Variability in Manufacturing Systems

Stanley B. Gershwin
gershwin@mit.edu
http://web.mit.edu/manuf-sys
Massachusetts Institute of Technology

Manufacturing Processes and Systems, 2.810

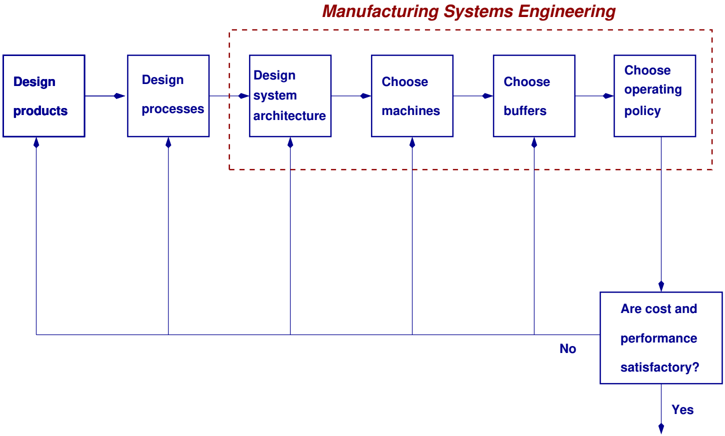
November 1, 2017

Manufacturing Systems
Definitions

- *Manufacturing:* the transformation of material into something useful and portable.
- *Manufacturing System:* A manufacturing system is a set of machines, transportation elements, computers, storage buffers, people, and other items that are used together for manufacturing.
  - ⋆ Alternate terms:
    - *Factory*
    - *Production system*
    - *Fabrication facility*
  - ⋆ Subsets of manufacturing systems, which are themselves systems, are sometimes called *cells*, *work centers*, or *work stations* .

- *Manufacturing Systems Engineering:*
  - ⋆ The design and operation of new manufacturing systems;
  - ⋆ The analysis and improvement of existing manufacturing systems.

Manufacturing Systems
Manufacturing Systems Engineering



Manufacturing Systems
Manufacturing Industry Challenges

- Short product lifetimes. Frequent factory reconfiguration or replacement. Limited time for real-time learning to optimize factory.
- Large product diversity. Factories must be flexible.
- Short lead times and impatient customers.
- Inventory is perishable. It loses value rapidly due to obsolescence.
- *Variability, uncertainty, and randomness*

Manufacturing Systems
Manufacturing Systems Engineering Objectives

- Manufacturing systems must be complex to manage or tolerate variability, uncertainty, and randomness.
  - ⋆ *Variability:* change over time.
  - ⋆ *Uncertainty:* incomplete knowledge.
  - ⋆ *Randomness:* unpredictability that has some regularity that can be described precisely.

- Improvements in the design and operation of manufacturing systems must
  - ⋆ ***reduce*** the variability, uncertainty, and randomness, or
  - ⋆ reduce the ***sensitivity*** of systems to variability, uncertainty, and randomness.

Manufacturing Systems
Manufacturing Systems Engineering Objectives

- Satisfy demand.
- Meet due dates.
- Keep quality high.
- Keep inventory low.

- ***Be robust.***
  - ⋆ Be insensitive to disruptions.
  - ⋆ Respond gracefully to disruptions.
  - ⋆ Respond gracefully to demand changes, engineering changes, etc.

- **Design Quality** – the design of products that give customers what they want or would like *(features).*

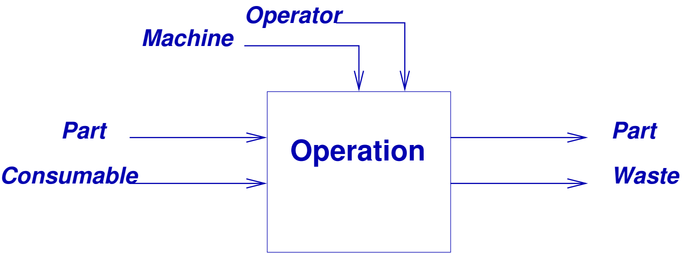
- **Manufacturing Quality** – the manufacturing of products to *avoid* giving customers what they *don't* want or *would not* like *(defects).*

- Quantity – *how much* is produced and *when* it is produced.
- Quality – *how well* it is produced.

In this session, we focus on *quantity*.

