Capacity Planning

Dr Feryal Erhun
fe251@cam.ac.uk

Recap of Last Lecture

- A number of rules (e.g., SPT, EDD and Moore’s Algorithm) and heuristics (e.g., MDD) can be used for **single machine scheduling**
- Johnson’s Rules is to minimise makespan when in a **two machine** flowshop
- **Push** systems schedule and issue orders centrally. **MRP** systems are computerised systems to support push scheduling
- In a **pull** system, processes are triggered by a replenishment signal, such as the **kanban** cards in a **JIT** scheduling
- **Assembly line balancing** involves grouping assembly tasks into workstations with as equal as possible workloads at each
- **Layout** is contingent on variety and volume

Some Key Questions

- How much can a given production operation make
  - Theoretically?
  - Practically?
- How does product mix affect capacity?
- How does changing demand influence capacity?
- How can capacity/demand tradeoffs be assessed?

What is Capacity?

What is Capacity?

The capacity of an operation is the maximum level of value-added activity over a period of time that the process can achieve under normal operating conditions (Slack *et al.*)

- Air-conditioner plant – Number of units per week (output)
- Brewery – Litres of beer produced per month (output)
- Steel Mill – Tonnes per hour (output)
- Electricity company – Megawatts of electricity generated (output)
- Hospital – Number of beds available (input)
- Theatre – Number of seats (input)
- University – Number of students (input)

… but capacity is also influenced by demand

Example: Calculating Capacity

- An AC factory produces three different models of AC units: the deluxe, the standard and the economy
- The deluxe model can be assembled in 1.5 hours, the standard in 1 hour, and the economy in 0.5 hours
- The assembly area in the factory has 800 staff hours of assembly time available each week
- What is the aggregate production capacity of the factory if the demand for deluxe, standard and economy units is in the ratio 2:3:2?
- What is the capacity if the ratio is 1:2:4?

Example: Capacity Calculation

- Roughly: capacity = hours available / hours per unit

- For ratio 2:3:2, 7 units take
  - 2 x 1.5 + 3 x 1 + 2 x .5 = 7 hrs [per 7 units] = 1 hrs/unit
  - Capacity = 800/1 = 800 units

- For ratio 1:2:4, 7 units take
  - 1 x 1.5 + 2 x 1 + 4 x .5 = 5.5 hrs [per 7 units] = 5.5/7 hrs/unit
  - Capacity = 800/5.5 x 7 = 1018 units

- Hence mix affects capacity

  … But this is theoretical capacity only …

Related Capacity Definitions

- **Capacity:** The capacity of an operation is the maximum level of value-added activity over a period of time that the process can achieve under normal operating conditions (Slack *et al.*)
- **Utilization:** Measure of the number or % of hours worked by equipment, line, staff, etc. (Hill)
- **Efficiency:** Comparing actual output to the level of output expected (Hill)
- **Capacity Planning:** The task of setting the effective capacity of the operation so that it can respond to demand

Challenges for Capacity Planning

- Capacity is a soft, malleable constraint
- Capacity is like "black art"; it depends on everything
- Capacity frictions: leadtimes, lumpiness, fixed costs
- Capacity requires large and irreversible investment
- Capacity decisions can be political
- Measuring and valuing capacity shortfall is not obvious
- Capacity investment involves long-run planning under uncertainty
  - Arguably the greatest challenge for capacity strategy

Capacity Decisions

- **Sizing:** How much capacity to invest in?
- **Timing:** When to increase or reduce resources?
- **Type:** What kinds of resources are best?
- **Location:** Where should resources be located?

Managing Capacity Sizing Drivers

Capacity Sizing

Forecasting

- Average demand
- Volatility
- Separate planning from other uses

Shortage Cost

- Volatility
- Standard loss
- Shortage penalty
- Waiting cost
- Holding cost

Safety Capacity Cost

- Discount horizon and rate
- Marginal cost of:
  - Capacity
  - Subcontracting
  - Overtime/Temp
  - Spot-market buying

Theoretical vs. Actual Capacity

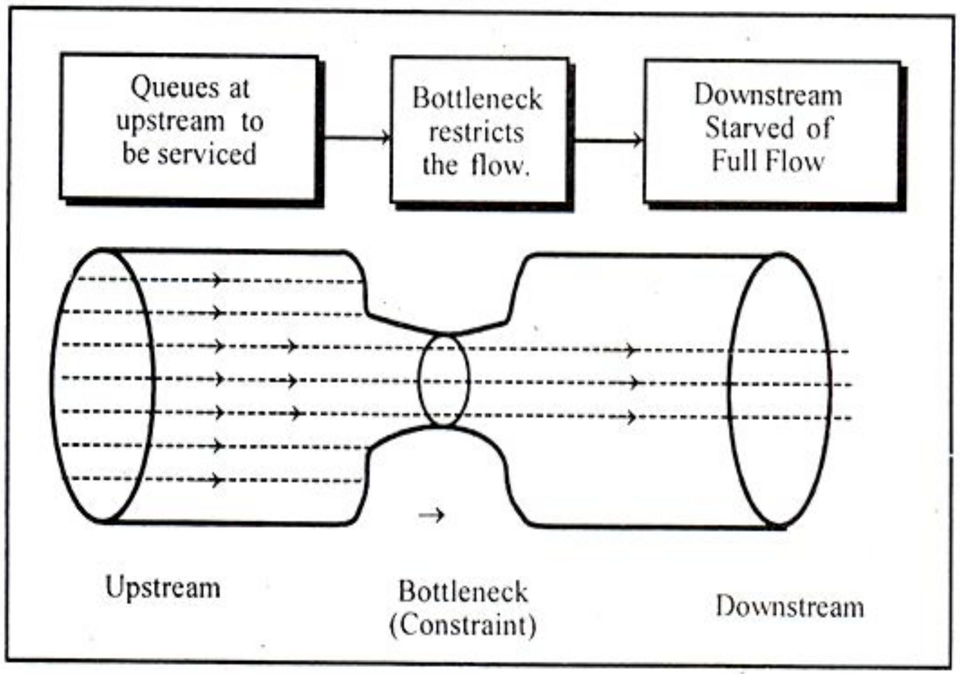
Theoretical vs. Actual Capacity

- Theoretical capacity is the maximum possible output rate, whereas the actual capacity is a realistic estimate of the achievable output rate
- The main difficulties which inhibit perfect utilisation of a manufacturing system, i.e., which restrict the capacity are:

MANAGE	**Lost Time (Planned)** • Setup times • Switchover delays	**Imbalances** •Bottlenecks •Imbalances in task times
MITIGATE	**Lost Time (Unplanned)** • Breakdowns • Co-ordination conflicts (of equipment and labour) • Supply shortages	**Reduced Yield** • Quality problems **Variable Conditions** • Variability in process times causing a build up of inventory • Variability in raw material arrivals • Variability in order arrivals • Unplanned downtime

Bottlenecks

- Capacity is always restricted by the slowest task, which is called the bottleneck
- When capacity at each stage is not balanced, the capacity of the total system is limited by that of the bottleneck stage



Set-up Times

- Set-up is
  - any action required to prepare a machine to accomplish an operation
  - required when type/size/colour of part being worked on is changed

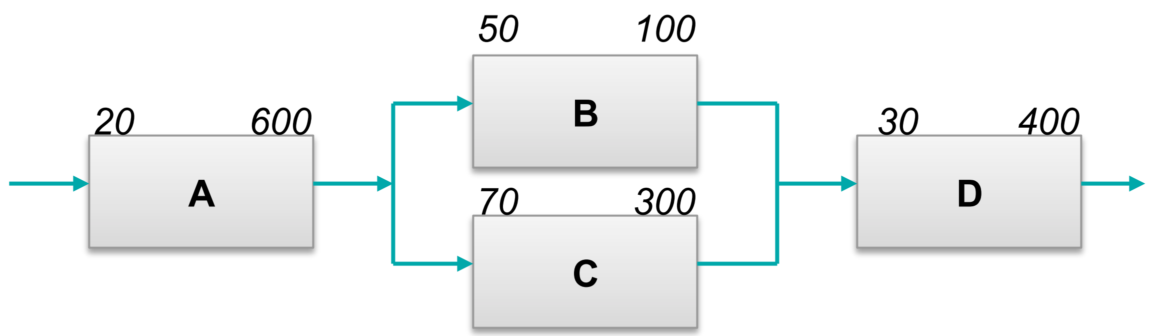
Example: Understanding the effect of setup times on capacity & bottlenecks

- A factory has four machines A, B, C and D
- Product 1 follows the route A-B-D and product 2 follows the route A-C-D
- Production is organised so that whenever a batch of product 1 is made, an identical sized batch of product 2 is always made immediately afterwards
- Data on process times and setup times are in the table below

Machine	Process time (seconds)	Setup time (seconds)
A	20	600
B	50	1000
C	70	300
D	30	400

- What is the long-run average capacity of the process assuming a batch size of 10 and 100?

