HARVARD | BUSINESS | SCHOOL

9-618-023
REV: MARCH 30, 2023

WILLY SHIH
MICHAEL W. TOFFEL

Production Processes

There are many ways to arrange production, ranging from job shops that are tailored to producing customized products, to assembly lines where items are processed in a linear sequence and often transported between steps with a conveyor belt, to continuous flow processing. The cases are set in factories that make different kinds of products. Factories are an appealing setting to assess and improve processes because producing physical goods makes it easier to visualize process steps and work flow.

This note describes four broad categories of process architectures and then examines the nature of task assignment that typically would be found in a factory organized along the lines of each process type. We will then delve more deeply into work flow policies, materials handling, and line pacing for the assembly line, since this process architecture is so widely used for the mass production of everything from smartphones to automobiles.

Production Process Types

Job shops, line flows, and continuous flow are some of the most common production processes deployed in factories around the world. [1]

1. *Job Shops* produce a large variety of products that require many different production steps and/or sequences between production steps. Job shops tend to be physically laid out by clustering workstations by equipment type, for example by placing all 3D printers in one area and all the drills in another. A product's routing through a job shop depends on the physical location of where specialized processing steps occur. Some job shops feature a similar flow of steps, though some steps might be skipped depending on the features of a given product. Most job shops feature product routings that are referred to as a *jumbled flow*, where each product's routing can be distinct depending on the particular set and sequence of processing steps it requires. (For example, one product might require steps performed at workstations A, B, then C, whereas another might require only steps at workstations C then A.) When an individual unit or batch of work arrives at a job shop workstation to have a particular task performed, setups are common (that is, the equipment must be adjusted to

[1] Robert Hayes and Steven Wheelwright, "Link manufacturing process and product life cycles," *Harvard Business Review* (January-February 1979): pp 133-140.

Professors Willy Shih and Michael W. Toffel prepared this note as the basis for class discussion.

work on that specific product). A job shop's atmosphere is akin to project work because each order (or job) is like an assignment with unique tailoring requirements in its production. Commercial printers, machine shops, tool and die shops, and paint shops are good examples of job shops that make physical products. Law offices, many consulting firms, medical practices, automobile repair shops, and custom tailors are examples in the services sector.

2. *Disconnected Line Flows* produce multiple units of the same product (batches) that proceed through a sequence of workstations that are not connected to each other by a fixed-route material handling system. Each workstation can encompass a complex process. Disconnected line flows can accommodate several different products that use a limited number of distinct routings through a factory. Although disconnected line flow processes may produce several different products, economies of scale usually lead them to produce several basic models with a variety of options. Thus, while the plant might also be described as having a "jumbled flow," the flow would tend to have a dominant direction. Because the goods are not passed between workstations by a machine-paced material handling system, work-in-process inventory can build up between workstations. Many manufacturing processes, especially of industrial equipment, are produced using disconnected line flows. Service examples include many office operations like processing invoices, classes within universities, and food preparation.

3. *Connected Line Flows* produce a limited variety of standardized discrete products, typically in batches. Work is conducted in a sequence of small tasks that are grouped into process steps, where each step is typically performed by a single worker at a workstation. Workstations and their associated equipment are therefore arranged according to the sequence of steps in which a product is made (or a service is delivered). The work is conducted in the same sequence of operations, following a rigid routing as it moves down the line; the sequence is fixed from batch to batch. Thus, the process steps are "connected" in a sequential flow. Work-in-process is often (although not necessarily) moved between steps by a machine-paced material handling system such as a conveyor belt. Workers can also manually pace a connected flow line. Connected flow lines are widely used to produce high volumes of a product because they are simple and inherently efficient. They can also be applied to service operations, for example at a car wash or a fast food restaurant.

4. *Continuous Flows* typically involve producing batches of product by having bulk materials flow through a sequence of connected steps that chemically, mechanically, and/or thermally convert them into the finished product. Such processes are common in food and chemical applications such as making beverages or converting crude oil into distillates like gasoline and jet fuel. These processes tend to be highly specialized, inflexible, and capital intensive—and typically produce a high volume of product through a standardized process at a low cost per unit.

For job shops, disconnected line flows, and connected line flows, output rates are typically measured as discrete *units* of output per time (e.g., widgets per hour), whereas output rates in continuous flow processes are typically measured via *volume* or *weight* (e.g., cubic meters per hour or kilograms per hour). All four of these production process types can also be characterized by *batch size*, the number of units or the measure of product quantity produced at a time. Thus, while job shops may have a batch size of one for making a custom tool, it might also be much larger, as in the case of printing a thousand business cards. Continuous flows also have batch sizes, but the batch is more often thought of as the length of a production run, with setup and maintenance being performed between batches. **Table 1** lists typical features of these production process types.