General Statement: Variability is the enemy of manufacturing.

Why care about variability?
What is an operation?



Nothing happens until everything is present.

Why care about variability?
What is an operation?

Whatever does not arrive last must wait.

- *Inventory:* parts waiting.
- *Under-utilization:* machines waiting.
- *Idle work force:* operators waiting.

Why care about variability?
Examples

- Factories are full of random events:
  - ⋆ machine failures
  - ⋆ changes in orders
  - ⋆ quality failures
  - ⋆ human variability
- The economic environment is uncertain:
  - ⋆ demand variations
  - ⋆ supplier unreliability
  - ⋆ changes in costs and prices

Why care about variability?

Therefore, factories should be

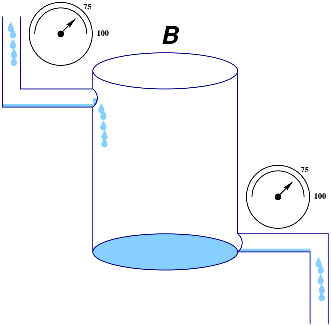
- *designed* and *operated*

to minimize the

- *creation,* *propagation,* or *amplification*

of *uncertainty,* *variability,* and *randomness.*

Variability and Inventory
No variability

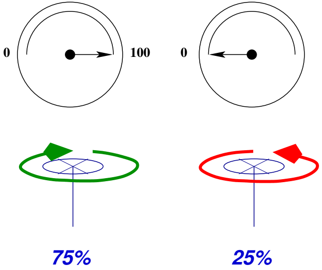


75 gal/sec in, 75 gal/sec out constantly

The tank is always empty.

Variability and Inventory
Variability from random valves

Consider a random valve:



- The average period when the valve is open is 15 minutes.
- The average period when the valve is closed is 5 minutes.
- Consequently, the *average* flow rate through it is 75 gal/sec.

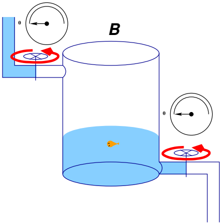
... as long as flow is not impeded upstream or downstream.

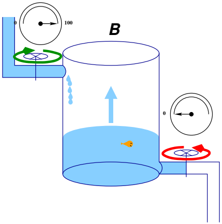
Variability and Inventory

Variability from random valves

Four of the possibilities for *two* valves and one tank:

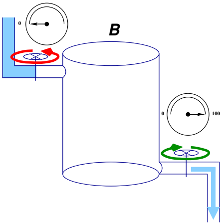
0 in, 0 out

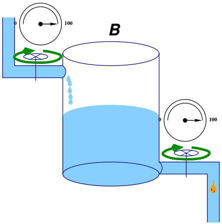




100 gal/sec in, 0 out

0 in, 100 gal/sec out





100 gal/sec in, 100 gal/sec ou

Variability and Inventory

Observation:

- There is never any water in the tank when the flow is constant.
- There is sometimes water in the tank when the flow is variable.

Conclusions:

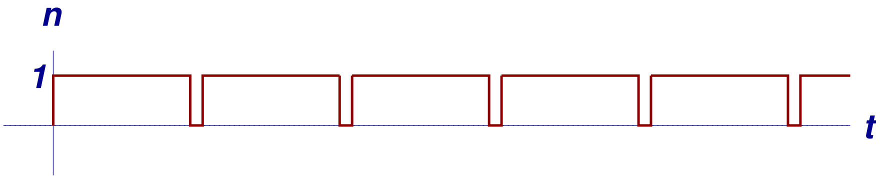
1. You can't always replace random variables with their averages.
2. *Variability causes inventory!!*

Variability and Inventory

- To be more precise, *non-synchronization causes inventory.*
  - ⋆ Living things do not acquire energy at the same time they expend it. Therefore, they must store energy in the form of fat or sugar.
  - ⋆ Rivers are dammed and reservoirs are created to control the flow of water — to reduce the variability of the water supply.
  - ⋆ For solar and wind power to be successful, energy storage is required for when the sun doesn't shine and the wind doesn't blow.

