An Overview of Recent Research in Design and Operation of Manufacturing Systems

by

Stanley B. Gershwin
Massachusetts Institute of Technology
Laboratory for Manufacturing and Productivity
Cambridge, Massachusetts 02139-4307
617-253-2149
gershwin@mit.edu
http://web.mit.edu/org/m/manuf-sys/www

Laboratoire MASI (URA 818 CNRS)
Institut Blaise Pascal
Université Pierre et Marie Curie
4, place Jussieu
75252 PARIS CEDEX 05
FRANCE

September 19, 1995
Outline

- Introduction
- Design
- Operation
- Heuristics
- Dynamic Programming
- Single Station
- Networks
- Hierarchy
- New Directions
Introduction

- Two major issues: design and operation
- Performance measures:
  - Capital cost
  - Production Rate or Capacity
  - Cycle Time
    \textit{the time a part spends in the system}
  - Lead Time
    \textit{the time between order and delivery}
  - Machine Utilization
  - Work in Process (WIP)
  - On-time Performance
Design

- Jackson-type networks:
  - infinite buffers
  - exact decomposition
  - CANQ by Solberg
  - FMS, multiple part types.

- Flow lines or transfer lines
  - exact results for 0-buffer lines, infinite buffer lines, two-machine lines.
  - approximate decomposition results for long lines with finite buffers.
  - Single part types.
  - Extended to tree-structured (acyclic) assembly/disassembly systems.
  - Finite buffers important to limit inventory – propagation of disruptions upstream.
Current work:

• loop networks
• optimization
• multiple machines per station
• variability of output
Difficulty with Loops

- **Correlation:** the total number of parts in the system is fixed. If a part is known to be at a certain location, there are fewer parts for all the other locations.

- However, this appears NOT to be a problem if the loop is sufficiently large.

- What will happen when there are multiple loops?
Operation

- We focus on real-time scheduling.

- Scheduling is the selection of times for future controllable events.

- Scheduling calculations must be fast to be useful in real time.
Heuristic approaches

• FIFO,
• SRPT
• kanban
• CONWIP
• LS
• many others

*important distinction: LS keeps track of when parts are due.*

• advantage: easy to implement
• disadvantage: hard to determine if or why policies are good
Dynamic programming approaches

• advantage: optimal

• disadvantages:
  – impossible to determine exactly except for trivial problems;
  – difficult to implement optimal even if we could get it;
  – no bounds for approximations;
  – most people believe that optimal is not much better than good heuristic.
Oldser-Suri, Kimemia-Gershwin formulation

- Single station — no buffers.
- Possibly multiple machines.
- Possibly multiple part types.
- Machine failures and operations are the only important events.
- Operations are much *more frequent* than failures.
- Operations are much *shorter* than failures.
- Changing operations causes no cost or delay.
- The time horizon is very long.
- A long term production rate target is specified, and *is feasible*.

- **essential role of surplus**
- **relationship between surplus and slack**
- **analytical solution for one machine, one part type**
Problem Formulation

- The goal is to meet the long-term specified production rate target in a way that keeps inventory and backlog costs small.

- Three-level hierarchy
  - Level 1 determines Level 2 parameters.
  - Level 2 observes failures and surpluses, and determines short-term production rates.
  - Level 3 determines loading times.
Cumulative Production and Demand

\[ x_j(t) \]

\[ \int_0^t u_j dt \]

CUMULATIVE PRODUCTION OF PART TYPE \( j \)
Three-Level Hierarchy

Level 1
Select hedging points

Level 2
Select production rates

 failures, surpluses

loading times

Level 3
Select loading times
Hedging Point Problem — variables

- **States**
  - $x = \text{surplus/backlog}$
  - $\alpha = \text{machine state}$

- **Controls**
  - $u = \text{short-term production rate}$

- **Parameters**
  - $\lambda_{\beta\alpha} = \text{transition rate matrix; failure and repair rates}$
    
    \[ \lambda_{\beta\alpha}\delta t = \text{probability of moving from } \alpha \text{ to } \beta \text{ in } (t, t + \delta t) \]
  - $d = \text{long-term demand rate target}$
  - $g() = \text{cost of surplus and backlog}$
  - $\tau_{ij} = \text{operation time of Part Type } j \text{ at Machine } i$
Hedging Point Problem — formulation

