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
Push vs. Pull Mechanics

**PUSH**

(Exogenous) Schedule

Job

Production Process

Push systems are inherently make-to-order.

**PULL**

(Endogenous) Status

Job

Production Process

Pull systems are inherently make-to-stock.

Pulling with Kanban

Outbound stockpoint

Completed parts with cards enter outbound stockpoint.

Production cards

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

Production card authorizes start of work.

Outbound stockpoint

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

Production card authorizes start of work.
Push and Pull Line Schematics

Pure Push (MRP)

Stock Point

Pure Pull (Kanban)

Stock Point

CONWIP

Authorization Signals

Full Containers

Push/Pull Interface

Eliminate: entire portion of cycle time by building to stock.

Requirements:

- Level demand.
- Relatively few distinct parts.
- Relatively constant product mix.

Implementation:

- kanban
- late customization (postponement)
Example - Custom Taco Production Line

Push/Pull Interface

Pull

Refrigerator

Cooking → Assembly → Packaging → Sales → Customer

Push/Pull Interface

Push

Example - Quick Taco Production Line

Push/Pull Interface

Pull

Refrigerator

Cooking → Assembly → Packaging → Sales → Customer

Push/Pull Interface

Push
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

Advantages of Pull Systems

Low Unit Cost:
• high throughput
• low inventory
• little rework

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

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

Flexibility:
• avoids committing jobs too early
• tolerates mix changes (within limits)
• encourages floating capacity
# Pull Benefits Achieved by WIP Cap

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

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

## Reduces Variability:
- reduces cycle time variability
- pressure to reduce sources of process time variability (e.g., long repair times)
- promotes improved customer service

## Maintains Flexibility:
- accommodates engineering changes
- less direct congestion
- less reliance on forecasts
- air traffic control analogy

## 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 \( m \) 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
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

   \[ \text{Profit} = \rho \text{Th} - \lambda \text{WIP} \]

   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 ($u=THt_e=TH$)
- exponential (moderate variability) process times

CONWIP System: looks like PWC, so

$$TH(w) = \frac{w}{w + W_0 - 1} = \frac{w}{w + 4}$$

Pure Push System: looks like series of M/M/1 queues, so

$$w(TH) = \frac{5u}{1-u} = \frac{5TH}{1-TH}$$

Comparison: WIP needed in CONWIP to match push throughput

$$w\left(\frac{w}{w+4}\right) = \frac{5(w/(w+4))}{1-(w/(w+4))} = \frac{5w}{4}$$

in this example, WIP is always 25% higher for same TH in push than in CONWIP

CONWIP Robustness Example

Profit Function: Profit = $pTH - hw$

CONWIP: Profit(w) = $p\left(\frac{w}{w+4}\right) - hw$ need to find “optimal” WIP level

Push: Profit(TH) = $pTH - h\left(\frac{5TH}{1-TH}\right)$ need to find “optimal” TH level (i.e., release rate)

Key Question: what happens when we don’t choose optimum values (as we never will)?
CONWIP vs. Pure Push Comparisons

Modeling CONWIP with Mean-Value Analysis

**Notation:**

- $u_j(w)$ = utilization of station $j$ in CONWIP line with WIP level $w$
- $CT_j(w)$ = cycle time at station $j$ in CONWIP line with WIP level $w$
- $CT(w) = \sum_{j=1}^{n} CT_j(w)$ = cycle time of CONWIP line with WIP level $w$
- $TH(w)$ = throughput of CONWIP line with WIP level $w$
- $WIP_j(w)$ = average WIP level at station $j$ in CONWIP line with WIP level $w$

**Basic Approach:** Compute performance measures for increasing $w$ assuming job arriving to line “sees” other jobs distributed according to average behavior with $w-1$ jobs.
Mean-Value Analysis Formulas

Starting with $WIP_j(0)=0$ and $TH(0)=0$, compute for $w=1,2,...$

$$CT_j(w) = \frac{t_j^e(j)}{2}[c_j^2(j) - 1]TH(w - 1) + [WIP_j(w - 1) + 1]r_e(j)$$

$$CT(w) = \sum_{j=1}^{n} CT_j(w)$$

$$TH(w) = \frac{w}{CT(w)}$$

$$WIP_j(w) = TH(w)CT_j(w)$$

Computing Inputs for MVA

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>STATION</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Process Time (hr)</td>
<td>$t_0$</td>
<td>0.090</td>
<td>0.090</td>
<td>0.090</td>
<td>0.090</td>
<td>0.090</td>
</tr>
<tr>
<td>Natural Process CV</td>
<td>$c_0^2$</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>Number of Machines</td>
<td>$m$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MTTF (hr)</td>
<td>$m_{\tau}$</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>MTTR (hr)</td>
<td>$m_{\tau}$</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Availability</td>
<td>$A$</td>
<td>0.990</td>
<td>0.990</td>
<td>0.962</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>Effective Process Time (failures only)</td>
<td>$t_0^e$</td>
<td>0.091</td>
<td>0.091</td>
<td>0.098</td>
<td>0.092</td>
<td>0.092</td>
</tr>
<tr>
<td>Eff Process CV (failures only)</td>
<td>$c_{0e}^2$</td>
<td>0.936</td>
<td>0.936</td>
<td>6.795</td>
<td>2.209</td>
<td>2.209</td>
</tr>
<tr>
<td>Jobs Between Setups</td>
<td>$N_s$</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
</tr>
<tr>
<td>Setup Time (hr)</td>
<td>$t_s$</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>Setup Time CV</td>
<td>$c_{e}^2$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Eff Process Time (failures+setups)</td>
<td>$t_e$</td>
<td>0.096</td>
<td>0.096</td>
<td>0.103</td>
<td>0.097</td>
<td>0.097</td>
</tr>
<tr>
<td>Eff Station Rate</td>
<td>$r_e$</td>
<td>10.428</td>
<td>10.428</td>
<td>9.731</td>
<td>10.331</td>
<td>10.331</td>
</tr>
<tr>
<td>Eff Process Time Var (failures+setups)</td>
<td>$\sigma_e^2$</td>
<td>0.013</td>
<td>0.013</td>
<td>0.070</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>Eff Process CV (failures+setups)</td>
<td>$c_{e}^2$</td>
<td>1.382</td>
<td>1.382</td>
<td>6.621</td>
<td>2.517</td>
<td>2.517</td>
</tr>
</tbody>
</table>

