Factory Operations Research Center (FORCe II)

SRC/ISMT 2004-OJ-1214

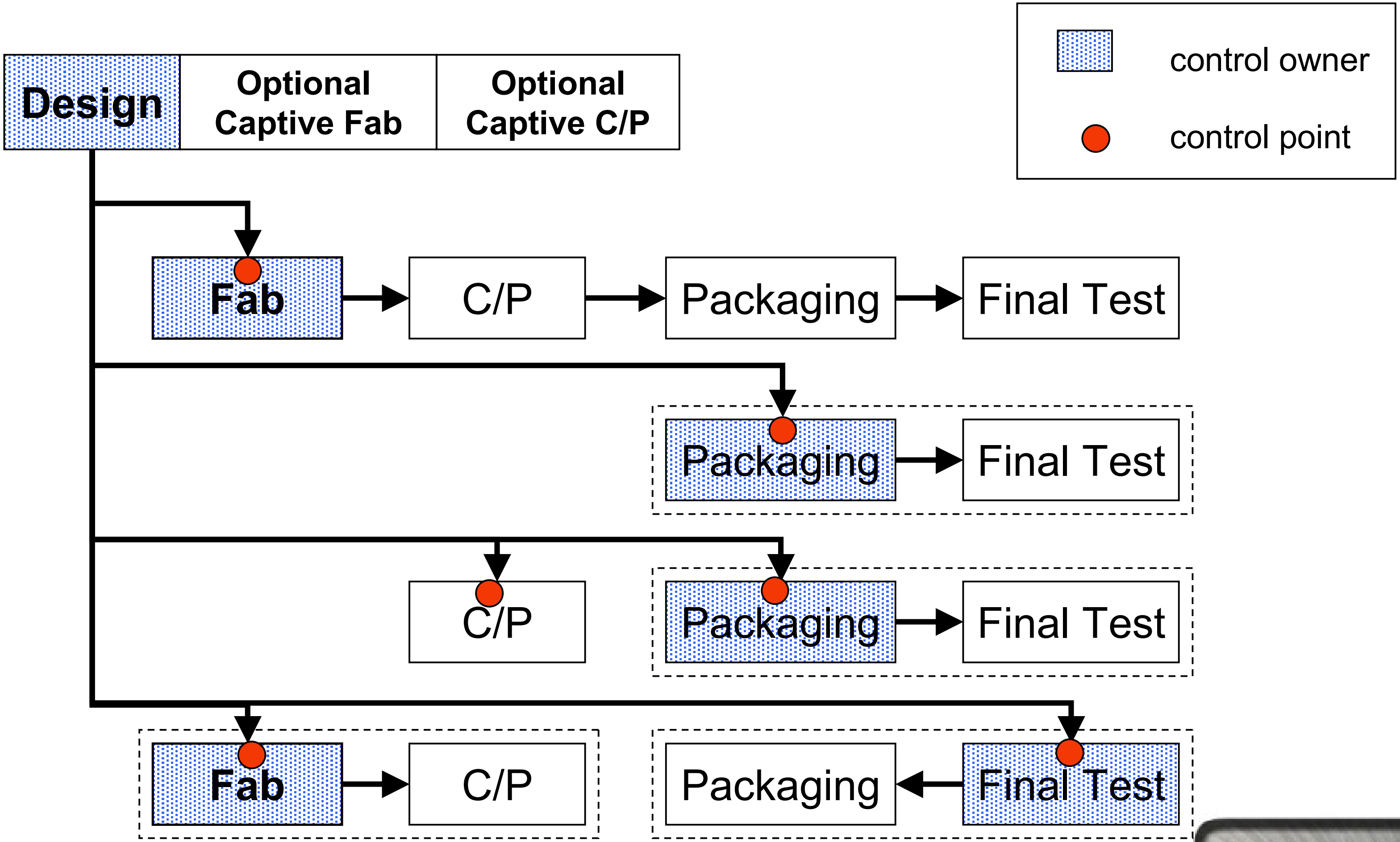
Configuration, monitoring and control of semiconductor supply chains

Jan. 7, 2005

Shi-Chung Chang (task 1)
Argon Chen (task 2)
Yon Chou (task 3)
National Taiwan University

INTERNATIONAL SEMAT

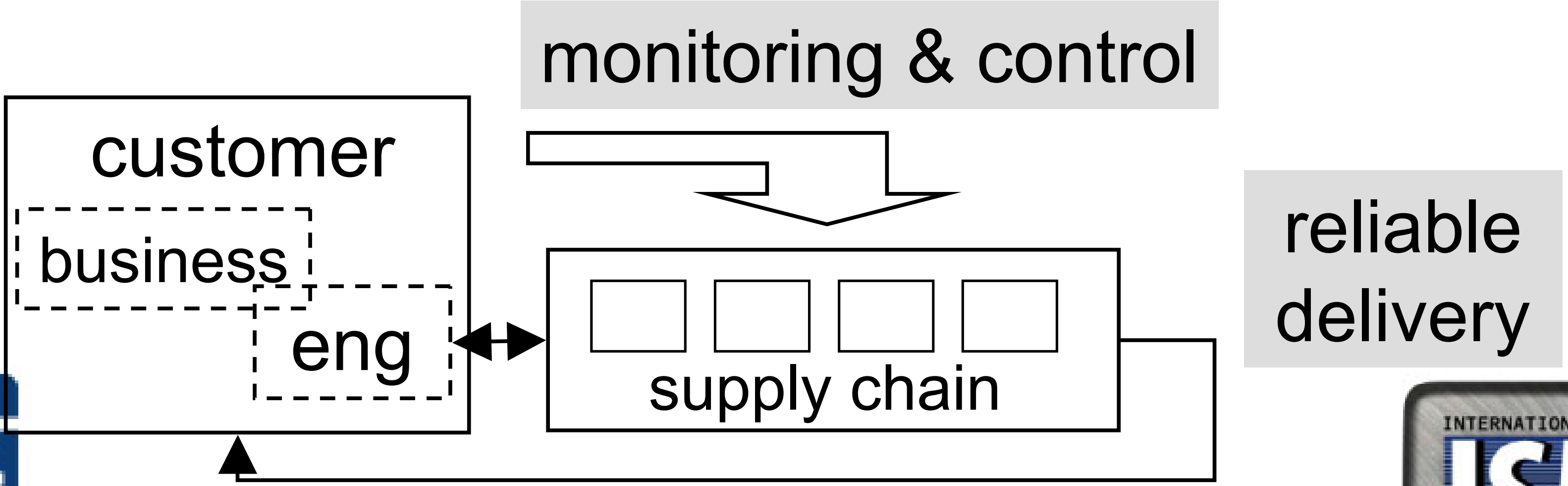
Multiple Threads of Manufacturing Services



Physical Layer

Challenges of Manufacturing Services

- Effective collaboration between engineering and manufacturing
- Reliable delivery
- Supply and service monitoring and control