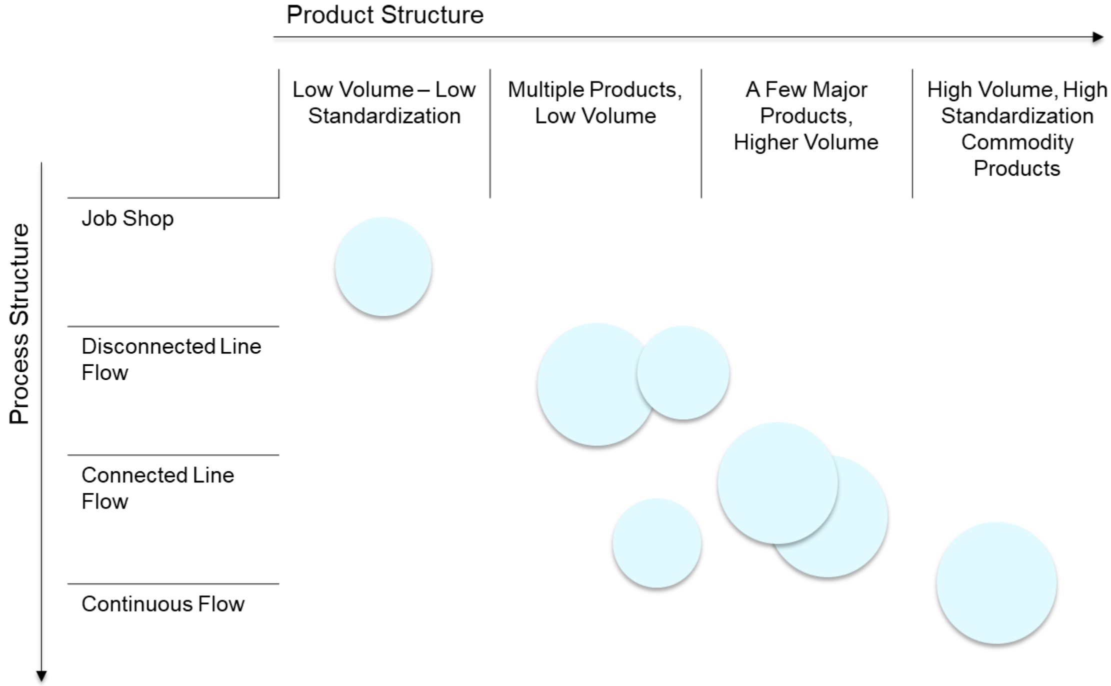
Table 1 Typical Process Characteristics

Process Type	Flow	Ability to Accommodate Various Production Configurations	Capital Investment	Variable Cost	Labor Content/ Skill	Volume
Job Shop	Jumbled, or with dominant direction	Very High	Low	High	High	Low
Disconnected line flow	Jumbled, with dominant direction	High	Moderate	Moderate	Moderate	Moderate
Connected line flow	Connected with rigid routing	Moderate	High	Low	Low	High
Continuous flow	Continuous	Low	Very high	Very low	Low, but with skilled overseers	Very high

Source: Adapted from http://www.netmba.com/operations/process/structure/

The type of production process typically used to manufacture a particular type of product is displayed on the product-process matrix shown in **Figure 1**. Most production processes in practice will lie close to a diagonal of cells in this matrix from the upper left to the lower right. For example, higher volume/lower mix products tend to be produced using a process closer to continuous flow on the lower right.

Figure 1 Hayes - Wheelwright Product - Process Matrix



Source: Adapted from Robert Hayes and Steven Wheelwright, "Link manufacturing process and product life cycles," *Harvard Business Review* (January-February 1979): pp 133-140.

Note: Each bubble represents a hypothetical company, and its shape reflects the range of structures of its products and processes.

Task Assignment

Different types of production processes tend to use different task assignment policies. In job shops, highly skilled, cross-trained workers (generalists) who can operate many different processes, tools, and machines typically are responsible for a single order, traveling with it from one functional area to another. The range of skills that job shop workers possess is driven by the almost artisanal nature of work conducted in job shops, in which each product or small batch of products require different inputs and tasks. Automobiles were initially produced in this manner, with skilled craftsman conducting a wide variety of tasks to fabricate them one at a time. Larger jobs shops dedicate workers to specific specializations, such as operating particular machine tools. In general, irregular and somewhat unpredictable demand favor workers with a broad range of skills, capable of performing a wide array of tasks.