$T_0 = 0.488$

$W_x = 4.750$
### Output of MVA

<table>
<thead>
<tr>
<th>w</th>
<th>TH Actual</th>
<th>CT Actual</th>
<th>CT1(w)</th>
<th>CT2(w)</th>
<th>CT3(w)</th>
<th>CT4(w)</th>
<th>CT5(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.049</td>
<td>0.488</td>
<td>0.096</td>
<td>0.096</td>
<td>0.103</td>
<td>0.097</td>
<td>0.097</td>
</tr>
<tr>
<td>2</td>
<td>2.928</td>
<td>0.683</td>
<td>0.118</td>
<td>0.118</td>
<td>0.185</td>
<td>0.131</td>
<td>0.131</td>
</tr>
<tr>
<td>3</td>
<td>3.644</td>
<td>0.823</td>
<td>0.134</td>
<td>0.134</td>
<td>0.245</td>
<td>0.155</td>
<td>0.155</td>
</tr>
<tr>
<td>4</td>
<td>4.185</td>
<td>0.956</td>
<td>0.149</td>
<td>0.149</td>
<td>0.303</td>
<td>0.177</td>
<td>0.177</td>
</tr>
<tr>
<td>5</td>
<td>4.629</td>
<td>1.080</td>
<td>0.163</td>
<td>0.163</td>
<td>0.357</td>
<td>0.198</td>
<td>0.198</td>
</tr>
<tr>
<td>6</td>
<td>5.000</td>
<td>1.200</td>
<td>0.176</td>
<td>0.176</td>
<td>0.410</td>
<td>0.219</td>
<td>0.219</td>
</tr>
<tr>
<td>7</td>
<td>5.317</td>
<td>1.317</td>
<td>0.189</td>
<td>0.189</td>
<td>0.462</td>
<td>0.238</td>
<td>0.238</td>
</tr>
<tr>
<td>8</td>
<td>5.592</td>
<td>1.431</td>
<td>0.202</td>
<td>0.202</td>
<td>0.513</td>
<td>0.257</td>
<td>0.257</td>
</tr>
<tr>
<td>9</td>
<td>5.834</td>
<td>1.543</td>
<td>0.214</td>
<td>0.214</td>
<td>0.563</td>
<td>0.276</td>
<td>0.276</td>
</tr>
<tr>
<td>10</td>
<td>6.049</td>
<td>1.653</td>
<td>0.226</td>
<td>0.226</td>
<td>0.614</td>
<td>0.294</td>
<td>0.294</td>
</tr>
<tr>
<td>11</td>
<td>6.241</td>
<td>1.763</td>
<td>0.237</td>
<td>0.237</td>
<td>0.664</td>
<td>0.312</td>
<td>0.312</td>
</tr>
<tr>
<td>12</td>
<td>6.414</td>
<td>1.871</td>
<td>0.249</td>
<td>0.249</td>
<td>0.714</td>
<td>0.330</td>
<td>0.330</td>
</tr>
<tr>
<td>13</td>
<td>6.572</td>
<td>1.978</td>
<td>0.260</td>
<td>0.260</td>
<td>0.763</td>
<td>0.347</td>
<td>0.347</td>
</tr>
<tr>
<td>14</td>
<td>6.715</td>
<td>2.085</td>
<td>0.271</td>
<td>0.271</td>
<td>0.813</td>
<td>0.364</td>
<td>0.364</td>
</tr>
<tr>
<td>15</td>
<td>6.846</td>
<td>2.191</td>
<td>0.283</td>
<td>0.283</td>
<td>0.863</td>
<td>0.381</td>
<td>0.381</td>
</tr>
</tbody>
</table>

### Using MVA to Evaluate Line Performance

![Graph showing MVA output](http://factory-physics.com)
Implementing Pull

Pull is Rigid:
- replenish stocks quickly (just in time)
- level mix, volume, sequence

JIT Practices
- capacity buffers
- setup reduction
- flexible labor
- facility layout

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 – 8 – 4 – 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 as possible of the internal setup to the external setup
3. Eliminate the adjustment process
4. Abolish the setup itself (e.g., uniform product design, combined production, parallel machines)

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..."
Cellular Layout

Advantages:
- Better flow control
- Improved material handling (smaller transfer batches)
- Ease of communication (e.g., for floating labor)

Challenges:
- May require duplicate equipment
- Product to cell assignment

Inbound Stock  Outbound Stock

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
Push/Pull Takeaways

Magic of Pull: the WIP cap

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

Overcoming Rigidity of Pull:
- capacity buffers
- setup reduction
- flexible labor
- facility layout
- many others (postponement, push/pull hybrids, etc… )