Qualitative Observations on Production Line Behavior —
Intuitive and Counter-Intuitive Explanations

Stanley B. Gershwin

Department of Mechanical Engineering
Massachusetts Institute of Technology

4th International Conference on Information Systems, Logistics and Supply Chain
Quebec, Canada

August 26-29, 2012
Outline

- Introduction
- General Observations
- Infinite Buffers
- Two-Machine Lines
- Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$
- Long Lines — Optimization
- Conclusions
Introduction

Objectives:

▶ To describe interesting behaviors, properties, and phenomena of flow line models.
▶ To propose explanations.
▶ To provide intuition.
Kinds of Systems

Focus:

- Flow lines
- Randomness due to unreliable machines
- Finite buffers
- Propagation of failures through starvation and blockage
Kinds of Systems

Performance measures:

- Production rate $P$, a benefit
  - Production rate increases with buffer sizes.
- Average in-process inventory $\bar{n}_i$, a cost
  - In-process inventory varies with buffer sizes.
- Other costs: machines and buffer space
Methodology

Models:

- Lines are modeled as Markov processes.
Methodology

Models:

- Lines are modeled as Markov processes.
- Material and time can be discrete or continuous.
Methodology

Models:

- Lines are modeled as Markov processes.
- Material and time can be discrete or continuous.
- Machines fail and are repaired at random times.
Methodology

Models:

- Lines are modeled as Markov processes.
- Material and time can be discrete or continuous.
- Machines fail and are repaired at random times.
- Models used here have geometric or exponential distributions of up- and down-time.

Notation:
- \( p = 1/\text{MTTF} \) is failure probability or failure rate.
- \( r = 1/\text{MTTR} \) is repair probability or repair rate.
Performance evaluation:

- Two-machine lines are evaluated analytically, exactly. We obtain the steady-state probability distribution and use it to determine the average production rate and buffer level.
Methodology

Performance evaluation:

- Two-machine lines are evaluated analytically, exactly. We obtain the steady-state probability distribution and use it to determine the average production rate and buffer level.
- *Short lines can be evaluated numerically, exactly.*
Methodology

Performance evaluation:

- Two-machine lines are evaluated analytically, exactly. We obtain the steady-state probability distribution and use it to determine the average production rate and buffer level.
- Short lines can be evaluated numerically, exactly.
- Long lines are evaluated analytically, approximately.
Methodology

Long lines are approximately evaluated by decomposition.
In discrete models, $N_i$ can be treated as continuous because $P(N)$ and $\bar{n}(N)$ change only a little when $N_i$ changes by 1.

Monotonicity and concavity have been proven for some 2-machine line models.

$P(N)$ appears to be monotonic and concave for longer lines.
Infinite Buffers

Single bottleneck
Infinite Buffers

- The system is not in steady state.
- An infinite amount of inventory accumulates in the buffer upstream of the bottleneck.
- A finite amount of inventory appears downstream of the bottleneck.
Infinite Buffers

Two bottlenecks
Infinite Buffers

- The *second bottleneck* is the slowest machine upstream of the bottleneck. An infinite amount of inventory accumulates just upstream of it.
- A finite amount of inventory appears between the second bottleneck and the machine upstream of the first bottleneck.
- Et cetera.
Infinite Buffers

Two bottlenecks

Qualitative Observations on Production Line Behavior

Copyright ©2012 Stanley B. Gershwin. All rights reserved.
Infinite Buffers

- Each slope is the *difference* between the production rate of the immediate downstream bottleneck and the next upstream bottleneck.
- The largest accumulation of inventory is located at the largest difference, not (necessarily) the first bottleneck.
Two-Machine Lines

Simulations

\[ r_1 = .1, \quad p_1 = .01, \quad r_2 = .1, \quad p_2 = .01, \quad N = 10 \]

Buffer size is the same as MTTR. Blockage and starvation are frequent.
Two-Machine Lines
Simulations

\[
\begin{align*}
  r_1 &= .1, \quad p_1 = .01, \quad r_2 = .1, \quad p_2 = .01, \quad N = 100
\end{align*}
\]

Buffer size is $10 \times$ MTTR. Blockage and starvation are infrequent.
\[ r_i = .1, \; i = 1, 2, \; p_1 = .02, \; p_2 = .01, \; N = 100 \]

*First machine is a bottleneck. Blockage is infrequent. Starvation is frequent.*
Two-Machine Lines
Simulations

\[ r_i = 0.1, \ i = 1, 2, \ p_1 = 0.01, \ p_2 = 0.02, \ N = 100 \]

*Second machine is a bottleneck. Starvation is infrequent. Blockage is frequent.*
Probabilities of the buffer being empty or full are much larger than probabilities of intermediate buffer levels.
As $N \to \infty$,

- Production rate approaches an upper limit.
- If the first machine is a bottleneck, average inventory $\bar{n}$ approaches an upper limit.
- If the second machine is a bottleneck, $N - \bar{n}$ approaches an upper limit.
- If the machines are identical, $\bar{n} = N/2$.