For large $T$, find $u(x, \alpha)$ to satisfy

$$\min E \int_0^T g(x(s)) ds$$

such that

$$\frac{dx}{dt} = u - d$$

Markov dynamics for $\alpha$

$$u(t) \in \Omega(\alpha(t))$$
Short-Term Capacity Set

\[ \Omega(\alpha(t)) = \text{capacity set:} \]

\[ \sum_j \tau_{i,j} u_j \leq \alpha_i \]

\[ u_j \geq 0 \]
Objective Function

\[ g(x) = \sum_{j} g_j(x_j) \]
Hedging Point Problem — solution

Bellman equation:

Let 
\[ J(x, \alpha, t) = \min E \left\{ \int_t^T g(x(s)) \, ds \mid x(t) = x, \alpha(t) = \alpha \right\} \]

Then

\[ -\frac{\partial J}{\partial t} = \min_{u \in \Omega(\alpha)} \left\{ g(x) + \frac{\partial J}{\partial x} (u - d) + \sum_{\beta} \lambda_{\beta \alpha} J(x, \beta, t) \right\} \]

If \( J \) were known, \( u \) would be determined by the linear program:

\[ \min \frac{\partial J}{\partial x} u \]

subject to

\[ u \in \Omega(\alpha) \]
Ω and the Partition of $x$ Space
One-Part-Type, One-Machine Problem

\[ g(x) = g_+ x^+ + g_- x^- \]

\[ \lambda_{01} = p; \lambda_{11} = -p; \]
\[ \lambda_{10} = r; \lambda_{00} = -r. \]

\[ \Omega(0) = \{0\}, \]
\[ \Omega(1) = \{u \mid 0 \leq u \leq \mu\}. \]

where

\[ \mu = 1/\tau \text{ is the maximum production rate of the machine.} \]

\[ d \text{ is feasible, i.e.,} \]
\[ d < \frac{\mu r}{r + p} \]
Hedging Point for
One-Part-Type, One-Machine Problem

If \( g_+ - b(g_+ + g_-) < 0 \), \( Z = \frac{\ln(Kb(1 + \frac{g_-}{g_+}) > 0)}{b} \).

If \( g_+ - b(g_+ + g_-) \geq 0 \), \( Z = 0 \).

where

\[
K = \frac{\mu p}{b(r + p)(\mu - d)}
\]

\[
b = \frac{r}{d} - \frac{p}{\mu - d}
\]
Cumulative Production and Demand
Optimal Behavior

\[ x(t) \]

\[ \int_0^t u dt \]

Hedging point Z

\[ dt \]
Multiple Part Types, Multiple Machines

• **Difficulty:** Bellman equation impossible to solve.

• **Approaches:**

  1. Modify a strategy for a failure-free system.
  2. Approximate the Bellman cost-to-go function with a quadratic:

\[
J(x, \alpha, t) \approx J^*(T - t) + \frac{1}{2}(x - Z)^T A(x - Z)
\]
Real-Time Controller for First Approach

1. Rank order the products.

2. Produce the most important product, until its surplus reaches its hedging point. The others fall behind.

3. Keep the most important product at its hedging point. Devote all remaining capacity to the second most important product, until it reaches its hedging point. The others fall further behind.

4. Keep the two most important products at their hedging points. Devote all remaining capacity to the third most important product, until it reaches its hedging point. The others fall still further behind.

5. etc.

