XJTU-IE, 2007.9-2008.9, M: xuchen.xuchen@163.com(徐琛), lucifer_tcl@126.com(李慰祖)

第十八章 产能管理

你不总能得到你想要的，
对，你不总能得到你想要的。
不过如果你有时试试看，你会发现
你得到了你需要的。
——滚石乐队

18.1 产能设定的问题

企业应该具备多少以及何种产能的选择对企业的净利润有直接而巨大的影响。而且，由于**产能计划（capacity planning）**位于设备计划层级的顶端（见图 13.2），产能决策对于其他所有的生产计划问题（如，集结计划、需求管理、排序与排程、车间作业控制）有着巨大影响。在这一章我们将利用工厂物理的概念将战略产能计划转化为详细的战术规则。我们的目的是提供一个产能计划的框架，它将详细地识别产能计划对于整个工厂管理过程的影响。

18.1.1 短期和长期产能设定

在制造业工厂的生命周期里，合理的产能调整会发生许多次。大多数情况下，改革的动机是为了迎合需求总量或者产品组合的变化。短期内，工厂可以利用加班、增加或者减少班次、外包和劳动力数量的变化来适应需求的改变。这些策略在第十六章关于集结计划的内容时已经讨论过了；它们在产能计划中也同样是明显的选择。

一些短期选择在长期内也是可能实施的。比如，我们可以在非永久性的基础上实施三班制或者外包全部或部分生产。当然，如果我们长期把生产外包出去，那么外包商最后就可能决定直接卖掉产品继而成为竞争者。幸好，这样的事情常常被进入壁垒所防止，比如，知名商标的使用权或者一个有效运输/服务网络的所有权这样的非制造业的因素，而这些是很关键的。（626|627）即使最终的竞争并不会造成严重的风险，如我们在第十二章讨论的，依赖外包商来制造部件或者产品会使他们成为质量管理过程中的一个重要的合作者。如果缺乏手段来保证外包商的质量，外包生产的决策就会严重影响企业控制自己命运的能力。

长期内，我们必须超越短期的选择并且考虑永久性的设备，或者结构上的变革。这包括现有工厂的重大变革与新工厂的建设。在某些情况下，企业可以通过使用面向制造的设计（DFM）方法重新设计产品以永久性地增加产能（见 Turino 1992，第七章的讨论）。然而，更常见地，这些变化只能来自新增的机器或工站，或者对于现有的设备和工艺流程的生产力的永久变革。

18.1.2 战略产能计划

在企业决定应该具备多少以及什么类型的产能之前，它应该清楚地阐述产能战略。这些战略依赖于一些和企业的核心经营计划接近的决策。比如，它可能需要决定是否要进入一个新市场，是否要继续留在原有市场，应该引领还是追随工艺改革，自制还是外包，应该追寻怎样的细分市场以及其他的一些问题。总而言之，这些问题等同于那些超出了工厂物理学范畴的类似于“我们经营什么”的战略基础问题。物理定律可以告诉我们一个特定的物理系统

会怎样运行，但不能告诉我们该对什么样的系统感兴趣。类似地，制造业的定律可以帮助我们为达到特定的目的而设计系统，但不能告诉我们这个目的应该是什么。因此，针对我们讨论的目的，我们将假设上述的战略决策已经被制定出来了，于是问题就变成了怎样制定产能计划来支持它们。

一旦我们决定了我们需要增加产能，那么有几个问题就应该提出来了：

1. *增加多少产能以及何时增加？*是不是只有在需求已经开发出来（当我们已经损失了一定销售额）的时候才增加，还是面向预期的未来需求？如果我们不预测需求，我们是否应该用短期的行为比如加班或者外包来满足超量的需求？如果我们决定预测需求，那么我们应该尽力满足未来多久的需求呢？产能的大幅度增加可以满足未来更久的需求，带来更少的结构分裂，实现规模经济。但是，产能大幅度的增加也意味着更低的设备利用率和面临更大的风险（如果预测的需求不出现将会怎样？）。方式是否合适也取决于所使用的生产技术。比如，炼钢厂只能以增加新的熔炉和轧钢机的方式来大规模增加产能，而金属加工厂就可以通过单个地增加机器来少量地增加产能。可参考 Freidenfelds（1981）有关此问题的分析。

2. *应该增加何种类型的产能？*我们增加产能的量也依赖于所选择设备的柔性。如果现在购买的机器可以适应将来引入市场的产品生产的需要，增加多于现在需求的产能的风险就会降低很多。在现在产品迅速变化的环境下，产品的生命周期常常比生产设备要短；因此，这样的柔性已经成为选择新产能的关键因素。（627|628）可参考 Sethi 和 Sethi（1990）关于制造系统中不同类型的柔性的概述。

3. *应该在哪里增加产能？*我们是不是应该通过扩大整个现有工厂来增加产能，还是应该新建一个？虽然新建一个工厂比扩大现有工厂要贵很多，但新的工厂常常能满足新市场和运输效率，比如离供应商或者消费者更近。可参考 Daskin（1995）关于选址问题的模型。

一个为大家所熟知的重要战略观念是**产量规模经济（production economics of scale）**。基本的思想是大工厂的单位成本通常（但不总是）比小工厂的少。Hayes 和 Wheelwright（1984）讨论了三种不同的规模经济：短期、中期和长期。

带来**短期规模经济（Short-term economics of scale）**的因素都是近期的，很多制造成本是固定的。虽然在较长期内有一定的可调节性，但生产设备、劳动力、管理、安全成本，资产税等等在任何确定的时刻都是*固定的*。这些成本不因产量的变化而改变。事实上，在近期，唯一真正的可变成本是原材料，一些公共物品，以及机器的磨损。我们可以这样表述单位成本

$$\text{单位成本} = \frac{\text{固定成本} + \text{可变成本}}{\text{产量}} = \frac{\text{固定成本}}{\text{产量}} + \text{单位变动成本}$$

所以，在短期内，产出增加时单位成本降低。

中期规模经济（Intermediate-term economics of scale）依赖于在生产中使用的生产周期——在工厂转换到另一种产品之前生产出的产量。已知转换成本和每个特定产品的生产周期，单位成本可以这样表述

$$\text{单位成本} = \frac{\text{转换成本}}{\text{运行单位}} + \text{单位运行成本}$$

在这样的情况下，劳动力是固定或不固定的。生产周期会受到更低频率的新建（通过减少设备的建造），提供设备（这样一些产品系列就可以在不转换生产的情况下持续运行），使用特殊设备（柔性制造系统）的影响。当然，其中的一些选择将导致更大的投资，如我们在第二篇所讨论的那样。