$\bar{n}$ increases as the first machine becomes faster (i.e., more productive).
Two models. Both have exponentially distributed repair and failure times.

- **Exponential**: discrete material, exponentially distributed operating times (mean operation time $= \mu$).
- **Continuous**: continuous material, constant flow through machines when operational (flow rate $= \mu$).

Parameters of the models are the same.
Third model, *No-variability limit*. 

Same model as continuous but perfectly reliable and $\mu$ adjusted so that isolated production rates of corresponding machines are the same.
Observations:

- Variability of exponential model > variability of continuous model
  > variability of no-variability limit.
- Production rate is greatest when variability is least.
- Average inventory is closer to 0 or $N$ when variability is least.
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Distribution of average inventory in a line with identical machines and buffers.

$P = .79$; Total average inventory $= 490$
“Lazy integral”

Inventory distribution is anti-symmetric: $\bar{n}_i = 20 - \bar{n}_{50-i}$.

More inventory at the upstream end because

- each buffer near the beginning of the line has few machines upstream and many downstream.
- This is like a 2-machine line with a faster upstream machine and a slower downstream machine.

Less inventory at the downstream end for a similar reason.

The buffers in the middle are close to half full because they have close to the same number of machines upstream and downstream.
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Analytical vs simulation

Time steps | Decomp | 10,000 | 50,000 | 200,000
---|---|---|---|---
Production rate | 0.786 | 0.740 | 0.751 | 0.750
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Effects of a Bottleneck
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Effects of Very Large Buffers

\[ \text{Average Buffer Level} \]

\[ \text{Buffer Number} \]

50 Machines; $r=0.1$; $p=0.01$; $\mu=1.0$; $N=20.0$ EXCEPT $N(25)=2000.0$

\[ \text{Average Buffer Level} \]

\[ \text{Buffer Number} \]

** A very large buffer breaks a line into two smaller lines.
This line is the same as the first half of **.
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Upstream same as **; downstream faster.
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Effect of one buffer’s size on the average inventory in others

8-machine line; we vary the size of Buffer 6, $N_6$

Observation: as $N_6$ increases, upstream inventories decrease, downstream inventory increase.

Explanation: Any Buffer $i$ divides the line in two parts.

- If Buffer $i$ is upstream of Buffer 6, the increase of $N_6$ makes the downstream part of the line faster.
- If Buffer $i$ is downstream of Buffer 6, the increase of $N_6$ makes the upstream part of the line faster.
Consider a three-machine line.

Break up the line at each buffer into a single machine and a two-machine one-buffer line and compare the production rates of each.

For the split at $B_1$, consider the average inventory in $B_1$ as the size of $B_2$ varies from 0 to $\infty$.

For the split at $B_2$, consider the average inventory in $B_2$ as the size of $B_1$ varies from 0 to $\infty$. 
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Effect of one buffer’s size on the average inventory in others

\[ \begin{align*}
\text{Definitions:} \\
\text{\quad} & P_1 \text{ is the isolated production rate of Machine 1.} \\
\text{\quad} & P\{2,3\}(N_2) \text{ is the production rate of the line formed by } M_2, B_2, M_3. \\
\text{\quad} & P_3 \text{ is the isolated production rate of Machine 3.} \\
\text{\quad} & P\{1,2\}(N_1) \text{ is the production rate of the line formed by } M_1, B_1, M_2. \\
\text{\quad} & \text{We consider } \bar{n}_1 \text{ when } N_2 = 0 \text{ or } \infty. \\
\text{\quad} & \text{We consider } \bar{n}_2 \text{ when } N_2 = 0 \text{ or } \infty.
\end{align*} \]
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Effect of one buffer’s size on the average inventory in others

There are five possible cases. Two cases are interesting.