6. If there is sufficient capacity, all parts eventually reach their hedging points.
Real-Time Controller for Second Approach

\( u \) is determined by the time-varying linear program:

\[
\min (x - Z)^T A u
\]

subject to

\[
\sum_j \tau_{ij} u_j \leq \alpha_j
\]

\[
u_j \geq 0
\]
Special Case: One Machine, Many Part Types

*Produce the part that is furthest behind its hedging point.*
Van Ryzin-Lou-Gershwin Formulation

- two-station (single-machine per station)

- optimal boundaries determined numerically
- approximate boundaries — two-boundary policy
Bai approach

- assumed two-boundary policy for multi-stage system
- simple approximate model to determine optimal buffer sizes and hedging points
- developed for multiple part type system
Networks for Real-time Control

- CONWIP

- two-boundary

- timeliness of information seems to be important issue
Motivation for Hierarchy

- **Difficulty** Systems are large and heterogeneous.

- **Goal** Good performance.

- **Constraint** Limited computation.

- **Approach**
  
  - Decomposition by frequency/duration and location.
  
  - Develop building blocks – solutions to limited problems.
  
  - Construct hierarchy from building blocks.
Building Blocks

Solve problems with limited frequency range and limited factory floor space.

*Input* Relatively long range targets and relatively short range disturbances.

*Output* Controllable events and relatively short term targets.

*Examples* Hedging point problem/solution.

Staircase policy.

Corridor policy.
# Building Blocks

<table>
<thead>
<tr>
<th>Problem</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous</td>
<td>Discrete (disturbances)</td>
</tr>
<tr>
<td>Hedging point</td>
<td>Long term production targets</td>
<td>Failures</td>
</tr>
<tr>
<td>Hedging point</td>
<td>Long term production targets</td>
<td>Failures, starvations, blockages</td>
</tr>
<tr>
<td>spatially decomposed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staircase</td>
<td>Long term production targets</td>
<td>—</td>
</tr>
<tr>
<td>Corridor</td>
<td>Long term production targets</td>
<td>—</td>
</tr>
</tbody>
</table>
Multiple Time Scale Production

Cumulative Production and Requirements for Type $j$ parts

- SET UP FOR TYPE $j$
- NOT SET UP FOR TYPE $j$
- SET UP FOR TYPE $j$
Four-Level Hierarchy

Level 1
Select hedging points, etc.

Level 2
Select production rates, setup changes

Level 3
Select production rates

Level 4
Select loading times

set up changes
surpluses

machine failures
surpluses

loading times
Capacity

Capacity is a complex concept because it is

• not a scalar: a set;
• not a constant set: a stochastic set;
• not a single stochastic set: a set of sets, one for each time scale.
Definitions

- A resource is any part of the production system that is not consumed or transformed during the production process.

- An activity is a pair of events associated with a resource. The first event corresponds to the start of the activity, and the second is the end of the activity. Only one activity can appear at a resource at any time.

- A disruption is an undesirable activity.

- Let $i$ be a resource and $j$ an activity. Define $\gamma_{ij}(t)$ to be the activity state of Resource $i$. This is a binary variable which is 1 if Resource $i$ is occupied by Activity $j$ at time $t$, and 0 otherwise; or

\[
\sum_j \gamma_{ij}(t) \leq 1.
\]
Production Rate

• Let $N_{ij}(T)$ be the total number of times that Resource $i$ is occupied by Activity $j$ in $(0, T)$.

• Then define the Activity $j$ frequency (or rate) by

$$u_{ij} = \frac{1}{T}N_{ij}(T).$$

(For example, $N(T)$ is cumulative production and $u$ is average production rate.)

• Let $\tau_{ij}$ be the average duration of Activity $j$ at Resource $i$ (e.g., operation time.) Then

$$\tau_{ij}u_{ij} = E\gamma_{ij}.$$  

• $E\gamma_{ij}$ is the fraction of time that Machine $i$ is busy doing operations on Part Type $j$. 
Capacity

\[ 1 \geq E \sum_j \gamma_{ij}(t) = \sum_j \tau_{ij} u_{ij} \text{ for all resources } i. \]

This is the fundamental capacity limitation: no resource can be occupied more than 100\% of the time.
Frequency Separation

Assumption

The events and activities can be grouped into sets $\mathcal{L}_1, \mathcal{L}_2, \ldots$ such that for each set $\mathcal{L}_k$, there is a characteristic frequency $f_k$ satisfying

$$0 = f_1 \ll f_2 \ll \ldots \ll f_k \ll f_{k+1} \ll \ldots$$

and

$$j \in \mathcal{L}_k \Rightarrow f_{k-1} \ll u_{ij} \ll f_{k+1}$$

That is, the activities and events in class $\mathcal{L}_k$ have frequencies that are much closer to each other, and to $f_k$, than to the frequency of any other activity or event.