长期规模经济（Long-term economics of scale）是设备自身的功能带来的。经济学家很

早就注意到设备的成本倾向于和它的表面积成正比，而产能更倾向于和它的数量成正比。为了解释这个推断，假设设备是一个边长为 l 的立方体。那么我们就可以把成本这样表示：

$$K = a_1 l^2$$

产能则是这样：

$$C = a_2 l^3$$

这里的 a_1 和 a_2 是比例常量。为了把成本表达成产能的函数，我们用 C 来表示 l，得出 $l = a_3 C^{1/3}$，这里 a_3 表示另一个常量；然后我们把它带入成本的表达式。这样有

$$K(C) = aC^{2/3}$$

这里，再一次地，a 是一个比例常量。（628|629）

对于大多数（非立方体的）设备，用产能的函数表示的成本大致是这样：

$$K(C) = aC^{b}$$

其中 b 应该在 0.6 到 1 之间。

我们现在可以把单位成本表述成这样：

$$单位成本 = \frac{K(C)}{C} = aC^{b-1}$$

因为 b 常常是小于 1 的，这表明单位成本会随着产能的提高而降低。也就是说，大的工厂比小的工厂更有效率。

实践中，规模经济常常用较大的工厂来实现较低的成本，但也不总是这样。当企业变得更大的时候，也可能会出现**规模不经济（diseconomies of scale）**，使得组织失去效率。一个可能出现这样情况的地方是产品递送。一个小的密集单元中的物料搬运少于一个大的由多个加工中心组成的工厂。虽然大工厂中的加工中心会比组成单元的单个工站效率高得多，但工件也需要移动更大的距离。这就增加了物料搬运和周期时间。而且由于大的制造工厂通常比小的工厂要求更大的区域，这样搬运成本通常也会更高。对于大体积的产品如砖块，最有利可图的工厂规模应该很小。

另一个导致规模不经济的原因在于官僚体制。随着经营规模的扩大，就需要更多的监控和支持。为了保证控制幅度的可管理性，大企业增加了管理层级，这就进一步减弱了交流的效果。这就带来了部门之间的隔阂和矛盾。如果不小心管理，这些不经济可能是致命的。

最后，大的工厂自然会带来更大的风险。自然灾害如地震、火灾、洪灾和飓风对于那些只运营一个大工厂的企业的负面作用要比那些依赖很多小的独立的工厂的企业明显得多。类似的，失败的管理、罢工等等诸如此类的事情对于产能集中的企业的损坏远远大于产能分散的企业。

在这一段很自然的提出了一个问题：那么怎样工厂规模是最优的？这是一个大的战略问题，超出了我们这本书的范围。而且，由于这牵涉到了很多企业自身特性的问题，一个通行的答案是不可能存在的。以上的讨论对于我们要研究的问题给出了一个初步的印象。更多详细的情况可以参考制造业战略文化（见 Hayes 和 Wheelwright 1984，Schmenner 1993）

为了把注意力集中在工厂管理上，我们假设工厂的大小已经在战略考量的基础上确定了。因此，我们将考虑关于在工厂内如何改变产能来完成一组特定的目标的问题。特别地，我们考察了两种情形：新建一个工厂和改造一个现有的工厂。

18.1.3 关于产能管理的传统和现代观点

为了把产能计划问题限定在工厂的级别，很必要区分在**传统**（**tradition**）和**现代**（**modern**）观点中产能所扮演的角色（Suri 和 Trville 1993）。传统观点基于对制造有效性的阐述，如图 18.1 的左边部分所示。这里，唯一的问题是是否拥有足够的产能来满足特定的产出目标，而答案非是即否。如果利用率低于产能，那么生产就是可行的；否则就是不可行的。

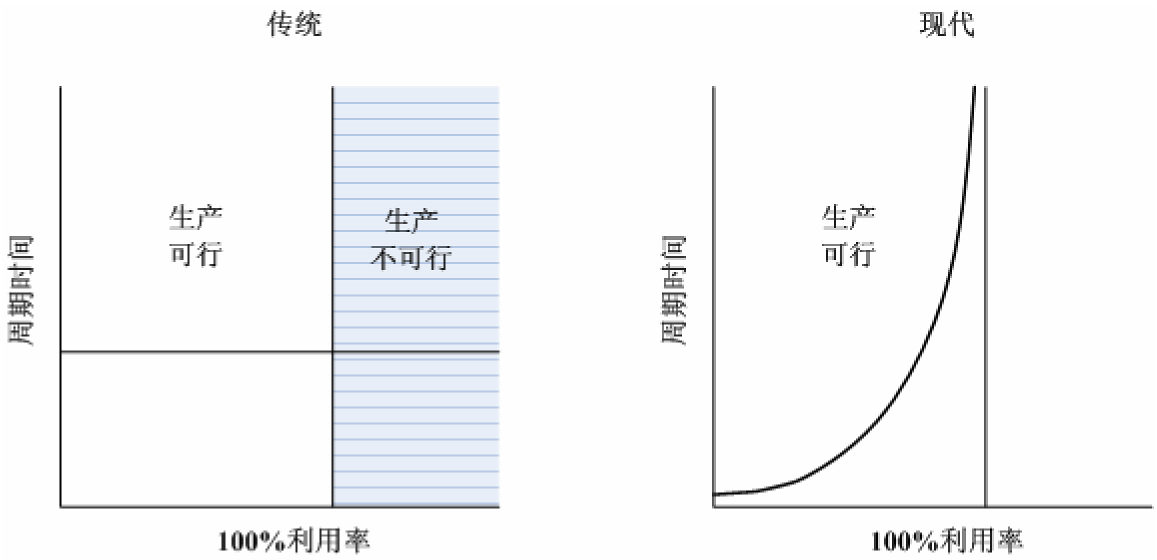

图 18.1 产能计划的传统与现代观点

现代的观点更加实际，并且符合工厂物理学的原理，它认为提前期和在制品水平会随着利用率的增加而持续增加；这可以参考图 18.1 的右边部分。按照这个观点，没有一个点上的生产力是不可行的。然而，响应度会随着产能利用率的提高而持续降低。

这两种观点揭示了产线设计的两种非常不同的方法。传统的观点建议在可能的最低成本下，选择一整套有足够生产力的设备。但是这样的做法常常在产线投入生产后带来很多问题。我们遇到过许多这样的企业，他们拥有一条产线的机器，每一台机器的设计产能都高于所需产能，但合在一起就远远达不到它们的产能目标。（已经在第二篇了解了相关工厂物理原理的读者应该对为什么这些产线完不成目标产量有清晰的认识）（629|630）

现代观点对产能问题给出了更加丰富的解释。因为产能不再是一个简单的是非问题，我们必须考虑到成本和产出以外的其他指标。在制品、平均周期时间、周期时间方差和品质都会受到产能决策的影响。如果我们能用这些量度来表示我们的目标，那么我们就能简单地把产能计划问题列成式子（然而，解式子则是完全不同的事），如下：

