Getting Started with Value Stream Mapping

Anders Nielsen
Gardiner Nielsen Associates Inc.
24-107 Atkins Rd.
Salt Spring Island, BC, V8K 2X6
250-537-0831
info@gardinernielsen.com
© 2008
1. Introduction: Why We Need to Map the Value Stream

Shigeo Shingo, co-developer of the Toyota Production System (TPS) with Taiichi Ohno, constantly stressed the difference between process and operation. A process is the transformation of raw materials into finished goods (what is now referred to as the internal value stream); an operation is the interaction of operator, machine and materials. While most members of an organization are involved in one or more operations, few have a complete understanding of the process. Consequently, much of the effort to improve is spent on operations. It was Shingo’s insight that more benefits could be derived from improving the process as a whole than from optimizing individual operations. He pointed out, for example, that it is better to eliminate transportation than to find ways to speed up the movement of materials. All improvements to operations must therefore fit into an overall plan for what the process should look like.

Perhaps Shingo didn’t get it entirely right. An operation not only includes the interaction of operator, materials and machine, but also their interaction with information. When studying the process, it is therefore also essential to consider the role of information in the internal value stream. Information comes, ultimately, from the customers; their requirements are translated into product specifications, production schedules, pull signals, and delivery directions.

Understanding the internal value stream is therefore critical both to serving the needs of the customer, and to improving the efficiency (and profitability) of the value stream. Value stream mapping, a means of representing the value stream with symbols and numbers, is the key to understanding the entirety of the transformation of raw materials into finished goods.

For most of the time that industry has existed on a large scale, the process has been devised as the push of materials through a set of operations. Production has occurred in large batches, almost as if this was a natural law of how things are made. Shingo and Ohno sought to bring back the pride, quality, and customer service of an earlier craftsman era. The craftsman made only what customers ordered. They found that it was possible to produce to demand more effectively than it was to produce in batches. The benefits of producing to demand will be further explored later in this book. Without a value stream map that thoroughly describes the current state of operations, and which provides key performance measures of the process, today’s organizations are likely to repeat some of these errors of the past, and fail to see opportunities in simple changes to the value stream.

A value stream, as the name implies, flows to some specific end. This end is a set of requirements, as expressed by the customers of the value stream. When the value stream fails to meet the requirements it is intended to serve, it must be improved. Customers have three simple requirements – price, quality and delivery. In today’s market, competitive price and quality are basic requirements for staying in business. On time delivery at a cost that allows for profitability is therefore the key to competitiveness.

“But we already have a process flow diagram – isn’t that good enough?”

Process flow diagrams and traditional value stream maps don’t tell the whole story. What is missing? In a word: **time**. All traditional maps, whether drawn by hand, or created using specialized software such as Microsoft Visio, lack this important element. Traditional maps can only present a “snapshot” of the sequence of steps in the flow. Pictures of actors and scenery (no matter how many there may be) are not the same as a movie, for the same reason – they lack the element of time. Time is essential to understanding how one operation affects another, how a particular resource may influence the entire process, and how the status of the queues and operations vary with time.

Discrete event simulation introduces a whole new dimension to value stream mapping - time. The value streams will actually “run”, and provide results for how each operation, queue, and resource performed over the duration of the run. With the element of time added to a value stream map, it is possible to test hypotheses about which changes will produce the best results, to see the full effects of apparently small changes, even to challenge the correctness of a map when it fails to produce
results that are compatible with the behavior of the real-life value stream.

Building and testing value stream maps with a computer simulation, whether of current processes or proposed improvements, invariably improves everyone’s understanding of these processes. This is because simulation provides the user with feedback in the form of run results. By looking at the performance of the whole system, of individual operations and queues, or the amount of time a resource (such as an operator or a lift truck) is busy, one can determine whether each element of the value stream has been characterized correctly, whether there are activities that have been left out, or whether an improvement idea will actually provide better results, and by how much. The activity of building a simulation-based value stream map in a team setting also gets the team committed to the improvements, and implementation therefore becomes much easier than with other approaches.

The first step of value stream mapping is to map the existing process – the result is a Current State Map. The validity of this map can be tested by comparing the simulation results to actual results. Once a valid current state map has been constructed, it is then possible to suggest improvements, with confidence that the results from the proposed system will be quite similar to those of the Future State Map. In this way, the lean initiative moves from solid footing to solid footing. There is no “leap of faith” or “trust me” required to convince everyone that there really is a better way to carry out the process (a more profitable one too, since simulation even supplies financial statements).

A word of caution: Lean has a bias for action, and it is important to move on to improvements and to not get stuck on analysis. Furthermore, lean proposes that “better beats best”, again warning us that we should not be obsessed with perfection, but get on with many small improvements. What we are suggesting here is that the value stream mapping process be done quickly, but accurately. With the ability to test proposals quickly, we ensure that we move on to action, but we also ensure that everyone is on board.
2. Getting Started: The Current State Map

The process for current state value stream mapping has six steps:

1. Select the product family that will be mapped
2. Decide what the goal for improvement will be
3. Form a team to collect data and map the selected value stream
4. Walk the flow and collect data on the value stream
5. Understand the value stream from the customer’s point of view and how scheduling tries to meet that need
6. Draw a map of the value stream (we’ll look at verifying the value stream map in Chapter 3).

Let’s look at each step, plus some tips and tricks for data collection:

**Step 1: Select the Product Family**

Value stream maps are created for a single product, or a family of products. A family is a group of products with similar routings, similar process times, and customers with similar needs and demand rates. ‘Similar’ means that while there may be some variation, it is recognizable that all members of the group have a core set of operations that are the same. Products may vary by color, size, minor features, or one or two steps in the production process. In *Creating Mixed Model Value Streams* (Productivity Press, 2002), Kevin J. Duggan has suggested that a product family shares at least 80% of the process steps downstream from the point where continuous flow becomes possible (i.e. where there are no more shared process steps), and that the total work content of the process steps of each member of the family should be within 30% of any other member. The reason for concentrating on a family of products is that lean improvements such as cells and kanban systems only work well if all products in the stream are similar in process times and routings, and changeover time is minimal. A family of products is selected through Part Quantity Process Routing (PQPR) Analysis, as shown:

<table>
<thead>
<tr>
<th>Part #</th>
<th>Demand</th>
<th>% of total</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>420</td>
<td>35%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
<tr>
<td>7</td>
<td>288</td>
<td>24%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
<tr>
<td>3</td>
<td>276</td>
<td>23%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
<tr>
<td>9</td>
<td>84</td>
<td>7%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>4%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>2%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>2%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>1%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>1%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>1%</td>
<td>A 1  B 2 C 3 D 4 E 5 F 6</td>
</tr>
</tbody>
</table>

While it is understood that there may be shared resources at various steps in the value stream, for the purposes of value stream mapping, they should be considered only with respect to the selected product family. Many products will have a multi-level bill of materials, with various model options and other features leading to multiple part numbers. In determining a suitable family of products, it is necessary to consider the main routing only. There may be a common chassis, or a main part,
to which parts are joined; or there may simply be a chain of operations which is the longest of the various chains that eventually become the finished product.

Bear in mind that ultimately, everything is connected to everything else. If in no other way, the ERP system tries to tie all production together. This connectedness can be a major hindrance to improving delivery. Upstream from the point where continuous flow is possible (which is often only the last few steps in the process), production is regulated by the push system, and batch sizes are based on equipment reliability and setup time just as much as on what is in the order book. In fact, the system deliberately hides these factors, in the way information is recorded and presented. Improving delivery therefore rests on changing a mindset just as much as it rests on changing the value stream. When operations are viewed in isolation, there may appear to be great value in improving, for example, the paint and assembly value stream; however, in reality the flow of this operation is entirely reliant on the upstream operations, such as molding, stamping, and extrusion, which are predicated, with their large and fast machines, on producing in large batches. In the rush to reap the benefits of small batch, continuous-flow production, it is easy to overlook or downplay the difficulty of both the culture change and machine modifications required to effectively pull from these shared resources. Another commonly overlooked factor is the distortion of true demand by the sales department, by offering volume discounts when in fact they should be offering incentives for predictable repeat orders. Later sections of this manual will show how to ensure that the product family has the right characteristics for value stream mapping and lean improvement.

**Step 2: Decide on the Goal for Improvement**

Improvement goals are essential to avoid an open-ended (and typically never-ending) mapping process. Without a goal, “paralysis by analysis” is the most likely outcome. That being said, there are different opinions about what the goal should be. Advocates of the TPS style of lean manufacturing will claim that production at takt (i.e. the rate at which the customer demands your products) is the only acceptable goal of the improvement process. Others, such as John W. Davis (in *Fast Track to Waste Free Manufacturing*, Productivity Press, 1999), will say that takt is an outcome, not a design parameter. Whether the goal is to reduce cost by eliminating waste, or to produce in time with customer demand, value stream mapping is an essential tool, and in that sense it is independent of any one production philosophy. As mentioned previously, improving delivery time should usually take precedent over reducing prices; on the other hand, a more appropriate goal may be to simply maintain the current delivery time, but to free up operators to take on additional tasks, or even just to eliminate overtime. Any of these goals can be met by carefully mapping the current state, and then finding the right changes to create a future state of the system that meets a clearly stated performance level.

One reason pull systems that deliver product at customer demand rate are in some respects easier to design, is that they don’t require verification. Section 3 of this book describes how to verify that a selected target state meets a given performance criterion. This is a direct outcome of designing a mapping process that can meet all goals, and not just the typical future state advocated by, for example, Rother and Shook (in *Learning to See*, Lean Enterprise Institute, 1999), or Duggan.

**Step 3: Form the Value Stream Mapping Team**

Value stream mapping is best done by a team, in a workshop format. This is because, as explained previously, few people really understand the entire value stream, or have the knowledge to anticipate the impacts of specific changes to it. A team with a variety of members is therefore required. The workshop format will also bring to light undocumented procedures and policies, and alternative routings used when the standard steps cannot be followed. Finally, it helps create buy-in from a wider array of functions, than if the mapping were to be done by one or a few members of the shop floor departments where the value stream exists, or if someone was appointed to do the mapping independently. The following people are useful to include in a team:

- Supervisors and other shop floor leaders from the departments that the product family is produced in.
- A Representative from the IT department.
- Someone from engineering or ERP, who sets up routings for products.
Activities of the VSM Team

The VSM Team is now ready to go to work; it has been given the task of improving a particular value stream, and it has a specific goal to guide their efforts. Many members may need some training in subjects such as the history of lean manufacturing, the various tools found in the lean toolbox, structured problem solving, and project management. It is important that the team champion is good at keeping team members focused and mindful of deadlines, and can meld members from very different disciplines into a team.

The tasks of the team are:

1. To collect data on the value stream.
2. To map the value stream as it currently exists.
3. To test the validity of the current state map.
4. To develop ideas for improving the value stream, both in the short and long term.
5. To ensure that all improvements can be carried out while maintaining delivery and quality.
6. To get buy-in from anyone affected by the changes, and ensure that the target state becomes the standard way of operating.
7. To put together a project plan and budget, and get management approval.
8. To manage the projects and kaizen events as they are undertaken.
9. To verify results, or go back to step 4. if they fall short.

Step 4: Walk the Flow

Once the basic routing of the main element of the product family has been determined, it is time for the team to go to the shop floor where the product is made, or, in the case of an office or administrative value stream, the areas of the office where the process is carried out. The team members should have a form for gathering data on each step (including work centers, queues, raw materials arrival, shipping to the customer, and operators). A rough map of the process should be drawn during the walk, with each step connected to the previous one with an arrow. This is often called a “spaghetti chart” since the arrows usually go all over the place, and the picture ends up looking like a bowl of spaghetti.

