，还有我们前面所担心的 w 字母（像上文的二级布莱叶盲文单词编码一样，表示 w 字母的编码也可以表示一个单词）。



例如，单词“about”使用二级布莱叶盲文可以记作：

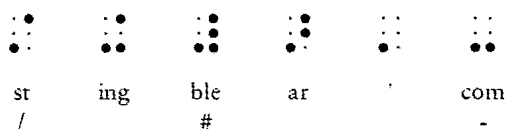


然后，我们取从 a 到 j 的码字，“降低”它们使用的点位，只用到编号为 2、3、5 和 6 的点，这样就得到了新的码字。根据上下文环境，它们将被用来表示一些标点符号或者字符串缩写。

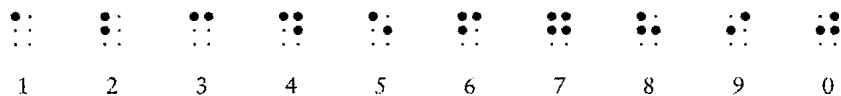


前 4 个码字分别表示逗号、分号、冒号和句号。需要注意的是，左右括号使用的是相同的码字，但是开闭引号使用的却是不同的码字。

目前为止我们已经定义了 51 个码字了。“占用 3、4、5 和 6 号点”的码字还有很多组没有被定义，下面的 6 组就是这些被遗漏的码字，我们也用它们来表示一些字符串缩写和其他的标点符号。



表示字符串缩写“ble”的码字很重要，因为它不是用来作为单词的一部分时，它的出现就意味着之后的码字应被译为数字。表示数字的码字与表示字母 a 到 j 的码字是相同的。



因此，下面这组编码：



表示的就是数字 256。

如果你还要继续下去，那么我们还需要定义 7 个码字才算完全用满原有的 64 个码字。它们是：



第一个码字（4 号点凸起）用来表示音调符号（**accent indicator**）。其他的码字则用来表示缩写的前缀或者其他的功能：4 号和 6 号点凸起的码字（上表的第 5 个码字），根据上下文用来表示小数点或者强调符号（**emphasis indicator**）。5 号和 6 号点凸起的编码，表示与“数字标识符”相对应的“字母标识符”（**letter indicator**）。

最后（如果你现在还为布莱叶盲文如何表示大写字母忧心忡忡的话），我们还有 6 号凸点——大写标识符。这个符号表示它之后的字母将都被译为大写。例如，我们可以把这套文字系统的原创者的名字记做：

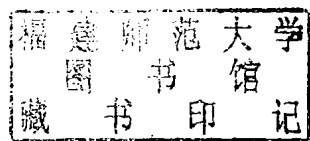


这串编码分别表示了一个大写字母标识符、字母 l、缩写字符串 ou、字母 i 和 s，以及一个空格，再加一个大写字母标识符，还有字母 b、r、a、i、l 和 e。（在实际应用中，

可能会把名字末尾不发音的 2 个字母省略掉，以进一步压缩编码)

综上所述，我们将 6 位二进制码（其实是 6 个点）所能表示的全部 64 种可能的编码都罗列了一遍。而且这 64 组编码中有很大部分，根据上下文的不同将有着双重身份。尤其值得注意的是数字标识符和取消“数字标识状态”的字母标识符。它们改变了后面编码的意义——从表示字母到表示数字，又从表示数字回到表示字母。像这样的编码通常被称作“优先码”（precedence codes）或者“换档码”（shift codes）。它们改变着作用域内编码的含义，直到作用域结束。

大写字母标识符表示紧随它的字母（而且仅仅是紧随它的字母）应该被译为大写。类似这样的编码被称为“逃逸码”（escape codes）。逃逸码让你“逃离”对编码串单调的、一成不变的解析，而转入一种新的解析方式中。在以后的章节中我们将看到，在使用二进制码对书面语言进行编码时，换档码和逃逸码是相当常见的。



4

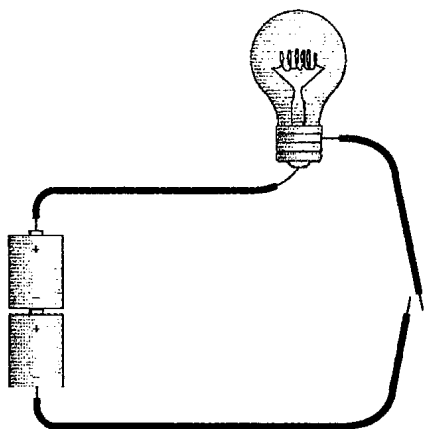
手电筒的剖析

手电筒的应用是极其广泛的，例如它可以帮助人们在背光的环境下阅读，也可以用于发送编码后的信息，这只是其中最显而易见的两个功能。普通的家用手电筒，还可以作为科普教育中的重要道具，引导人们走进神秘的电的世界。