对于给定的预算，设计尽可能“最好”的工厂。

由于我们通常有一个以上的目标，难以定义“最好”，所以这个式子是含糊不清的。比如，一条低产出、短周期时间的产线与一条高产出、长周期时间的产线相比是好些还是差些呢？像我们在第六章所讨论的那样，我们采用**满意**（**satisficing**）技术逼近来处理多目标问题，也就是说选择一个量度作为目标，把其他的固定为约束。用这样的方式，问题可以分解为一个拥有一个或者多个**战术**（**tactical**）问题的**战略**（**strategic**）问题。这个战略问题可能

是应该具备多大的产能，多长的周期时间，什么类型的产能，以及多大的产量等等问题。战术问题则是最小化成本或是其他服从战略问题所决定的约束的数值。这种高层级的问题的解决给低层级的问题提供约束的方法，在第六章已经讨论过了。

在预算约束，也许还有在制品和周期时间的约束之下，一个表达式会使产量最大化。另一个表达式会使周期时间在预算和产量约束下最小化。（630|631）还有一个会使在产出、周期时间和在制品约束下达到成本最小化。哪一个最好取决于具体环境。一方面，如果我们打算改进一条现有的产线，而预算是给定的，那么优化某个受到预算约束的指标的式子（最大化产量或者最小化周期时间）就能够完美地解决问题。另一方面，如果我们正在设计一条新的产线以达到给定的绩效要求，在像产量和周期时间这样的约束下最小化成本就很合适。

不管我们选取的是哪个式子，我们都可以使用结果模型来检验重要的取舍。比如我们选择了一个在产出和周期时间约束下最小化成本的模型，我们可以改变产量和周期时间约束的水平来看看成本是怎么变的。结果将是产出与成本以及周期时间与成本的曲线，它们将在决定我们的初步战略特征是否合理时都很有用。

除了专注于产能决策的优化（*optimality*），还必须对它们的稳健性（*robustness*）行进敏感性分析。我们今天提出的需求可能与将来的需求有很大不同。有时，投入比计划多一点的金钱（用在像产能缓冲，或者购买贵一点但更有柔性的系统上）以满足将来的偶然需求会是个好主意。我们可以考查模型中不同需求的状况来考虑这些选择。但是，我们需要注意不要超过稳健性的界限。晶圆设备之所以异常地贵，一个原因就是因为它被设计来希望能够制造在近未来可能需求的任何产品。由于在半导体工厂中技术的不确定性极高，这要求我们必须安装非常前沿（或“刀锋（Bleeding-edge）”）的设备。

在这章的剩余内容里，我们将关注在安装或改造一条产线时，在各种绩效约束下的成本最小化的问题。我们选择这个特别的式子是由于以下的原因：（1）这对设计新产线而言是一个最自然的框架，以及（2）适合于产生一个成本与绩效的的曲线。不过，你可以轻松地使用这里提到的工具和技术来分析其他的式子（在产出和成本的约束下使周期时间最小化）。

18.2 建模与分析

在整本书中我们都严重地依赖模型，这主要是因为模型强迫我们认真考虑所研究的系统并且帮助我们发展关于系统怎样运行的直觉认识。但是在实践层面上，要是没有一些模型，无论是清晰的还是模糊的，就根本不能分析了。会计、营销、财务，质量控制以及事实上所有其他的经济功能都依赖于模型来解析数据、预测绩效和评估行为。令人高兴的是，我们用来解决产能计划问题的模型和我们在第二篇用来解释工厂物理概念的模型大致是相同的。特别地，我们将把排队网络作为制造产线中的代表来开发产能分析的工具。虽然我们要坚持第二篇介绍的基础公式，但这些工具都是很有学问的，有兴趣的读者可以参考 Buzacott 和 Shanthikumar（1993），Suri 等（1993），以及 Whitt（1983、1993）以获得更多的细节。

为了说清楚，我们将分析集中在一个单一的产线并且把工厂的其他部分当作是固定的。我们假设产线由 M 个工站组成并且“制造订单”已经给定——也就是说，每个工站制造部件或者产品所需的运行是预先设定的。（631|632）我们这里只考虑产线生产单一产品的情况，尽管我们可以用将工站上不同产品不同加工时间的变动性分散为加工中心的自然变动性（将有效加工时间的变异系数扩大）。我们给每个工站编号为：$1, 2, ..., M$，工件首先到达 1 号工站，并传递给 2 号工站，然后是 3 号，诸如此类。在这个讨论中，我们没有涉及到重工和分支路径，尽管它们可以通过使用更精巧版本的排队网络模型（见 Suri 1993） 得到解决。

对于每一个工站都有一组不同的**技术设置（technology options）**供我们选择，包括特定

的生产准备和/或运行策略。这些设置可能包括从不同的设备供应商那里获得的不同的机器模型。它们可能包括一些有可替换部件和一些没有的机器；而那些拥有可替换部件的机器与那些没有可替换部件的机器相比较，修理时间更短但是成本更高。要注意，这一定义使得识别一套合适的技术设置不仅仅是从设备供应商那里收集数据。我们必须使用第二篇的有关工厂物理的直觉来认清这些像可替换部件一样吸引人的设置。我们假设可以计算出一套合理的技术设置，并且每套设置的成本，产能以及变异系数都能估计出来。

为了保持技术设置的数量以及分析的可控性，我们假设在多机器的工站中不允许不同类型的机器混合。也就是说，如果一条产线需要三台车床并且我们选择了 South Bend X-14 作为我们的模型，我们就得用三台 South Bend X-14。我们不能用两台 South Bend X-14 和一台 Peoria P10000。这个限制在一条新的产线中似乎自然就满足了，因为我们不大可能在可以只对付一个设备供应商的时候对付两个。在设备改进时，这可能不能完全被满足，但是从建模的角度看一般不是大问题。

每个工站的设置是用以下的五个参数来描述的：

t_e = 机器的平均有效加工时间，包括断供、生产准备、重工以及其他日常扰动

c_e = 机器的有效变异系数，也考虑了断供、生产准备、重工以及其他日常扰动

m = 工站中（相同）机器的数量

k = 每台机器的成本

A = 生产准备的固定成本

每套设置的安装成本总额用 $A + km$ 的形式给出，也就是说，如果安装一台机器要花 \$75,000 而安装两台要花\$125,000，那么 A = \$25,000，k = \$50,000。这里的意思是允许我们表示那些无论安装多少台机器都只需进行一次的活动的成本，比如改造电力或者通风系统，或者加固地板。

我们从第八章的更加基础的参数开始描述如何计算 t_e 和 c_e^2。这里我们假设对于每套设置，这些都计算过了。不过，检验这些更加基础的参数（MTTR、MTTF 等）来建议其他技术设置可能很有用。

为了清晰地表达模型中的约束，我们假设考虑整条产线绩效的战略决策已经确定了，也就确定了下面这些：

TH = 所需产出

CT = 最大量的总周期时间

接着，使用上面的参数和产线的过程到达时间的描述，我们为产线上的每个工站计算以下的参数。

$u(m)$ = 拥有 m 台机器的工站的利用率

$CT(m)$ = 拥有 m 台机器的工站的周期时间 （632|633）

c_a = 到达工站的变异系数

c_d = 离开工站间的变异系数

计算 u 和周期时间的公式与第二篇中的相似，可以表示成这样

$$u(m)=\frac{r_a t_e}{m} \quad （18.1）$$

$$CT(m)=\left(\frac{c_a^2+c_e^2}{2}\right)\left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)}\right)t_e+t_e \quad （18.2）$$