Look for the following at each stage of the process:

Shipping

- Start at the end of the process (where the finished product leaves the system) and move upstream. “Cause and effect” relationships will be easier to identify using this approach.
- Gather information on customer requirements, including shipping frequency and shipping quantities. Medium term quantities are probably best to collect, since they will have less variation. For example, take the average monthly amount and divide by the number of times shipping occurs in the course of a month. Average this
value over all customers for the entire product family, to create a single “customer” with a single demand.

- Team members from the ERP or sales department may have valuable information to contribute here as well.

**Operations**

- Walk along the routing, gathering standard data on each step. Note where subsidiary streams join the main stream (i.e. where parts, subassemblies, or purchased components are used in the specific step; for a full mapping, these streams will be treated as separate value streams, with the main product as the customer, and will need to be walked separately for data gathering.)
- For each operation, gather the machine cycle time. See below in the discussion of operators whether the machine cycle time will be the operation cycle time used in the mapping.
- Find out about how often the products of the product family are produced at this step. This is called EPE, or every part every time period. For example, if the product is produced every two weeks, then EPE = 10 working days (in the case where there are no weekend shifts scheduled).
- Determine the standard batch size. Different models within the family may have different batch sizes. For the purposes of mapping, this step is simplified by averaging: take the total of all batch sizes and divide by the number of models in the family. For example, if a family contains left and right sides for a large and small part, and the batch sizes are 800 and 400 for small and large, then the total is 2400, and the average batch size is 600 (2400/4 = 600).
- Changeover time, yield, and downtime are not always available, and may be difficult to observe during the walkthrough. Ask operators and supervisors how long a typical setup takes. Be prepared to make adjustments during the testing phase, since it is common to get an “optimistic” answer. Make sure the setup is expressed in terms of “good-part-to-good-part”, and not just the machine changeover. Part-to-part includes removing the old material, changing the machine over, and any time spent making adjustments to get an acceptable part.
- When calculating how many parts a certain amount of raw material will yield, subtract material wasted during adjustments and lost to scrap, though not what can be reworked.
- Use the same approach for downtime as was used for changeovers – the true downtime value is likely higher than what is reported. Calculate the uptime of the machine in percent of scheduled operating time. Find out the scrap percentage, and convert it to lost cycles. Add it to the downtime for a final uptime calculation.

**Operators**

- Determine how many operators are working at each operation. Is this the standard number, or are there extra operators today (e.g. doing rework)? Only use the standard number for the mapping.
- Make note of how operators interact with the machine. If the machine waits for the operator, then it is the operator’s cycle time that governs the operation. If the operator waits for the machine, then use the machine cycle.
- In cases where there is manual loading or unloading, in addition to the machine cycle, use the complete part-to-part cycle time (= load time + run time + walking to get material + unload time).
- If the operator occasionally carries out other functions that result in the machine not running (such as periodic fixture checks), find out how many minutes of lost production results.
- Find out how many shifts are scheduled, the actual time the operator runs the machine (as opposed to the total length of the shift), and if overtime is used, and how much.
- Machine availability will be equal to the length of a shift minus scheduled stoppages (breaks and scheduled meetings), whereas the operator’s activities may mean that the machine actually runs for a shorter period.
- If not captured in the above downtime calculation, add the lost cycles from the lack of an operator.

**Queues (Inventory)**

- Make note of the amount of inventory between operations (both stored at the work centre and in work in process inventories in other parts of the plant). All inventory will need to be converted into the equivalent quantity of finished goods. For example, a roll of steel of a certain length, say, 500 feet (or of a certain weight), is the equivalent of a certain number of finished parts, say 1000 pieces.
• Where the product family has several members, it is useful to distinguish between them. However, it is the total quantity of material at each stage of manufacturing that will be used for mapping purposes.
• Assuming that the process contains parts that are part of the family of products, as well as some that aren’t, distinguish the two kinds.

**Suppliers**

• The value stream will have a single supplier, as discussed above. Data to collect includes frequency of delivery, and the quantity of product per delivery.
• Since medium term quantities are being used to determine what is shipped, and this tends to be fairly steady, the supplier delivery rate should have a similar steadiness.

**Shared Resources**

• Shared resources (i.e. machines used by several product families) can be found anywhere in the process. They can also include outsourced steps, such as coating, heat treating, or plating. They may need to be treated in different ways, according to how the work from the product family is being prioritized:
  o **Case 1: all work has equal priority.** It is not certain exactly when the work will be done, but in general, it gets processed within a given period, perhaps every two weeks.
  o **Case 2: certain jobs have priority over others.** In this case a specific processing date will be set, and when this date is reached, the job will be first in line.
  o **Case 3: Expediting.** This case can be thought of as last-minute prioritizing. It may result from unscheduled downtime or yield losses, either at the shared resource or downstream (when additional product needs to be made to make up for lost product), or from sudden changes in customer demand.
• It is also necessary to understand the variability of shared resources when collecting data, as well as their medium and long term stability. Changeover time is usually more variable in shared resources, since there may be significant variation in the sequence of jobs performed. However, a medium term or long term average will still exist. The same will also hold for uptime and yield. Of course, it is always important to find out if there have been specific changes recently (e.g. major repairs, or material changes) which may have shifted the average.

**Special cases**

• Outsourced operations are not always possible to document in detail. It is advisable to model such steps as a “black box”. This means that the outsourced operation is treated as a step where material is held for a specific period of time, but no cycle time, changeover time, or other operation specifications are required. A batch simply enters, stays for a period of time, and is returned to the process flow. Outsourced operations will usually have a transportation step before and after the process step.
• Note that this modeling approach is not appropriate for conveyor-based ovens and furnaces. Although these operations work on complete batches, they are seldom loaded as a batch, and each item can be accessed for further processing when it emerges from the process. For these cases, a normal process operation should be used. It may be useful to think in terms of processing time, and drop-off rate: For example, in a conveyorized process, an item could stay in the process for, say, 90 minutes (processing time), but a new item is loaded, and a finished one unloaded, say, every minute (drop-off rate). The processing time is the time to use in the value stream map.

**Step 5: Understand Customer Value and Scheduling**

**Value**

The key to lean is to understand the product or service from the perspective of the customer. Value stream maps are drawn to reduce waste and improve the rate of flow so that production can be done in the most cost-effective manner; but ultimately, the goal is for the customer to receive the right goods at the right time and at the right price. The value stream
mapping exercise does not engage in value engineering. There may therefore be cases of complicated flows producing a product that is not really what the customer wants. Nevertheless, it can be useful to consider competitive pressures and advantages possessed by the value stream. When target states are considered, the targets should be set with input from sales, in order to keep the cost of improvement consistent with the customer’s service expectations.

**Scheduling**

All processes need to be scheduled; how this is done varies greatly from company to company. As was explained above, for most of the industrial age, scheduling has been done for each work center, resulting in a push system. Product is produced only with regard to a schedule or order, and not with regard to whether the next step in the process is actually ready to use the material being produced. Furthermore, the schedule or order has called for a batch of material to be produced. This might be acceptable if the schedule is revised frequently (i.e. at least every day). Unfortunately, the complexity of calculating what is required, and the fact that batches often take more than a day to produce, means that schedules are usually updated on a weekly basis at best. Even the most modern ERP systems are typically updated only every week. Since the VSM team is on a journey to a lean production system, it will need to understand how scheduling is done. While it may be relatively simple to change the scheduling process, attention must also be paid to how materials are replenished. This is usually a function of the same system that sets the schedule. With an appropriate bill of materials, it is possible to calculate how much material will be consumed to produce a given quantity of product. With a list of orders on hand (or a production forecast) and supplier lead times, it is also possible to calculate when to order raw materials. As the production lead time is reduced through lean initiatives, and customers start to place orders closer and closer to their required date, the forecast becomes more and more important in determining when to replenish stocks. Beware that forecasting in many companies consists of educated guessing, or worse!

**Step 6: Draw the Current State Map**

The Current State Map is simply a set of connected operations and queues, starting with a supplier and ending with a customer. The most important guideline to follow when drawing the map is that it should be done by the whole group, in a workshop format. When the entire group participates in the drawing, the discussion and the questions raised regarding “what really happens” leads to greater insight than when it is the project of only one or two individuals. The group should ignore any process flow charts already in existence, and rely solely on the data gathered during the walk through the process.

[INSERT DRAWING]

If a simulation is being created of the value stream, a member of the group who has had training in using simulation should set up a computer, connected to a projector, and construct the map from the suggestions of the rest of the group. Once data has been entered, the value stream simulation should be run, and the results reviewed.

**Data Collection: Tips and Tricks**

The current and future state maps are only as good as the data on which they are based. Collecting good, useful data can be a challenge. It can be a very time consuming step if everything needs to be measured for the first time. In addition, where no measures are consistently collected, it will be necessary to obtain information about the operation from those directly involved with the operation, and this kind of data is often unreliable. Consider also that what is in the standards and the ERP system is often out of date or only estimates.

The following tips may help you get a better picture of what happens:

- Follow the manufacturing process from start to finish to get the actual routing. Walk the route and confirm that at each step the product came from the previous step; or better still, start at the end and work towards the start, ensuring that product travels to the step just visited.
- Note levels of WIP at each step. Ask if there is more material stored elsewhere, and include that in the count as...
well. Ask if this is the normal routing, or if it is an alternative step (due, for example, to a machine being unavailable as a result of breakdown, or a quality problem with another product).

- Observe some setups to get times and batch sizes. Don’t take anyone’s word for how long a setup takes; most people are optimistic, simply don’t know, or don’t want anyone to know. Use a stopwatch or other method of getting a value, and do it several times; people usually do a faster job if they know they are being watched. Note examples of disorganization, and setup people leaving to get parts, tools, and so on. The future state will often require reduced setup times, and in most cases, setup times can be cut in half through better organization (as outlined by the 5S approach).

- Ask how often a particular product is run. Is it every week, twice a month, or some other period (called EPE, or Every Product Each period)?

- Find out what the downtime percentage is, and how frequently the machine goes down for repair. Distinguish preventative maintenance from breakdowns. Note that in many cases, setup has a standard allowable time, and that anything longer may be recorded as breakdown. Again, observe how the breakdown is dealt with. The same problem can have large variability in repair time, due to the time it takes for anyone to start working on the problem, how well organized the repair crew is, and whether or not spare parts are available.

- Find out what the scrap rates are. Count the scrap periodically. Find out why there is scrap. Improved yields are necessary for a pull system to work, so there will be a quality improvement initiative required in most cases in order to implement the future state.

- Collect data on cycle times. Note that manual processes and manual load/unload of machines can have large variation. Make a note of the time of day, since operators will slow down once they know that they will meet their production quota for the day. Again, cycle time reduction may be an aspect of the future state implementation, so it is important to note any opportunities to do so, such as redesign of the process, the layout, or better organization and ergonomics. Also note that even automated processes can have variability in cycle time from setup to setup, as setup people may be dealing with material variation by slowing a machine down, or accommodating operators who need to do additional inspection or in-line rework due to tooling problems such as wear.

- Check over the course of the run, and for a few days, what the actual WIP is (count it!), and try to find out what the average, maximum and minimum amounts are.

- Observe how many operators there are for a given process step, and what exactly they are doing. Note whether they are following the standard operating procedure. Again, look for simple opportunities, such as imbalance in the amount of work, excessive walking, or poor ergonomics, and so forth, that can be exploited to reduce cycle time, or which can be used when line balancing is required.

- In general, it is important to understand what the current state is. This is also a good opportunity to understand the reasons for variation, since all aspects of the process will need to be standardized for a pull system to work well. This will be one of the biggest challenges faced during the lean initiative, so don’t underestimate the work required. Variation has many sources, and it is important to get a handle on them at an early stage in the process.

- There may be cases where details on the operation are hard to obtain. Outsourced steps, for example, are often better treated as a “black box”, and the only information required is the batch size and the time the product leaves and returns.

- Remember to distinguish in all cases the cycle time of the operation from transportation time, and also from the time one product waits while the batch is being completed.