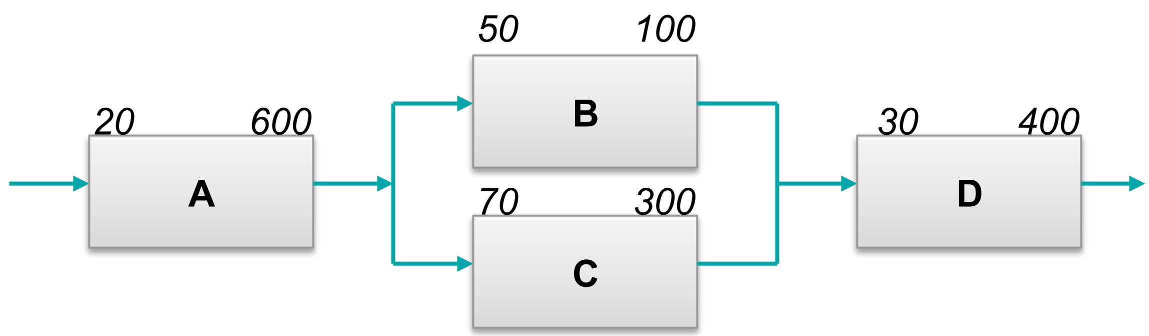
Solution for Example (Batch Size = 10)



	A (one batch of each product)	B	C	D (one batch of each product)
No. products to process for one batch of each product	20	10	10	20
Setup time	1200	1000	300	800
Run time	400	500	700	600
Time for one batch of 10 of each product	1600	1500	1000	1400
Average time for one pair of products	160			
Capacity in "no of pairs (one of each product) per hour"	22.5			

Bottleneck

Solution for Example (Batch Size = 100)



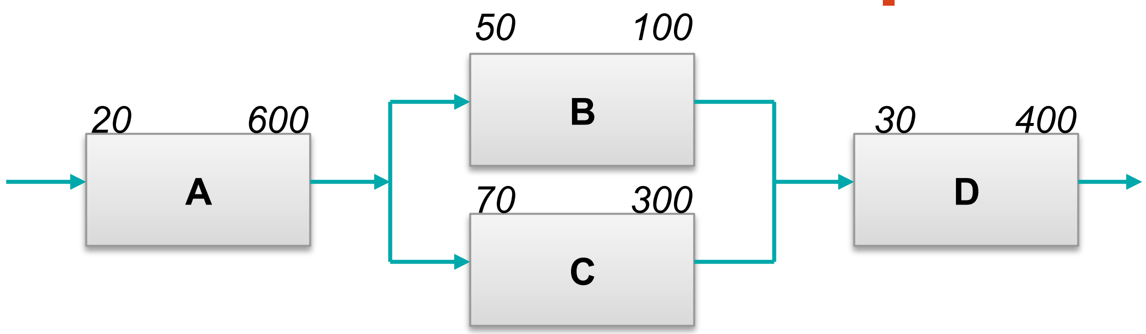
	A (one batch of each product)	B	C	D (one batch of each product)
No. products to process for one batch of each product	200	100	100	200
Setup time	1200	1000	300	800
Run time	4000	5000	7000	6000
Time for one batch of 10 of each product	5200	6000	7300	6800
Average time for one pair of products			73	
Capacity in “no of pairs (one of each product) per hour”			49.3	

Bottleneck

Impact of Reducing Set-up Times

- Smaller batch sizes become economical
- Reduced cost of setup labour required
- Increase production capacity [on bottlenecks]
- Reduce scale of potential quality problems, and hence waste

Solution for Example (Batch Size = 10) with 50% Reduction in Set-up Times



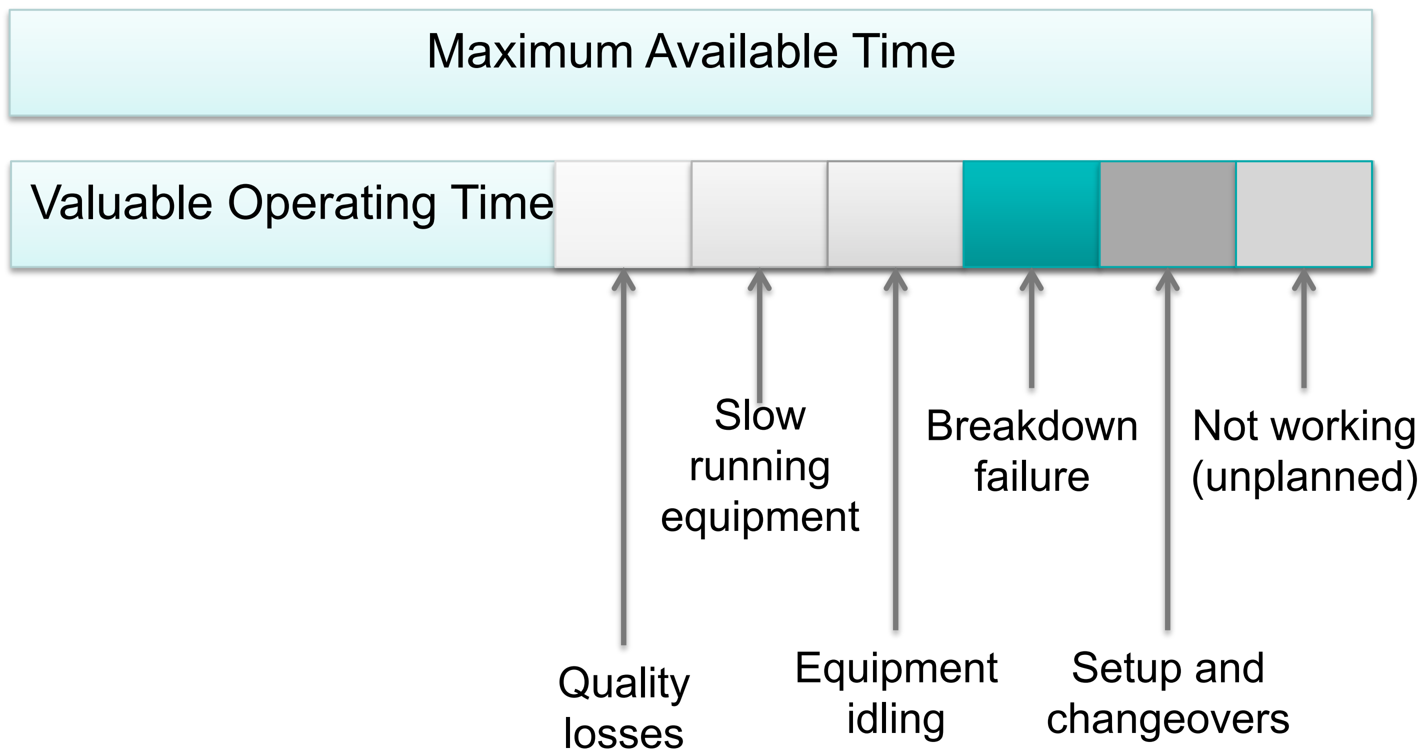
	A (one batch of each product)	B	C	D (one batch of each product)
No. products to process for one batch of each product	20	10	10	20
Setup time	600	500	150	400
Run time	400	500	700	600
Time for one batch of 10 of each product	1000	1000	850	1000
Average time for one pair of products	100			
Capacity in "no of pairs (one of each product) per hour"	36			