Supply Chain Configuration, Monitoring and Control

Objectives: to enhance

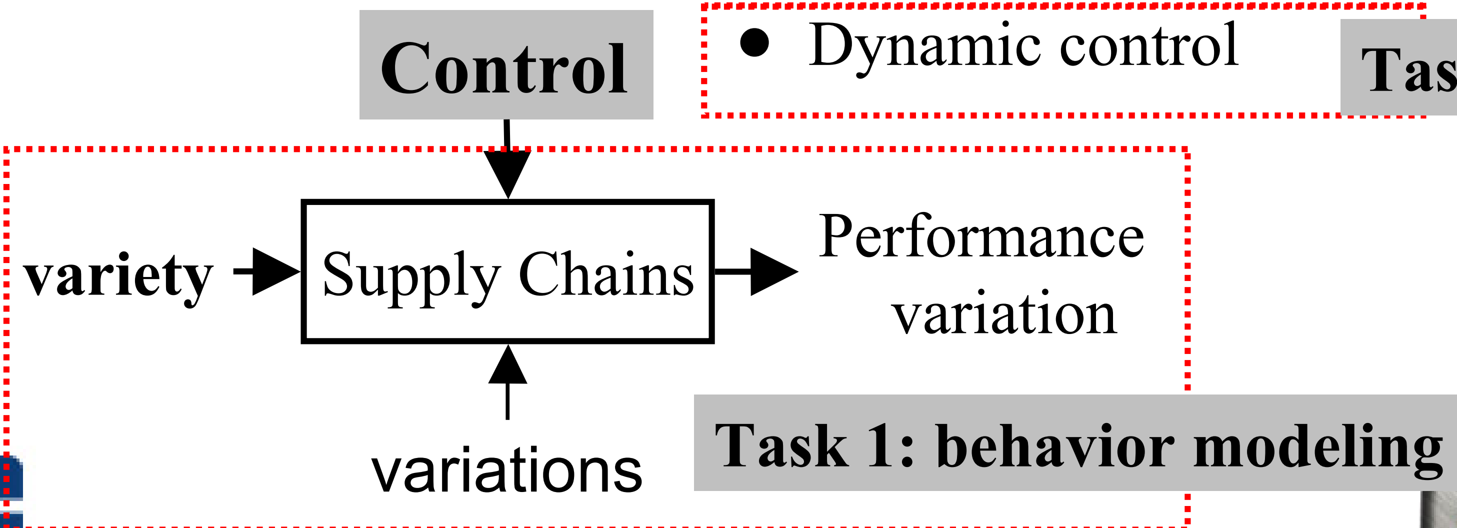
- Predictability
- Scalability
- Service differentiability

- Robust configuration
- Monitoring

Task 2

- Dynamic control

Task 3

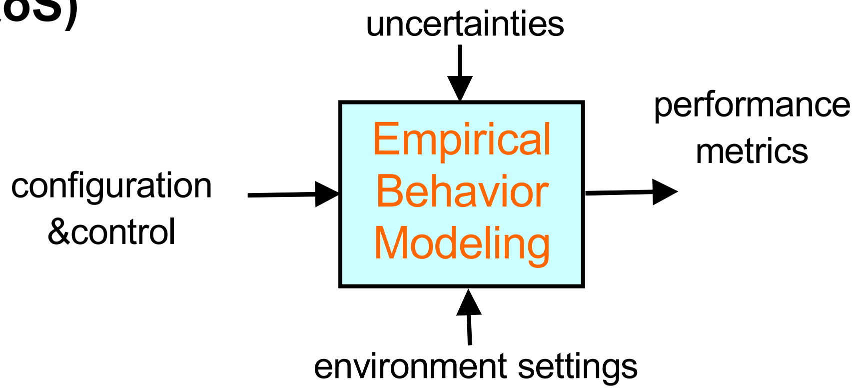


Task 1: Empirical Behavior Modeling

PI: Shi-Chung Chang
Co-PIs: Da-Yin Liao, Argon Chen

To develop methodology:

1. **Definition of quality of service (QoS) metrics**
   - Scalability
   - Controllability
   - Service Differentiability



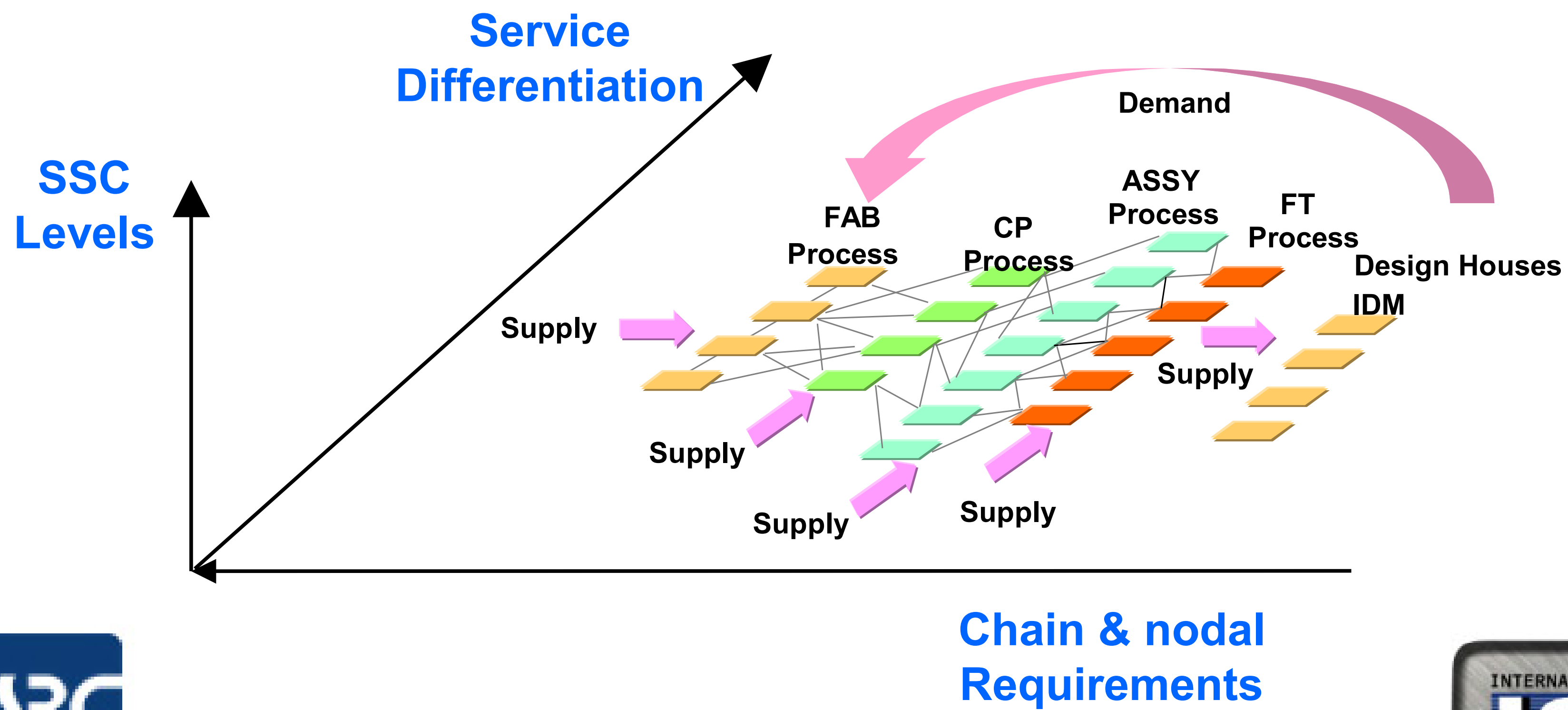
2. **Modeling and simulation**
   - Performance
   - Variability (engineering and business)
   - Capacity allocation & control

Mid-year Progress – Task 1

- Definition of Performance Metrics
  - **Have identified reduced set of QoS metrics from SCOR**
  - **Have defined the QoS translation problem among nodes of supply chain**
  - **Is developing a queueing network-based QoS translation method**
- Fab Behavior Modeling
  - **Have developed a baseline model**
    - **open queueing nework with priority**
    - **mean and variability**
  - **Have defined a response surface fitting problem**
  - **Is developing response surface modeling method**
- Simulation
  - **Have developed a baseline Fab simulator based on QN models**

Performance Metrics Definition

- SCOR-based six categories
  - quality, cost, cycle time, delivery, speed, and service