- Team members should be allocated to each of the documentation tasks required:

  - Routing (process or queue name; routing step; number of machines allocated to step)
  - Cycle times (process c/t; queue minimum wait time)
  - Setup time (time; frequency)
  - Yield (scrap or yield %)
  - Availability / frequency of utilization of an operation in the particular value stream (allocation % to value stream; stoppages for breaks; how often the process is used for the value stream’s products)
  - Shifts and number of operators (overtime requirements; regular shifts; number of operators per machine)
- Breakdown (uptime or downtime %)
- Lot sizes (production lot size; transfer lot size)
- Work in process quantities (counted units; maximum queue capacities)

After the data collection is complete, all of the data can be collected in a single spreadsheet, or transferred directly to the value stream map from each team member’s notes.

ACME STAMPING
CURRENT STATE VALUE STREAM MAP
STEERING BRACKETS

<table>
<thead>
<tr>
<th>Process</th>
<th>C/T (sec)</th>
<th>C/O (min)</th>
<th>Uptime (%)</th>
<th>5 days</th>
<th>7.6 days</th>
<th>1.8 days</th>
<th>2.7 days</th>
<th>2 days</th>
<th>4.5 days</th>
<th>Process Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stamping</td>
<td>1 sec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Weld #1</td>
<td>1 sec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Weld #2</td>
<td>1 sec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ass'y #1</td>
<td>1 sec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ass'y #2</td>
<td>1 sec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shipping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steel Co.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PLT: 23.6 days
3. Validation: Testing the Accuracy of the Current State Map

This section is applicable to both maps and simulation models, in that validation is required to move forward with confidence. However, static maps are more difficult to validate, since all results are calculated manually, and lack the element of time. Typically, static maps only provide lead time and cycle time data, whereas computer simulations can provide data on inventory fluctuation, process waiting time, and other time based values.

Take a Reality Check

With the current state map drawn, and results available, the team should tackle the question “Is this actually what happens?” It is not uncommon for the results to differ from what people think they should be. This may be because something (a step in the process, for example, or an undocumented variation from the official process) was missed, because people have a faulty understanding of how the process actually works (they may, in fact, want it to be different, usually in a more favorable way), or simply from errors in constructing the map (in entering incorrect data, for example, or in mistakes made in the process of simplification). If performance data is available for the product family being mapped, they can be compared to the data from the value stream map. Make sure that they measure the same thing, that the data is in the same units, and are otherwise comparable. This is a critical point in the study, and great care must be taken to get it right. In fact, it may be necessary to set up a data collection project to ensure that good, reliable numbers are used for comparison. A value stream map must be verified (i.e. it must map what actually happens, and produce comparable data to the system being mapped). If it fails in this aspect, it can’t be used as a basis for planning the future. Again, bear in mind the “bias for action” of lean, and avoid “paralysis by analysis”.

A Second Aspect of Getting the Map Right

There are two aspects to ensuring that the model is right: verification, and credibility. Verification, as was just described, means that the map has been shown to have the right operations, delays, resources, and so forth, that they are linked correctly, and that the values used in characterizing each component are correct. However, in the case of a system being considered for implementation, or a system that will come into being at some future date, there is no real-life system to check the results against. Here, the second test of validity is very important. Credibility means that users of the map trust it when making decisions based on it. This is a function of a number of factors: verification of input numbers and assumptions; involvement of decision makers in the process of selecting assumptions and laying out the basic structure of the model; and in the case of future state maps, a clear path from the current state to the proposed future state.

The Process of Verification

Verification is a systematic process. Value stream maps are usually built quickly, with the main emphasis on getting the flow correct. It is not even necessary to use any numbers at this point. All that is important, is to see the whole process that converts raw materials into finished goods. Once this stage is complete, the detailed work of characterizing each step begins. Then testing begins. If a simulation is used, it will provide a complete set of results, and testing consists of comparing these results to performance measures from the actual system. For the most important aspects of the model (i.e. where accuracy is important to decision making), it may also be prudent to carry out a sensitivity analysis. This involves making small changes to system parameters, and then seeing if the results change slightly, or a lot. If the latter occurs, then extra care must be taken to get everything absolutely correct, since a small error will be magnified, and a consequent decision may be wrong. “Delta” reports (comparison of the current run with the previous or a saved run) are convenient. They support sensitivity analyses by allowing multiple runs (based on the same random numbers) to be compared.

Documentation of the model is another important aspect, both during verification and in subsequent validation. Not
everyone will have the time or training required to examine the map itself in detail. Good documentation helps other users, and especially decision makers, to understand and gain faith in the map, and the target or future state maps based on it.

**What to Do if the Model Fails the Validation Tests**

Getting a value stream map right is an iterative process. There are a number of common mistakes that are made, which will quickly come to light when the model is run as a simulation. Assumptions may be incorrect or not close enough to the real values; simplifications may have been made incorrectly (either by getting the values wrong, or by leaving out the controlling process); and finally, the map may have been built incorrectly (by using wrong values, or incorrect order of process flows).

Errors are usually found in current state maps when the results don’t compare with the results from the real world system. The first place to look is at results such as production lead time and number of units produced in the run period. If these are off, then the queues of the map should be examined, to see if they behave as expected. A queue that grows continuously, or one that quickly drops to zero, indicates a problem with an operation. Usually the cycle time or the uptime is wrong; it may also be the case that overtime is used, but was not factored into the available time for production. Once these errors are corrected, there may still be problems in getting the correct lead time. It can be harder to determine the cause, but generally, it is because errors were made in simplifying one or more steps. It can help to isolate the problem by substituting a simplified process with a simple batch process (since it only requires setting the batch size and the time to complete). If this corrects the problem, then a further effort should be made to recalculate the values.

It should also be borne in mind that many numbers gleaned from production reports and other shop floor data collection systems may simply be wrong. Downtime of less than, say, five minutes may not be recorded at all; setups lasting longer than the standard may be recorded as downtime. This may mean that the results of the model are being compared to an incorrect set of results, leading to bad decisions. This reinforces the importance of spending time on the floor, observing what actually takes place, and understanding how data is collected.

**Why the Need for Accuracy Regarding a Situation We Want to Change Anyway?**

As explained above, the purpose of mapping is to assist in making decisions about how to change an existing process, in order to reduce lead time and eliminate waste. To arrive at a proposal for a future state, start with the system as it exists today, and then develop specific improvement projects that will lead to the future state. Often, the current state map is merely a straw dog: destroyed as soon as it has been constructed. But as we also saw, this ploy requires a “leap of faith”. If the future state is so easy to construct, there would be no need for the current state map. We suggest that the path to the future state is not so clear cut, and that assumptions about what can be done to change need more attention than is usually given. Otherwise, buy-in and willingness to change will not come easily.

So the current state map, as the basis for analysis of the performance of the future state, must be accurate. Without a verified current state map, the changes that are built upon it may turn out to have less impact than was believed. And without a validated current state map as the basis for future state proposals, decision makers would be foolish to give their approval for an improvement program.

Approval will be given if the results are better. Therefore, a baseline must be established for calculating the improvement that a proposal can deliver. Here is yet another reason for the need for accuracy in the current state model: it is the baseline. If the baseline values are wrong, then it cannot be used to judge the value of an improvement proposal. Unfortunately, it has often been the case that much effort has gone into change that in the end proved to have no real impact or which fell short of the expected return on investment.
Learning from the Validation Stage

One of the most important outcomes of the validation stage is a greater appreciation of how things actually work today. This is worth knowing, since it provides insight into the kinds of things that are accepted, are taken for granted, and have become a part of the way things are. Think especially hard about that last point – why are things the way they are? Is it a result of a deliberate decision? Is it something inherited from the past that has never been questioned? Do people in fact believe it is the best or right way to do things? The answer to these questions will have a bearing on how difficult it will be to change to a new way of doing things. The hardest part of change is learning to think in new ways. Once that is accomplished, the organization will change very easily, because organizations are primarily the result of the thought patterns of its members.

Another important outcome of the validation process is a new appreciation of what gets measured, how it is measured, and how it is presented. One of the first steps in change is changing what is measured. This is because what is measured is what is important to the organization. In order to reduce waste and lead time, knowing what they are today is an important first step. Chances are that inventory turns, days on hand, transportation distance, and lead time are not even being measured today, because they are not considered important in push systems.
4. Waste Elimination: Constructing Target State Maps

What are Target Maps?

With the current state drawn and tested, it is very likely that many team members have started asking the questions: “Why do we do it that way?” and “Wouldn’t it be better to do it this way?” At this point in the process, there are many opportunities to make improvements, and to get a wide variety of people involved in improvement (kaizen) projects. The fact that a fully fledged pull system, designed to meet customer takt time, can be developed at this point, does not entail that this is necessarily the right step to take next. There are many reasons for looking for a better version of what exists today, but which falls short of the ultimate future state. For starters, an organization may not have the money, time or skills required to realize the future state in one step. Reducing setup time to less than 10 minutes, or getting operators to work in cells with new cross-trained skills, may be too expensive or time consuming. More importantly, improvement projects should not get in the way of serving the customer right now. Getting there in several stages generally suits companies best. With the ability to design and obtain performance data on any value stream, discussed above, it is possible to test out ideas for improvement that the company feels comfortable about, and which are affordable and achievable. Target maps outline suggestions for such improvements, in the form of a value stream map.

Considerations for Waste Elimination

Value stream mapping is a process designed to reduce lead time, to make product flow, and to eliminate waste (non-value added operations or activities), all for the purpose of meeting customer demand at the lowest cost, and with the highest quality. Lean thinking relies on recognizing the “seven wastes” – over-production, over-processing, inventory, motion, scrap, waiting, and transportation. Target maps reveal which of these wastes can be eliminated now, and where. With simulation, it is easy to avoid the traditional problem of eliminating waste at an operation where there is no net gain. That is because the revised system’s performance can be compared to the current state, to see the impact of the proposed change.

The key to producing useful target maps is to look for low-cost improvements that encourage flow, reduce inventory, and test the organization’s ability to manage in a lean environment. The challenge of developing the attitudes, systems and communication necessary for a true pull system operating at customer takt should not be underestimated. A high inventory system hides a multitude of problems, which will slowly be exposed as batch sizes and WIP are reduced. The level of organization and standardization required for one-piece flow are rarely found in companies with traditional production planning and traditional management.

1. Over-production

Over-production is the production of material which is not needed now. It usually occurs in the form of large batches, produced faster than the rate at which they can be consumed (and ultimately shipped). In job shops, it means working on something before it can be used by the next step in the process, or before it is required by a customer. In either case, the result is product that sits in work in process queues, or in a finished goods stock, but is not needed today. Over-production is caused by a number of factors, such as long setups, poor quality, machine unreliability, avoidance of setups in order to make performance measures look better, or the desire to keep an expensive resource working.

Lead time is, of course, directly related to inventory and over-production. For operations that are easily able to produce at a faster rate than demand, it is typical that one machine produces a variety of products. This means that the machine must be changed over periodically. Traditional cost accounting has ways of calculating the batch size appropriate for a given length of changeover (such as the economic order quantity, or EOQ). Changeover time is usually set as a standard, and therefore there is no argument about how much must be produced each time the machine is set up. Contrary to Lean thinking, if quality is poor or the machine is subject to breakdowns, the batch size will be increased. Furthermore, it is not uncommon for operators and supervisors to decide to produce even more, when things are going well, because “we’ll use it
some time”, or “just in case…” Lean thinking challenges the notion of a standard changeover time. Simple industrial engineering will easily point out where changeover time can be reduced. Good organization and changeover planning, by themselves, are often capable of reducing setup time by 50%. Standardization and integration of changeover components will often account for another 25%. This kind of relatively low-cost improvement will allow setups to take place much more frequently, thus allowing smaller batches to be run economically. Of course, the kind of machine being considered here (a stamping press, an injection molding machine, a large mixer, a high volume printing press, etc.) is shared among a number of product families, so the impact of change will be felt beyond the product family being considered in the particular mapping exercise at hand; this may limit which improvement suggestions can be implemented. Furthermore, if demand is erratic in some product families, it is unlikely that batch sizes will be reduced by even as much as 50%. It may be necessary to include demand profiles for these shared products in the target map, in order to test the feasibility of reducing batch sizes for the product family being considered. With simulation, try mapping all the value streams that use a particular resource, and add up operating and setup times for the resource, to see if it can all fit into the available time.