Balanced operation – 3 Bottlenecks

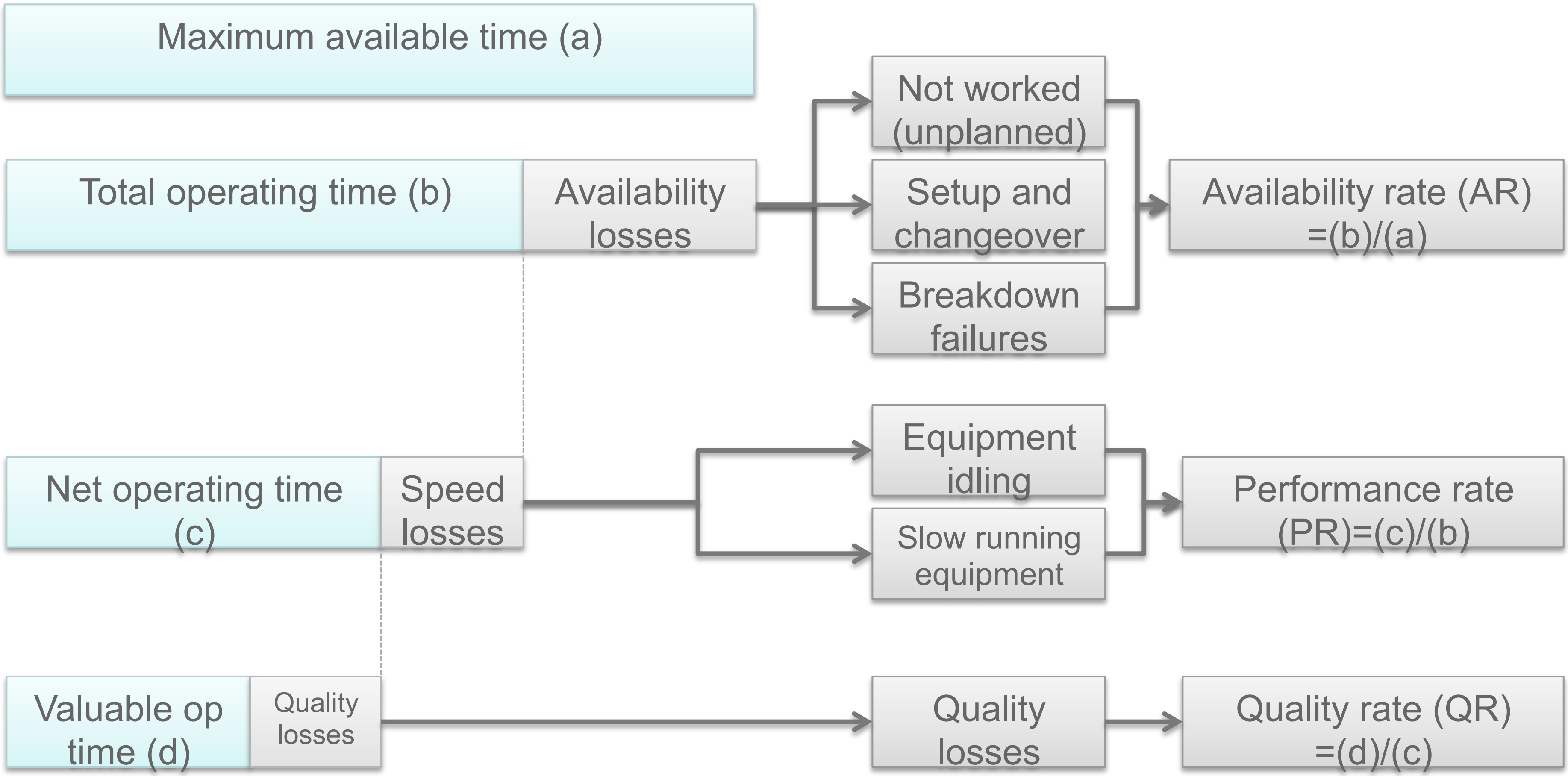
Breakdowns

- Assume that a process with 100 workstations is making 1000 products, and that each process has a probability of 1/1000th of breaking down at any given time:
  - P(completing a product) = (1-P(breakdown))100
  - P(completing a product) = (1-0.001)100
  - P(completing a product) = 0.999100
  - P(completing a product) = 0.905
- Thus 9.5% of production will be delayed or even lost if it is assumed that the products are wasted as a result of breakdown
- Hence the importance of breakdown free production and routine maintenance
- The effect is exaggerated if a new setup is required after an interruption [a similar but more compounded analysis applies for defective parts]

Theoretical vs. Actual Capacity: Loss of Capacity



Theoretical vs. Actual Capacity: Overall Equipment Effectiveness



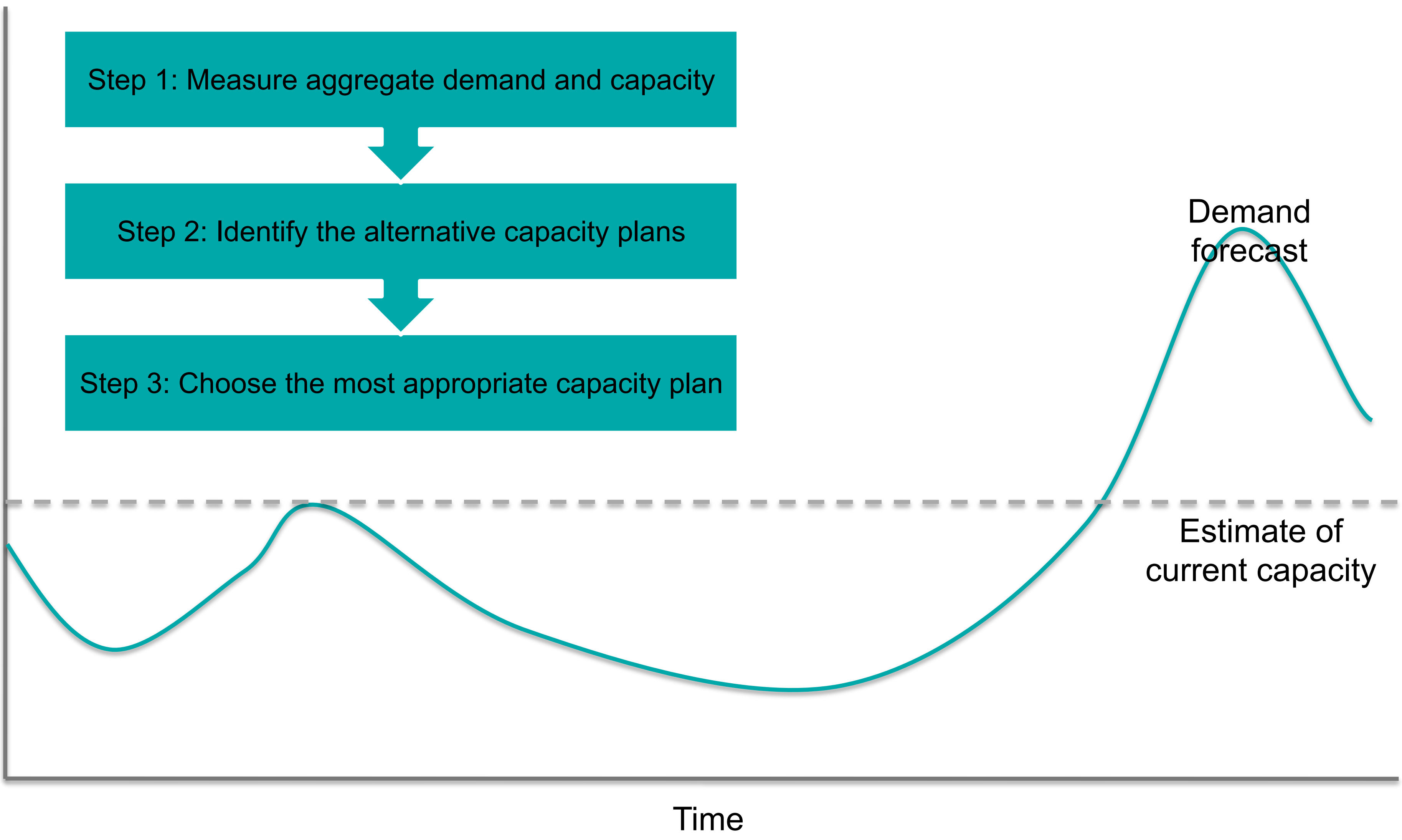
Overall Equipment Effectiveness (OEE) = AR x PR x QR