到达时间的变异系数的平方（SCV） c_a^2 被规定为 1 号工站的一个参数，它在后继的工站要与前一个工站的离开的 SCV 相同。也就是说让 $c_a^i(i)$ 和 $c_e^i(i)$ 分别表示在工站 i 的到达时间的 SCV 以及有效加工时间。我们就得到了以下的公式：

$$c_a^i(i)=\begin{cases}\text{某个常量} & i=1\\ c_d^i(i-1) & i>1\end{cases} \quad （18.3）$$

这里 i = 1, …, M，

$$c_d^2(i)=1+[c_a^2(i-1)-1][1-u^2(m)]+\frac{u^2(m)}{\sqrt{m}}[c_e^2(i)-1] \quad （18.4）$$

对于一套给定的机器配置（即，每个工站技术设置的选择），我们用（18.2）式来计算 $CT(m)$ 并检验总周期时间的约束。如果超出了约束，我们就要考虑更多的产能或者更低的可变设置。技巧是在成本最有效的地方改变配置。

不过，在这些问题可以被解决之前，我们必须有一个拥有足够产能的起点。我们称之为**产能可行（capacity-feasible）**解，下面我们将给出一个如何找到它的范例。

18.2.1 范例：成本最小的产能可行产线

考虑一个拥有四个工站的产线，它的目标产出为每小时 2.5 个工件或一天 60 个工件（实行每天三班轮流制）。假设到达时间的 SCV 等于 1.0（回想在第二篇我们将它称作中度变动性的状况）。所以，对于第一个工站，TH = 2.5 个工件每小时，c_a^2 = 1.0。设定产线的目标周期时间 CT = 16。开始时，假设每个工站只有一种类型的机器可用（虽然我们允许在每个工站安装不同的机器）。表 18.1 给出了四个工站的数据。

首先，我们演示一个产能检查来决定在每个工站我们需要机器的最小数目。我们解（18.1）式来得到能保证利用率低于 1 的 m 的最小值，也就是

$$u(m)=\frac{r_a t_e}{m} \qquad u<1$$

或者 $m>r_a t_e$。（633|634）

表 18.1 一个产线设计问题的基本数据

工站	固定成本（$000）	单位成本（$000）	t_e（小时）	σ_e^2
1	225	100	1.50	1.00
2	150	155	0.78	1.00
3	200	90	1.10	3.14
4	250	130	1.60	0.10

对于工站 1

$$r_a t_e = 2.5 \text{ 件/小时} \times 1.5 \text{ 小时} = 3.75$$

这意味着我们至少需要 4 台机器。表 18.2 总结了其他的机器需求以及相应的利用率。

表 18.2 成本最小的产能可行解

工站	机器数量	利用率	成本（$000)
1	4	0.94	625
2	2	0.98	460
3	3	0.92	470
4	5	0.80	900
总计			2,455

注意对于工站 4

$$r_a t_e = 2.5 \text{ 件/小时} \times 1.6 \text{ 小时} = 4.00$$

然而，这会导致一个几乎是 1.0 的利用率。由于工厂物理学的利用率法则规定利用率必须总是严格低于 1.0。我们必须为工站 4 安装 5 台机器，把利用率降低到 0.8。

注意表 18.2 的解是拥有足够产能的成本最低的配置。这就叫做**成本最小的产能可行**（**minimum cost, capacity-feasible，MCCF**）配置，在这个例子里面成本为$2,455,000。

当每个工站有不止一个技术设置时，很容易扩展这个分析来找到 MCCF 配置。对于每个工站，我们决定每种设置需要多少台机器来达到产能目标并选择总成本最小的那种设置。对每个工站进行这样的处理，将得到一条 MCCF 配置的产线。

18.2.2 强迫周期时间服从约束

一旦我们有了一个产能可行的配置，我们就可以利用（18.2）和（18.4）式来检查周期时间。（634|635）

工站 1：

$$CT(4) = \left(\frac{1.0+1.0}{2}\right)\left(\frac{0.94^{\sqrt{2(4+1)}-1}}{4(1-0.94)}\right)1.5 + 1.5 = 6.72 \text{ 小时}$$

$$c_d^2 = 1 + (1-1)(1-0.94^2) + \frac{0.94^2}{\sqrt{4}}(1-1) = 1.0$$

工站 2：

$$CT(2)=\left(\frac{1.0+1.0}{2}\right)\left(\frac{0.98^{\sqrt{2(2+1)}-1}}{2(1-0.98)}\right)0.78+0.78 = 15.82 \text{ 小时}$$

$$c_d^2=1+(1-1)(1-0.98^2)+\frac{0.98^2}{\sqrt{2}}(1-1) = 1.0$$

工站 3：

$$CT(3)=\left(\frac{1.0+3.14}{2}\right)\left(\frac{0.92^{\sqrt{2(3+1)}-1}}{3(1-0.92)}\right)1.1+1.1 = 8.87 \text{ 小时}$$

$$c_d^2=1+(1-1)(1-0.92^2)+\frac{0.92^2}{\sqrt{3}}(1-1) = 2.0$$

工站 4：

$$CT(5)=\left(\frac{2.0+0.1}{2}\right)\left(\frac{0.80^{\sqrt{2(5+1)}-1}}{5(1-0.80)}\right)1.6+1.6 = 2.59 \text{ 小时}$$

这些周期时间之和为 34 小时，它远远大于目标的 16 小时。很明显，产线需要改进设计以使其符合战略特征。

这里有三个基本的改进选择：（1）改造现有的机器，（2）改变机器的设置，或者（3）增加更多的机器。第九章描述了怎样使用工厂物理学的原理来诊断产线中的问题。这个方法同样可以发现周期时间过长的原因（如，一次时间长而发生频率低的停电事故）以及由此导致的怎样的机器改进会是最有效的。减少变动性或者加快机器的运行速度比购买一台新机器更划算。当然，如果我们要设计一条新的产线，那就没有“现存”的设施，因此要改变一个是不可能的。

为了追求缩短周期时间的目标而改变生产准备可能需要购买不同的，或许更贵而具有更好运行特性的机器（如，更高的效率或更小的流程变动性）。然而，经常地，特别是在高科技时代，不同的机器类型相当有限。在某些情况下可能只有唯一一个供应商能提供可用设备。在这种情况下，大多数能用于缩短周期时间的技术设置是一个给定机器类型上的改变。改变包括给机器提速，减少设备调整时间，减少 MTTR 等等。