Working ahead is, unfortunately, very common. It is also a significant reason for long lead times. Working ahead happens for two main reasons. If work lists are available, operators will tend to put together similar orders, and do them together. This avoids setup and feeds the natural tendency to gravitate towards repetitive work. Secondly, not every machine has a full schedule every day, but everyone wants to look busy, so they tend to overproduce (or slow down). Simple ways to avoid these problems are to put out only what the next product to be produced is, and to help operators stay busy by cross-training them, and then moving them to where the work is when their first task is complete. In the target map, this is accomplished by putting operators into groups. When the situation is one of keeping a fast, expensive machine (henceforth called a “super-machine”) going, it is easy enough to say that it should only be used when required. However, the reality is that managers and cost accountants want to see it run. It takes a real attitude change to admit that super-machines are not always the answer. Some other potential solutions might be to sell it and buy more appropriate technology (i.e. smaller, more flexible machines), to shut it down when not needed (and absorb the overhead elsewhere), or to find more work for it (after solving the setup problem).

Finally, scrap and downtime can be decreased (though not completely eliminated) relatively cheaply through standardization. Standardization means doing things the same way each time. Setup is the key here, and standardization of setup means that the settings and materials are standardized. This leads to less scrap and better reliability, since each run will have almost identical characteristics to all other runs. Once again, just being consistent can reduce scrap and increase reliability by about 50%.

A lot of what has been discussed above can be put under the heading of 5S. 5S is an approach to shop floor cleanliness, organization, and discipline that is considered the foundation of lean manufacturing. The 5S system consists of five standardized activities, implemented through five sets of activities:

1. Activity number one (called “seiri”, or clearing up; popularly “Sort”) gets rid of all unnecessary items in the workplace. It creates space and flexibility to do what is required, without hindrance. The outcome is a standard that states what is allowed in the workplace, and how often the workplace needs to be reviewed for unused items.
2. The second activity (called “seiton” or organizing; popularly “Separate”) finds a place or role for everything that remains after clearing up. It ensures that everyone knows where to find what they need with a minimum of delay. The outcome is a standard that states where everything is to be found at all times, and systems for laying out work areas so that the most frequently used items are closest at hand.
3. The third activity (called “seiso” or cleaning; popularly, “Shine”) ensures that everything works well and is properly adjusted, through operators checking and cleaning the workplace regularly. The resulting standard specifies how often the cleaning and checking activity should take place, and includes it in the scheduling of work (much like breaks and meetings are scheduled).

The fourth and fifth aspects of the 5S system are not so much transformative steps as they are activities designed to maintain the state of affairs created through the first three steps.
4. The fourth step (called “seiketsu” or standardizing; popularly “Standardize”) makes sure that as the right way of doing things is discovered, it is turned into a standard practice, through development of policies and procedures. The standard for standardizing spells out review periods, and the data to be collected to ensure that the policies and procedures are working as expected.

5. Finally, the fifth step (called “shitsuke” or training and discipline; popularly “Sustain”) gets to the personal level, and demands that each member of the group is aware of what the rules are, and follows them. Without this level of standardized behavior, the 5S system will not be effective. Standardizing this aspect means ensuring that there is continuous improvement in what each person knows and is able to do, and in adherence to ever more stringent standards.

2. Over-processing

There are two aspects to this kind of waste – (1) overdoing it in the sense of doing too much, too soon, and beyond what is necessary; and (2) using inappropriate equipment, especially equipment that is much larger, faster, or more complicated than necessary. It can be difficult to distinguish between over-processing and over-production, because the first often leads to the second.

Over-processing is usually associated with going beyond what the customer requires. Examples are reports and presentations that have more information than the audience is looking for, and therefore are difficult to understand and act on. Products may be designed with more features than the customer needs, which end up being difficult to learn to use, and which cost more than necessary. In the rush to outdo its competition, a company may offer far more features than the market demands. In doing so, they add unnecessary complexity to the layout, process, and product, and subsequently suffer from poor quality, longer lead times and higher costs.

Over-processing, in the second sense, can also be associated with super-machines. These are machines built for mass-production, and are capable of production rates far exceeding customer requirements. Many problems are associated with these production centers. They tend to be difficult to repair, and since most factories only have one, they can actually cause shortages when they are out of commission. It can be difficult to determine the source of quality problems, due to their complexity. It is very difficult to incorporate them into schedules, since they usually have long setups. In other words, they rob a plant of flexibility. When starting the journey to Lean, the first action should be to get rid of super-machines, and replace them with appropriately sized machines (usually several of them) that can be dedicated to individual product families. Since super-machines are usually only replaced at long intervals, using appropriately sized machines will also ensure that up-to-date technology is constantly flowing into the factory, as the smaller machines will be replaced more frequently.

The final reason for over-processing has to do with excessive processing in the form of removal of material, or requiring several assembly steps, when a “near net shape” piece of material would have required less. Examples include using two steps to assemble a metal part to a plastic part, when insert molding could have accomplished this in one step; operators trimming flash from plastic parts, when a well-maintained mould could eliminate this operation altogether; or having a cutting department, when steel could be purchased already cut to size.

3. Inventory

Whether in the form of work in process (WIP) or finished goods, inventory is considered the great evil of production. With material always available, the focus is taken away from the process, quality, and the rate of work. Inventory thus actually hides problems that exist in the production system. Over-production leads to waste associated with inventory, in requiring extra space for storage, time and effort spent controlling inventory, money tied up in purchased materials, the potential for damage and obsolescence rendering the inventory unfit for use, the need for larger material handling systems to move larger quantities of goods, and increase in lead time for delivery, to name only a few direct costs! In addition, inventory has an impact on waste that is indirectly caused by having more than needed.

Inventory leads to a lack of attention to the process. This means that processes are designed with cycle times well
outside of the average. By buffering the process with inventory, the wide variance in cycle times is not noticed until an attempt is made to set up a continuous flow cell or line. Equipment must then be replaced, or great effort expended trying to balance the flow to the rate of customer pull. The reliability of the machine can also be overlooked when there is plenty of inventory. In a system with reduced inventory, reliability must be very high, or everything comes to a quick halt. Lean factories achieve 100% uptime through 5S, productive maintenance, and simple machines. The same holds for quality. While mistakes will be made (as Shingo noted in Zero Defects), control must be 100% at the source. In getting to 100% defect-free production, rapid problem solving (at the machine) is a must; quick development of mistake-proofing devices and the use of simple, capable machines is also a must. Finally, standardization of work is necessary to achieving smooth flow and reducing inventory to a minimum. All activities should have a standard time, and all personnel must know and follow the standard procedure. This goes for assembly, loading and unloading, changeover, machine operation, and other activities.

4. Transportation

When a facility layout extends over a large area, the movement of inventory from operation to operation becomes necessary. It is thus another result of over-production. It also results from laying out production equipment by function. Functional layout places each type of machine (stamping presses, welders, injection molding machines, etc.) in its own “department”, for a variety of reasons, mainly to do with the perceived benefits of specialization. The result, however, is usually over-production. When looked at from the point of view of uninterrupted flow of production, a functional layout is counter-productive. Focused factories, and cellular layouts, keep the equipment required for producing a family of products together. This is done in order to balance flow from one operation to the next, to provide rapid feedback on quality from one operation to the previous, to balance the number of operators to production requirements, and to allow pride in customer service. Creating a focused factory (a small space devoted to a product family, with all the necessary equipment for producing the products of the family), or setting up a cell (a group of machines which have one or at most a few pieces of WIP between operations, and usually laid out in a U-shape) brings the issue of over-processing to the fore, since in most circumstances the various pieces of equipment are not matched in production rate. It does, however, solve the issue of wasteful transportation, since the operations are now in close proximity. Movement of material can be accomplished using small containers, small hand-carts, gravity flow conveyors, or even taking a step or two from one operation to the next operation with the workpiece. Additional benefits of eliminating large material handling machinery include less damage to facility and WIP, the option of using narrower aisles, improved safety, and lower costs. In a cell or focused factory, visual control is much easier as well.

5. Motion

Motion is a waste associated with both operators and equipment. In the case of operators, wasted motion includes bending, walking to get or place parts, lifting, and taking more than one step to reach or view machine interfaces. In setups, it includes moving around the machine repeatedly to carry out the steps in the changeover in an unplanned fashion. Motion can add significantly to cycle time, and must therefore be considered separately when creating and balancing cells and focused factories. The waste of motion is reduced through ergonomics, work planning, standardization of work, 5S, and using smaller containers.

In the case of equipment, wasted motion is associated with long strokes, “air cut”, and other non-production movement of machine parts. In designing machines, the emphasis is often on versatility. This is associated with functional layouts and batch production. A general purpose machine is designed to handle a variety of tools. But from the point of view of continuous flow, this is not necessary, and waste therefore results. The solution is to customize the machine to its purpose, which is most easily accomplished when the machine is simple to start with.

6. Scrap

Scrap and rework are obviously wasteful. In batch production, scrap is rarely visible, since there is always more material available, and the run can be extended for a short while to produce the required quantity. In a continuous flow system, scrap is a serious problem, since every machine loses a cycle when a piece is rejected. This destroys balance, and
when producing to customer takt, results in a missed shipment. When perfect quality is required, 100% source inspection is necessary. This is achieved through mistake proofing (poka-yoke), as Shingo has so elegantly shown. It also, of course, rests on good maintenance, equipment improvement to achieve greater reliability, and simplification of production machinery. 5S (especially cleaning and checking) and standardization of work are also significant in reducing mistakes and defects. Design for manufacturability and simplification of processing can also help considerably to reduce scrap and rework. It should be noted that rework is as serious a problem as scrap, since, from the point of view of time, both are lost cycles.

7. Waiting

Waiting takes a number of forms. Operators wait for machines to complete their cycle, or for material to arrive so they can work on it. Machines wait for work, and also for operators to load and unload work pieces or other production material. The kinds of waiting that are common in batch production facilities are different from the waiting that is wasteful in a continuous flow system. Most batch systems strive to keep equipment working at all times. This requires buffers of inventory to be placed in front of all machines. By assigning operators to specific machines, they are consequently kept busy. In the progress to single piece flow, keeping all machines busy is not a goal (equipment is a sunk cost). The goal is to produce what is required by the customer, and no more. If a machine is capable of doing more, it is considered the wrong machine for the job. Over-production results from keeping the machine operating. It is considered important, however, to keep operators busy at all times. This is accomplished by moving operators from operation to operation, as the work flows through the process. This starts by completing the needed work (for example one day’s or week’s worth) at one station, and then moving on to the next step. With better balancing and training, as well as reduced setup time and improved reliability, it is possible to construct cells, where the number of operators is balanced with the required work, and there is only a small amount of work in front of each machine (an hour’s worth, or even only a single piece). Spare time should be used for continuous improvement activities and extra 5S operations.

Setting Targets, and Applying Appropriate Waste Reduction Methods

There are many waste elimination methods that can be applied in a given situation, but selecting the appropriate method is not easy. Consideration must be made of cost, disruption to other product families and parts of the production system, operator skills, maintenance, reliability, space, and so forth. In an ideal world, it would always be possible to go directly to a pull based single piece flow system, but this is seldom the case. There are therefore numerous choices. The solution starts with targets for improvement, i.e. specific intermediate goals. Ancient Rome, as the saying goes, was not built in a day. Nor did Toyota achieve its much heralded production system in one fell swoop. Rather, as the title of a famous study of Toyota stated, it took 40 years and 40 million suggestions.

