Push and Pull Production Systems

You say yes.
I say no.
You say stop,
and I say go, go, go!

– The Beatles

The Key Difference Between Push and Pull

Push Systems: schedule work releases based on demand.
• inherently due-date driven
• control release rate, observe WIP level

Pull Systems: authorize work releases based on system status.
• inherently rate driven
• control WIP level, observe throughput

What Pull is Not!

Make-to-Order:
• MRP with firm orders on MPS is make-to-order.
• But it does not limit WIP and is therefore a push system.

Make-to-Stock:
• Pull systems do replenish inventory voids.
• But jobs can be associated with customer orders.

Forecast Free:
• Toyota’s classic system made cars to forecasts.
• Use of takt times or production smoothing often involves production without firm orders (and hence forecasts).

Push and Pull Mechanics

Push systems do not limit WIP in the system.
Pull systems deliberately establish a limit on WIP.

Push and Pull Line Schematics

Push systems are not limited by WIP in the system.
Pull systems deliberately establish a limit on WIP.
Pulling with Kanban

Production cards authorize start of work.

Completed parts with cards enter outbound stockpoint.

When stock is removed, place production card in hold box.

Outbound stockpoint

Inventory/Order Interface

Concept:
• Make-to-stock and make-to-order can be used in same system.
• Dividing point is called the inventory/order interface.
• This is sometimes called the push/pull interface, but since WIP could be limited or unlimited in both segments, this is not a strictly accurate term.

Benefit: eliminate entire portion of cycle time seen by customers by building to stock.

Implementation:
• kanban
• late customization (postponement)

Example – Quick Taco Production Line

Example – IBM Panel Plant

Notes:
• Moving I/O interface closer to customer shortens leadtime seen by customer.
• Small number of core blanks presents opportunity to make them to stock.

Example – Custom Taco Production Line

Notes:
• I/O interface can differ by time of day (or season).
• I/O interface can differ by product.

Example – HP Deskjet Supply Chain

Notes:
• I/O interface located in markets to achieve quick response to customers.
• Delayed differentiation of products (power supplies for different countries) enables pooling of safety stocks.
I/O Interface Conclusions

Basic Tradeoff:
- responsiveness vs. inventory (time vs. money)
- moving PPI closer to customer increases responsiveness and (usually) inventory

Optimal Position of I/O Interface:
- need for responsiveness
- cost of carrying inventory ⇒ product diversification

Levers:
- product design ( postponement)
- process design (quick response manufacturing)

Pull Benefits Achieved by WIP Cap

Reduces Costs:
- prevents WIP explosions
- reduces average WIP
- reduces engineering changes

Improves Customer Service:
- reduces cycle time variability
- pressure to reduce sources of process variability
- promotes shorter lead times and better on-time performance

Improves Quality:
- pressure for higher quality
- improved defect detection
- improved communication

Maintains Flexibility:
- avoids early release (like air traffic control)
- less direct congestion
- less reliance on forecasts
- promotes floating capacity

Advantages of Pull Systems

Low Unit Cost:
- low inventory
- reduced space
- little rework

High External Quality:
- high internal quality
- pressure for good quality
- promotion of good quality (e.g., defect detection)

Good Customer Service:
- short cycle times
- steady, predictable output stream

Flexibility:
- avoids committing jobs too early
- encourages floating capacity

CONWIP

Assumptions:
1. Single routing
2. WIP measured in units

Mechanics: allow next job to enter line each time a job leaves (i.e., maintain a WIP level of n jobs in the line at all times).

Modeling:
- MRP looks like an open queueing network
- CONWIP looks like a closed queueing network
- Kanban looks like a closed queueing network with blocking

The Magic of Pull

Pulling Everywhere?

You don’t never make nothin’ and send it no place. Somebody has to come get it.

= Hall 1983

No! It’s the WIP Cap:
- Kanban – WIP cannot exceed number of cards
- "WIP explosions” are impossible

CONWIP Controller
CONWIP vs. Pure Push

Push/Pull Laws: A CONWIP system has the following advantages over an equivalent pure push system:

1) Observability: WIP is observable; capacity is not.
2) Efficiency: A CONWIP system requires less WIP on average to attain a given level of throughput.
3) Robustness: A profit function of the form
   \[ \text{Profit} = \frac{p}{\text{TH}} - h \text{WIP} \]
   is more sensitive to errors in TH than WIP.

CONWIP Efficiency Example

Equipment Data:
- 5 machines in tandem, all with capacity of one part/hr (\(u=\text{TH} \cdot te = \text{TH}\))
- exponential (moderate variability) process times

CONWIP System:
\[ \text{TH}(w) = \frac{w}{w + W_a - 1} + \frac{w}{w + 4} \]

Pure Push System:
\[ w/(TH) = \frac{5 - w}{1 - w} = 5 \frac{\text{TH}}{1 - \text{TH}} \]

CONWIP vs. Pure Push Comparisons

Key Question: what happens when we don’t choose optimum values (as we never will)?

Implementing Pull

Pull is Rigid: replenishing stocks quickly (just in time) requires level mix, volume, sequence.

JIT Practices:
- Support Rigidity: production smoothing/mix stabilization
- Mitigate Rigidity in Production System
  - capacity buffers
  - setup reduction
  - flexible labor
  - facility layout
  - product design ( postponement, etc.)
- Mitigate Rigidity in Organization
  - TQM
  - vendor management,
Capacity Buffers

**Motivation:** facilitate rapid replenishments with minimal WIP

**Benefits:**
- Protection against quota shortfalls
- Regular flow allows matching against customer demands
- Can be more economical in long run than WIP buffers in push systems

**Techniques:**
- Planned underutilization (e.g., use $u = 75\%$ in aggregate planning)
- Two shifting: $4 \rightarrow 8 \rightarrow 4 \rightarrow 8$
- Schedule dummy jobs to allow quick response to hot jobs

Setup Reduction

**Motivation:** Small lot sequences not feasible with large setups.

**Internal vs. External Setups:**
- External – performed while machine is still running
- Internal – performed while machine is down

**Approach:**
1. Separate the internal setup from the external setup.
2. Convert as much internal setup as possible to external setup.
3. Eliminate the adjustment process.
4. Abolish the setup itself (e.g., uniform product design, combined production, parallel machines).

Focused Factories

**Pareto Analysis:**
- Small percentage of SKU's represent large percentage of volume
- Large percentage of SKU's represent little volume but much complexity

**Dedicated Lines:**
- for families of high runners
- few setups
- can use pull effectively

**Job Shop Environment:**
- for low runners
- many setups
- poorer performance, but only on smaller portion of business
- may need to use push

Flexible Labor

**Cross-Trained Workers:**
- float where needed
- appreciate line-wide perspective
- provide more heads per problem area

**Shared Tasks:**
- can be done by adjacent stations
- reduces variability in tasks, and hence line stoppages/quality problems
- work can float to workers, or workers can float to work...

Push/Pull Takeaways

**Magic of Pull:** the WIP cap

**MTS/MTO Hybrids:** locating the I/O interface

**Logistical Benefits of Pull:**
- observability
- efficiency
- robustness (this is the key one)

**Overcoming Rigidity of Pull:**
- capacity buffers
- setup reduction
- flexible labor
- facility layout, etc.