压缩过长的周期时间的最明显方法就是单纯地购买更多的机器。如果产能只是稍稍增加，这可能就是最经济的方法。

依照需要降低的周期时间的水平、可行技术设置的范围，以及产能增量的成本和规模，最优的方法可能包含任何数目的这些类型的改变的组合。（635|636）

18.3 改造现有产线

我们现在提供一个探寻式方法来决定一个满足产出和周期时间的约束的成本最小的配置。探寻式方法始于 MCCF 配置，并在改进周期时间方面寻求最能够把钱“用在刀刃上”的改变。

为了解释这个方法，我们再次考虑表 18.1 的例子。回忆成本最小的产能可行配置（表

18.2）并不满足周期时间约束。尤其是所需的周期时间为 16 小时，而成本最小配置的总周期时间结果是 34 小时。我们现在考虑怎样以一种高性价比的方式把配置变得服从周期时间约束。注意这正是公司所面临的问题：努力在现有的工厂中贯彻压缩周期时间或基于时间的竞争的方法。

为了使这个例子更加真实，假设我们可以以在每个工站增加机器的方式进行改造。个别地看，假设通过在第三个工站的每台机器上花费 10,000，我们可以将时间长、频率低且随机发生的断供转化为总可用量相等然而时间短、频率高的断供（回忆我们在第八章的讨论，说明了为什么这是值得的）。我们可以利用安装可替换的部件和/或更多的预防性维护措施来达到这一目的。我们假设这样做不会改变 t_e，但是会将 c_e^2 由 3.4 减少到 1.0。利用这些成本和绩效数据，我们可以考虑用这种降低变动性的设置达到和增加机器一样的效果。

因此，这些都是可行的选择：在任何一个工站，我们可以增加一台机器；在工站 3，我们可以增加一台机器，或者改变机器的特性以降低机器变动性。对于每一个选择，我们可以计算工站中周期时间的变化以及成本的变化。[1]一个合理衡量变化的有效性的量度应当是周期时间改变量与成本改变量的比率。“最好的改变”应该是比率最低的。我们在表 18.3 中计算了每套设置中的这些比率。

表 18.3 改善选项的成本与周期时间影响

工站	现有机器数量	改变	**成本增加（\$000）**	**CT下降（小时）**	**比率（\$000/小时）**
1	4	增添机器	100	4.63	21.61
2	2	增添机器	155	14.73	10.52
3	3	增添机器	90	7.20	12.49
3	3	削减变动性	30	4.49	6.67
4	5	增添机器	130	0.71	183.10

在表 18.3 中我们注意到的第一件事情是，没有任何单个的改变可以降低总周期时间到满足周期时间约束的程度——我们需要降低 18 个小时。最小的比率是工站 3 通过改造机器得到的（通过降低修理时间的变动性），花费$30,000 使周期时间降低了 4.49 小时。这使得周期时间降低到了 29.51 小时，仍然比规定的 16 小时要长很多。如果我们重复这样的分析，最低的比率将通过在工站 2 增加一台机器得到，这花费了$155,000 并且把周期时间进一步降低了 14.7 个小时。我们这时把周期时间降低到了 14.81 小时，这就满足了 16 小时的约束。

虽然我们不能保证这样重复利用单个改变能够用最小的成本将周期时间降到约束之内，但这个方法通常很好用。在任何情况下，它都会产生一个产出和周期时间可行的配置。对于这个例子，表 18.4 中给出了最终的解决方案。

[1] 我们忽略了在这个点的下游可能发生的事情，所以我们的计算结果事实上是整条产线循环时间的改变量的近似值。回去检验某一特定设置的产线循环时间是很容易的，在估算每一单个改进的效果时包含在下游的效果并不很困难。然而，如果我们这样做，我们一次只能估算一个改进——两个设置并行导致的循环时间的降低并非必然是每个单独设置的循环时间减少量的和。

表 18.4 产能与周期时间可行的配置

工站	机器数量	利用率	成本（$000）
1	4	0.94	625
2	3	0.65	615
3	3（改变后）	0.92	500
4	5	0.80	900
总计			2,640

总成本为$2,640,000，比 MCCF 配置多出$185,000。不但如此，注意这条产线远远没有达到平衡。令人惊奇的是，最昂贵的工站（4）利用率的最低。这是因为工站 4 固定成本和单位成本很高，而且四台机器的话就会造成 100%的利用率。（636|637）

18.4 设计新产线

设计新产线的问题与改造现有产线区别很大，这里会有多得多的典型设置需要考虑。在一条新产线中，我们不会受到现有的机器、设施，甚至结构的约束。确实，我们拥有了这么多的自由，以至于要以最优的方式解决这个问题变得几乎不可能。

18.4.1 传统方法

在十八世纪，当设计第一个工厂时，主要的考虑是如何安排这些装置才能利用能量的单一来源——水轮来推动它们。于是，装置被安排为一个沿水车轴的线性布置，每台都与合适尺寸的齿轮上的传送带相联接以从水轮获得所需的旋转速度。今天，很难找到依照传统设计安排的工厂，它们的加工中心在矩形设施中布置成一条直线。

我们很惊奇地发现制造业已经有 150 年不利用水能了，我们还询问了几个设计复杂工厂（如，晶圆厂）的建筑工程师和在工厂工作的制造工程师。我们发现设计新工厂和新产线时有如下典型的过程：

1. 确定新设施的基本规模和形状。
2. 决定支持设施（电、集汽联箱、天然气等等）的走向以最小化设施的成本。（637|638）
3. 决定工站在工厂中的位置以最小化成本。
4. 决定产品流程。

知道了这些，那种线性布置的倾向就不足为奇了。由于设计的过程始于设施的规模和形状，传统的设计思想对设计结果发挥了巨大的影响。但是这个方案中存在明显的问题。最严重的一个是，几乎在整个工厂都设计好了之后，产品的流程才被给予了很少的关注。

18.4.2 工厂物理学方法

一个好的替换方法是从顾客的角度来看待这个问题。这使一条产线或一座工厂存在的主要意义变得很清楚：以及时而有竞争力的方式提供高质量的产品。包含这一目的的设施设计过程与传统方法几乎是相反的：

1. 顾客决定产品。产品组合、批量和周期时间是事先设定的。

2. 产品决定工艺。对于大多数产品，要生产一个单位都会有一些基本的步骤要做。

3. 工艺决定一套基础的机器。对机器的描述会在对计划工艺的细化过程中从很概括的程度开始，逐渐变得细节丰满。

4. 机器决定用于支持它们的基础设施。

5. 基础设施决定了整体的架构和规模。

当然，如果我们逐字逐句地按照上面的程序做，我们会得到一个装备得很好的工厂，它能够以所需的批量生产产品，但建造成本太昂贵。单纯地关注产品流程以最小化周期时间，可能导致我们在应该安装机器时安装多功能的昂贵的机器。比如，在一个晶圆厂，光刻设备就是典型的较昂贵的机器之一。它对设施的需求巨大，并且让问题更加糟糕的是，晶片必须遍历加工中应用的设备的每一层（常常有十层或更多）。纯粹的最小化周期时间的视角也许要求以巨大的成本为代价而安装 10 台设备。纯粹的最小化成本的角度就只要求一台设备。“最好的”设置只能通过考虑在其他机器影响下的光刻设备和比较满足绩效目标的不同配置的相对成本之后才能决定。