Suppose the current state map is of a system that has a three week production lead time, and carries 15,000 pieces of WIP. A reasonable improvement target might be a 50% reduction in lead time, and a similar resultant reduction in WIP. But how is this accomplished, and how does one know that a given set of suggestions will meet the target? The value stream map will show how much inventory sits in each queue at any given time, and for how long. It is usually the case that the closer to the beginning of the value stream a piece of equipment is, the faster it is, the longer the setups take, and consequently the larger the batch sizes. But, it is equally likely that these upstream machines are shared by a number of product families. Since setup reduction is the key to smaller batch sizes, it must be considered whether this is economically feasible, and how quickly it can be accomplished. If the upstream machine is a stamping press, for example, and the daily requirement of any particular part is 1000 pieces, then it is likely that 50 to 75 active dies can be run in the press. To standardize these to the point where a 10 minute die change can be achieved is very expensive. But perhaps a 5S program will achieve sufficient organization and standardization of the setup, and better maintenance will allow predictability of the length of a run, so that the setup can be cut to 30 or 45 minutes, from the usual hour or two. This will allow a run every week for the parts of the family, instead of every two or three weeks. This intermediate reduction will have a significant impact on lead time. Moving two or three pieces of downstream equipment together into a cell, and reducing the WIP through line balancing and local pull, will also show up quickly in reduced lead time. These changes will usually be easier to accomplish with some training and a willingness from management to see equipment idle at times. Changing the way performance is measured, to bring it in line with Lean thinking, will present a much greater challenge.
It is thus possible to conduct improvement planning through “what-if” methods, or through a more formal approach. The ability to obtain results from the value stream map similar to the actual improved system, is the key to finding the right solutions for each business situation. By selecting appropriate solutions, and testing them, it is possible to see if a given target can be met, or if expectations or budgets need to be revised.
What Is Pull?

Most target improvements retain the traditional push system. This is especially true of upstream, shared, resources. The lean future state is based on pulling work through the system, at the required rate, rather than pushing.

The first question to ask about this future state is, “How does anyone know what to produce?” In a push system, everyone has a schedule, a work list, or some other kind of information that tells what and how many. In the simplest version of a push system, the requirement is simply to carry out a standard operation on whatever is first in the WIP queue. The pull system is really not that different. Instead of working constantly on what is in the upstream queue, or following a schedule, the operator fills containers that arrive from the downstream operation, or produces a quantity equal to the number of work orders (called pull or “kanban” cards) that arrive from a downstream operation. If the operation is separated from the downstream operations by a supermarket (a storage area with only limited space for particular items), then as these operations take parts from the storage spots, the upstream operation fills the empty spots. If there is a lack of balance in the system, then one or more upstream operations will be idle at times. By remembering that the objective of the system is to produce what the customer needs, and not to absorb overhead or simply keep operators and machines busy, this should not present a significant problem (although most supervisors will feel uncomfortable for a while). Of course, good planning will help to minimize such idleness, as idleness is, after all, waste.

The next question to ask is, “How does the whole system know what to produce?” A pull system may extend all the way to the customer, so that the customer issues cards or containers that serve to authorize production. But somewhere (since at some point the chaos of the marketplace meets the production system) there is a need to take a forecast or a set of customer orders, and decide how to schedule the required work. In many pull systems, this point of contact is simply shipping. Each day, orders arrive - perhaps in a batch, or throughout the course of the day - and when the appropriate shipment size is reached or a shipping window closes, the shipment is sent to the customer. As soon as this happens, a signal (be it a set of containers, a card, or an electronic signal) is sent upstream, authorizing replacement of what was shipped. This leads to a cascade effect, with each operation signaling the immediate upstream operation to produce a given amount of product. In a cell, the authorization to produce will be an empty spot for WIP at all machines except for the last one, which will get a card, signal, or container. Eventually, the process of pulling reaches a machine or storage area that is shared with other value streams. At this point, the ability to supply parts needs to be at least equal to the inverse of the combined takt times of the downstream processes (i.e. if a process serves three value streams each with a takt time of one minute, then the shared resource must have a cycle time of no more than 20 seconds times the standard pull quantity, and sufficient time for changeover and other scheduled downtime). If the system lacks adequate balance, then the bottleneck operation may be the point of scheduling, pulling from upstream and pushing downstream (actually, flowing, since there is by definition more capacity than required). If there are no shared resources downstream, this will work well, with continuous flow through to shipping. When there are shared resources between the bottleneck and shipping, then there will some other point downstream of the bottleneck where continuous flow becomes possible, and this “pacemaker” operation is then the logical point to schedule the entire process.

An additional complication might exist when demand varies over time. This means that the calculated rate of flow, or the takt of the system, needs to vary. There are a number of solutions to this that can be used to make the pull system work, and they are generally familiar. They include putting more inventory between operations or in finished goods, using overtime to complete the day’s requirements, adding operators to cells, using additional pieces of equipment at operations that are at capacity, and so forth. Since the situation is not expected to be permanent, these changes must be implemented in a flexible manner. Below, the concept of load leveling is also discussed.

Some Advantages of Pull over Push

Of course, converting traditional manufacturing to demand-based manufacturing is not quite as simple as pulling instead of pushing. While a push system bases production authorization on a calculation of what will be needed in the future, if everything goes according to plan, pull authorizes production only when material is actually needed (with a calculation of
when this material should be made, which assumes that the producing operation will function as planned). Neither system deliberately tries to create too much inventory. But whereas a push system frequently creates too much, and in some places too little, a pull system is most likely to produce too little (or equivalently, too late) when it fails.

As indicated above, balance and reliability are important to a successful pull system. As well, shared resources must be able to serve all value streams at the required rate. As we have seen, this requires substantially faster setups. Consider the steps that need to be taken to convert a system from push to pull, and simultaneously to reduce the cost of production (since pull by itself is not necessarily a lower-cost mode of production).

In trying to pull, it will quickly become apparent that predictability and reliability are key factors in meeting demand. It is, of course, just as important for push systems to be reliable, if they are to be cost effective. However, it is because of the larger amounts of inventory usually contained in push systems that reliability receives less attention. Given sufficient inventory, and a tendency to ignore the costs of inventory, the sporadic replenishment of inventory appears to be adequate in push systems. As soon as inventory is only produced to replace what has already been consumed, reliability becomes much more important, since there will be constant shortages under a system that produces sporadically and unpredictably. With a high degree of reliability, it is possible to calculate lead times, and hence the number of production authorizations (“kanbans”) that the system requires to function acceptably.

Predictability will not, by itself, reduce costs sufficiently to justify instituting a pull system (because a predictable push system holds the same level of inventory, and may even function more smoothly). The next requirement for a pull system is balance. This means that each step in the process takes the same time to complete a unit of production. When we say “the same time”, there is some room for variation if the system produces a “family” of parts. The accepted guideline for variation is a 30% spread from the shortest to the longest. A family of products is produced by a group of production centers when there is insufficient business for just one product. In a pull system, product design and sales will try to achieve, as close as possible, a much narrower variation in time than the outside limit of 30%. Here a significant distinction between the push approach and the pull approach is evident. Push measures utilization, whereas pull measures flow. This explains the two control systems’ very different approaches to what is wasteful, and ultimately the lower cost of pull systems.

The final requirement of cost reduction through pull is stability. While the market may be chaotic, it is possible to impose a large amount of order through leveling the production requirements. This leads to a highly stable load on the system, and furthermore, a repeatable pattern of demand. A pull system becomes highly responsive when it is capable of short runs. This is achieved through quick setup, productive maintenance, multi-skilled operators, and so forth. In this way, all products will be available at all times, and whatever the customer uses will be replaced in quick order. On the other side of the equation, there will be an attempt to impose some sort of order on the chaos of the market through rewarding steady demand (rather than looking for larger orders than the market really needs through volume discounts, which the push system strives for).

Assuming that everything works exactly as planned, there ought not to be any significant difference, from the point of view of inventory levels, between pull and push. In fact, if the planning horizon is short enough, push would appear to turn into a roundabout pull system. But this assumption is rarely correct. The reason push systems tend to create excess inventory is that the penalty for failure is not immediately apparent. Few managers complain of too much inventory; too little, on the other hand, is cause for serious concern, because shipments will be late, and customers dissatisfied. So why does anyone want to move to a pull system? The clear cost advantages outlined above are one reason. There is also an advantage in quality level, because the system simply will not work without flawless quality (since scrap will throw off the timing of the pull signal). Therefore, assuming that a pull system can be set up that delivers unfailingly on time, there is a clear competitive advantage. So one of the key advantages of pull is that it forces a producer to constantly strive for perfection. A producer using a push system is only forced to improve if it stands to lose significant business due to high cost and low quality. One of the ironies of push systems is that they are also poor at on time delivery, despite an excess of inventory – because the inventory on hand is often not what the customer wants.
The Role of Scheduling in a Pull System

Few supplier-customer relationships use a pull system exclusively between companies to authorize production (though it is possible). Typically, customers provide at least two levels of information about their needs: long term and short term. Long-term needs are expressed in forecasts, and short term needs in firm orders (with firm shipping dates, though customers can be fickle and change what they want at the last minute). There may also be a middle term need, expressed as authorization to produce, but without authorization to ship. The supplier must therefore translate this information into authorization to produce. While the kanban authorizes production, there still has to be a basis for calculating the number of cards (or other signals). Ideally, the middle term forecast is used to determine the daily quantity that the customer will need, but short or long term forecasts can be used as well. The downside of using these latter forecasts is that safety stocks or overcapacity may be necessary in order to meet the daily (or other periodic) shipment quantity, if the plan falls short of actual requirements. The number of signals in the system as a whole is based on production lead time, and the period of planning.

With a plan in place for production, a schedule is developed for one process in the value stream. This is most often shipping, but if there is single-piece flow from a point upstream of shipping right to the dock, then the process just prior to the start of single-piece flow can also be the scheduling point. Given a schedule, the scheduling point issues a signal to the upstream process, which in turn does the same. At the beginning of each scheduling period (a shift, a day, or a week, whichever has steady demand), a new signal is issued by the scheduling point. The schedule is reviewed periodically (every month or two, for example, if the medium term forecast is being used), and the authorization quantities revised if necessary (for example, to adjust for seasonal demand, or if the economy changes pace). What is not seen in a pull system is constant change in the schedule.

Pull and the Family of Products

Within the family of products, a further refinement of the pull system is often implemented, called production leveling. To “level production” means to reduce the batch sizes of each of the part numbers in the family of products below the requirements of the planning period, and repeat production of all the part numbers in a regular manner. Assume that the planning period requires the following batches: AAAAAAAA BBBB BAB CCC (where each letter represents a standard kanban quantity, say 50 pieces, and the planning period is three days). Leveled production will set up the following production plan: AAA BB C AAA BB C AAA BB C. Each day now looks the same as the rest. The advantage of this is that setup, 5S, maintenance, and so forth become easier, because these activities happen at the same time each day. Furthermore, it becomes easy to notice any irregularity, so reaction and correction happen almost as soon as something goes wrong. This is one of the ways that visual control is built right into a pull system.
6. Planning for Improvement: The Future State Map

Hints for achieving pull

Kanban

The purpose of Lean is to flow product to the customer at the rate of consumption. Since it is rare for processes to be balanced so that all steps take exactly the same time, or located so that one process can directly control the activation of another, a system of signals is used to authorize processes to make what is needed in the quantities required. This authorization is referred to as a “kanban”.