Key Level 1 Metrics for SSC

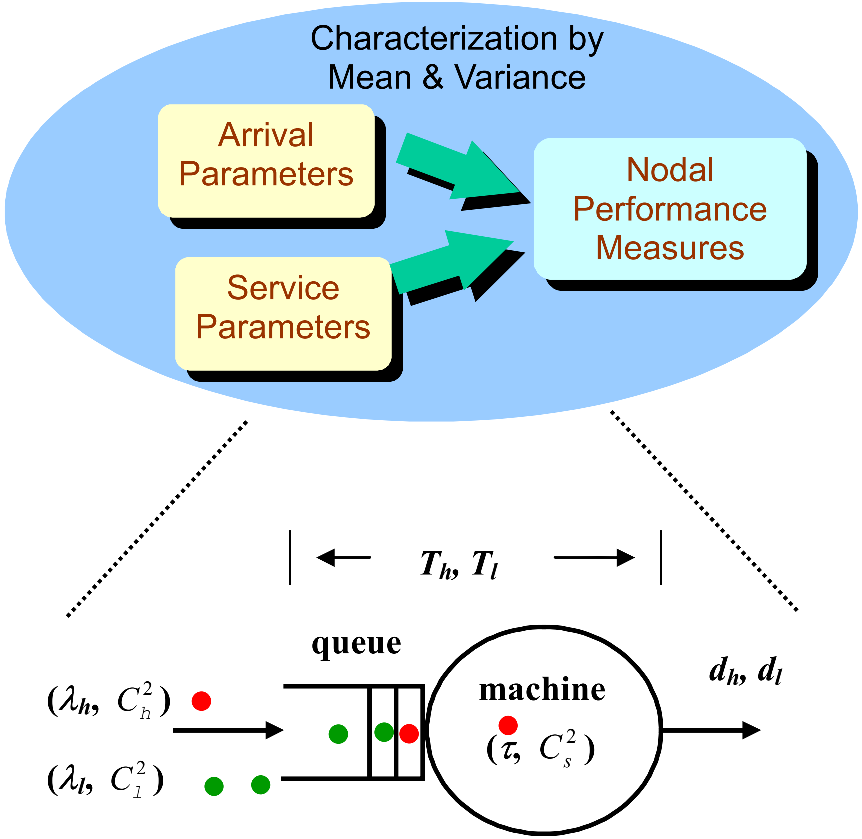
Q	Yield
	Yield Variability
Co	% of Downtime due to Non-availability of WIP
	Capacity Utilization
	Finished Goods Inventory Carrying Costs
	In-process Failure Rates
	Machine Wait Time
CT	Actual-to-theoretical Cycle Time
	Make Cycle Time
	Ratio of Actual to Theoretical Cycle Time
D	% of Orders Scheduled to Customer Request
	Average Days per Schedule Change
	Delivery Performance to Customer Request Date
	Forecast Accuracy
	Forecast Cycle

Sp	% Schedules Changed within Suppliers' Lead Time
	% Schedules Generated within Suppliers' Lead Time
	Intra-Manufacturing Re-Plan Cycle
	Schedule Achievement
	Schedule Interval
	Re-Plan Cycle Time
	Responsiveness Lead Time
Sv	% Orders/Lines Received On-Time to Demand Requirement
	%Orders/Lines Received with Correct Shipping Documents
	Delivery Performance to Customer Request Date

INTERNATIONAL SEMAT

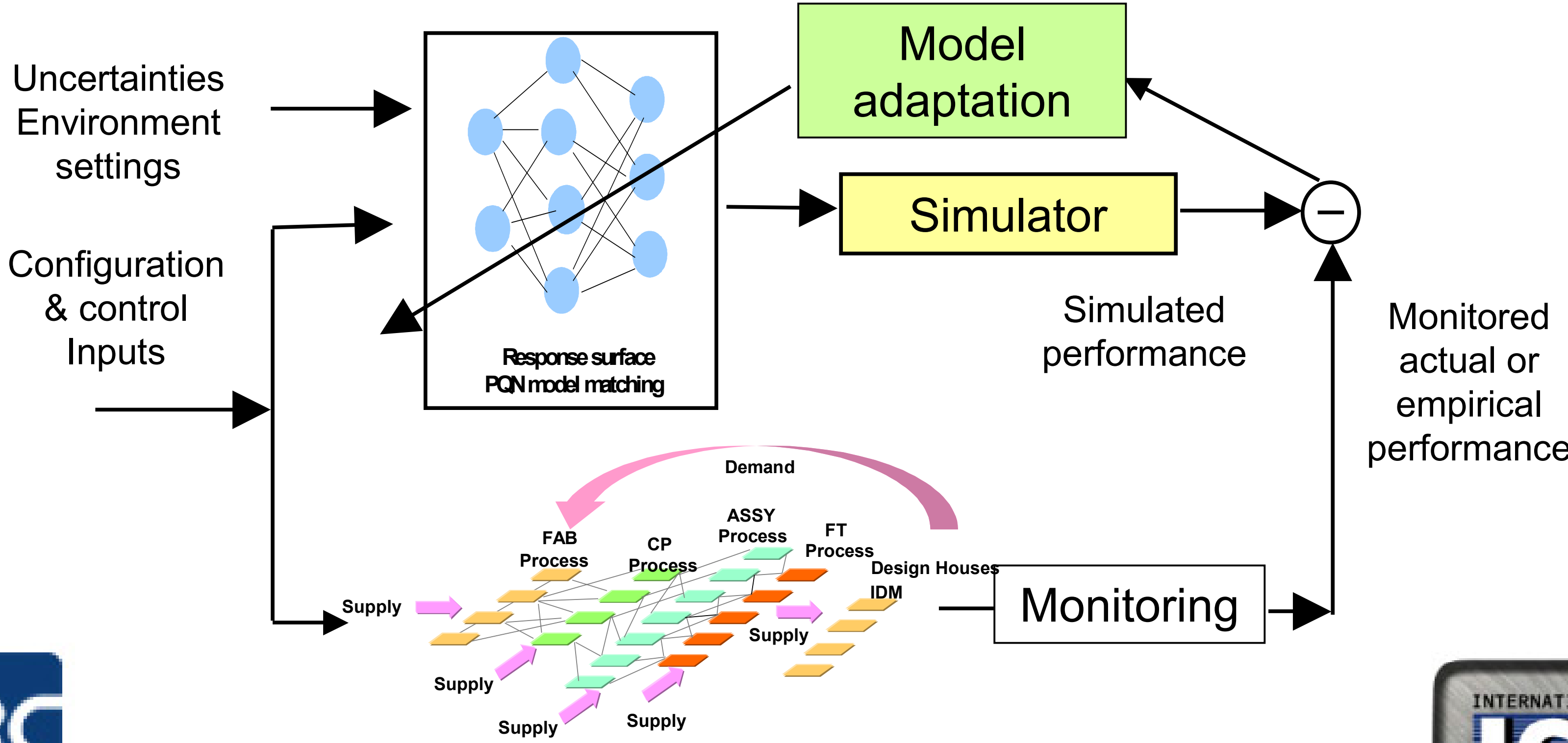
Characterization of Variability

- Sources
  - Process varieties
  - Engineering changes
  - Operation excursions
  - Demand plan
- Hybrid models
  - Response surface
  - Priority queueing
  - Simulation