Capacity Planning and Control

Ideal Demand

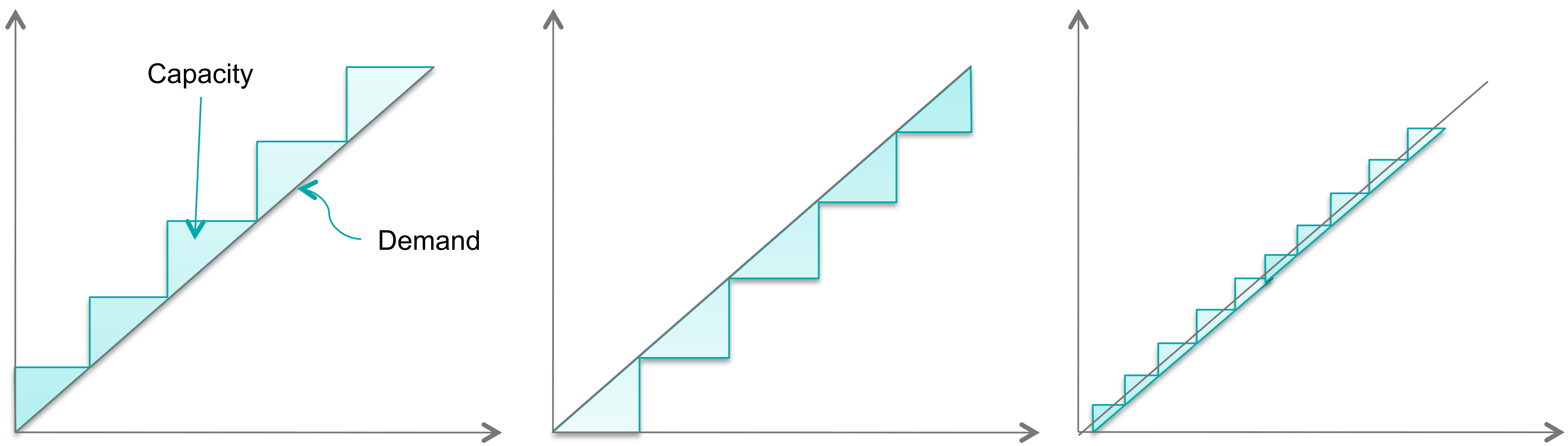
- Ideal demand is smooth and predictable:
  - Total demand = maximum output capacity of resources
  - Any changes are perfectly forecast in sufficient time to allow capacity change
- But …
  - Real demand is usually not predictable
  - Demand has peaks - lunchtime/Saturday/summer, etc.
  - Demand varies through product life cycle & competition
- Long, medium, short term capacity planning challenges!

Capacity Planning and Control



Long-term Capacity Planning

Three basic strategies



(a) Capacity lead

- Sufficient capacity to meet demand
- Capacity cushion
- Low impact of start-up problems
- Low utilisation
- Risk of over-capacity
- Early capital spending

(b) Capacity lag

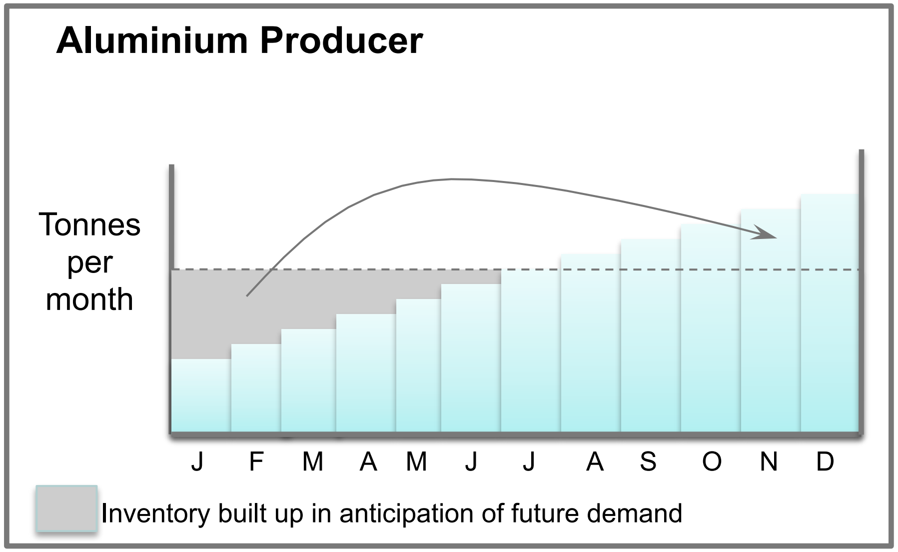
- Sufficient demand for full working capacity
- No over-capacity risk
- Capital spending is delayed
- Insufficient capacity to meet demand
- High impact of start-up problems
- No capacity cushion