There are a number of ways to use kanban. Typically a card (although it could be an empty inventory location, an empty parts container, a ball, or perhaps a fax or email or other electronic signal), a kanban provides authorization to produce or ship a certain quantity of material. There are several kinds:

1. A SIGNAL kanban is sent upstream, to an operation which is shared by a number of product families. Since the shared resource gets kanbans from several streams, it produces in a batch. The batch size is equal to the amount consumed by the sending operation during the time it takes to receive the material after sending the card. The time it takes to receive the material can be expressed as (production interval) + (transportation time), where production interval = (sum of changeover times) + (production times for all products that are produced regularly). There are a number of possible refinements to this, such as prioritization of product families, variation in the quantity per batch when demand changes, and so forth. If the shared resource has quality and downtime problems, additional material may be added to the batch to cover this. Additional safety stock may also be kept to allow for unforeseen quality problems and downtime (since there may be times when the usual quantity is exceeded). A signal kanban should be sent when an equivalent quantity or more is left at the sending operation (so that the material just runs out, at worst, when the new material arrives). Since signal kanbans usually go to shared resources, a consideration of prioritization must be made. The system can be first come, first served, or some products can have priority over others. In either case, of course, the shared resource must have sufficient capacity to serve all pulling streams; prioritization just allows some streams to be leaner than others. Prioritization may also be necessary, in an emergency, when demand suddenly goes up, or a quality problem appears. (Note, in general, that kanban systems work best under steady demand, and high quality and uptime).

2. PRODUCTION kanbans are like signal kanbans, except that they are sent only to dedicated upstream resources.

3. WITHDRAWAL kanbans are used to authorize movement of material from a supermarket or storage location, usually in small quantities (e.g. one small basket).

4. BATCH ARRIVAL is used if withdrawal happens in a large quantity, such as when a truck is loaded and sent to a customer.

5. SEQUENCED PULL BALL is a method of sending a requirement to a subassembly station (physically or electronically) for a particular subassembly to be built, needed in a given amount of time, when the subassembly has too many option variations to simply replenish a supermarket (because the supermarket would take up too much space, or material would sit ready for too long due to low demand). In in-sequence delivery from just in time suppliers, this kind of kanban is replaced by a “broadcast”, sent to all subassembly stations, with subassembly specifications and delivery time requirements.

6. FIFO LANES are a kind of kanban square. This means that an empty storage spot is itself authorization to produce a product for that square. The FIFO lane holds only identical products, and when one is used up at the downstream operation, the upstream operation produces another. The lane has a set allowable number of products that are used in a first in, first out order. When the FIFO lane only holds one item (or even none), we call this situation “single piece flow”. This is considered the ideal production system.

7. A SUPERMARKET is another form of kanban square, where one or more items are held in each opening (cubby hole), and a signal kanban is sent when the minimum quantity is reached.
Quick Setup

Shingo and Ohno credited their success in developing TPS in part to discovering how to reliably reduce changeover time (good part to good part) to an insignificant amount of time. Shingo called this SMED, or “single minute exchange of die”, since it was originally concerned with reducing changeover time on large stamping presses used for producing body side panels for cars. The idea was that a changeover of less than ten minutes (hence single minute, or a number of minutes expressible with a single digit) was insignificant in the calculation of the economic order quantity for a production run. This meant that just what was required could be produced economically, independently of what the setup time was.

The process that Shingo discovered for reducing setup time is an example of elegant industrial engineering. There are four steps to achieving SMED.

1. The entire process employed for changing over is analyzed, and individual steps are classified as “internal” or “external”. An external step is one that can be carried out independently of the state of the machine, i.e. it doesn’t require that the machine be stopped. For example, obtaining a tool, say a wrench, can be done whether the machine is running or not, and is therefore classified as external. Using the wrench to undo a bolt holding a die in the press, on the other hand, can only be done if the machine is stopped, and is therefore internal. Having thus classified all steps, the changeover is replanned so that all external steps are done before or after the machine is stopped, and only internal steps are carried out during the changeover. This is of course pure common sense, but rarely common practice. In many cases, setup time is reduced by 50% just by separating the external from the internal.

2. The next step in the setup reduction process is to look for internal steps that can be redesigned to make them external. These are usually found when a setup consists of building up a number of components, or cleaning needs to be done before a new part can be made. The usual approach is to replace a larger part of the machine during changeover. For example, it is faster to replace a tire already mounted on a rim than it is to take the old tire off a rim, and put a new one on, and then remount the rim. Building a setup on a sub-plate off-line thus requires only the exchange of sub-plates rather than exchanging each piece. Once the cost of long changeovers is factored in, the additional cost of having more than one of a setup component is negligible.

3. Conversion of internal to external is further carried out in the third step, elimination of adjustment. Here the emphasis is on standardizing those aspects of a setup that require machine adjustment, so that a single setting can be left alone. Typically, adjustments could be made to height, location, pressure, temperature, and so forth. It may be possible to redesign the process so that common tooling can be used for most operations. Redesign of the tooling, process, or materials looks to eliminate the need to adjust, and thus significantly reduces changeover time. By eliminating adjustment, it usually follows that much less trial time is required in order to get a good part (first off), and overall quality improves. Steps 2 and 3 generally reduce changeover time by a further 30%.

4. The final step in reducing changeover time is automation and streamlining of the remaining internal elements. Removal and replacement, fastening, material change, and similar aspects are easily automated since they have already been standardized. It is important to leave automation to the last, since there will otherwise be a tendency to automate unnecessary or externalizable steps.

Quality at the source

Shingo also developed the theory of quality at the source (which he called “zero quality control”), and mistake proofing (often referred to as “poka yoke”, the Japanese for mistake proofing). Shingo proposed that mistakes will happen, but that it is unnecessary for them to find their way to the customer, or even the next operation. He criticized sorting (i.e. final inspection), because it is wasteful, and doesn’t lead to correction, as well as statistical process control (SPC), because it merely states the likelihood of a certain number of defects being present in a lot of a given quantity. He felt that the logical place to deal with mistakes is where they occur, by inspecting 100% of product as it is being produced, and providing immediate feedback to the operator when a mistake is detected. This should be done either by preventing mistakes from happening, or by alerting the operator to the fact that a mistake has just occurred. His approach was to build into production
equipment devices to prevent mistakes (for example, incorrect placement of the work piece can be prevented through the use of locating pins in a jig), or sensing mechanisms to detect that a mistake has occurred (for example, a switch will trigger a warning sound if a part is missing on an assembly).

Problem solving teams with the attitude that all mistakes are opportunities to do better, and that are in possession of good analytic skills, will over time eliminate most sources of mistakes, if they go immediately to the operation where a mistake happens, and, working with the operator, develop counter-measures.

**Breakdown elimination**

Pull production systems strive for zero breakdowns. The way to achieve this is through total productive maintenance (TPM). TPM consists of three approaches to breakdown prevention: (1) inspection and basic maintenance by operators (often called autonomous maintenance), (2) constant equipment improvement to eliminate weaknesses in design and materials, and (3) preventive and predictive maintenance, to ensure that equipment operates at an optimal level at all times.

Autonomous maintenance usually consists of training operators to carry out basic inspection and maintenance tasks, under the supervision of the maintenance group. These tasks are called C.L.A.I.R., or cleaning, lubricating, adjusting, inspecting, and repairing. This approach to breakdown prevention has the advantage of freeing highly skilled maintenance personnel from mundane tasks, and simultaneously getting operators more involved with the equipment they use. This makes the operators more attentive to problems, and they can alert maintenance to have the problems corrected before they cause a breakdown. There is an obvious overlap with the 5S system here (as there is with quick setup and quality at the source).

Equipment improvement often starts right with the purchase of new production equipment (to the consternation of managers who insist that it is up to equipment suppliers to fix all problems, and who try to push warranties to the limit). Competition often pushes equipment builders to use cheap bearings and other inferior components, and where the particular use warrants it, they will be replaced. Older equipment will be rebuilt, and design problems corrected.

Preventive and predictive maintenance bases maintenance activities on usage and testing of equipment condition, rather than just basing them on elapsed time. This kind of data based maintenance has as its goal to repair or replace components before they cause a breakdown, and to plan maintenance activities to ensure optimum performance rather than just operability.

Again, it is the attitude and focus on a different set of measurables that sets the pull system apart from the push mentality. A pull system cannot afford to lose even a single machine cycle, since this would destroy balance and flow. A push system typically assumes that inventory is a good thing, and doesn’t see a problem with (normal levels of) downtime, since with sufficient inventory, a breakdown will not affect other parts of the process.

**Cells**

Cells are groups of machines that perform quick, incremental, changes to material (but don’t do all operations that transform raw material to finished goods – that would be a focused factory). There are a number of characteristics of these machines and their relationship that make for successful cells:

- The cycle time of all but the last operation should be as nearly equal as possible, with the last operation faster than the rest.
- Cycle time (plus changeover time) should be less than or equal to demand frequency (takt time).
- Inventory between operations is generally limited to one piece.
- The number of operators may be less than the number of operations (with operators moving with the work, and tending several machines).
• Where there are fewer operators than operations, the combined cycle time divided by the number of operators must be less than the takt time, and each cycle time must also be less than the takt time.

• Uptime and yield should be 100% or nearly so.

• A safety stock will be required if uptime percentage or quality is low (and overtime is used to replenish the safety stock; many lean operations are designed to be utilized at no more than 80% of capacity, with spare time used for 5S, improvement activities and training).

• A ‘U’ shaped layout is often used, to make movement of operators easier, and to provide a single point of material handling in and out of the cell.

Suppliers

Many suppliers have already adjusted to the requirement for more frequent delivery. Often, to get small batch delivery, all one has to do is ask. But to make this change successful, suppliers need something in return: They need good information, steady schedules, and a non-adversarial relationship. It is a fact that companies don’t compete – supply chains do. Work with suppliers to extend the Lean value stream and its benefits.

Creating the Future State Map

Taking the above ideas for creating a flow system into account, the question becomes how to apply them in an actual situation. It has already been established that by selecting a planning period, requirements for lead time reduction are automatically imposed on each operation in the value stream.

Consider a planning period of one week. This means that all products in the family will be made each week, and that the quantities produced will be equal to the customer’s demand for a week. Now, setup times, quality levels, downtime, and so forth must be accommodated within this same planning period, while still leaving time for production. In the case of shared resources, other products may have a different planning period, but all the products from the family must be fitted into the production plan for the resource. If there are problems downstream, and product is not needed quite when it was thought it would be, the resource must be flexible enough to allow production any time during the planning period. For portions of the value stream that are exclusive to the family of products, determine if cells can be constructed out of operations that are nearly balanced. Where balance is not possible, the solution is to create a focused factory, by moving equipment together in one area, combined with the use of supermarkets to reduce inventory levels. It may also be necessary to set up some safety stocks, especially at the beginning of the transformation, to balance out those machines prone to breakdown and quality problems.

Start the construction of the future state map by deciding where the value stream will receive the schedule. If this isn’t obvious, choose shipping. Shipping usually works in a batch mode, i.e. it processes large quantities in a short period of time. It will need to pull from a supermarket to fulfill the shipping requirement. After the pull has taken place, replenishment signals are sent to the immediate upstream operation. Now, determine successively, moving back to raw materials, how each operation will let its supply operation know what to produce. What is the smallest batch that this upstream operation can reasonably make, given its characteristics? If it is one or just a few, there is an opportunity for constructing a cell. If there is only imbalance in setup, but the upstream operation is faster, then a standard pull signal will work when there are only a few part numbers in the product family. If there are more part numbers, a supermarket will probably work better.

With a first cut of the map constructed, it can be tested to see if the proposed setup times, downtimes, kanban quantities, and operating times yield a satisfactory result. The same iterative process used to test the current state map can be employed. This time the goal is to make the map conform to the target production lead time, and overall inventory in the system. Look for the same problems as before – queues that don’t behave as one would expect, for example – to pinpoint operations that need to be reviewed. Carry out sensitivity analysis. Make sure to document all assumptions and decisions made in getting to the final value stream map.
Finally, don’t expect to get it right the first time. It takes time to plan the future, and the final results may be surprising. The future state map will certainly be better planned and tested than if it had simply been drawn on paper, with a few calculations about the relationship of one operation to the one next to it.
7. Managing the Transition: From Push to Pull

Where Do You Start?