Line flows are characterized by increasing levels of specialization among workers. As production volumes increase, there is often a greater division of labor, with workers focusing on a narrower set of tasks. These tasks often become simpler as well. The main innovation in Henry Ford's early efforts to produce automobiles was the decomposition of the vast majority of jobs in the factory from complex tasks to a series of simple routine steps that did not require specialized artisan skills.[2]

Continuous flows employ the highest level of specialization, but that specialization is often embodied in the machines themselves. Trained technicians are usually needed to service the machines and keep them running, and low skilled workers sometimes feed raw materials into the process.

Once a production process type is selected and the workflow is laid out, workers need to be assigned to tasks. In a *fixed assignment system*, each worker is assigned a sequence of tasks that constitute the step, and this does not vary between batches of different products. Workers perform their assigned tasks within a limited space or physical zone. The workstations do not overlap, and the work proceeds in assigned sequence. Coordination between workers is done by a simple handoff of work-in-process from one worker who completes her tasks to the next worker to begin his tasks and so on down the line. With a *work-sharing system*, task assignments may change for different units or batches. Faster workers might perform more of the work, and a shift of work between upstream and downstream workers can lead to higher output rates and shorter throughput times.

The variability in the time that a worker takes to complete a task has different implications depending on the process type. In a job shop or in batch production, variability in the time needed to complete each process step leads to longer throughput times. This might motivate the deployment of more resources to increase the capacity of bottleneck steps, and it might drive efforts to reduce setup times. In continuous flow processes, worker variability – such as in the speed of loading or unloading equipment at a particular step – might govern the pace of the overall factory.

Connected line flows warrant additional study because they afford an opportunity to improve performance through balancing across tasks, something that is often more difficult to do with the other production process types.

Connected Line Flows: Special Considerations

Variability can seriously impact the output rate and quality of output in connected line flows. It often stems from differences between workers with varying skill levels, but can also occur within a single worker/workstation. Factors like machine adjustments or setups might be needed for individual pieces that come down the line, for example, loading a piece into a specialized holder prior to processing, or plugging a part into a testing machine. Variability in materials might also influence handling and therefore the rate of processing. An example is the variation in the stiffness of a hose that needs to be attached, or other factors unique to the production environment.

Variability is overcome by line pacing. Pacing sets the cycle time of the line, and is typically set to match the bottleneck step, although it can be slower if the line designer wants to include some slack time. This can be helpful if a bottleneck step has variability – the worker assigned to the step will always have an opportunity to finish.

Pacing is intimately linked to the way work-in-process material is transferred between workstations. Manual transfer is the simplest material handling scheme. Workers simply pass the product to the next workstation after they complete their process step. In manual transfer lines, the

---

[2] David Hounshell, *From the American System to Mass Production, 1800-1932* (Baltimore: Johns Hopkins University Press, 1984), p. 221.

pacing is set by the workers; factory managers usually post electronic signboards that have daily output targets and up-to-the-minute progress towards that target.

Automated (mechanical) transfer is usually accomplished with a conveyor belt or another mechanism for moving the workpiece from one workstation to the next. Using a conveyor reduces the amount of non-productive labor by compelling operators to match an enforced pace, which results in a predictable output rate. This is meant to reduce the natural variability in the speed with which a worker completes repetitive tasks.

Implementation of a conveyor-paced line requires decomposition of tasks into balanced sets of tasks, as well as thoughtful product design that is compatible with this style of assembly. Conveyor-paced lines used with the assembly of small products usually employ a central belt with workstations positioned on either or both sides, with workers manually placing and removing items as the product passes by their workstation. Larger items such as cars or large industrial equipment like construction machinery or airplanes use a moving transfer line with a variety of conveyance mechanisms. In both cases, zones are marked for each workstation, and the pace of movement is set to the *takt time*,[3] which is defined as the *cycle time at which a finished product needs to be completed to meet customer demand*. The work is divided into a sequence of indivisible tasks, and then one or more of these tasks are assigned to each workstation, with an attempt to *balance the line* by assigning tasks so that each workstation takes roughly the same amount of time to complete all of their tasks. Time studies are often used to estimate the completion time for each task and workstation.

Establishing the appropriate pace of a conveyor-paced line is a challenging but vital decision. Operating the line at too fast a pace risks quality problems, as some items might pass through a workstation without being processed completely or accurately. Operating too slowly adds unnecessary cost, as the workers will experience additional idle time.

Buffers placed between workstations may accommodate slight variation in a worker's task completion time, as long as on average each worker's total task time and workstation size are matched to the conveyor pace. In practice, such buffers are simply a small accumulation of work-in-process that workers set aside from the conveyor (and draw from when they have time); such buffers are typically small due to space constraints.