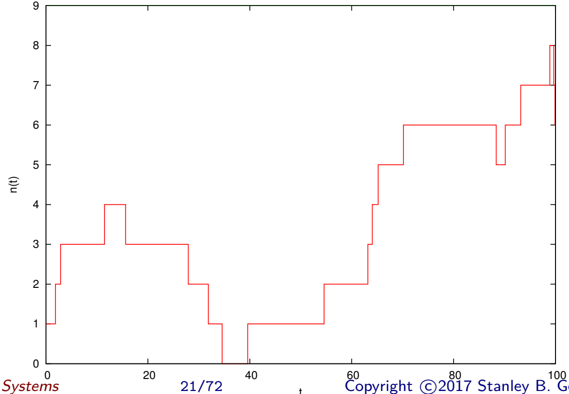
Variability and Inventory
Queues

- Suppose customers arrive at time 0, time 1 minute, 2 minutes, etc. and that service time is always exactly 54 seconds (.9 minutes).
  - ⋆ Then there is 1 customer in the system for 54 seconds of every minute and 0 customers for 6 seconds every minute. Therefore there are *54/60=.9* customers in the system on the average.

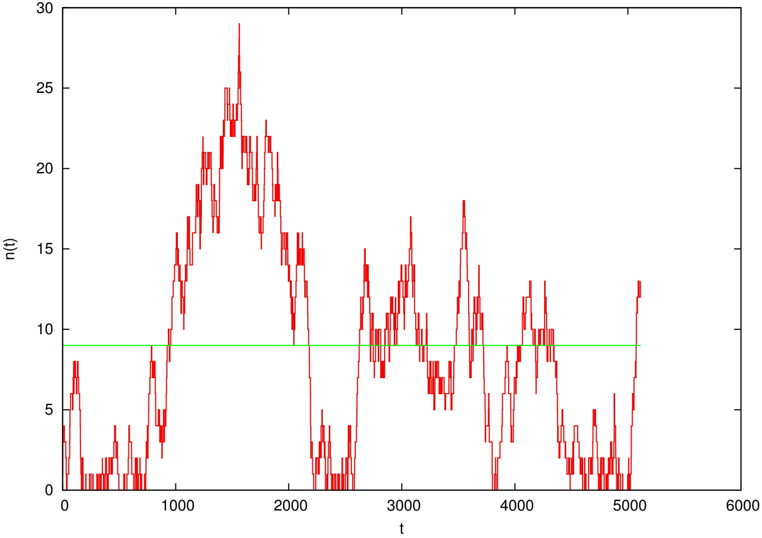


Variability and Inventory
Queues

- Suppose customers arrive with exponentially distributed inter-arrival times with *average* inter-arrival time 1 minute; and that service time is exponentially distributed with *average* service time 54 seconds.
  - ⋆ Then the average number of customers in the system is *9.*



Variability and Inventory
Queues



Let $\rho = \lambda/\mu$. Under the assumptions of exponential inter-arrival time, exponential service time, and infinite buffer (called *M/M/1 queue*), it can be shown that the average number of parts in the system is

$$\bar{n} = \sum_n nP(n) = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}.$$

ρ is called the *utilization* of the server.

Little's Law: $L = \lambda W$, where L is the average number of customers in a system, W is the average customer waiting time, and λ is the average arrival rate.

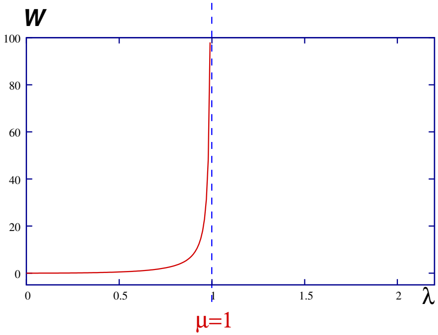
This is true for almost all queuing systems in steady state.

In the $M/M/1$ queue, $L = \bar{n}$, so the average waiting time is

$$W = \frac{1}{\mu - \lambda}.$$

ρ is called the *utilization* of the server.

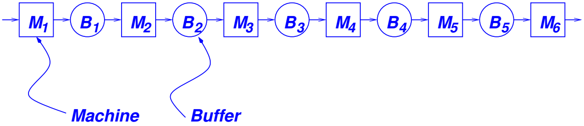
Variability and Inventory
Queues



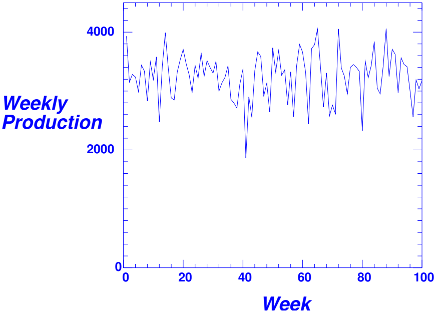
- μ is the *capacity* of the system.
- If $\lambda < \mu$, system is stable and waiting time remains bounded.
- If $\lambda > \mu$, waiting time grows over time.