结果是，综合传统的和工厂物理学的观点来解决工厂设计问题会很有意义。我们从一个关于工厂的基本流程和布局的看法开始。用最基本的布局，我们安排加工中心，并设定它们的规模来满足所需的产出和周期时间水平。如果最终的配置导致成本过高，我们就再次考虑基本布局。另一方面，如果周期时间太长，我们就考虑安装更多的支持设施来改进流程。

作为分析的一部分，我们可能也想对产品组合做一个帕累托分析来决定“工厂中的工厂”这一概念是否适用。如果占批量中大多数的是数量相对比较少的产品，在工厂中设置两条流程就很有意义。工作流较紧张的一条产线配置用于生产数量较少但代表了产出中较大部分的产品。另一条更多地安排成工作车间的配置，以较低的利用率或较高的周期时间为代价，使得柔性达到最大。因为批量（被设计得）很低，这一部分的利用率很低就是理所当然的。（638|639）

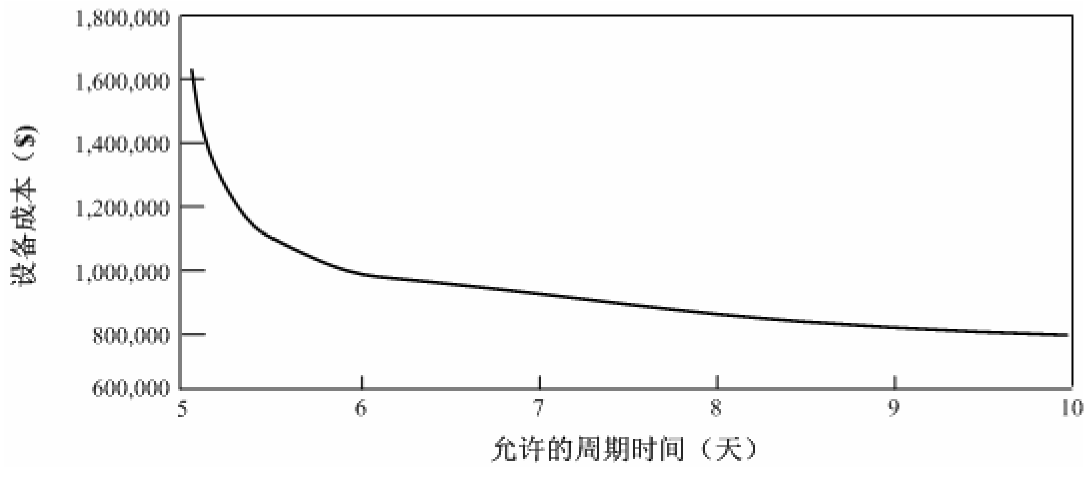

图 18.2 总的设备成本与总的周期时间

一旦我们选择使用一个基本布局，我们将转向一套特定设置的细节的选择和机器的数量。一个相对简单的过程是从 MCCF 配置开始并成功地选择最优的单个改进，就像前文描述的那样，使得产线服从周期时间约束。为了使这一过程更有效，我们应该算上我们可以提出的在决策者能力范围内的尽量多的可行技术设置(包括购买更多的机器和调整已确定位置的机器和/或过程)。我们不能忽略那些便宜的改进，它们能改善绩效问题并消除对昂贵的新

机器的需求。工厂物理学的诊断过程（第九章）对于识别有希望的设置是很有用的。

当然，如我们所知道的，绩效的要求（如，产量和周期时间目标）是它们自身决定的变量。虽然我们可以指定一些看似有理的数值开始分析，检验成本和绩效之间的权衡仍是有意义的。例如，如果我们能以$100,000 的花费把周期时间减少 5 天，我们就能很容易地决定这样做。我们可以用我们的模型做到这一点，在不同数值的周期时间或者产量约束下对它求解以产生成本与绩效的曲线。典型的成本和总周期时间的曲线如图 18.2 所示。虽然模型不能指出曲线上哪一点是最优的，但它确实为帮助决策者做出理性决策提供了有用的信息。

18.4.3 工厂设计的其他考虑因素

这些讨论提供了一些如何把成本、产出、周期时间以及其他因素整合到一个顾客导向的设施规划过程中去的观点。然而，工厂设计的问题比我们在这里解决的要多得多。事实上，存在一个广泛的学科，被宽泛地叫做**工厂布置**（**plant layout**）或**设施规划**（**facilities planning**），解决从各种加工中心的布置到生产流的最小化、决定工人的工作岗位数量等等的问题。这个学科致力于物料搬运、物理布置规划、库存与仓储、办公室规划、设施服务以及设施规划的开发与维护等重要事务。我们建议把 Tompkins 和 White（1984）的著作作为这个领域的一个很好的入门介绍。

18.5 产能分配和产线平衡

像上一个例子描述的那样，设计产线的工厂物理学的过程不太会产生一条平衡的产线。原因如下：（639|640）

1. 一条有明显的瓶颈的不平衡的产线比起相应的平衡的产线更容易管理，并且表现出更有逻辑的行为（拥有和最优更加相近的特性）。
2. 每个工站的产能成本一般是不同的，因此在某些工站保持过剩的产能比在其他一些更便宜。
3. 产能通常是以离散形式增加的（如，我们可以购买一或两台车床，但是不能购买一台半），所以让一个给定的工站刚好达到一个特定的目标产能是不可能的。

当合理地考虑了以上几个因素之后，大多数产线的最优配置都会是一条不平衡的产线。

18.5.1 定速装配线（Paced Assembly Lines）

尽管存在那些支持不平衡产线的观点，有时产线平衡还是有意义的。确实，产线平衡（line-of-balance，LOB）问题是工业工程的一个经典问题。然而，它只在**定速装配线**（**Paced Assembly Lines**）而非**流水线**（**flow lines**）才是适用的。在流水线上，每个工站本质上是独立的。每个工站以它自己的速度运行，所以瓶颈就是产线中最慢的那个工站。在一条定速装配线中，部件以恒定的速度在传送带或链上流动，通过通常有一个或更多操作员的**区域**（**zones**）。产线的设计使得每个操作员差不多总能在部件处于自己的区域时完成他们的任务。如果不能，产线就会受到干扰，因为工人会到下一个工人的区域尝试完成任务。因此，在定速装配线中瓶颈并不是产线中最慢的工站，而是产线运作机制本身。

另外，在定速装配线中产能的增量往往比流水线中的小得多。在一条定速装配线中，任务会被安排给产线上的每个工人，并被分割成细小的增量。比如，在一个手工电子装配的操作中，每个工站都“填满”了拥有大量组件的电路板。因为有许多组件，产线可以通过调整