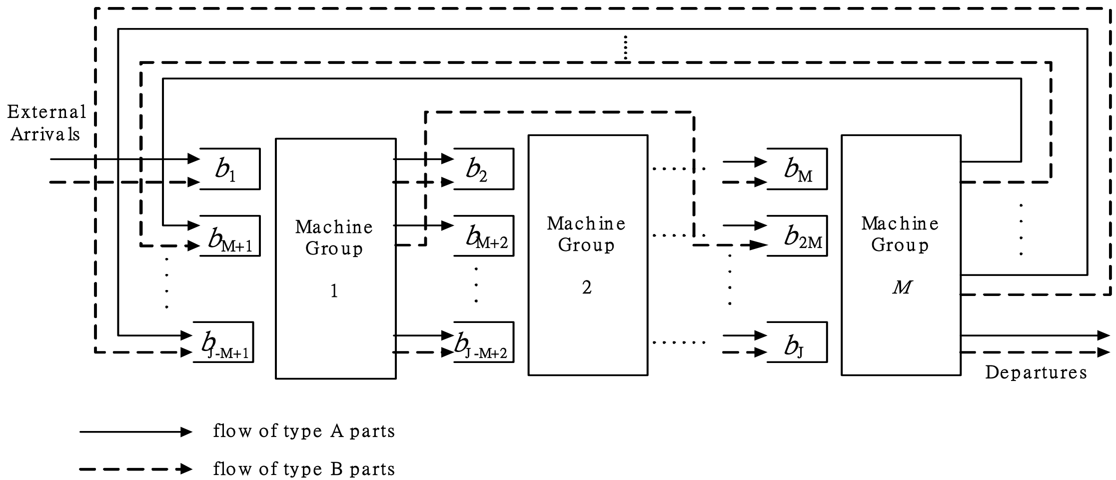


Behavior Modeling Methodology

- Priority open queueing network (OQN)
  - Nodal and system characterization by mean and variance
  - Response surface matching with empirical data
- Simulation for performance prediction/model adaptation



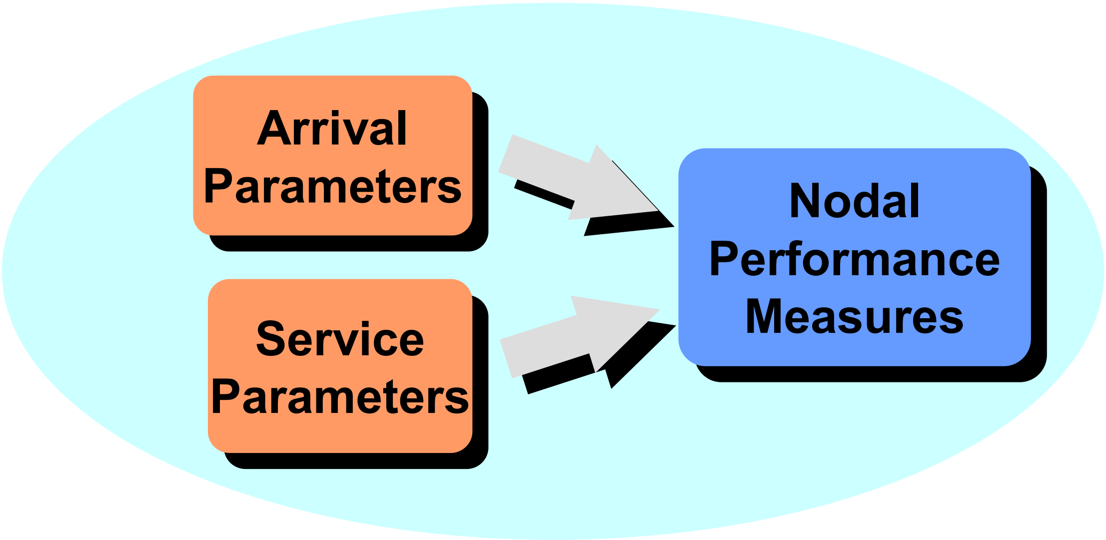
Priority OQN Model: Fab Example



- Node : Group of identical failure prone machines
- Queue : Infinite buffer for each step
- Job Class : Part type
- Arrival : General independent processes
- Service : General time distribution (single/batch, failure)
- Routing : Deterministic with feedback
- Discipline : Priority

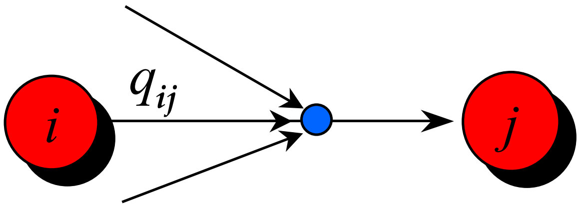
Decomposition Approximation

- Two Notions
  - Each Network Node as an Independent GI/G/*m* Queue
  - Two Parameters, Mean & SCV, to Characterize Arrival & Service Processes



Three Flow Operations

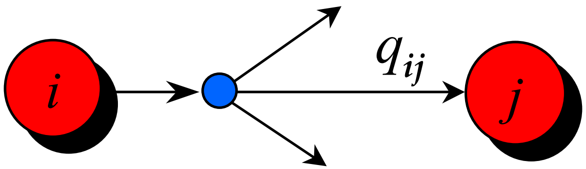
- Merging



$$\lambda_j = \lambda_{0j} + \sum_{i=1}^{n} \lambda_i q_{ij}$$

$$C_{aj}^2 = a_j + \sum_{i=1}^{n} C_{ai}^2 b_{ij}$$

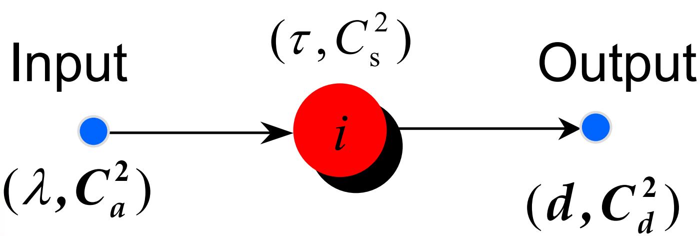
- Splitting



$$\lambda_{ij} = \lambda_i q_{ij}$$

$$C_{ij}^2 = q_{ij} C_{di}^2 + 1 - q_{ij}$$

- Flow Through a Queue



- Departure Rate $d = \lambda$
- Inter-departure Time SCV

$$C_d^2 = \left(1 - \rho^2\right) C_a^2 + \rho^2 C_s^2$$

Traffic Equations: $\underline{F}(\cdot)$

- **Traffic Rate**

$$\lambda_{an} = \lambda_e \delta_n + \sum_{m=1}^{M} \lambda_{mn} = \lambda_e \delta_n + \sum_{m=1}^{M} \lambda_{am} q_{mn}$$

- **Traffic Variability**

$$C_{an}^2 = a_n + \sum_{m=1}^{M} C_{am}^2 b_{mn}$$

λ_{an}

(τ_n, C_{sn}^2)

C_{an}^2

Performance/ QoS Measures

: WIP, Cycle Time, ...