Flow Lines

... also known as a Production or Transfer Line.



- Machines are unreliable.
- Buffers are finite.
- In many cases, the operation times are constant and equal for all machines.



Production output from a simulation of a transfer line.

Single Unreliable Machine

Failures and Repairs

- Machine is either *up* or *down* .
- MTTF = mean time to fail.
- MTTR = mean time to repair
- MTBF = MTTF + MTTR

Single Unreliable Machine

Production rate

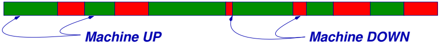
- If the machine is unreliable, and
  - ⋆ its average operation time is τ,
  - ⋆ its mean time to fail is MTTF,
  - ⋆ its mean time to repair is MTTR,

  then its maximum production rate is

$$\frac{1}{\tau}\left(\frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}}\right)$$

Single Unreliable Machine

Proof



- Average production rate, while machine is up, is $1/\tau$.
- Average duration of an up period is MTTF.
- Average production during an up period is MTTF$/\tau$.
- Average duration of up-down period: MTTF + MTTR.
- Average production during up-down period: MTTF$/\tau$.
- Therefore, average production rate is (MTTF$/\tau$)/(MTTF + MTTR).

- *Assumptions:* Operation time is constant (τ). Failure and repair times are *geometrically* distributed.

- Let p be the probability that a machine fails during any given operation. Then $p = \tau/\text{MTTF}$.

- Let r be the probability that M gets repaired during any operation time when it is down. Then $r = \tau/\text{MTTR}$.

- Then the *average production rate* of M is

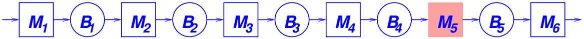
$$\frac{1}{\tau}\left(\frac{r}{r+p}\right).$$

- *(Sometimes we forget to say "average.")*

Infinite-Buffer Lines

Bottleneck

- The production rate of the line is the production rate of the *slowest* machine in the line — called the *bottleneck* .
- Slowest means *least average production rate* , where average production rate is calculated from one of the previous formulas.



- Production rate is therefore

$$P = \min_i \frac{1}{\tau_i}\left(\frac{\text{MTTF}_i}{\text{MTTF}_i + \text{MTTR}_i}\right)$$

- and M_i is the bottleneck.

Infinite-Buffer Lines

Bottleneck

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

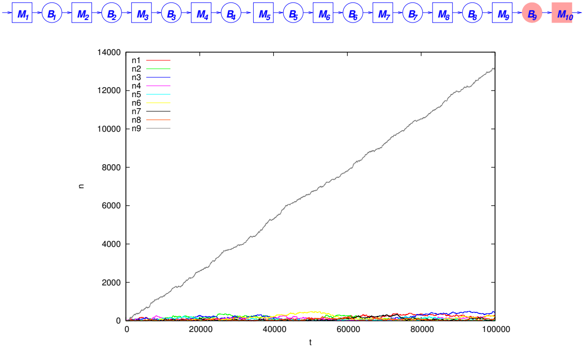
Infinite-Buffer Lines

Example 1

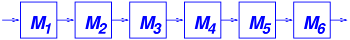
- Parameters: $r_i = .1, p_i = .01, i = 1, ..., 9; r_{10} = .1, p_{10} = .03.$
- Therefore, $e_i = .909, i = 1, ..., 9; e_{10} = .769.$

Infinite-Buffer Lines

Example 1



Zero-Buffer Lines

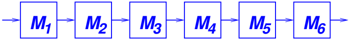


- If any one machine fails, or takes a very long time to do an operation, *all* the other machines must wait.

- Therefore the production rate is usually less — possibly much less – than the slowest machine.

Zero-Buffer Lines

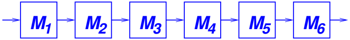
Constant, equal operation times, unreliable machines



- *Assumption:* Failure and repair times are *geometrically* distributed.
- Define $p_i = \tau/\text{MTTF}_i =$ probability of failure during an operation.
- Define $r_i = \tau/\text{MTTR}_i$ probability of repair during an interval of length τ when the machine is down.