每个工站的任务来达到平衡。附录 18A 给出了一个解决 LOB 问题的讨论和示例技术。

另一种平衡装配线的调整方法是人员管理。没人愿意长期处在比同事干更多的活却拿同样工资的环境里。由于现在大多数装配线都是由人来操作的（尽管有些装配线用机器人），公平就成了一件很重要的事情。尤其是在那些每个工站需要完成的任务几乎相等的产线中。

相反地，在流水线中，任务大多数是依靠机器本身，因此很难划分清楚。要增加一个特定工站的产能，我们就得要么给工站增加机器，要么对现有的机器进行提速。不幸的是，产线平衡的概念变得如此的根深蒂固，常常在不适合的时候也被应用了。它和希望拥有高的利用率的想法是导致人们生常常遇到接近平衡的流水线的原因。

18.5.2 不平衡的流水线

以上关于不平衡的流水线的原因指出，能以小而廉价的增量增加产能的工艺流程决不该是瓶颈。这样一个工艺流程可以容易而便宜地以很小的增量增加产能，直到不再出现由产能不足引起的问题。另一方面，把产能只能按昂贵的大块增长的工艺流程作为瓶颈则是一个好的选择。（640|641）

作为一个例子，考虑电路板工厂中的两个不同的加工中心：铜板蚀刻和人工检验。人工检验的操作发生在铜板蚀刻之前。[2]铜板蚀刻采用了一种机器，包括一个会耗费巨额电量的化学溶解槽。每台机器都有大约每天 2,000 块电路板的产能。增加一台铜板蚀刻的新机器大约要花费二百万美元左右的机器和设施成本，还需要大量的厂房空间。铜板蚀刻代表了工厂里最大、最昂贵的机器中的一种。与之相对，每个人工检验工站只需要一个半技术熟练的操作员，一个照明放大镜以及修整工具。每个工站每天能检验大约 150 块电路板。没有任何一个工站的成本超过 100 美元，而且工厂面积的需求也很少。

如果这是产线上仅有的两个工站，情况就很容易分析了。如果我们将铜板蚀刻设定为瓶颈，那么我们可以简单而便宜地增加人工检验操作的产能以避免瓶颈的等待。人工检验操作的利用率不太可能是不满的。相反，如果把人工检验设定为瓶颈，[3]那么我们必须大幅增加昂贵的产能给铜板蚀刻工站。因此，把铜板蚀刻设定为瓶颈并据此来管理它是很有意义的。

18.6 结论

这一章的主要议题是把工厂物理学的框架应用在关于产能的新产线设计和已有产线改进中。我们的主要观点总结如下：

1. *产能决策对于企业的制造竞争力有战略性影响。*产能战略对于成本有巨大而直接的影响，也通过影响其他一些计划和控制问题，比如集结计划、排班和车间作业控制，从而对绩效产生间接的效果。决策包括要增加多少产能、什么时间、在哪里，以及要增加什么类型的产能。其他的战略问题包括各种规模经济和不经济。

2. *工厂物理学的式子可以为产线设计和改进程序提供基础知识。*通过在一个给定的配置下计算产出、周期时间和在制品，这些式子可以让我们把将产线设计和改进的问题放进一个受给定产出、周期时间和/或在制品约束的成本最小化问题的框架里。通过改变约束条件，我们就能得到成本对绩效的约束。

3. *增加产能与改进设备或程序可以相互替代和/或相互补充。*比如，减少对现存机器的

---

[2] 这里的产能，生产率甚至工艺描述都被更改了，与以前请作者做过咨询的电路板工厂不同。

[3] 回忆在一条 CONWIP 的产线中，事实上是没有前端（front）的。因此，在拉动信号（如，CONWIP“卡偏”）没有及时返回的情况下，产线中早些的工作站会被较晚的工作站拖后腿。

修理时间与以增加机器的形式增加一个工站的产能有时会产生类似的逻辑效果。在其他条件相同时，程序上的改进的价值通常比增加设备更大，因为从改进产线上获得的知识和规律可以应用到其他的产线上，而简单的产能增加提供不了这样的学习机会。（641|642）

4. *流水线应当被设计为不平衡的*。工站之间的物流和成本差异使得将流水线配置为不同工站拥有不同利用率有意义。

5. *定速装配线应该设计为平衡的*。在定速装配线中，节律机构（如，传送带和传动链）通常是瓶颈。为了让工人在给定的节律时间内完成分配给他们的工作，如同分工要公平一样，根据优先权和离散性的需要把工站之间的任务划分得尽可能公平非常有意义。

要注意的是，根据工厂物理学程序设计的产线可能比用传统的 MCCF 方法设计的产线昂贵。然而，它们更可能按照设计时的要求运行。考虑到如由于不能满足目标产量而造成的销售损失，由于不能满足周期时间目标而导致失去忠诚的顾客，以及试图运作一条在持续的混沌状态下的产线导致的困惑等等因素，较昂贵的工厂物理学产线可以在长期内表现出更大的盈利可能性。

附录 18A 产线平衡问题

在定速装配线中，应当把任务安排到工站，以保证每个工站有几乎相同的工作量。这里有两个很好的原因：更有效地使用劳动力，以及避免当一个工站的工作比另一个辛苦得多时产生的公平问题。

假设在产线上运行的每个工件上都有 n 个任务需要执行，第 i 个任务所花费的时间为 t_i。这些任务被分配给 k 个工站且 $k \le n$。如果分配给每个工站的时间都是 t_0（即，传送带通过工站的时间），那么产线的速率为 $r_b = 1/t_0$。

由于任务会花费随机的时间，我们需要承认变动性的存在。我们定义 $c < t_0$ 作为被分配任务所花费时间的上限。为了使平均任务时间小于等于 c，我们为每个工站提供一些额外时间以适应任务的固有的变动性。注意，$u = c/t_0$ 是产线上任意工站的最大利用率，这个数字总是小于 1 的。

在很多处理 LOB 问题的文章中，c 被称作周期时间。但是，由于我们使用这种形式代指通过整个流程的时间，我们把 c 当做**传送带时间（conveyor time）**（即，它是传送带通过每个工站的时间）。

大多产线平衡算法的目标是为了使空闲时间最小，我们将其表示为

$$\text{总空闲时间} = kc - \sum_{i=1}^{n} t_i$$

其等效量度是**平衡延迟（balance delay）**

$$b = \frac{kc - \sum_{i=1}^{n} t_i}{kc}$$

它表示空闲时间的总比例。

更进一步地，我们应该考虑一些其他的约束。最常见的是**优先序（precedence）**约束，也就是说一些任务必须在其他任务之前完成。我们将只考虑优先序约束，不过建议读者阅读 Hax 和 Candea（1984，第 5.4 节）关于 LOB 问题的更加完整的讨论以及对相关文献的研究。