Interaction among Nodes

QoS Translation

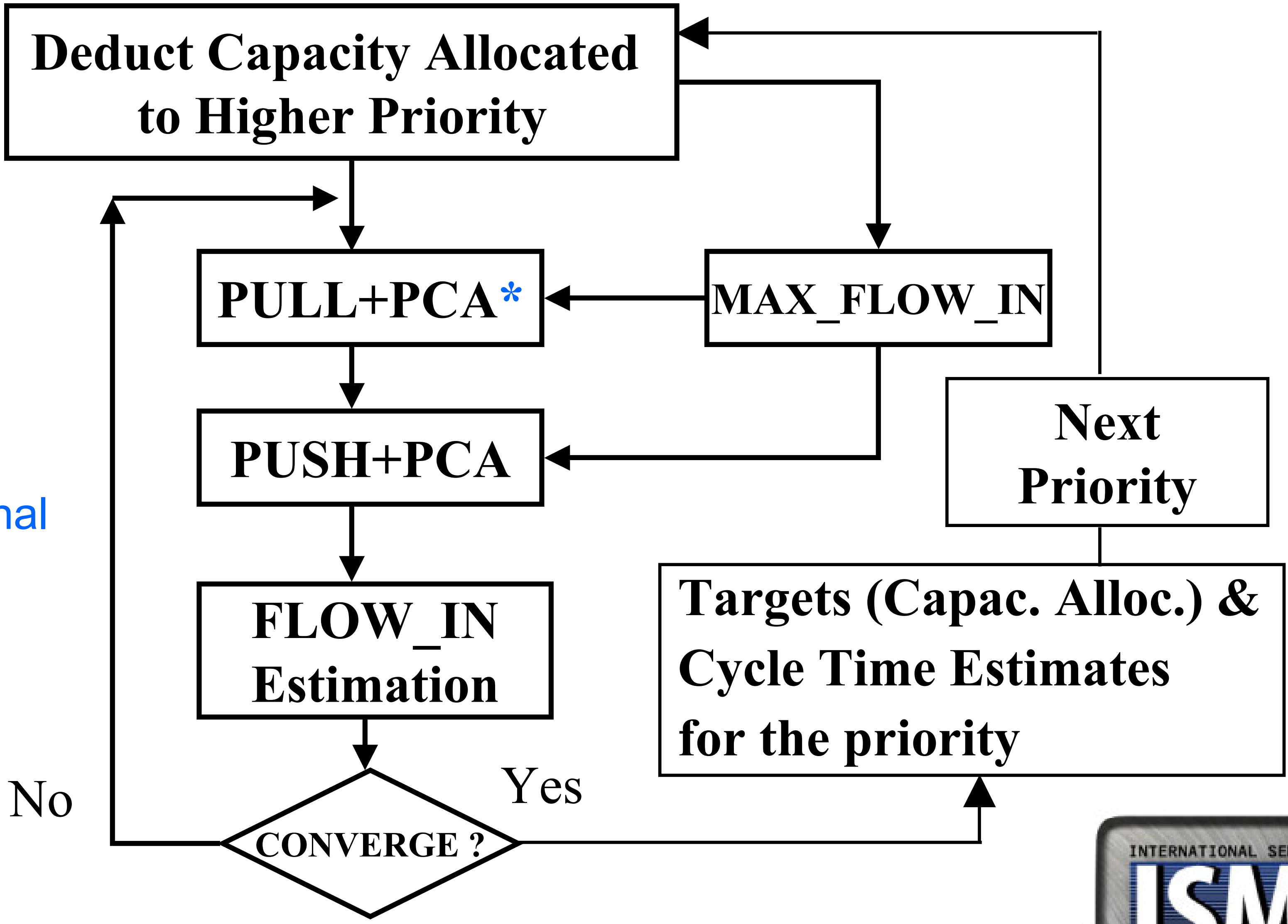
- Given
  - **Higher Level/Coarse QoS spec.** $\underline{\mathrm{Y}}$
  - **Service Node Parameters** $\underline{\theta} = (\tau_n, C_{sn}^2)$
  - **Flow Routing Information** $\underline{\mathbf{Q}}$
  - **Priority OQN model** $F(\underline{\alpha}, \underline{\theta}, \mathbf{Q})$
  - **FCFS Discipline for Each Priority**
- Derive by solving $\underline{\mathrm{Y}} = \underline{F}\left(\underline{\alpha}, \underline{\theta}, \underline{\mathbf{Q}}\right)$
  - **External control specs.** $\underline{\alpha} = (\lambda_e, C_e^2)$
  - **Nodal Level QoS reponsibility** $\underline{\mathbf{y}} = \underline{G}\left(\underline{\omega}, \underline{\theta}\right)$

Response Surface Modeling

- Given
  - **Empirical I/O Characterization** $(\underline{I}, \underline{O})$
  - **Service Node Capacity** m_n
  - **Flow Routing Information** $\underline{\mathbf{Q}}$
  - **Priority OQN model** $F(\underline{\alpha}, \underline{\theta}, \mathbf{Q})$
  - **FCFS Discipline for Each Priority**

- Fit $F(\underline{\alpha}, \underline{\theta}, \mathbf{Q})$ to $(\underline{I}, \underline{O})$ and derive
  - **Node characteristic parameters** $\underline{\theta} = (\tau_n, C_{sn}^2)$

Modeling Capacity Allocation



*P.C.A: Proportional Capacity Allocation

Deliverables – task 1

- July 2005
  - Selection and definition of key QoS metrics
  - Translation algorithm of QoS from chain to nodes
  - Fab behavioral model
    - Priority, capacity allocation, source of variation
  - Fab behavioral simulator
- July 2006
  - Methodology generalization to the service thread from design house, fab to circuit probe
  - Methodology and tool integration with control (task 3) and optimization (task 2)

Task 2: Robust Allocation and Monitoring

PI: Argon Chen Co-PI's: David Chiang, Andy Guo

Will develop:

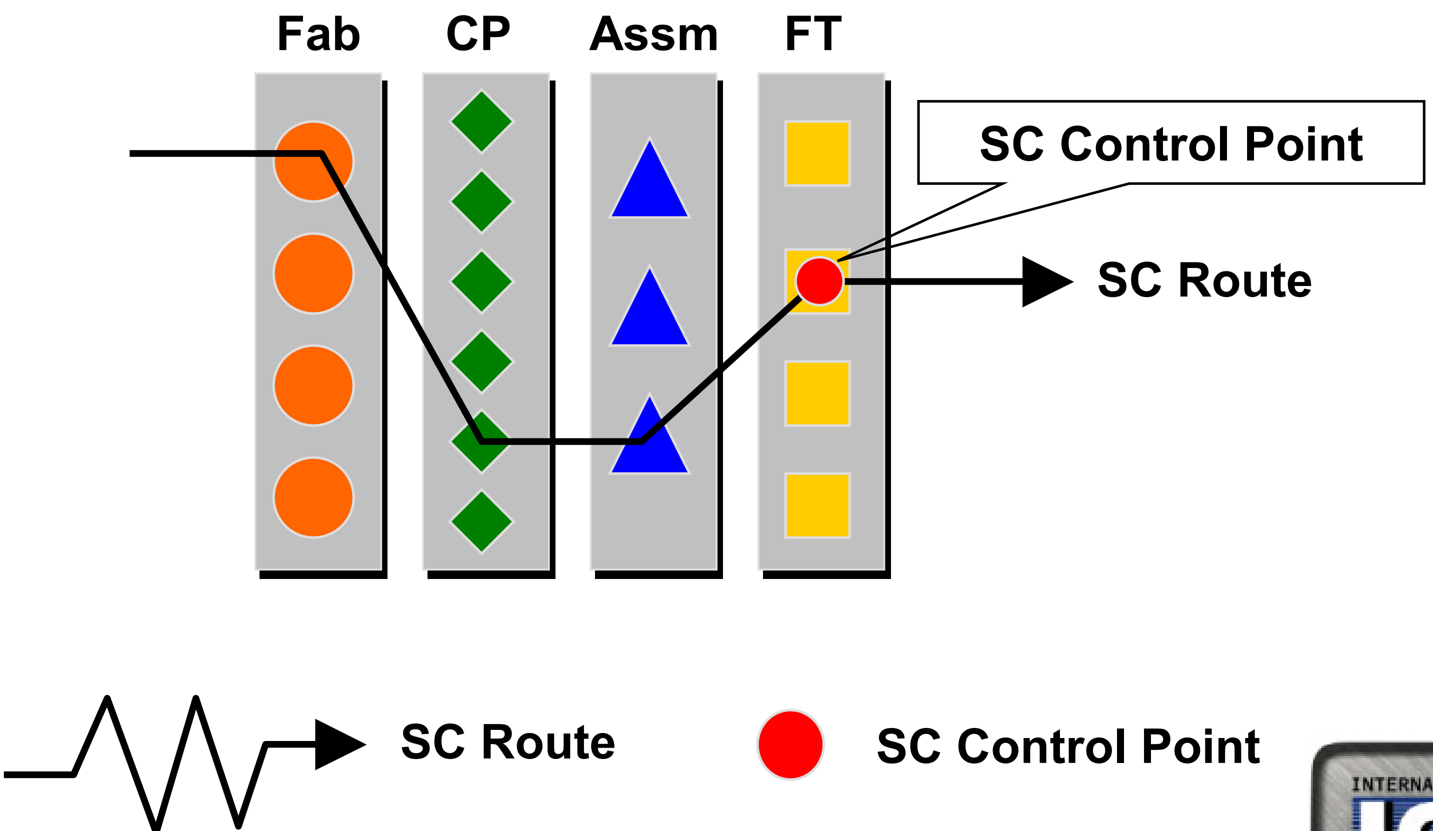
1. A baseline supply chain allocation strategy
   - Robustness on performance
   - Robustness on performance variability
   - Quadratic approximation