Zero-Buffer Lines

Production Rate



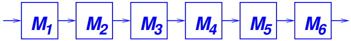
Buzacott's Zero-Buffer Line Formula:

Let k be the number of machines in the line. Then

$$P = \frac{1}{\tau} \frac{1}{1 + \sum_{i=1}^{k} \frac{p_i}{r_i}} = \frac{1}{\tau} \frac{1}{1 + \sum_{i=1}^{k} \frac{MTTR_i}{MTTF_i}}$$

Zero-Buffer Lines

Production Rate



Same as the earlier formula when $k = 1$.

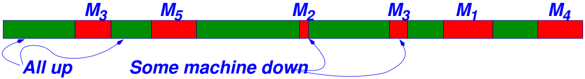
- The *isolated production rate* of a single machine M_i is

$$\frac{1}{\tau}\left(\frac{1}{1+\frac{p_i}{r_i}}\right) = \frac{1}{\tau}\left(\frac{r_i}{r_i+p_i}\right).$$

Zero-Buffer Lines

Proof of formula

- Let τ (the operation time) be the time unit.
- *Approximation:* At most, one machine can be down.
- Consider a long time interval of length $T\tau$ during which Machine M_i fails m_i times $(i = 1, \ldots k)$.



- Without failures, the line would produce T parts.

Zero-Buffer Lines

Proof of formula

- The average repair time of M_i is τ/r_i each time it fails, so the total system down time is close to

$$D\tau = \sum_{i=1}^{k} \frac{m_i \tau}{r_i}$$

where D is the number of operation times in which a machine is down.

Zero-Buffer Lines

Proof of formula

- The total up time is approximately

$$U\tau = T\tau - \sum_{i=1}^{k} \frac{m_i \tau}{r_i}$$

- where U is the number of operation times in which all machines are up.

Zero-Buffer Lines

Proof of formula

- Since the system produces one part per time unit while it is working, it produces U parts during the interval of length $T\tau$.
- Note that, approximately,

$$m_i = p_i U$$

because M_i can only fail while it is operational.

Zero-Buffer Lines

Proof of formula

- Thus,

$$U\tau = T\tau - U\tau \sum_{i=1}^{k} \frac{p_i}{r_i},$$

or,

$$\frac{U}{T} = E_{ODF} = \frac{1}{1 + \sum_{i=1}^{k} \frac{p_i}{r_i}}$$

Zero-Buffer Lines

and

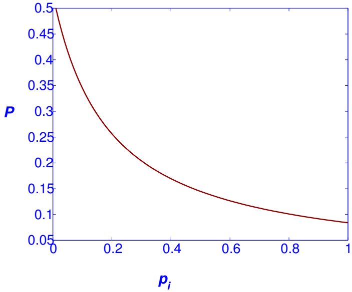
$$P = \frac{1}{\tau} \frac{1}{1 + \sum_{i=1}^{k} \frac{p_i}{r_i}}$$

- Note that P is a function of the *ratio* p_i/r_i and not p_i or r_i separately.
- The same statement is true for the infinite-buffer line.
- However, the same statement is *not* true for a line with finite, non-zero buffers.

Zero-Buffer Lines

P as a function of p_i

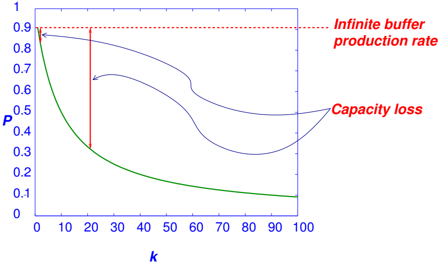
All machines are the same except M_i. As p_i increases, the production rate decreases.



Zero-Buffer Lines

P as a function of p_i

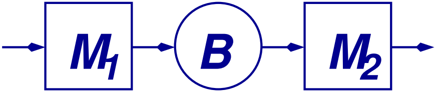
All machines are the same. As the line gets longer, the production rate decreases.