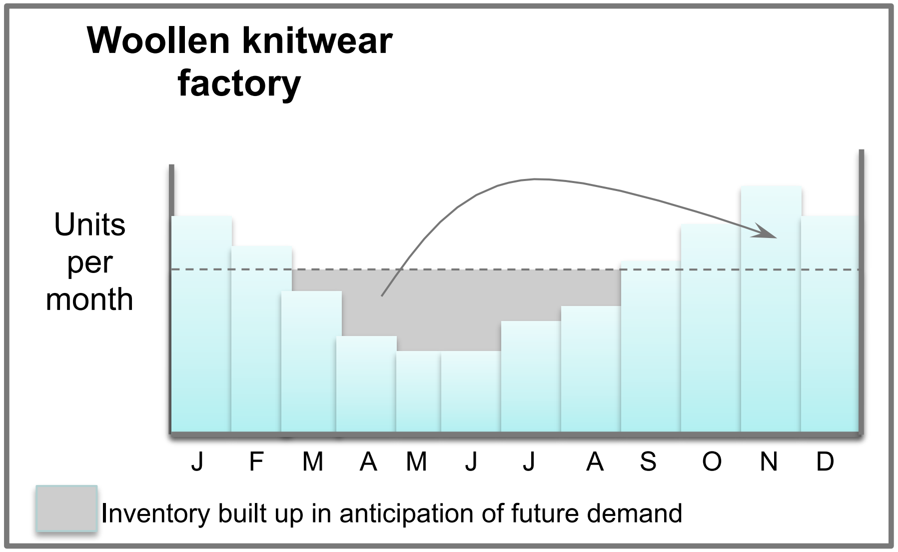
(c) Smoothing with inventory

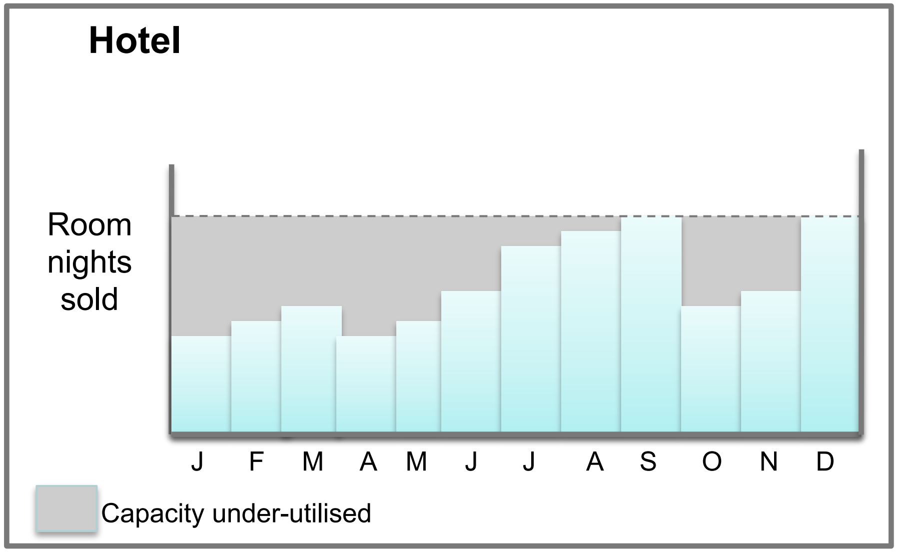
Medium-term Capacity Planning

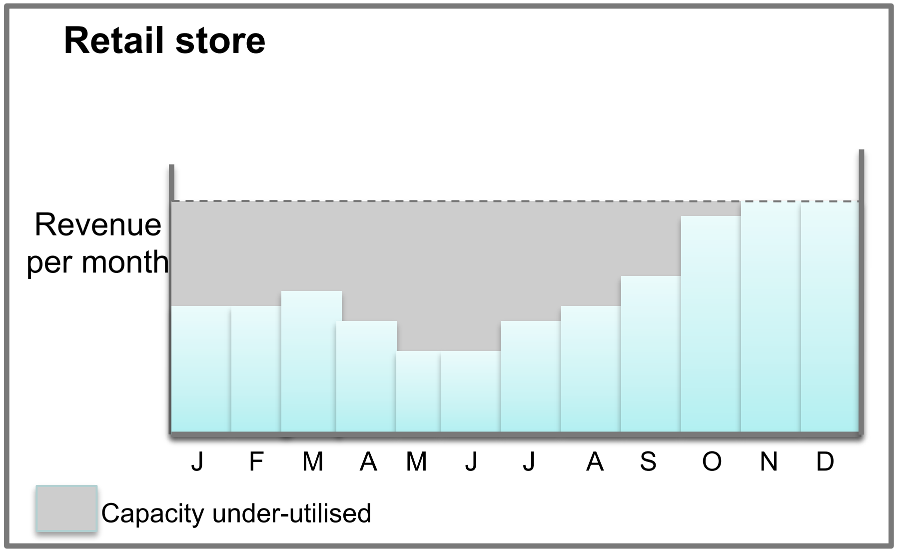
- There are three options available for coping with variations in demand:
  - **Level capacity plan:** processing capacity is set at a uniform level throughout the planning period, regardless of the fluctuations in forecast demand
  - **Chase demand plan:** attempts to match capacity closely to the varying levels of forecast demand
  - **Demand management:** change demand to suit capacity

Level Capacity Plan









Level Capacity Plan

Capacity at uniform level throughout the planning period

- Same number of staff operate the same processes
- Finished goods transferred to inventory in anticipation of sales at later time
- Suitable for non-perishable goods

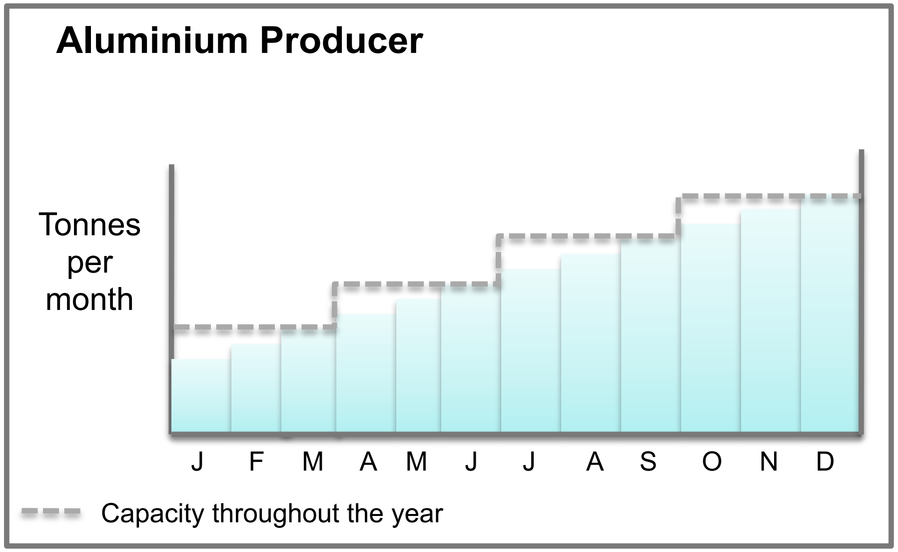
Advantages:

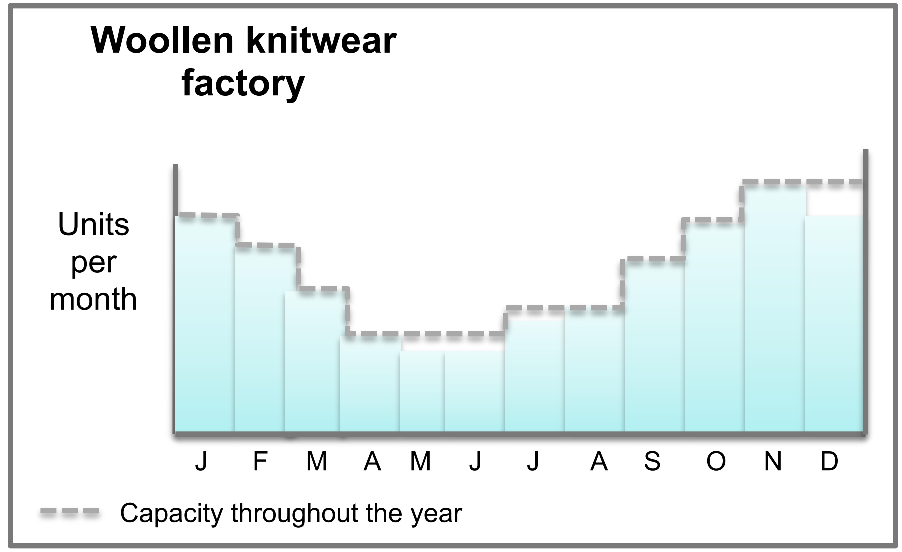
- Stable employment patterns
- High process utilisation
- High productivity with low unit costs

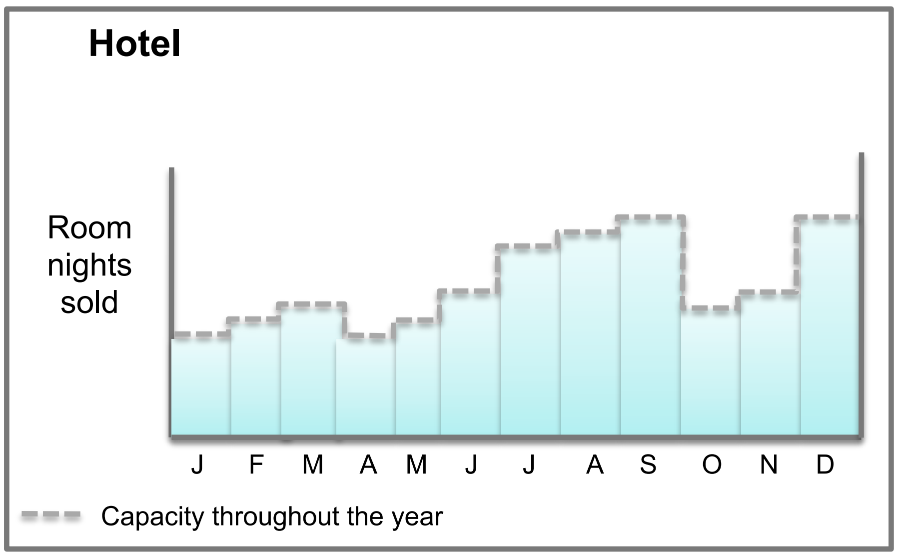
Disadvantages:

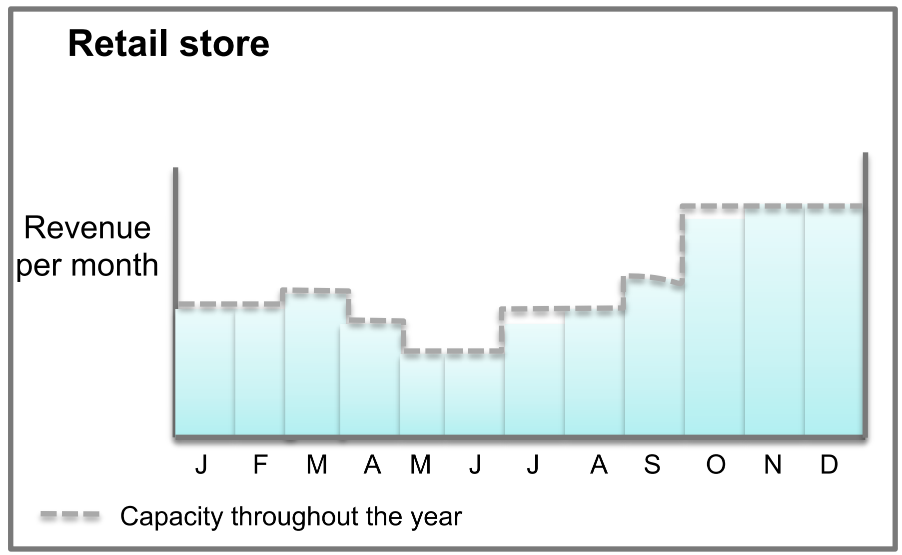
- Considerable inventory costs
- Decision-making: what to produce for inventory vs immediate sale
- High over/under utilisation levels for service operations

Chase Demand Plan









Adjusting Capacity

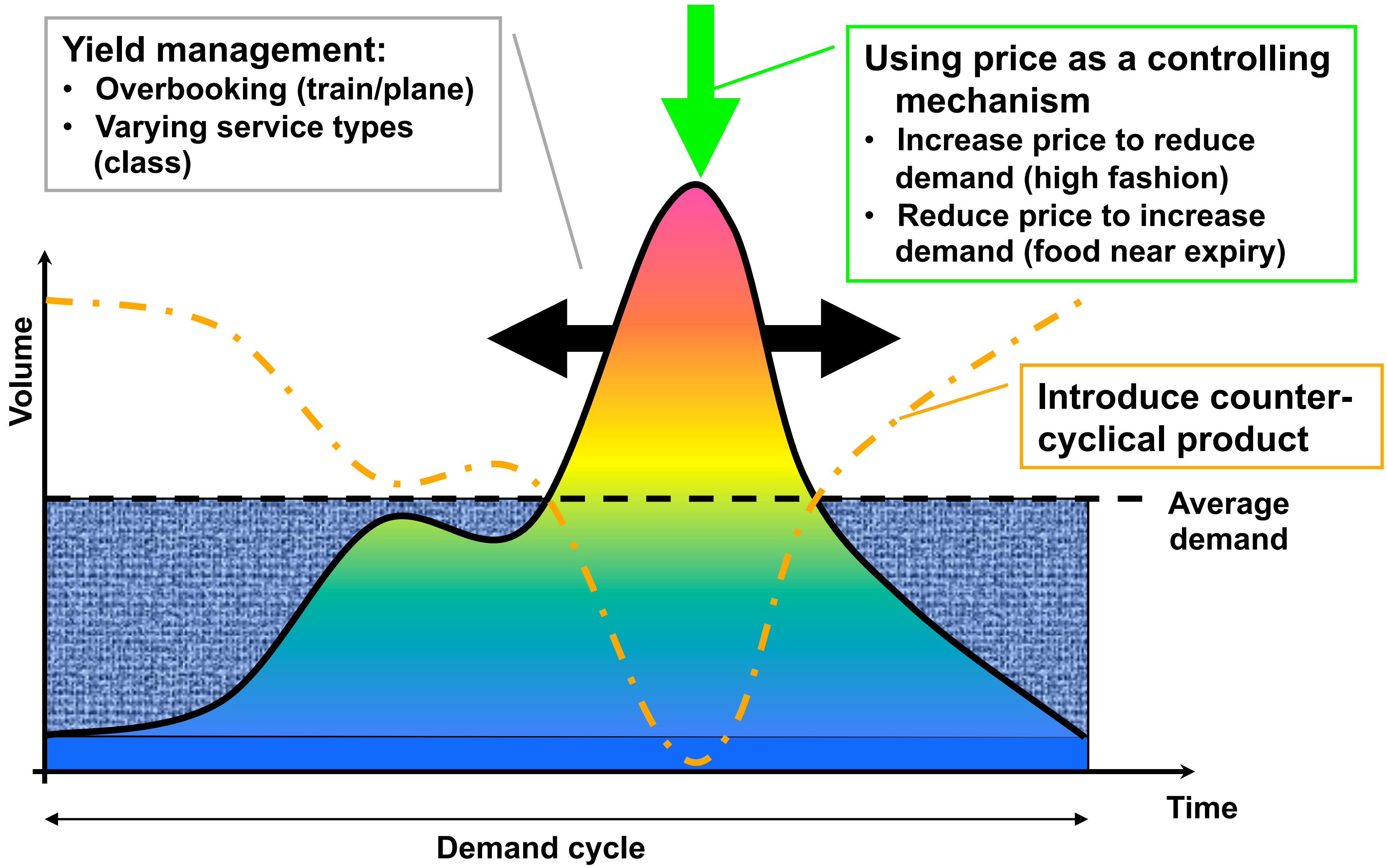
Common methods

- Overtime and idle time
- Varying the size of the workforce (hire and fire)
- Using part-time staff
- Subcontracting

Trade-offs

- Inventory cost vs. cost of changing capacity
- Flexibility vs. quality
- Customer satisfaction vs. employee satisfaction

Demand Management



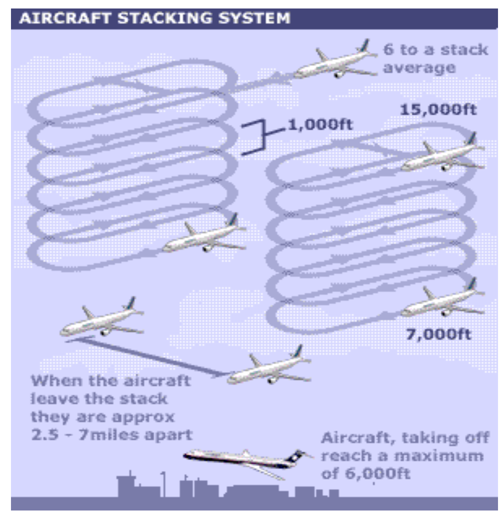
Short-term Capacity Planning

- Manage order mix
- Schedule downtime appropriately
- Overtime planning
- Outsource
- …

Queueing Modelling

Example of Queues!