One of the choices to be made in initiating the actual transformation from the current production system to the system designed in the Future State Map, is whether to improve individual processes first, or create flow. There are many good arguments for creating flow first. The key one is that this is the aim of the whole process. Secondly, if it isn’t started right away, there will be a tendency to keep improving individual processes, and implementation of flow will never happen. It should be said, on the other hand, that an insistence on starting with the creation of a pull system has led to chaos in some instances, and a poor opinion of this kind of production control.

An assessment of the level of reliability of the system, the response time to correct problems, and the resultant length of the planning period should guide the choice. The worse the reliability, of course, the longer the planning period needs to be, and the more overcapacity will be required in order to meet requirements.

Starting with Flow

If attempting to create a flow system first, the first choice is the period of planning. This is not the takt, in the strict sense. As we saw, the takt time is the time it takes for the customer to use one item of production. For example, if the customer makes 500 cars in an eight hour shift, then the takt time is 57 seconds (8 hours x 3600 seconds per hour / 500 cars per eight hours = 57.6 seconds per car). For the manufacturer of the engine blocks, the takt time is 57 seconds (since there is one engine block per car); for the tire manufacturer, the takt time is 14 seconds (57 seconds per car / 4 wheels per car = 14 seconds per wheel). The planning period is the time allotted to produce what the customer will require for that period. Usually, at the start of a lean implementation, the rate of consumption (the takt) differs from the rate of production, and the level of reliability will not allow continuous production at the rate of consumption. A system might be able to produce a tire in 10 seconds, but if its changeover time and downtime are high, then a tire is not necessarily produced every 14 seconds.

For example, in a two week period, the same customer will need 40,000 tires. But with a five hour changeover, an uptime of 80%, four different part numbers, and a quality level of 95%, a two week planning period is required in order to meet requirements and still have time for planned maintenance and continuous improvement activities. So in this case, a two week period is chosen as the kanban quantity. Kanbans can be set up either with the quantity per part number required in two weeks, or more or fewer kanbans of a standard quantity, depending upon the requirements (i.e. one kanban can be for 4,000 of a particular part number, or each kanban can be for 500 tires, and there will be eight for the part number in question, and perhaps other quantities for other part numbers). With a long setup, and a two-week planning period, it probably makes sense to create a single kanban for the quantity required. The key point to remember is that once the required quantity has been produced, any time remaining must not be used for production, but should be used for continuous improvement activities, or preventive maintenance. Exceptions are if safety stocks were depleted and need to be replaced.

As mentioned above, the scheduling point must also be determined. Unless there already is one-piece flow from some point in the value stream all the way to the end of the process, the scheduling point should be shipping. The warehouse being shipped from must of course have at least one period’s worth of material, as it can take a full period to replenish.

In essence, a pull system has now been achieved. The key to the system is that it makes it impossible to overproduce. It will, in addition, be as lean as it can be, given the level of performance of the value stream. Chances are, however, that at this stage shipments will be missed, and the situation will be no better than with a traditional system. That is because no action has yet been taken to deal with variation. Statistics being what they are, a spike of high downtime or low quality will eventually appear.
With this in mind, waste elimination activities must start soon after the pull system has been implemented. In the meantime, safety stocks and a full warehouse can be used to prevent inconveniencing the customer. Waste elimination will come from the various activities already examined, such as 5S, standard work, quality improvement programs, maintenance (TQM) programs, and changeover reduction programs. As these programs prove effective, the planning period will decrease, and the shorter the planning period, the higher the profitability. At some point, suppliers will become the barrier to getting leaner.

**Who Does What?**

Many organizations will take on a Lean initiative, while already having continuous improvement groups in place. These groups are a kind of “quality circle”, often dealing with issues arising from “business operating system” measurables. When lean manufacturing is adopted as a strategy for competitiveness, it is important to review all measurables. There must be a coherent set of measures, supporting all aspects of the lean strategy (the so called “balanced scorecard”), if the organization is to be successful in becoming lean. With this in place, it is clear that the continuous improvement groups are the implementers of waste elimination requirements. The lean implementation steering group decides on the rate of planning period reduction, and the continuous improvement groups work to meet the target levels for the various barriers to the goal. They will thus have goals for unscheduled downtime (reduce setup time, improve mean time between failure, reduce mean time to repair, etc.), for cycle time losses (implement standard work, 5S, and standard materials, etc.), for quality levels, and for machine cycle times and line balance. Often, improvements are implemented in kaizen events – quick, intense periods of improvement.

**The Case for Waste Elimination First**

In some cases, especially with a long value stream that includes a number of shared resources, it makes sense to spend a certain amount of time reducing waste and improving standardization before introducing flow. This is especially true for the first value stream to be converted. This is because the shared resources will be difficult to schedule so as to ensure that all part numbers are produced within the planning period, while serving all other value streams that still work in the traditional manner. It is not a good idea, from the point of view of creating a positive image for Lean manufacturing, if the products of the Lean value stream get priority, and other value streams end up causing customer complaints due to late deliveries. As indicated above in the section on target maps, it may also be necessary to spend some time reducing waste, in order to get the savings required to fund a lean implementation.

**Project Management**

Regardless of the specific approach, it is important to use good project management principles to ensure successful implementation. This means:

1. Start with a plan, and ensure the entire organization supports it. This includes:
   a. A thorough assessment of the organization, to determine the barriers and aids to adopting lean, and current status of the production system.
   b. An implementation plan with goals, objectives, and milestones.
   c. An organization structure and policies to support lean manufacturing.
   d. Most importantly, communication, communication, communication!
2. Select the initial application area, and move quickly to put lean manufacturing in place. The approach is “ready, fire, aim” - action is better than analysis. With good support, the approach can be refined along the way, to do an even better job next time.
3. Move on to new areas when the current initiative starts to get results. This is the subject of a more complete study of Lean, whereas this text deals with the basic understanding of lean, and the role of value stream mapping.
Appendix 1: Case Studies

What Shape is Your Value Stream?

Value streams come in three basic shapes, funnel, pipe, and pyramid, according to the volume or batch size entering and exiting the value stream. Most lean future state value streams are pipe shaped, in order to promote flow. The size of the pipe (the daily production capacity) is set according to customer demand for the family of products. The examples below will show you how to create current and future states for each of these kinds of value streams.

**FUNNEL**

Funnel shaped value streams are those with large incoming amounts of material, and fast initial processing. They narrow progressively, as batches become smaller due to part number diversity and manual processing. Common examples are plastic injection molding, metal stamping, warehousing, custom assembly, and multiplex movie theatres. In all cases, the incoming work starts out through common processes (ticketing, admission, resin drying, receiving, etc.), and then gets differentiated (separation or assembly into specific SKUs, distribution to specialists, selection of a specific movie, etc.). Most funnel value streams are straight line flows after the initial common processes. To create lean value streams from funnels, we identify families of products/processes, and separate these out from other products/processes as early as possible. A common approach is to dedicate resources to a specific family, usually by replacing large, fast processes with smaller, right-sized processes. This creates a pipe, with balanced capacity throughout the value stream. Where this is not possible, the large, fast processes are separated from the balanced flow in the pipe by a supermarket pull system, which activates the upstream resource only to the extent required by the pacemaker process. With effective setup reduction, this works well.

Our example of a funnel shaped value stream is a molding and assembly plant producing brake light and turn signal lenses for a motor cycle assembler. Resin arrives in bulk, and is dried in bulk, every two weeks. Both products are made on the same press, which runs one week’s worth of brake light lenses, and then one week’s worth of turn signals. The molds are multi-cavity (i.e. producing more than one part per cycle – two brake light lenses, or four turn signal lenses per cycle). After molding, lenses are assembled to make the brake light assembly or the turn signal bodies, which are then tested and packed for shipping. There is a 10% failure rate (scrap) at testing. The manufacturer runs MRP once a week, to determine what materials to order.
This is a straight line flow with no need for prior simplification in order to map it. It is necessary to convert pounds of resin into units of output. Each lens uses ½ lb. of resin. Finally, since this press produces both brake light and turn signal lenses, we set the EPE to one week, since one week’s worth of these lenses are made per run. Based on cycle time, we know that the press can handle two products. There is scrap produced at testing. With a cycle time of 10 seconds and 5 seconds respectively, testing and packing only require one shift per day. In mapping the assembly process, on the other hand, overtime is required to meet the target daily quantity. Of course, overtime may be scheduled in specific hourly quantities, according to regulations or union rules – in which case the appropriate value should be entered.

From a lean standpoint, there are a number of issues with this production system. Funnels have large starting batches, and we would like to change them to pipes, since they have continuous flow. The usual causes of funnels are batch processes, long setups, and weekly MRP runs. In this case, the delivery and drying of resin is carried out in a large batch. To change the shape of the funnel to a pipe, we implement a technology change (right-sized technology) by integrating drying and molding (thus eliminating the need for bulk delivery, which served bulk drying), and set up a cell for assembly, testing, and packing, which replenishes what is shipped from a finished goods supermarket. With a takt time of 57.4 seconds, we can assemble in a day what we will ship in a day. With uptime improvement and setup reduction at the press, we can also mold in a day what we need to assemble in a day. We also replace MRP with a “go see” reorder process, using simple min/max indicators. There is a bit of a problem with testing and packing, since this only requires 15 seconds per unit. We need to build a second tester, to avoid changeover, but this allows us to use one operator to test and pack both brake lights and turn signal lenses.
lights, thus getting the benefits of a cellular layout, while keeping all operators busy. We will run assembly of both kinds of lamps in parallel, feeding the test and pack operator.
PYRAMID

Pyramid shaped value streams are the opposite of funnels. They have slow, steady incoming materials and orders, and intermittent, large outputs. Common examples are accounts payable departments, outgoing flights, postal collection, custom assembly of large systems, university admissions, lottery systems, and film making. These processes may consist of many separate value streams that merge (e.g. props, location scouting, costumes, story boarding, etc., for a feature film that will only exist once all the film has been edited, or all the various activities necessary for a Space Shuttle launch), or many similar value streams that merge (processing of identical applications for a limited number of jobs, placements, awards, that culminate in a single announcement, or payables processing with a weekly cheque run). A common approach to making this kind of process lean is to separate out the jobs that can be automated or standardized, and replace the typical managerial approval steps with rules-based approval. This allows for continuous flow, or makes it possible to pull, since the time to complete a step is now standardized.

Our example of a pyramid shaped value stream is an accounts payable process. Invoices arrive daily by mail from vendors (the average number received is 50). They are recorded in the ERP system, matched to purchase orders and delivery records, and sent to the original requisitioner for approval. Once approved, cheques are printed in a weekly batch, sent for a signature, and mailed out. The whole process takes a little over three weeks. One saving that this denies the company is the early pay discount (2%, 10 days) which most vendors offer. The long process even leads on occasion to missing the 30 day terms, and incurring late payment penalties. Paying its bills is not a problem for the company, and it wishes to have the best possible relationship with its suppliers.

While we know that the number of invoices received daily is not exactly 50, but varies somewhat, we use the average number. Since we normally study long term performance of a value stream, this does not present a problem.

This is a straight line flow. There are constraints in the processing of invoices due to several factors, such as waiting for a batch to complete, lack of availability of the resource (e.g. managers who travel, or are occupied with other tasks), and scheduling. We think of this in terms of batch processes, with set receiving and return intervals, to model processes requiring a varying quantity of material, from a minimum to a maximum. This contrasts with a process component, which has a fixed transfer lot size, and releases the material as soon as it has been processed, though its activity may also be controlled by the advanced scheduling method. The third means of controlling activity of a process is through the availability determination. It is used mainly when a process needs an operator to function, and this operator is doing other things (such as taking a break, or working elsewhere). The key to choosing which approach to use, is determining which means of constraining the map best matches reality.