Finite-Buffer Lines

- Difficulty:
  - ⋆ No simple formula for calculating production rate or inventory levels.
- Solution:
  - ⋆ Simulation
  - ⋆ Analytical approximation
  - ⋆ *Exact analytical solution for two-machine lines only.*

Inventory and the Propagation of Variability
Two-machine production lines



- A simple production line consisting of two machines and one in-process inventory buffer, making discrete parts.
- Both machines have unit operation time.
- Both machines are unreliable with geometric uptimes and downtimes. The parameters are
  - ⋆ $\text{MTTR}_1 = 1/r_1$, $\text{MTTR}_2 = 1/r_2$
  - ⋆ $\text{MTTF}_1 = 1/p_1$, $\text{MTTF}_2 = 1/p_2$
- The buffer can hold a maximum of N parts.
- The first machine is never starved; the last machine is never blocked.

- The *efficiency* e_i of Machine i is the fraction of time the machine is operational, *not counting idle time* .

- Because the time unit is the operation time, the efficiency of a machine is the same as its production rate in isolation, i.e., not counting idle time.

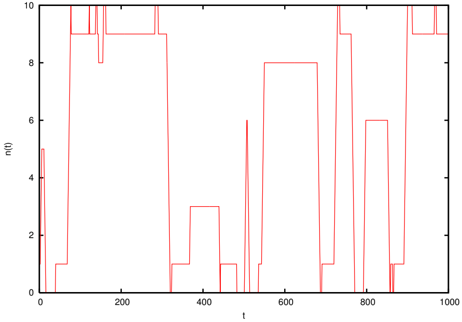
- The efficiency of Machine i is given by

$$e_i = \frac{\text{MTTF}_i}{\text{MTTR}_i + \text{MTTF}_i} = \frac{\text{MTTF}_i}{\text{MTBF}_i} = \frac{r_i}{r_i + p_i}$$

Simulations: Buffer Level vs. t

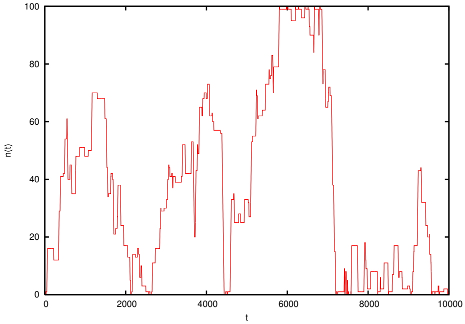
- The next four slides show simulation results. They show how the buffer level ($n(t)$) varies with time t as a result of random failures and repairs of the two machines.
- Compare the slides in order to relate the buffer level behavior to the machine parameters.
- Think about the average buffer level.

Simulations: Buffer Level vs. t



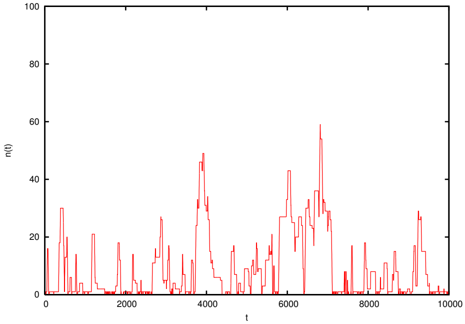
$$\text{MTTR}_i = 10, \text{MTTF}_i = 100, i = 1, 2; N = 10$$
$$(r_1 = .1, p_1 = .01, r_2 = .1, p_2 = .01)$$

Simulations: Buffer Level vs. t



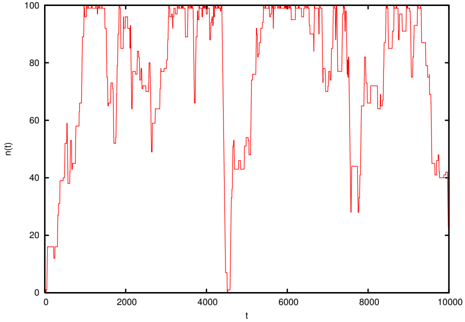
$$\text{MTTR}_i = 10, \text{MTTF}_i = 100, i = 1, 2; N = 100$$
$$(r_1 = .1, p_1 = .01, r_2 = .1, p_2 = .01)$$

Simulations: Buffer Level vs. t



$$\text{MTTR}_i = 10, i = 1, 2, \text{MTTF}_1 = 50, \text{MTTF}_i = 100; N = 100$$
$$(r_1 = .1, p_1 = .02, r_2 = .1, p_2 = .01)$$