Since each workstation in a flow line is assigned the same set of tasks for each product that comes down the line, workers at each workstation tend to learn over time how to complete their work in a more efficient way. Consequently, task repetition tends to decrease the amount of time a worker requires to perform each task (that is, workers exhibit a learning curve), affording an opportunity to rebalance and further optimize the line. There are downsides, as well, including worker boredom or loss of engagement, which may lead to quality problems or staff turnover.

Line Planning with Precedence Requirements

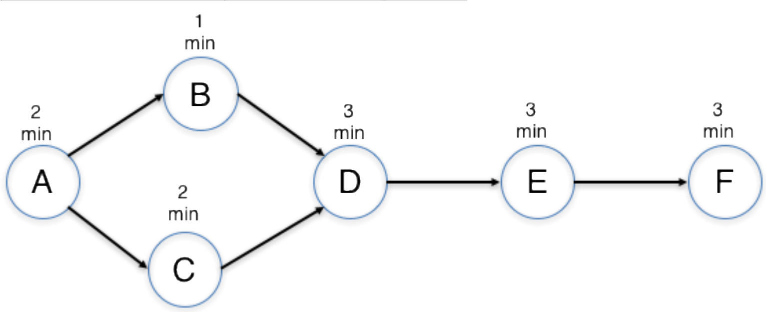
An important consideration in organizing a line are *precedence requirements*, physical restrictions on the order in which operations must be performed. Laying out the sequence of events and identifying the precedence requirements are usually the first step in designing a line flow. Consider the following example:

[3] The term comes from the German work "takt" which means "pulse."

Brian's frozen fruit pies has a demand for 120 fruit pies to be delivered over one 8-hour (480 minute) shift. The sequence of events, precedence, and time for each step are provided in **Figure 2**.

Figure 2 Pie Production Process: Tasks

Work Element	Precedence	Time
A: Measure crust, blend flour and shortening	--	2 min
B: Roll out crust, fit to pie pan	A	1 min
C: Mix pie filling	A	2 min
D: Assemble pie	B, C	3 min
E: Flash Freeze	D	3 min
F: Package in box	E	3 min



If the demand is 120 units over one 480-minute shift, the *takt time* (the cycle time required to meet demand) is calculated as 480 minutes/120 units, or 4 minutes per unit.

One can estimate the theoretical minimum number of workstations required by adding up the task times for all of the steps, and then dividing by this takt time. This assumes that the tasks can be further decomposed into sub-steps and then uniformly divided:

$$Theoretical\ minimum\ number\ of\ workstations = \frac{2+1+2+3+3+3\ min}{4\ min/workstation}$$

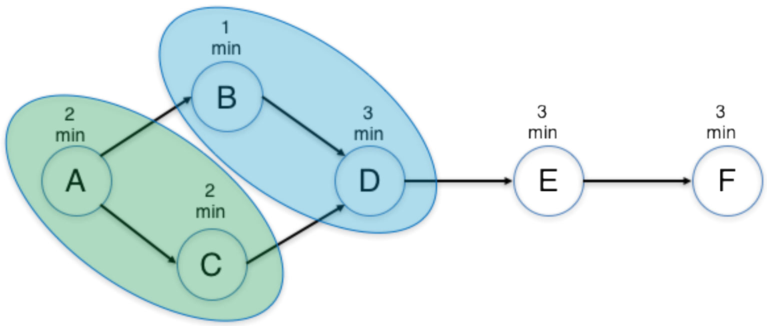
$$= \frac{14\ min}{4\ min/workstation}$$

$$= 3.5 \text{ workstations}$$

$$\Rightarrow 4\ workstations$$

In reality, these tasks are indivisible; each must be completed by one person and cannot be subdivided across two people. But the tasks can be grouped so that one worker can do more than one. **Figure 3** illustrates the process if one worker were assigned to both tasks A and C, another worker to B and D, a third to E, and a fourth to F.