$k = L(j)$ is the level of event $[i,j]$.

Example: In one factory, operations are high frequency events, failures are medium frequency events, and setup changes are low frequency events.
Level $k$ Observer

- A Level $k$ observer or manager has a precise model only of events that occur with frequencies near $f_k$.
- The observer has only simplified models of events that occur at frequencies far from $f_k$.

**Example:** Low level line managers dispatch repair personnel in reacting to machine failures, and reschedule production on surviving machines.

- They do not know or care when each operator performs each operation; they only care about hourly or daily quotas.
- They know that failures and repairs occur with exponential distributions; they know the means of the distributions.
- They know *nothing* about when the next major setup change will occur except that setup changes are much less frequent than failures.
Level $k$ States

- **Example:** Low level managers know the current repair and setup state of each machine.
- They know the current production rate target of each part type. This target changes as machines go up and down and setup changes occur.
- $\gamma_{ij}^k(t)$ is the *Level k activity state of Resource i*. This is defined, for activities $j$ whose level $L(j)$ is $k$ or higher, as

$$
\gamma_{ij}^k(t) = \gamma_{ij}(t) \text{ for } L(j) \leq k.
$$

That is, $\gamma_{ij}^k$ is the part of $\gamma_{ij}$ that changes more slowly than $f_{k+1}$.
Level $k$ Rates

- $E_k$ is the Level $k$ expectation operator. It is the conditional expectation, given that all Level $m$ quantities $(\gamma_{ij}^m(t), m \leq k)$ remain constant at their values at time $t$.

- For any random variable $z$,

$$E_k z(t) = E(z \mid \gamma_{ij}^m = \gamma_{ij}^m(t))$$

for all $i, j$ such that $L(j) \leq m \leq k$.

- We define $u_{ij}^k$, the Level $k$ rate of Activity $j$ at Resource $i$, as

$$u_{ij}^k(t) = \frac{E_k \gamma_{ij}(t)}{\tau_{ij}} \text{ for } L(j) > k.$$
Important Theorem

\[ E_{k-1} \underline{u}_{ij}^k = \underline{u}_{ij}^{k-1} \]

This relates rates at different levels of the hierarchy.
Control Problem

• **Problem 1:** Given $u_{ij}^{k-1}$, find $u_{ij}^k$ (for all $j$, $L(j) > k$) satisfying ** (and other conditions).

• The superior level specifies the goal ($u_{ij}^{k-1}$) and the lower level must find $u_{ij}^k$ to be as close as possible, while satisfying capacity constraints and suffering from high frequency disruptions.
Statistics Problem

- **Problem 2**: Given $u_{ij}^k$, find $u_{ij}^{k-1}$ (for all $j$, $L(j) > k$) satisfying **(and other conditions)**.

- The lower level specifies the observation $(u_{ij}^k)$ and the higher level must find an average $u_{ij}^k$. 
Research strategy

• Use dynamic programming methods to suggest policy structures.

• Develop decomposition methods to evaluate policies with specified parameters.

• Develop optimization methods as extensions of evaluation methods.
Extensions

- more general networks and policies setups
- quality – yield, quality strategies
An Overview of Recent Research
in Design and Operation
of Manufacturing Systems

by

Stanley B. Gershwin
Massachusetts Institute of Technology
Laboratory for Manufacturing and Productivity
Cambridge, Massachusetts 02139-4307
617-253-2149
gershwin@mit.edu
http://web.mit.edu/org/m/manuf-sys/www

Departement Automatique et Productique
Ecole des Mines de Nantes
4, Rue Alfred Kastler - La Chantrerie
F44070 Nantes Cedex 03
FRANCE

September 22, 1995