2. Supply chain sensitivity and monitoring
   - 2^{nd} moment performance of priority queueing network
   - Decomposition of supply chain performance
   - Ranges of optimality and feasibility
   - Trigger of supply chain control actions

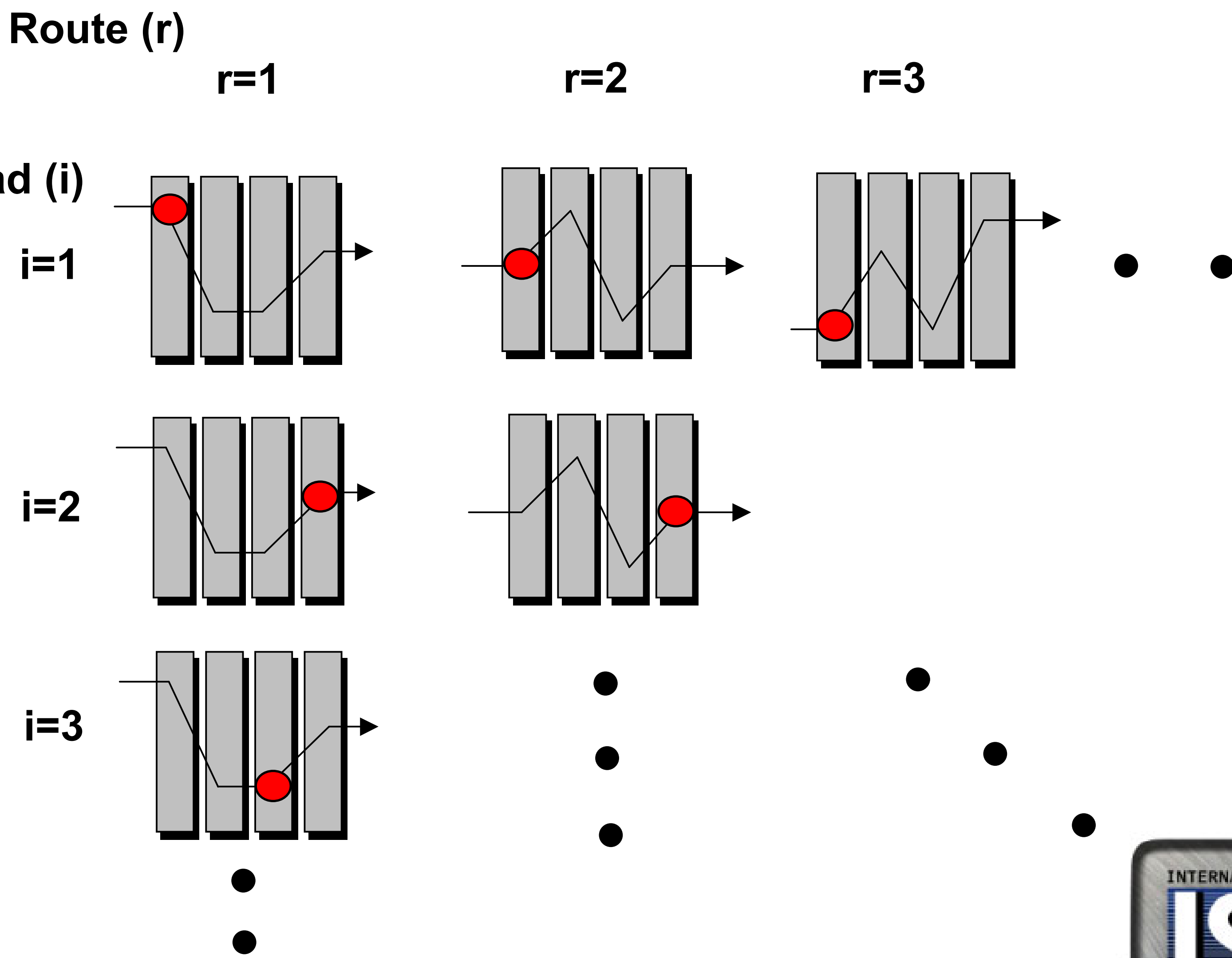
Mid-year Progress – Task 2

- Supply chain simulation model
  - **Have defined environment variables and variability sources**
  - **Have defined control policies for various supply chain threads**
  - **Have built a preliminary simulation model using ARENA**
- Supply chain allocation programming
  - **Have defined allocation decision variables**
  - **Have formulated constraints**
  - **Have started development of implementation strategies**
- Supply chain allocation optimization
  - **Have studied quadratic programming methodologies**
  - **Have studied Wolfe-dual based algorithm**
  - **Have studied piecewise linear programming methodologies**