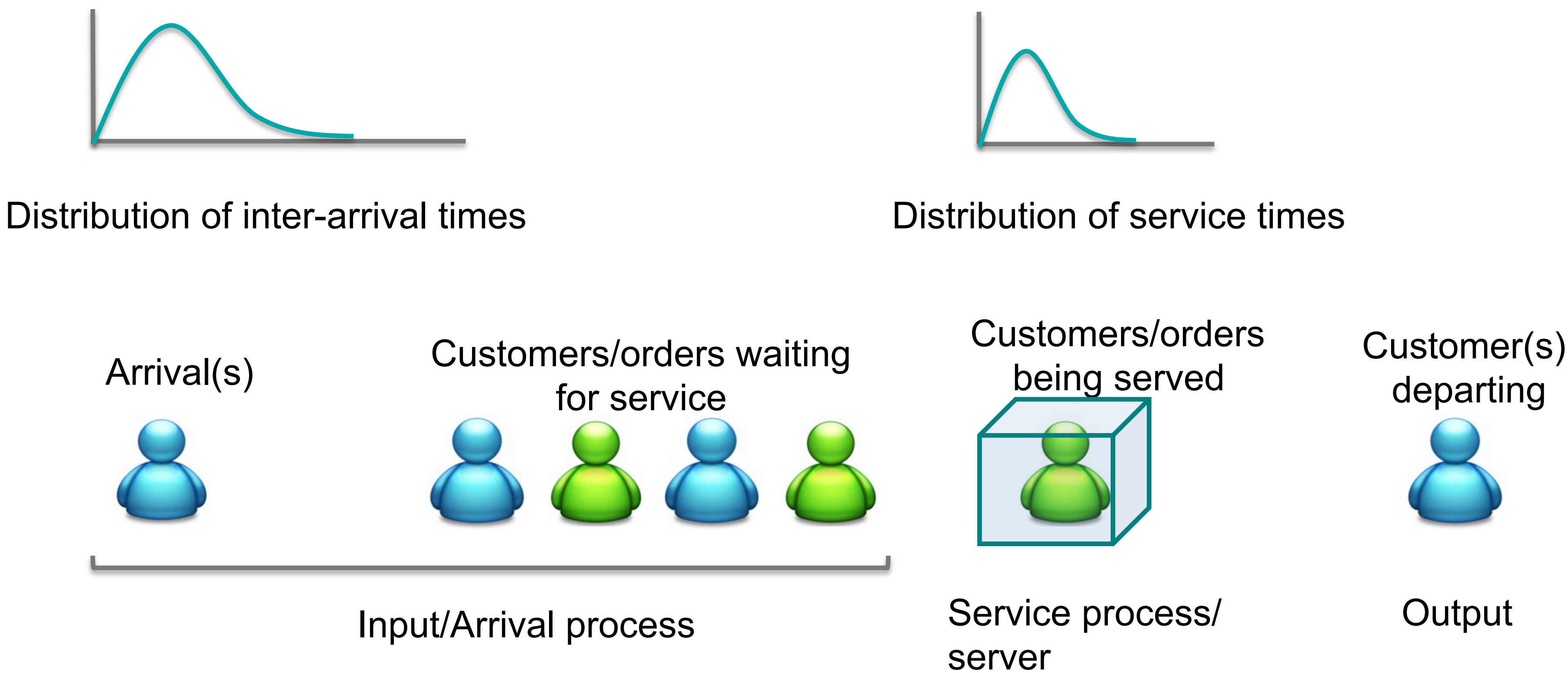
Operation	Arrivals of … in a queue	Processed or served by …
Car Paint Plant	Mercedes	Paint Station
Supermarket	Shoppers	Checkouts
Aircraft Landing	Aircraft	Flight controller
Packet Switching	Data Packets	Switches



Discussion

- How are queues and capacity related?
- Who cares about queues?
- Who cares about capacity?

Components of a Queuing System



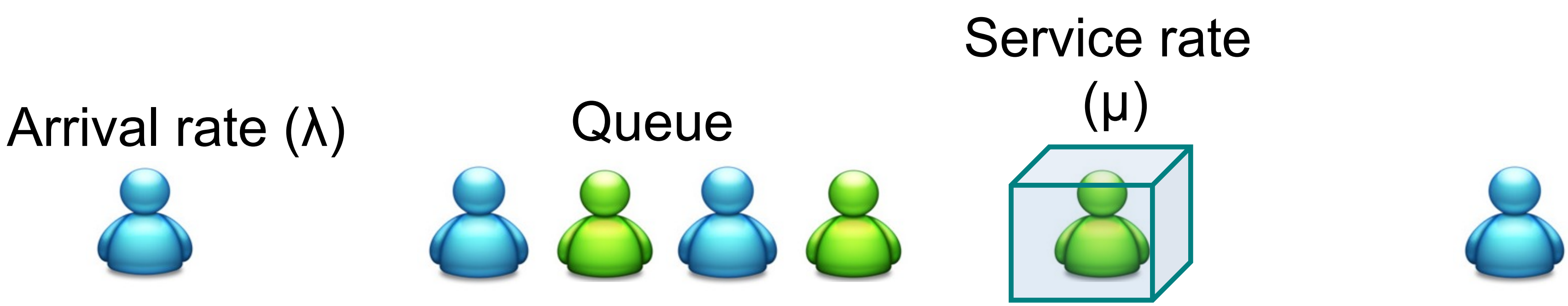
Balancing Capacity and Demand: Key Issues

- Unacceptable queueing times vs. unacceptably low utilisation of servers/machines [capacity] – trade-off
- Affect of variation in inter-arrival and service/processing times – capacity and demand rarely match
- It is important to be able to predict the expected waiting times and average utilisation of a processing system

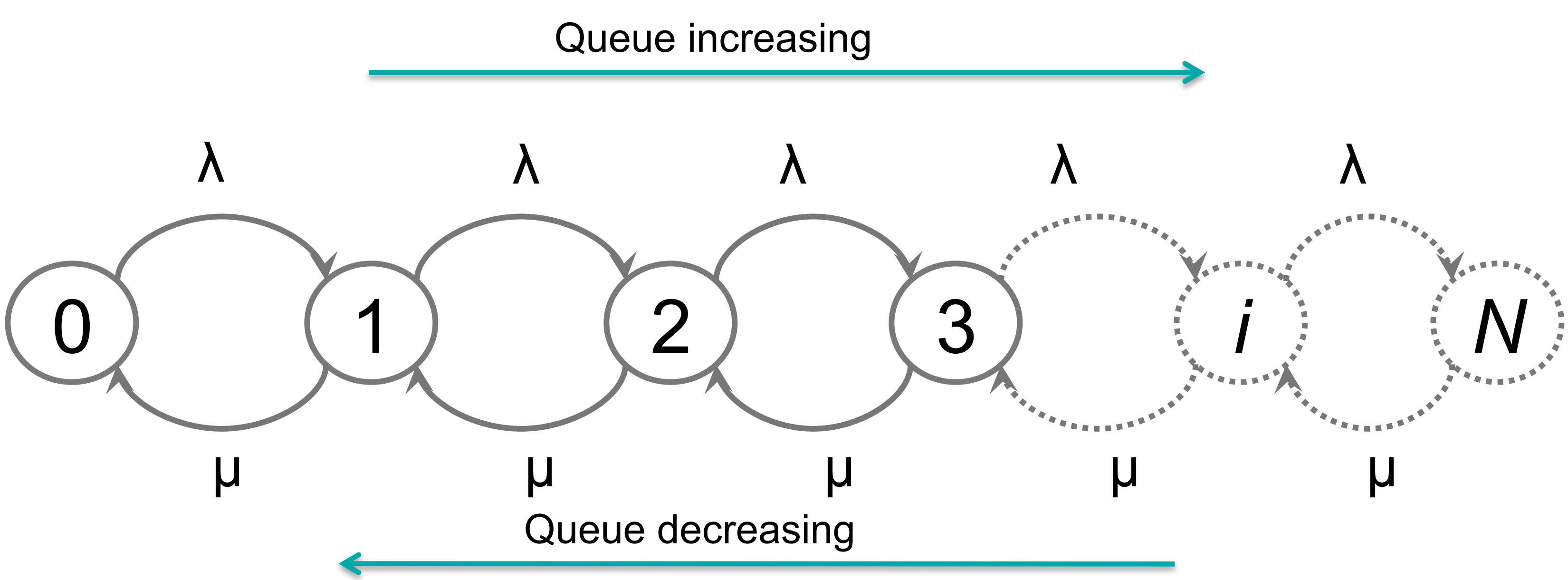
Analysis of Queues

- What is the average time an order [or a customer!] spends
  - in the system?
  - in the queue?
- What is the average length of the queue?
- What is the average resource utilisation?
- When is it justified to increase resources?

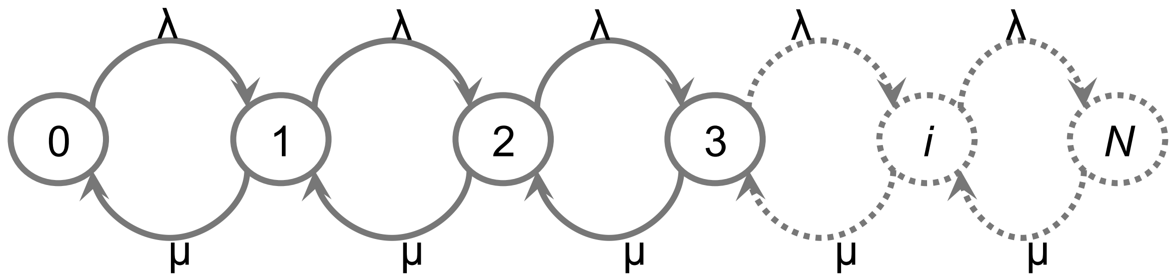
Modelling a Queue-Server System



Assume λ/μ < 1



Modelling a Queue-Server System



Let $\rho = \lambda/\mu$

Let p_i = probability of *i* orders being in the system after reaching equilibrium. Then <u>at equilibrium – p_i not changing:</u>

$\boxed{\mu p_1 = \lambda p_0}$ or $p_1 = \rho p_0$ for Node 0

Hence $p_2 = \rho p_1$ for Node 1.