Figure 3 Pie Production Process: Task Assignment Version 1



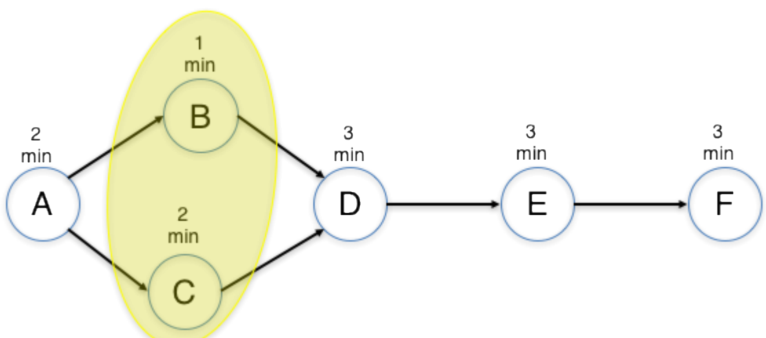
Under this task assignment, labor utilization would be calculated as labor used for productive work (in minutes) divided by labor available (in minutes):

$$Labor\ utilization = \frac{2+2+1+3+3+3\ min}{4x4\ min} = \frac{14\ min}{16\ min} = 87.5\%$$

With this task assignment, the cycle time would equal the takt time of 4 minutes, and capacity would be 120 units per 8-hour day (that is, 480 min per day/4 min).

Another approach is illustrated in **Figure 4**, where tasks would be allocated to five workstations (and thus five workers), assigning one worker to each of A, D, E, and F, and a fifth worker to both B and C.

Figure 4 Pie Production Process: Task Assignment Version 2



The bottleneck would now be 3 minutes, which means the line could be sped up to a cycle time of 3 minutes. This would increase capacity by 33% to 160 pies per day (calculated as 480 minutes per day / 3 minutes). The labor utilization for five operators is also higher than the prior configuration:

$$Labor\ utilization = \frac{2+2+1+3+3+3\ min}{5x3\ min} = \frac{14\ min}{15\ min} = 93.3\%$$

Line Balancing to Improve Performance

When designing any kind of connected flow line, managers initially have to distribute tasks across all the workstations. Usually they want to minimize the total number of workstations, as this will minimize the number of workers needed and reduce labor costs. Minimizing workstations also tends to reduce material handling costs and space utilization. With connected flow lines, tasks are assigned such that each workstation has approximately the same workload measured by task time. The sum of the task times for each workstation determines the total task time for that workstation. Since each workstation is performing a different set of tasks, these sums are likely to be different, and because the cycle time for the line (and thus its pace) is set by the bottleneck step, some workers will have more idle time than others, which means that lines are usually *unbalanced*. For instance, the example in **Figure 5** illustrates a line with three workstations, in which the total task time of workstation 2 is much longer than those of workstations 1 and 3. This line could operate with a 60 second cycle time (constrained by the total task time of workstation 2, the bottleneck), which would result in 15 seconds of idle time per cycle for workstation 1, and 20 seconds per cycle for workstation 3.

Figure 5 Initial Task Assignment

Workstation	Task	Task Description	Task Time (Seconds / unit)	Total Task Time for Workstation (Seconds)
1	1	Task 1	30	45
	2	Task 2	15	
2	3	Task 3	5	60
	4	Task 4	50	
	5	Task 5	5	
3	6	Task 6	5	40
	7	Task 7	20	
	8	Task 8	15	

To improve line balance, tasks can sometimes be reallocated to adjacent workstations in a manner that maintains the same overall task sequence. For example, consider shifting the assignment of task 3 from workstation 2 to workstation 1, and task 5 from workstation 2 to workstation 3, as shown in **Figure 6**. The line could now operate faster, with a 50 second cycle time (constrained by the total task time of workstations 1 or 2, the bottlenecks), which would result in only 5 seconds of idle time per cycle for workstation 3, and no idle time for workstations 1 or 2. These changes yield a more uniform distribution of work, with higher labor utilization (due to less idle time) and a lower overall cycle time.

Figure 6 Task Assignment Reallocated to Improve Line Balance

Workstation	Task	Task Description	Task Time (Seconds / unit)	Total Task Time for Workstation (Seconds)
1	1	Task 1	30	50
	2	Task 2	15	
	3	Task 3	5	
2	4	Task 4	50	50
3	5	Task 5	5	45
	6	Task 6	5	
	7	Task 7	20	
	8	Task 8	15	

While the line in this example is still slightly unbalanced, it lowers the cycle time substantially therefore increasing the output rate. Line balancing is an iterative process. As individual workers gain more experience at their respective workstations, their task completion times generally shorten, but usually at differing rates.

Summary

In summary, the steps to good line planning are:

1. Calculate the takt time, or the cycle time needed to meet customer demand.
2. Lay out the sequence of events and precedence requirements.
3. Calculate the theoretical minimum number of workstations.
4. Assign tasks to workstations, recognizing precedence constraints and the advantages of a balanced line.
5. Calculate the efficiency of the line.

Managers usually pilot test a line to assess their planning assumptions. Managers then begin production and do rebalance as workers learn to become more efficient. Well-designed production lines are an essential component of delivering on a firm's customer promise.