<table>
<thead>
<tr>
<th>Type</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>$P_3 \geq P_{{1,2}}(\infty)$ and $P_1 \geq P_{{2,3}}(\infty)$</td>
</tr>
<tr>
<td>Type 2</td>
<td>$P_3 \geq P_{{1,2}}(\infty)$ and $P_{{2,3}}(0) &lt; P_1 &lt; P_{{2,3}}(\infty)$</td>
</tr>
<tr>
<td>Type 3</td>
<td>$P_3 \geq P_{{1,2}}(\infty)$ and $P_1 \leq P_{{2,3}}(0)$</td>
</tr>
<tr>
<td>Type 4</td>
<td>$P_{{1,2}}(0) &lt; P_3 &lt; P_{{1,2}}(\infty)$ and $P_1 \geq P_{{2,3}}(\infty)$</td>
</tr>
<tr>
<td>Type 5</td>
<td>$P_3 \leq P_{{1,2}}(0)$ and $P_1 \geq P_{{2,3}}(\infty)$</td>
</tr>
</tbody>
</table>
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Effect of one buffer's size on the average inventory in others

- If $N_2$ increases, $\bar{n}_1$ decreases.
- $\star$ means that
  - when $N_2$ is small, $M^d(1)$ is a bottleneck.
  - when $N_2$ is large, $M_1$ is a bottleneck.

<table>
<thead>
<tr>
<th>Type 2</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_3 \geq P_{{1,2}}(\infty)$ and $P_{{2,3}}(0) &lt; P_1 &lt; P_{{2,3}}(\infty)$</td>
</tr>
</tbody>
</table>
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Effect of one buffer’s size on the average inventory in others

An example of $\bar{n}_1(N_2)$ in Type 2

- $r_1 = .8$, $p_1 = .096$, $r_2 = .1$, $p_2 = .01$, $r_3 = .1$, and $p_3 = .01$.

- These parameters satisfy $P_3 \geq P_{\{1,2\}}(\infty)$ and $P_{\{2,3\}}(0) < P_1 < P_{\{2,3\}}(\infty)$.
Consider profit $J(N_1, N_2) = 1000P(N_1, N_2) - N_1 - N_2 - \bar{n}_1 - \bar{n}_2$.

In spite of the non-convexity of the profit, we ran 5000 randomly-generated examples and observed that they all had a single local maximum.
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Effect of one buffer’s size on the average inventory in others

Longer lines: Effect on $\bar{n}_4$ of varying $N_2$ or $N_6$.

(a) $P(1) = .8990$
$P(2) = .9122$
$P(3) = .9097$
$P_{(2),(3)}(0) = .8614$
Type 2

(b) $P(1) = .9092$
$P(2) = .9092$
$P(3) = .9014$
$P_{(2),(3)}(0) = .8535$
Type 4
Long Lines — Behavior of $P(N)$ and $\bar{n}_i(N)$

Effect of one buffer’s size on the average inventory in others

- In this example, $\bar{n}_4$ is affected differently as $N_2$ and $N_6$ vary.
- The effect of each buffer’s size on each other buffer’s average inventory in a long line can be treated as one of the five types.
- $\bar{n}_i(N_j)$ can be non-convex and non-concave.
Allocation of buffer space that maximizes production rate. $P = .79$; Total average inventory $= 469$
Buffers are smaller at the ends of the line because no disruptions are considered that originate outside of the line.

Larger buffers in the rest of the line are all about the same size because they are affected by local disruptions more than remote disruptions.

Similar to the “bowl phenomenon” of Hillier and Boling.
BUT, there seems to be a paradox.

The distribution of inventory in a line with the same machines has more inventory at the upstream end and less at the downstream end.

 Doesn’t that mean there should be more space at the upstream end and less at the downstream end?
No! The level of buffer $B_1$ rarely gets below 10.

**Thought experiment:** Suppose the first buffer was Last-In-First-Out. Then the bottom 10 parts would stay there for a very long time.

Therefore we would get almost the same production rate if we replaced the bottom 10 parts with *bricks*.
50 Machines; r=0.1; p=0.01 EXCEPT p(10)=0.02; μ=1.0; Total Buffer Space = 980

Qualitative Observations on Production Line Behavior
Problem: pick buffer sizes so that $P \geq .88$ to minimize total buffer space.

Three cases:

- **Case 1** MTTF = 200 minutes and MTTR = 10.5 minutes for all machines ($P = .95$ parts per minute).

- **Case 2** Like Case 1 except Machine 5. For Machine 5, MTTF = 100 and MTTR = 10.5 minutes ($P = .905$ parts per minute).

- **Case 3** Like Case 1 except Machine 5. For Machine 5, MTTF = 200 and MTTR = 21 minutes ($P = .905$ parts per minute).
Long Lines — Optimization

Effect of a bottleneck

Bottleneck

Case 1 −− All Machines Identical
Case 2 −− Machine 5 Bottleneck −− MTTF = 100
Case 3 −− Machine 5 Bottleneck −− MTTR = 21

Buffer Size

Buffer
Other Phenomena

- Reversibility/equivalence
- Buffers as low-pass filters.
Conclusions

- Buffers absorb variability. Buffers are needed most where
  - variability is greatest (largest MTTR); or
  - vulnerability to variability is greatest (a bottleneck).

- Two-machine lines can tell us a lot about the behavior of larger systems.