There are many reasons for pyramid processes, some valid and some not. To make them lean, we look for ways to turn them into pipes. For example, intercity flights by air have this shape. The batch portion (the actual flight) is an economic necessity. Travel by local bus is much closer to a pipe, since more vehicles are economically feasible. There may also be an element of fairness involved. Lottery ticket sales take place a few at a time, but the draw is from the entire batch sold. The same happens with other selection processes. It is at the smaller scale that pyramids can be turned into pipes.
In the case of the accounts payable process, there is no economic or fairness reason to process in a batch. We find that 80% of invoices are for routine purchases, and always get approved by the requisitioner, so there is no need for the cumbersome process of approval. Invoices below a threshold amount also waste the signing manager’s time, and checks can be stamped with a signature instead. The idea of a weekly check run is a leftover from old technology and old management processes. The new process sorts the invoices into the 80% that can pass through a quick, semi-automated process, and the ones that need management attention. A cellular work area records, matches, and approves the invoice, which can then be paid immediately (thus taking advantage of suppliers’ discounts for early payment).
We map the cell and FIFO lane using the work cell component, and restricting the quantity in the downstream queue. Since we are mapping only the 80% of invoices that are routine, we now have only 40 per day, instead of 50. A separate mapl could be drawn to investigate the 20% of invoices requiring management approval, to see if additional improvements could be made to this process (such as, for example, putting approved invoices below a second threshold limit into the same print lane as the routine invoices, and thus eliminating the final hand written signature).
Pipe shaped value streams are already well on their way to becoming lean, because they have the basic shape required for flow. Common examples of pipe shaped processes are job shops, insurance claims processing, and many personal services. Unfortunately, most flows are “lumpy”, caused by attempts to get the supposed benefits of large batches and a low number of setups. A better approach is the “Runner, Repeater, Stranger” strategy, where separate pipes are created for common jobs that are received every day (e.g. business shirts for dry cleaning; “fender-bender” accident claims), jobs that are received in repeating patterns or for which the volume is less than daily (e.g. business suits; storm damage claims), and those that have no predictable pattern and require a longer or different process from the norm (e.g. wedding dresses; flooding and other major disaster claims). This eliminates setup from the Runner line, or standardizes it to the extent that batch sizes are no longer an issue. Using standardization and appropriately sized process equipment, Runners can be turned around in very short order (e.g. same day shirt service; on the spot insurance claims).

Our example of a pipe shaped value stream is a machine shop making mold components. Bars of steel are cut into billets, the billet is turned on a lathe to get the right diameter, and a number of features created on a mill, and long holes requiring extra precision are gun drilled. The part is then sent out to a subcontractor for heat treatment. On return, it is ground to an exact diameter and surface finish, and shipped to the mold maker. One of the parts made accounts for 60% of the total. There are a total of four different part numbers. The plant is laid out by function, and a batch of eight pieces is standard. The MRP system is used to schedule, and product is produced in lots of 32. It takes 25 days for an order to move through the production system.
To map this system, we first determine that all four components are part of the same family of products (they all have the same routing, and the only process where the cycle times are different for each part number is milling). We average out the cycle times (using the weighted average based on what percent of the total production each part number makes up). Since there are multiple machines at each step, we set the number of stations accordingly.

Since we already have essentially level flow, improvement of this value stream is accomplished by dedicating resources to the most popular part number, the Runner. We get orders for this product every day. The MRP system would lump all orders together, and try to achieve efficiency through setup avoidance based on long runs. We will avoid setup by not doing it at all. Once setup has been eliminated, we can reduce lot sizes to one, and strive for continuous flow. The actual value added time is just over two hours (plus heat treating), but it is taking us several weeks to complete an order. With continuous flow, throughput time will approach value added time. At the same time, there will be fewer setups for the other three part numbers, and it is likely that this value stream will improve as well. We also realize that we are not in the business of cutting up steel bars – this is only a necessity, and not really a value added activity. There is bar feed technology that we can add to the lathe, so that we can turn and separate the turned part on one machine (the so-called process step integration). Alternatively, we could purchase the parts already cut to size, if we did not wish to invest in this technology. In either case, we eliminate a queue from this part number’s value stream. Dedication of resources substantially reduces lead time, and building the same quantity each day, based in the mid-term forecast, stabilizes production so that each day looks like any other. This eliminates the need for daily scheduling and MRP runs, thus reducing cost further. Should the customer order more than normal, we will use overtime, or spare capacity on the remaining equipment, to ensure on-time delivery. Working
with the steel supplier to increase shipping frequency from a weekly delivery to a daily delivery substantially reduces lead time. With the increased prevalence of daily delivery of product, smaller manufacturers can often benefit from the work done by larger companies in setting up milk runs from their suppliers, without incurring increased delivery costs. We may also find that the heat treater can run two loads per day, if the loads are smaller. This further reduces throughput time, since we get back later in the day what went out in the morning. Quality control becomes much more important, since one piece of scrap would stop everything.

**Another Look at Acme Stamping**

We built a simulation of the “Acme Stamping Steering Bracket” value stream found in “Learning to See” by Rother and Shook. We were surprised when the simulation didn’t produce the same results as the map in the book. On closer examination, we found several problems. The map states that Assembly Station #1 has a cycle time of 62 seconds, and is available for 27,600 seconds per shift, on two a shift operation, five days per week. It is required to produce an average of 18,400 pieces per month, or about 920 pieces per day, 600 of which are left hand, and 320 are right hand. There is no changeover. But with 55,200 (27,600 x 2) seconds of production, only 890 pieces can be produced. Probably one of the above statements is wrong (maybe there is daily overtime of 30 minutes, or an extra shift on Saturday). That isn’t the point – what is important is that there is no immediate check on the data, to catch this kind of problem (even though the map looks very precise, timing availability in seconds). Another issue that baffled us was the amounts of inventory recorded for each queue. Our simulation never came close to duplicating these quantities. Again, this isn’t a big deal, but points to the fact that it is possible to count product on an abnormal day (right before the end of the month, or just after a holiday, for example), and then base improvement suggestions and benefits calculations on this abnormal data. The simulation gave us a very different picture of what is typically in queue, and actually showed a much longer lead time than the static value stream map.

Let’s look at the current state first. We have reproduced the value stream maps from “Learning to See” below. The current state of the steering bracket value stream at Acme Stamping is a funnel shape, and the first process is stamping. As expected, the press can make in a few hours what will be consumed at the next step, spot welding, over a two week period. The EPE (“every part every…”) method is used to describe how the press is scheduled. The rest of the value stream is already relatively well balanced, but the small variation in timing of the four work centres (39 seconds per part to 62 seconds per part) is enough to create “lumpiness”. In the current state, the solution is to buffer each step with several days inventory. Each machine receives a schedule (so many left, so many right, per week), and pushes work to the next operation. We are not given a batch size for spot welding. Shipping receives a daily shipping schedule, and ships from finished goods stock.
The future state suggested by “Learning to See” takes the spot welding and assembly processes, and creates a cell with three work stations. The cell replenishes a finished goods supermarket (instead of being scheduled by MRP), with kanban card activation, and in turn pulls stampings from the press via kanban cards. The press now runs steering bracket stampings daily, as setup time has been reduced significantly. It is unclear if the other value streams served by the press have also gone to kanban card scheduling, and if prioritization of value stream service is being used (as it must be with only an 85% uptime). In fact, the conversion of the press to a ten minute or shorter setup time would be quite expensive, since dies will have to have standard clamping heights, and locating points. We don’t know how many dies are normally run in the press, but it is probably 30 to 50, which is a large number of dies to standardize (otherwise, a large amount of idleness is tolerated, and the work ought to be outsourced). Once again, the future state looks great, but at what cost will it be achieved?
In our simulation, we built the value stream map using the supermarket and cell components. If we knew the capital cost to convert the value stream, we could enter it in the financial report, and then use various run lengths to determine the payback period. There are a number of other figures not provided by the original future state map, which we have made assumptions about, such as the exact cycle time of each of the elements of the cell, and whether the uptime of the press was improved.

As expected, the future state is financially better than the large-batch current state (as long as the capital cost of conversion is not included in the P&L statement), making a profit on an 8 week operation, as opposed to a loss. There are two sources of this profitability, lower inventory costs, and lower operation costs. Inventory costs are lower, because the future state lead time is about 4.5 days compared to the present 23 days. Operation costs are lower, because there are fewer operators required to run the revised system, and less time spent on value added activities (probably as the result of less material handling internal to the process). Interestingly enough, the profit contribution of improved operations is four (4) times that of the lower inventory costs. This decrease in cost could have been achieved without the expensive setup reduction program required to achieve reduced lead time. The reader may want to experiment with different versions of Acme Stamping’s future state, to see how much setup reduction, and how frequently to run the steering bracket stamping press, making reasonable assumptions about the cost of setup reduction, to get the best results financially, as well as in customer service.
Appendix 2: Simplification Strategies for Complex Flows

Once analysis of product families has been carried out, upwards of 70% of value streams are straight line flows (even if they include shared resources). For value streams that are complex, a great majority can be simplified, using the strategies outlined below.

The simplification strategies presented are:

1. Converging flows
2. Parallel flows with different cycle times
3. Diverging flows
4. Loop-back flows
5. Multiple parts per cycle

### 1. Converging Flows

**Examples**

- Two different raw materials combined (1+2)
- A sub-assembly added to a main part (1+2)
- Material from two streams with identical process steps after the converging process (i.e. they are not combined) (1,2)

**Simplification**

(1+2): Determine which branch sets the pace, and ignore the other one. Build a second value stream with the main flow as its customer, to see if improvements are required. Use a combined material value and adjusted process cost after the point of convergence.

(1,2): If the two flows are of separate products which will be run on the same line, then ignore one, use EPE scheduling according to the amount of time allowed for the product in question.
2. Parallel Flows with Different Cycle Times

Examples

- Several machines packaging identical product, but with different cycle times
- Lathe department handling a variety of incoming product, with a variety of cycle times

Simplification

Set the cycle time (C/T) of a single work centre to one over the sum of the normalized cycle times of each work centre \((1/(1/C:T:1A + 1/C:T:1B))\); set the changeover time (C/O) of a single work centre to one over the sum of the normalized changeover times of each work centre \((1/(1/C:O:1A + 1/C:O:1B))\). If one machine produces one piece in 1 minute and the other produces a piece in 2 minutes, then combined, they produce 3 parts in 2 minutes, or 1.5 part per minute, giving an effective cycle time of 2/3 minute (as shown above \(1/(1/1 + 1/2) = 2/3\))

3. Diverging Flow

Examples

- Material is packaged in different size containers for shipping
- One stream has additional processing required for additional features

Simplification

Ignore one branch, and use EPE scheduling to limit the rate of flow upstream of the point of divergence. Use a second value stream map to model the second flow. Combine the results for a complete look at the process.
4. Loop-Back Flow

Examples

- Milling is done both before and after heat treating
- Welding is done before drilling, and afterwards

Simplification

Insert the repeated operation twice in the straight-line flow. Use the cycle time for the operation where it occurs (if they differ). Put all changeovers at the first occurrence, and none at the second. Calculate the percent availability for each operation according to the cycle time divided by the combined cycle time (e.g. if they are the same, then the availability at each occurrence is 50%) and express as EPE.

5. Multiple Parts Per Cycle

Examples

- Multiple cavity molds, multiple parts per die
- Transportation of several people or parts in one trip
- Grinder with several parts mounted on bed

Simplification

Treat this like the case of multiple machines. The cycle time will be the same for all parts processed during one cycle, so the cycle time will just be divided by the number of items per cycle to give the effective cycle. If two pieces can be produced in one cycle of one minute, then the effective production rate is 30 seconds per part.