电是一种神奇的现象。尽管电已经被普遍地应用到各个领域，但当人类自豪地宣称已经理解了电的工作机制时，真实情况是，关于电仍存在着大量的未解之谜。但是，在本书中恐怕我们无论如何都得仔细钻研一下电学了，幸好，我们只需要一点基本概念就可以理解电在计算机中的工作原理。

手电筒无疑是大多数家庭里最简单的家用电器。拆开一个标准的手电筒，你会发现里面有一对电池、一个灯泡、一个开关、一些金属片，还有一个可以容纳这些元件的塑料外壳。

你只需要电池和灯泡就可以自己来做一个简单的手电筒了。当然，你还需要一些短的绝缘导线（末端剥去绝缘皮），此外你还必须把这些元件连接起来，如下图所示。



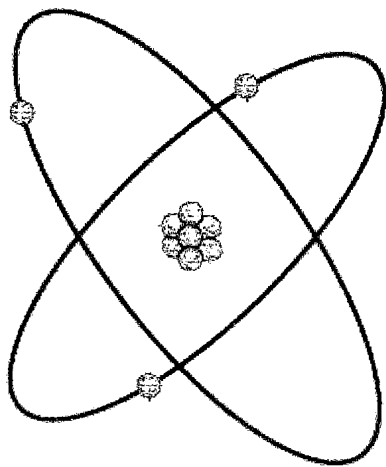
注意，在上图的右边，我们可以看到两个断开的线头，那就是开关。假设这两个电池没有问题，灯泡也没有烧坏，让这两个线头互相接触，就可以点亮灯泡。

刚刚完成的就是一个简单的电路。首先要注意的是，一个电路就是一个环路。只有从电池到导线，再到灯泡和开关，然后再通过导线回到电池的整个回路是连通的，灯泡才能被点亮。电路中出现任何的中断都将使灯泡无法正常发光。开关的作用就是控制这个断开的过程。

电路这种环状回路的特性说明电路中有某种东西在循环流动，或许有些像水流过水管那样。“水和水管”这个比喻经常被用来解释电流工作的原理，但是这种模型就如同所有其他的类比一样最终也无法自圆其说。在茫茫宇宙中，电不同于任何其他的事物，我们必须用它的术语来解释它。

研究电流工作原理的、最主流的科学理论叫做“电子理论”（electron theory），这套理论认为电流是由电子的运动而产生的。

众所周知，所有的物质——我们所能看到和感觉到的事物——（通常情况下）都是由叫做原子（atom）的极其微小的东西组成的。每一个原子又由三种粒子构成：它们分别是中子（neutron）、质子（proton）和电子（electron）。你可以把原子的结构画成一个小太阳系，其中中子和质子被束缚在原子核内，而电子则围绕着原子核旋转，犹如行星围绕太阳旋转一样。



我要提醒大家的是，如果你能搞到一个放大倍数很大的显微镜足以看清原子结构的时候，你会发现这张图与原子实际的样子并不一样。但是我们可以把它当做一个研究模型，方便我们进行一些探讨。

上图所表示的原子包含有 3 个电子、3 个质子和 4 个中子，这表明它是一个锂原子。锂是 112 种已知元素之一，这些元素都有一个 1 到 112 之间的特定原子序数（atomic number）。原子序数表明了这种元素一个原子的原子核中所含的质子数，（通常）同时也是一个原子所含的电子数。锂的原子序数是 3。

原子之间可以通过化学的方式结合形成分子（molecules）。分子的性质通常与组成它的原子大相径庭。例如，水是由水分子组成的，每个水分子由两个氢原子和一个氧原子（即 H_2O ）构成。很明显，水跟氢气或氧气都是截然不同的。同样，食盐的分子是由钠原子和氯原子构成的，显然这两种东西都没法让我们的炸薯条变得更加可口。

氢、氧、钠和氯都是元素。水和盐都是化合物（compound）。不过盐水是混合物（mixture）而不是化合物，因为水和盐都各自保留着它们自己的性质。

一个原子中电子的数目一般情况下与质子数目相同。但是在某些情况下，电子可能从原子中脱离。这就是电流产生的原因。

电子（electron）和电（electricity）这两个单词都是起源于古希腊文“ $\eta\lambda\epsilon\kappa\tau\rho\nu$ （elektron）”，关于这个词，你可能会猜测它是表示“微小而又无形的东西”之类的意思。