LOB 问题是一个很复杂的问题（NP 难问题），所以一个实际规模问题（100 项任务或更多）的最优算法往往需要多得过分的计算机时间。因此，大部分商用软件包依赖探寻式方法。

我们用一个简单的过程来描述 LOB 问题的探寻式方法，它与 Kilbridge 和 Wester（1961）使用的约翰逊和蒙哥马利的例子相似。考虑在图 18.3 中给出了优先序的九个任务。这些任务的工作时间和后继任务的编号在表 18.5 中给出。注意任务 5 有最大的平均运行时间，为 10。因此，$c \geq 10$。还要注意总运行时间为 $\sum_t t_t = 48$。

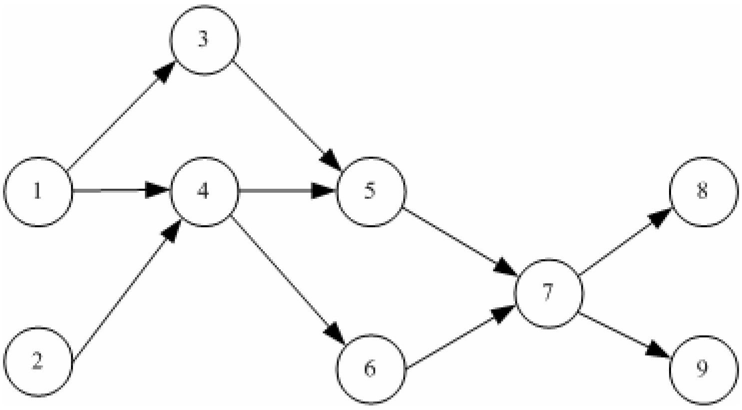

图 18.3 *LOB* 例子中的优先序

表 18.5 *LOB* 问题的数据

任务号	平均作业时间	后续任务号
1	5	7
2	3	6
3	6	4
4	8	5
5	10	3
6	7	3
7	1	2
8	5	0
9	3	0

为了达到零空闲时间，比率 $\sum_{i=1}^{n} t_i \big/ c$ 必须是个整数。然而，由于优先序约束可能妨碍工站所需的任务安排，这就不能保证零空闲时间。尽管如此，这一点与

$$\max_t (t_t) \leq c \leq \sum_{t=1}^{n} t_t$$

对给 c 决定一个合适的值有所帮助。如果我们定义 $\sum_{t=1}^{n} t_t = 48$，我们得到

$$2\times2\times2\times2\times3 = 48$$

这些因数的乘积在 10（最大运行时间）和 48（总运行时间）之间的组合如下

$$2\times2\times2\times2\times3 = 48$$
$$2\times2\times2\times3 = 24$$
$$2\times2\times3 = 16$$
$$2\times3 = 12$$

因此我们也许可以在 48/48 = 1 个工站（明显且不太有用），48/24 = 2 个工站，48/16 = 3 个工站，或者 48/12 = 4 个工站时达到完美的平衡产线。让我们考虑c =16 的三个工站的情况[4]。

为了描述我们的过程，定义现有工站的数量为 N，分配到当前工站的任务组合为 T，分配给当前工站的有效时间为 A，待分配的任务组合为 S，在这样的情况下，这些任务的优先序约束已经得到了满足，且其运行时间满足剩余时间。算法运行如下：

第一步：设定现在的工站号 N 为 1。

第二步：设定时间 c 有效，$A\leftarrow c$，$T = \Phi$，表明现在还没有任务。

第三步：决定组合 S 的候选任务安排。候选安排必须满足两个条件：

1. 所有前继任务都已经安排好了，也就是说候选安排没有前继任务。
2. 运行时间不超过有效时间：$t_i \leqslant A$。

第四步：使用以下两个原则从组合 S 中选择任务 j：

1. 选择后继数总和最大的任务。
2. 选择运行时间最长的任务以打破僵局。

将这个任务放在 T 里。

第五步：更新有效时间 $A\leftarrow A-t_i$。将任务 j 移出组合 S。

第六步：重复第三四五步直到没有剩余的候选任务（即，组合 S 是空集）。

第七步：如果还有剩余的任务，那就增加工站并跳到第二步。否则停止。

在我们的例子中应用这个算法：

$N = 1$　　$A = 16$　　$S = \{1, 2\}$　　$T = \Phi$

组合 S 只包含任务 1 和 2，因为它们是唯一没有前继操作的任务。由于任务 1 拥有最多的候选继承者，所以我们首先将它分配给工站 1。我们现在有

$N = 1$　　$A = 11$　　$S = \{3, 2\}$　　$T = \{1\}$

注意由于任务 3 的唯一前继任务——任务 1 已经被安排了，它就成为了候选任务。因为任务 2 拥有最多的后继且运行时间符合有效时间，我们接着安排它。

$N = 1$　　$A = 8$　　$S = \{3, 4\}$　　$T = \{1，2\}$

任务 3 和 4 都是下一次安排的候选任务了。这里我们看到了探寻式方法的重要性（以及随意性）。因为我们的规则就是选择拥有最多候选继承者的任务，所以我们选择最合适的任务 4（使用剩下的所有 8 个时间单位）。如果我们选择了任务 3，那么在任务分配完后我们仍有剩余时间。更加完善的 LOB 算法会尝试所有的剩余任务的组合方案以寻找一个最好的组合方式。当然，这需要更多的计算机时间。现在算法的状态如下：

---

[4] 当然，通过选择 c =16，我们就已经设定了生产线的产出。如果我们需要更高的产出，我们最好设定 c =12，尽管生产线并没有达到完美的平衡且空闲时间更多。这些状况在通常 LOB 软件中都没有考虑。

$N = 1$　　$A = 0$　　$S = \Phi$　　$T = \{1, 2, 4\}$

由于剩余的时间为 0，这里就没有候选任务了。我们现在必须开始开始安排工站 2。我们重新设定 A = c，发现这里有两个候选任务：

$N = 2$　　$A = 16$　　$S = \{3, 6\}$　　$T = \Phi$

任务 3 有最多的后继，所以在工站 2 中它第一个被安排，现在的状态为：

$N = 2$　　$A = 10$　　$S = \{5, 6\}$　　$T = \{3\}$

任务 5 和 6 都还有三个后继。但是，任务 5 是耗时最长的任务，而且刚好符合剩余的时间，所以我们这样结束工站 2 的安排：

$N =2$　　$A = 0$　　$S = \{6\}$　　$T = \{3, 5\}$

剩余的任务都符合工站 3 的传送带时间 c。

$N = 3$　　$A = 0$　　$S = \Phi$　　$T = \{6, 7, 8, 9\}$

由于 $b = 0$，所以这个安排是最优的。

注意在算法中我们有许多次幸运地使任务“刚好满足”剩余时间。这并不是典型的状况，事实上，当 $c = 12$ 或者 $c = 24$ 的时候是不会发生的。大多商用算法会尝试 c 的不同的值以及不同的打破僵局的规则。