Semiconductor Supply Chain

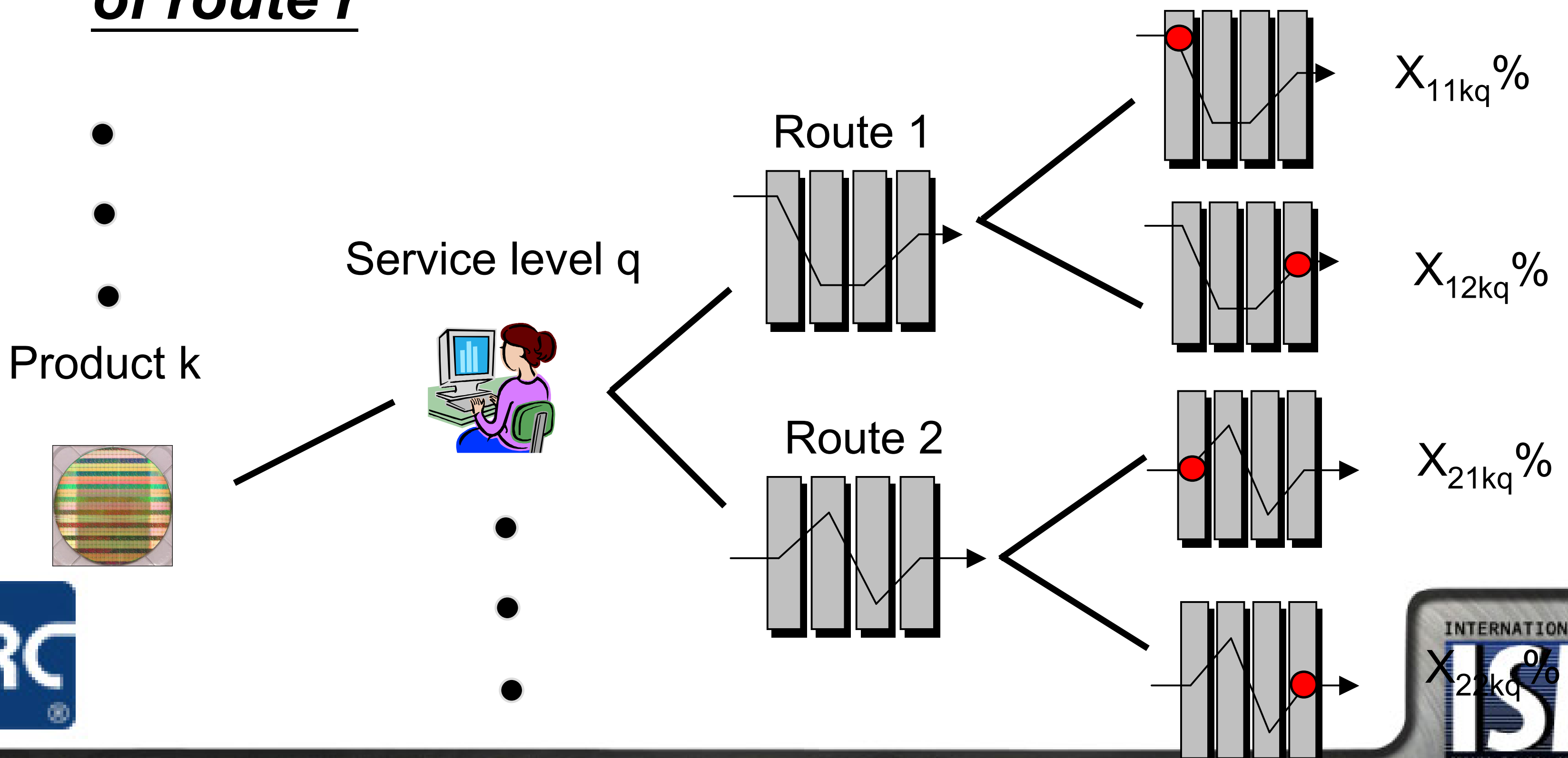


Supply Chain Routes and Threads

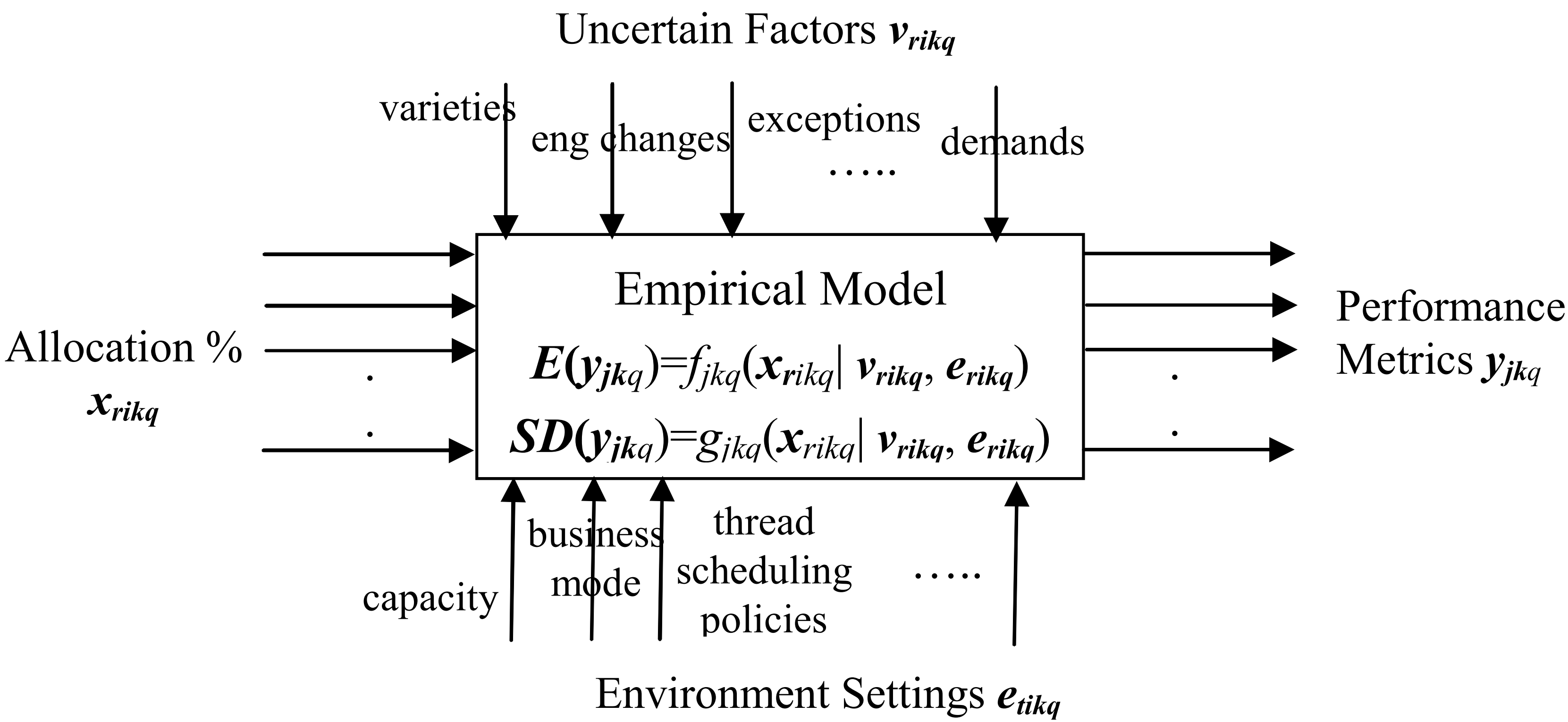


Supply Chain Allocation

- X_{rikq} (%)
  - ***Proportion of production for product type k at service-level q allocated to supply chain thread i of route r***



Supply Chain Behavior Model



y_{jkq} : **the *j*th performance index for product *k* at service level *q***

Supply Chain Constraints (I)

Product Mix Constraints

- ***The proportion of product type k to total production***

$$\sum_r \sum_i \sum_q X_{rikq} = \rho_k \quad \forall k$$

Priority Mix Constraints

- ***The proportion of service-level q production to total production***

$$\sum_r \sum_i \sum_k X_{rikq} \leq \phi_q \quad \forall q$$

Supply Chain Constraints (II)

Demands Fulfillment Constraint

– ***The total production is equal to or less than the demand***

$$\sum_{r}\sum_{i}\sum_{k}\sum_{q} X_{rikq} \leq 1$$

Example:

$$\sum_{r}\sum_{i}\sum_{k}\sum_{q} X_{rikq} = 0.95$$

meaning 95% of demand will be fulfilled

SRC

INTERNATIONAL SEMAT

ISM

Supply Chain Constraints (III)

Route Mix Constraints

- ***The proportion of production allocated to route r can not exceed a predetermined limit***

$$\sum_{i}\sum_{k}\sum_{q} X_{rikq} \leq \alpha_r \qquad \forall r$$

Thread Mix Constraints

- ***The proportion of production allocated to thread i can not exceed a predetermined limit***

$$\sum_{r}\sum_{k}\sum_{q} X_{rikq} \leq \beta_i \qquad \forall i$$

Supply Chain Constraints (IV)

Resource (Capacity) Constraints

– *The proportion of capacity consumed by route r cannot exceed a given proportion of route r capacity to the total capacity*

$$\sum_{k}\left[\left(\sum_{i}\sum_{q} X_{rikq}\right) * m_{rk}\right] \le C_r \quad \forall r \qquad \sum_{r} C_r \le 1$$

Where

– $m_{rk}=U\omega_{ki}$: the percent use of route *r* by one percent of production for product type *k* allocated to route *r*