但是事实并非如此—— $\eta\lambda\epsilon\kappa\tau\rho\upsilon\nu$ 的意思是“琥珀 (amber)”，它是树的汁液硬化后变成的一种玻璃状固体。这两种风马牛不相及的东西会被联系到一起的原因源于古希腊人所进行的试验，他们曾经通过琥珀摩擦羊毛，产生了我们称之为“静电”的东西。用琥珀在羊毛上摩擦使得羊毛掠夺了琥珀的部分电子。结果羊毛会因为电子数超过了质子数而卷曲，而琥珀的电子数比质子少了。后来在现代的其他一些实验中，人们发现，毛毯也可以从我们的鞋底中带走电子。

质子和电子都具有带电荷 (charge) 的性质。质子有一个正电荷 (+)，电子有一个负电荷 (-)。中子是中性的，因而不带电荷。尽管我们使用了加号和减号来标识质子和电子，但是其实这些符号并不带有任何的数学方面的含义，也就是不能说明质子拥有电子所没有的性质。使用这些符号仅仅表示质子和电子在某个方面的性质是相反的。这个相反的特性也正表明了质子和电子是相互关联的。

当质子和电子在相同数目的条件下共存时，它们都处于最和谐、最稳定的状态。如果质子和电子之间出现失衡现象，它们就会试图进行自我修复。当地毯偷偷摸摸地从你的鞋子上挖走电子后，一切会在你触摸到一些东西而感到被电了一下时，又回归到平衡状态。静电火花是电子运动引起的，是电子通过一个回路——从地毯传到你的身体，再回到鞋子中的过程所造成的。

质子和电子之间的关系还可以这样描述，就是异性电荷相吸引，同性电荷相斥。不过关于这一点，我们仅仅靠观察原子结构图是观察不出来的。原子核中的质子被一种力量束缚到一起，这种引力要强过同性电荷之间的斥力，我们称之为“强力” (strong force)。强力有可能会引起原子核的分裂，而核能就是由此产生的。在本章中，我们仅仅讨论通过电子得失来获得电能的问题。

静电不仅仅存在于手触摸门把手时产生的小小火花中。在风暴中，底层的云积聚了大量电子而顶端的云失去电子；最后，一道闪电划破长空，使这一切又回到平衡。闪电是大量电子从一端快速地移动到另一端所形成的。

手电筒电路中的电流显然要比火花或者闪电更容易驾驭得多。灯泡之所以能稳定而持续地发光，是因为电子并不是简单地只是从一点跳到另一点。电路中，某原子所含有的一个电子逃逸到它相邻的下一个原子中，与此同时，这个原子又从相邻的上一个原子中获取一个电子，而失去电子的原子又会从与其相邻的一个原子获得电子，如此循环。

电路中的电子不断地从一个原子移动到下一个原子，就形成了电流。

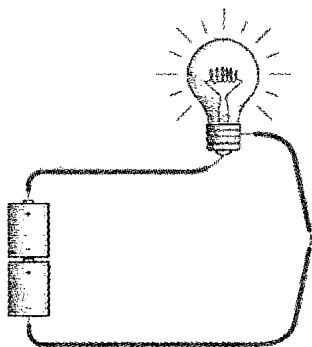
这一切都不是自发产生的。我们不能仅仅只把一堆破旧的电路材料随便连接起来后就开始祈求电流的出现。我们需要一种设备来促成电路中电子的流动。回头再分析一下前面所画的简单手电筒的线路图，我想我们能很有把握地断定，使电子流动的既不是导线，也不是灯泡，那么或许这个关键的设备就是电池。

关于手电筒中用到的电池类型，大部分人应该都有所了解。

- ✱ 它们是筒状的，并且有不同的大小，例如 D、C、A、AA 和 AAA 等型号。
- ✱ 无论电池的大小如何，都标有“1.5 伏”。
- ✱ 电池的一端是平的，标有一个负号（-）；另一端有一个小的凸起，并标有一个正号（+）。
- ✱ 如果你想电器正常工作的话，装电池的时候，就要注意让它的“+”端朝着正确的方向。
- ✱ 电池的电能终将被用尽。有些电池可以进行再次充电，但是有些却不可以。
- ✱ 最后，我们猜测，通过某种很神奇的方式，电池能够自己产生电能。