Then $p_3 = \rho p_2$ for Node 2. …

So $p_2 = \rho^2 p_0,$ $p_3 = \rho^3 p_0$ and we can generalise this to:

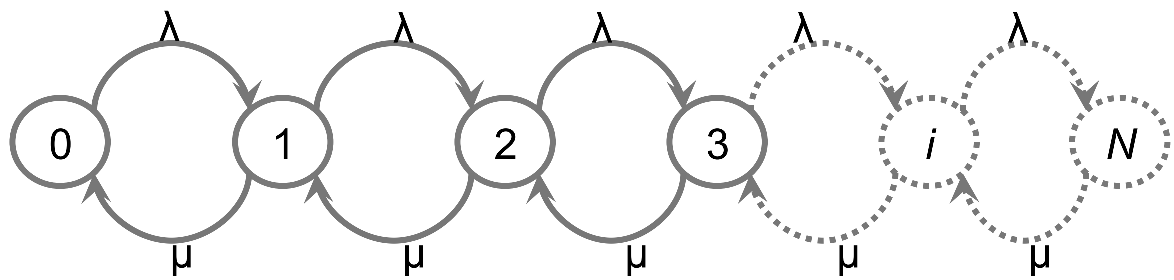
$$\boxed{p_{N+1} = \rho p_N = \rho^{N+1} p_0}$$

Noting that $p_0 + p_1 + p_2 + \ \dots + p_i + \ \dots = 1$

Then $p_0 [1 + \rho + \rho^2 + \rho^3 + \dots + \rho^i] = 1 \ \rightarrow \ p_0/(1 - \rho) = 1$

Hence, we have $\boxed{p_0 = 1 - \rho \ \text{ and } \ p_i = \rho^i (1 - \rho)}$

Modelling a Queue-Server System



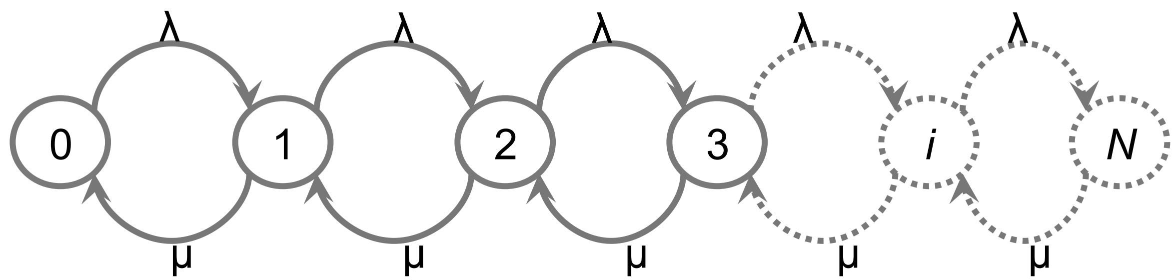
Average number of customers in the system, *N*:

$$\boxed{N = \sum_{i=0}^{\infty} i p_i = \sum_{i=0}^{\infty} i\rho^i (1-\rho) = \frac{\rho}{1-\rho}}$$

Average queue length, N_q:

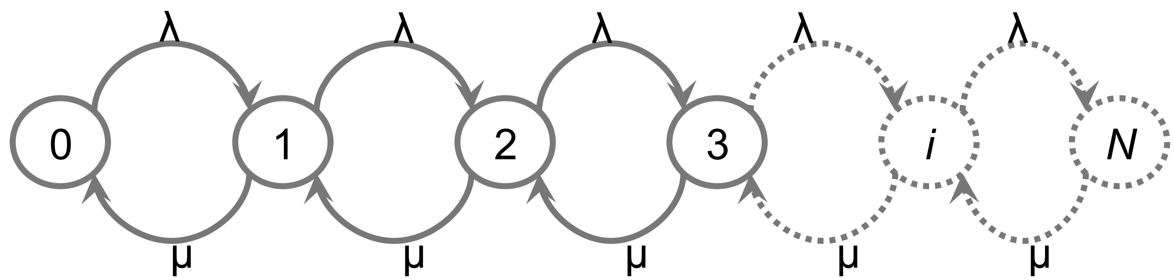
$$\boxed{N_q = \sum_{i=1}^{\infty} (i-1) p_i = \frac{\rho}{1-\rho} - (1-p_0) = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}}$$

Modelling a Queue-Server System



ρ - *Arrival rate / service rate*	*N – number of customers in system*	N_q *– number of customers in queue*
0.1	0.11	0.011
0.5	1	0.5
0.8	4	3.2
0.9	9	8.1

Modelling a Queue-Server System



Average time customers spend in the system, *W*:

From Little's Law: N= λW

$$\boxed{W = \frac{\rho}{(1-\rho)\lambda} = \frac{1}{\mu(1-\rho)}}$$

Average waiting time, W_q:

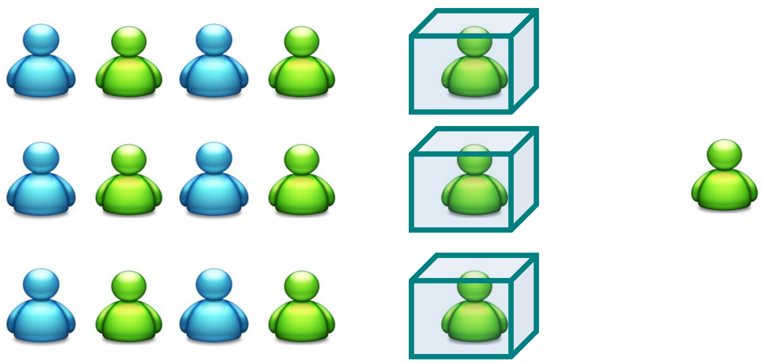
$$\boxed{W_q = W - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}}$$

Server utilisation: $\boxed{\rho = \lambda/\mu}$

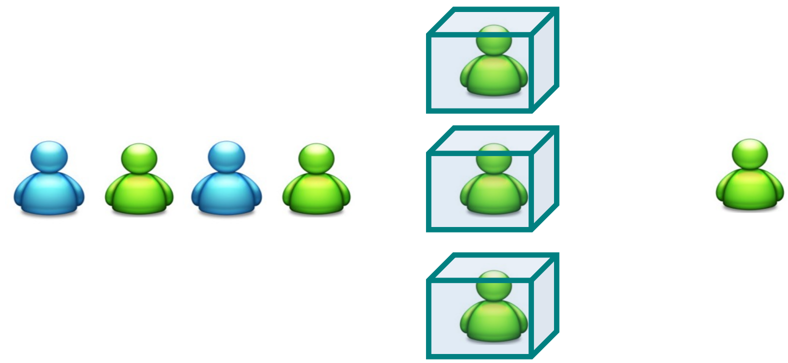
Using a Queueing Model

Can expand analysis to other cases …

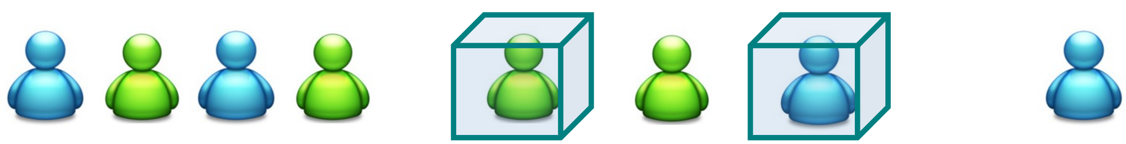
Multi queue/Multi channel



Single queue/Multi channel



Multi stage



Takeaways from Today

Capacity can be effected by various operational variables, and need to be managed strategically

Operations Management

Dr Feryal Erhun
fe251@cam.ac.uk