Simulations: Buffer Level vs. t



$$\text{MTTR}_i = 10, i = 1, 2, \text{MTTF}_1 = 100, \text{MTTF}_i = 50; N = 100$$
$$(r_1 = .1, p_1 = .01, r_2 = .1, p_2 = .02)$$

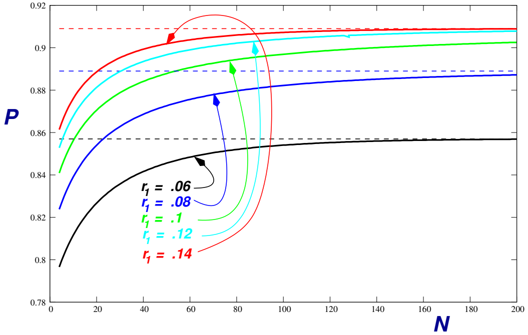
Inventory and the Propagation of Variability
Two-machine production lines

Scenario:

- We are designing a line. We have already chosen the second machine.
- Two decisions are left: the choice of the first machine and the size of the buffer,
- Five vendors produce versions of the first machine.
  - ⋆ All have the same operation time as the second machine.
  - ⋆ All have the same MTTF.
  - ⋆ They differ in MTTR. The vendors put different amounts of design effort into making the machines easy to repair.

Inventory and the Propagation of Variability
Two-machine production lines

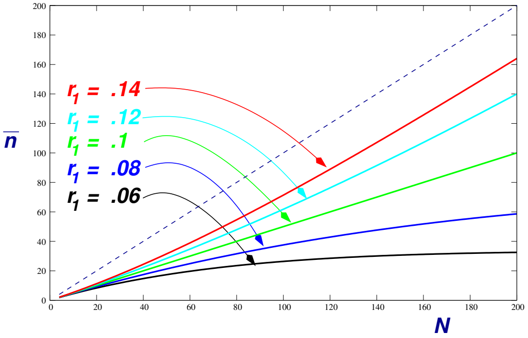
Production rate vs. Buffer Size



$p_1 = .01$
$r_2 = .1$
$p_2 = .01$

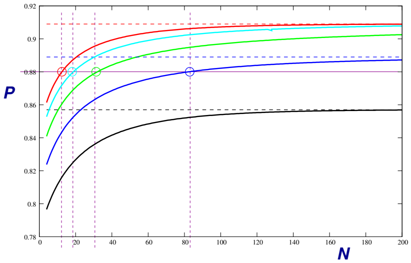
Inventory and the Propagation of Variability
Two-machine production lines

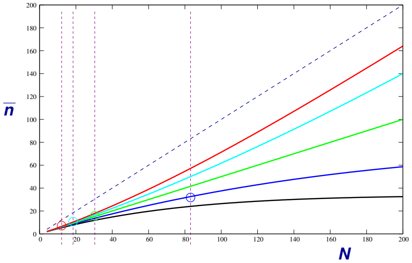
Average Inventory vs. Buffer Size



Inventory and the Propagation of Variability

Two-Machine Line Behavior



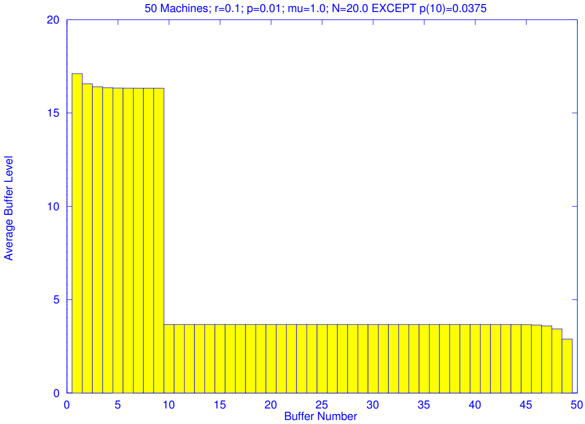


Problem: Select M_1 and N so that $P = .88$.

Solution:

r_1	N	$\bar{n}$
.14	13	7.0819
.12	19	10.1153
.10	32	16.0000
.08	82	32.2112

Inventory and the Propagation of Variability
Long production lines



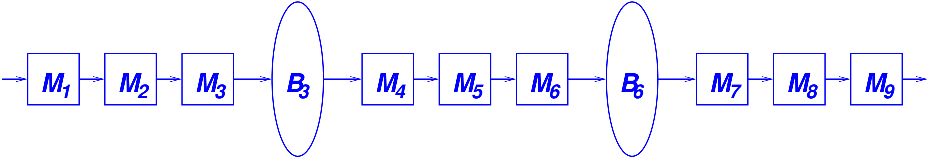
Effect of a bottleneck. Identical machines and buffers, except for M_{10}.