所有电池的內部都会发生化学反应，也就是说一些分子被分裂形成其他的分子，或者分子间互相结合形成了新分子。电池内的化学物质是经过研究精心选择的，它们之间的化学反应能够使多余的自由电子聚集到标负号“-”的那端（称为负极或者阴极），而在标有正号“+”的那端（称为正极或者阳极）则变得急需额外的电子。于是，化学能就被转化成了电能。

这种化学反应不会一直进行下去，除非我们有办法能使负极多余的自由电子回到正极去。因此如果我们没有给电池连上任何东西，什么事情也不会发生（实际上电池内仍有化学反应发生，但是非常缓慢）。电路可以将负极的电子带走，以补充正极缺失的电子，只有当存在电路的时候，化学反应才会发生。如下图所示，电子在电路中的这次旅行是沿逆时针方向的。

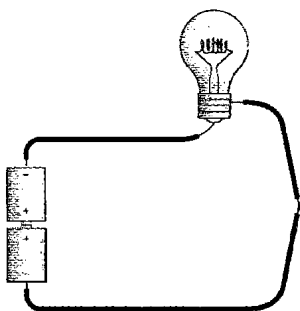


在本书中，红色¹用来表示在导线中有电流经过。

如果不是基于这样一个简单的事实——所有的电子，不管它们在哪里，都是完全一样的，或许电池内化学物质中的电子就无法如此自由地与铜导线中的电子混合在一起。铜的电子与其他物质的电子没有任何差别。

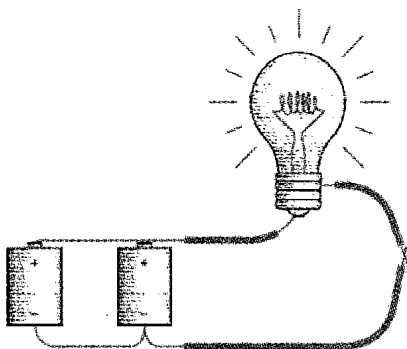
注意，所有的电池都是按相同方向放置的。下面的电池的正极，从上面电池的负极获得电子。就好像两个电池组合成了一块更大的电池，它的正负极分别在两块小电池的两端上。组合后的电池不再是 1.5 伏，而是 3 伏。

如果我们把电池组中的一块电池倒置，电路就不能工作了。



两个电池的正极都要为化学反应提供电子，但是由于这两个正极相互接触，电子根本无法通过某种途径到达它们。如果两个电池的正极连到了一起的话，它们的负极也应该连在一起，如下图所示。

¹ 由于黑白印刷的原因，在书中红色只能显示为淡灰色。——译者注



现在电路能够正常工作了。与前文我们提到的“串联”方式不同，这种连接电池的方式被称做“并联”。并联后的电压是 1.5 伏，与每个电池的电压是一样的。或许灯泡仍能发光，但是肯定没有串联电池的灯泡亮。不过电池的使用寿命也将延长一倍。

一般情况下，我们认为电池为电路提供了电能。但是我们也同样可以反过来想，电路为电池内的化学反应的进行提供了条件。电路把电子从电池的负极运走，然后转移到电池的正极。电池中的化学反应将持续进行，直到所有的化学物质被消耗完，那时你就可以把电池扔掉或者重新给它充电了。

从电池的负极到正极，电子流经了导线和灯泡。但是为什么我们需要导线呢？电流不可以直接被空气传导吗？哦，可以说能，也可以说不能。电流可以通过空气传导（特别是潮湿的空气），不然的话我们就看不到闪电了。但是电流也不能轻易地穿过空气。

一些物质在导电能力上要比其他一些物质明显更好一些。元素的导电能力与它原子内的结构有关。电子围绕原子核运行的轨道分为不同等级，我们称其为“电子层”。如果原子在最外电子层中只含有 1 个电子，那么这个电子很容易逃逸，这就是易导电物质所需具备的特性。这些物质对于电流来说是“导通”的，因而被称做导体（conductor）。最好的导体是铜、银和金。这三种元素位于元素周期表的同一列不是巧合。其中铜是用来制作导线的最常见的原料。

与导电性相反的是阻抗性。有一些物质与其他的物质相比更不容易让电流通过，我们称其为电阻。如果一种物质有着很强的阻抗性——也就是说它几乎不能传导任何电流——它就被称为绝缘体（insulator）。橡胶和塑料都是很好的绝缘体，因而它们经常被用来包裹金属导线。布料和木头在干燥的空气中也是很好的绝缘体。不过事实上只要有足够高