Inventory and the Propagation of Variability
Long production lines

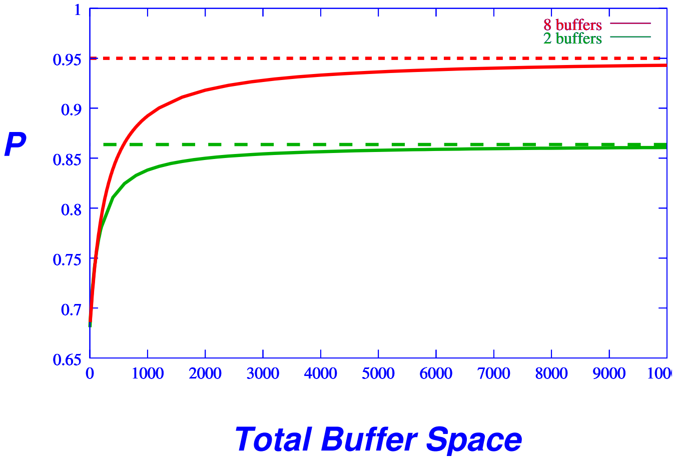
Which has a higher production rate?

- 9-Machine line with two buffering options:
  - ⋆ 8 buffers equally sized; and

  - ⋆ 2 buffers equally sized.



Inventory and the Propagation of Variability
Long production lines



Optimal buffer space distribution

- Design the buffers for a 20-machine production line.
- The machines have been selected, and the only decision remaining is the amount of space to allocate for in-process inventory.
- *The goal is to determine the smallest amount of in-process inventory space so that the line meets a production rate target.*

- The common operation time is one operation per minute.
- The target production rate is .88 parts per minute.

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

- *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).

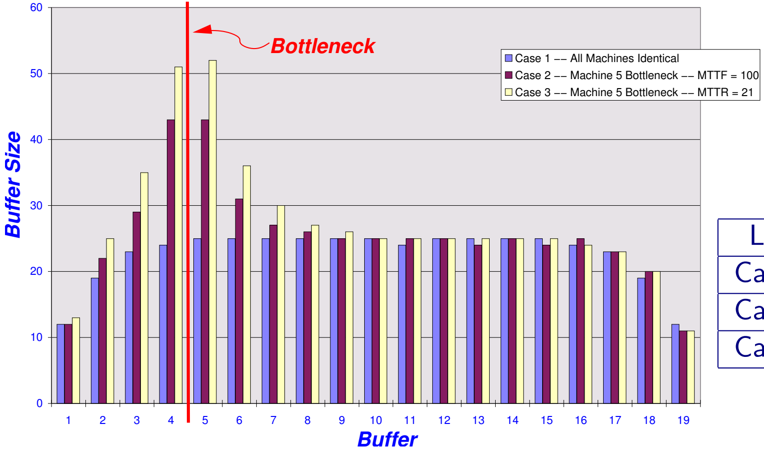
- *Case 1* MTTF= 200 minutes and MTTR = 10.5 minutes for all machines ($P = .95$ parts per minute).
- *Case 2* Like Case 1 except Machine 5. For Machine 5, MTTF = 100 and MTTR = 10.5 minutes ($P = .905$ parts per minute).
- *Case 3* Like Case 1 except Machine 5. For Machine 5, MTTF = 200 and MTTR = 21 minutes ($P = .905$ parts per minute).

Are buffers really needed?

Line	Production rate with no buffers, parts per minute
Case 1	.487
Case 2	.475
Case 3	.475

Yes. *These numbers came from a zero-buffer formula.*

Solution



Line	Space
Case 1	430
Case 2	485
Case 3	523

- Observation from studying buffer space allocation problems:
    - ⋆ *Buffer space is needed most where buffer level variability is greatest!*

Other issues

- Setup changes
  - ⋆ Controllable disruption
  - ⋆ Reduces production time and creates inventory.
- Deterministic scheduling and MRP
  - ⋆ Recalculation required when an unanticipated event occurs
  - ⋆ This can cause instability and confusion