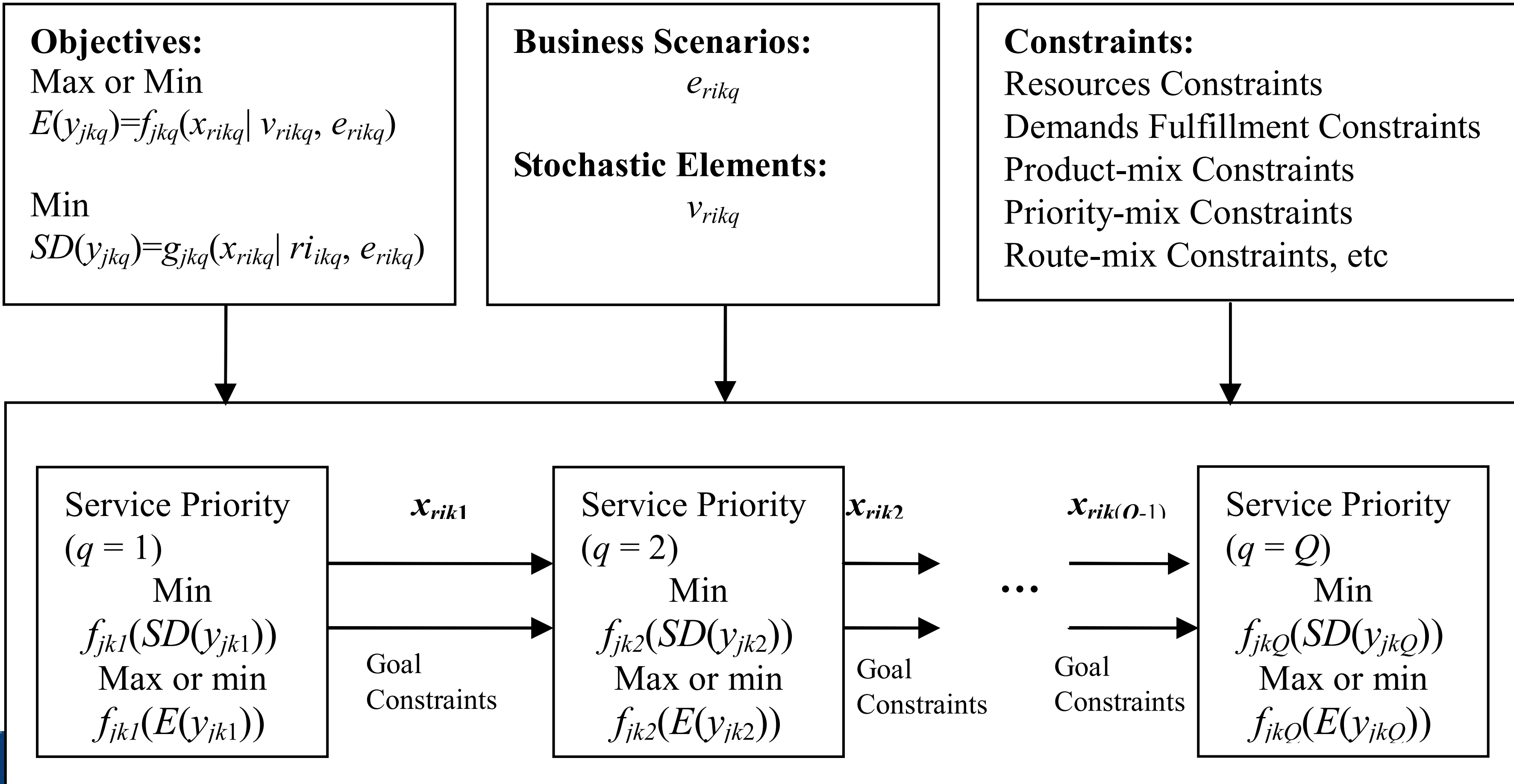
– C_r: the proportion of route r available capacity to total capacity

– $\sum C_r$: the proportion of available capacity to total capacity

– *U*(%): capacity utilization (production to total capacity ratio)

– ω_{ki}: the capacity of route *r* consumed by one unit of product type *k*

Supply Chain Allocation Optimization – Goal Programming

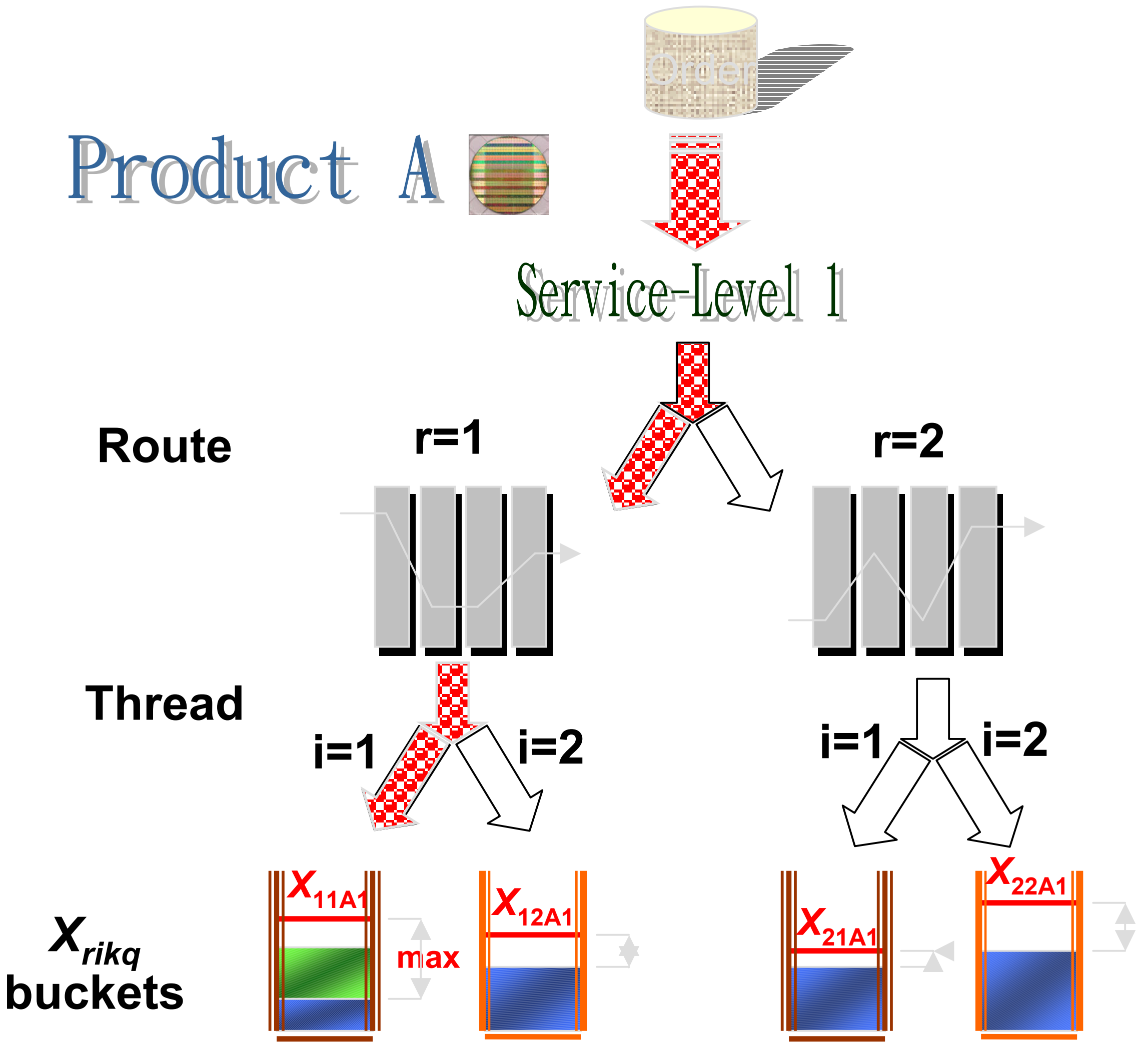


Solution Methodology

- Quadratic Stochastic Goal-Programming
  - Transform the model to a piecewise quadratic programming model
  - Construct Wolfe-dual based algorithm for the piecewise quadratic programming model
  - Develop a preemptive goal programming approach for differentiable service priorities
  - Perform sensitivity analysis through parametric quadratic programming

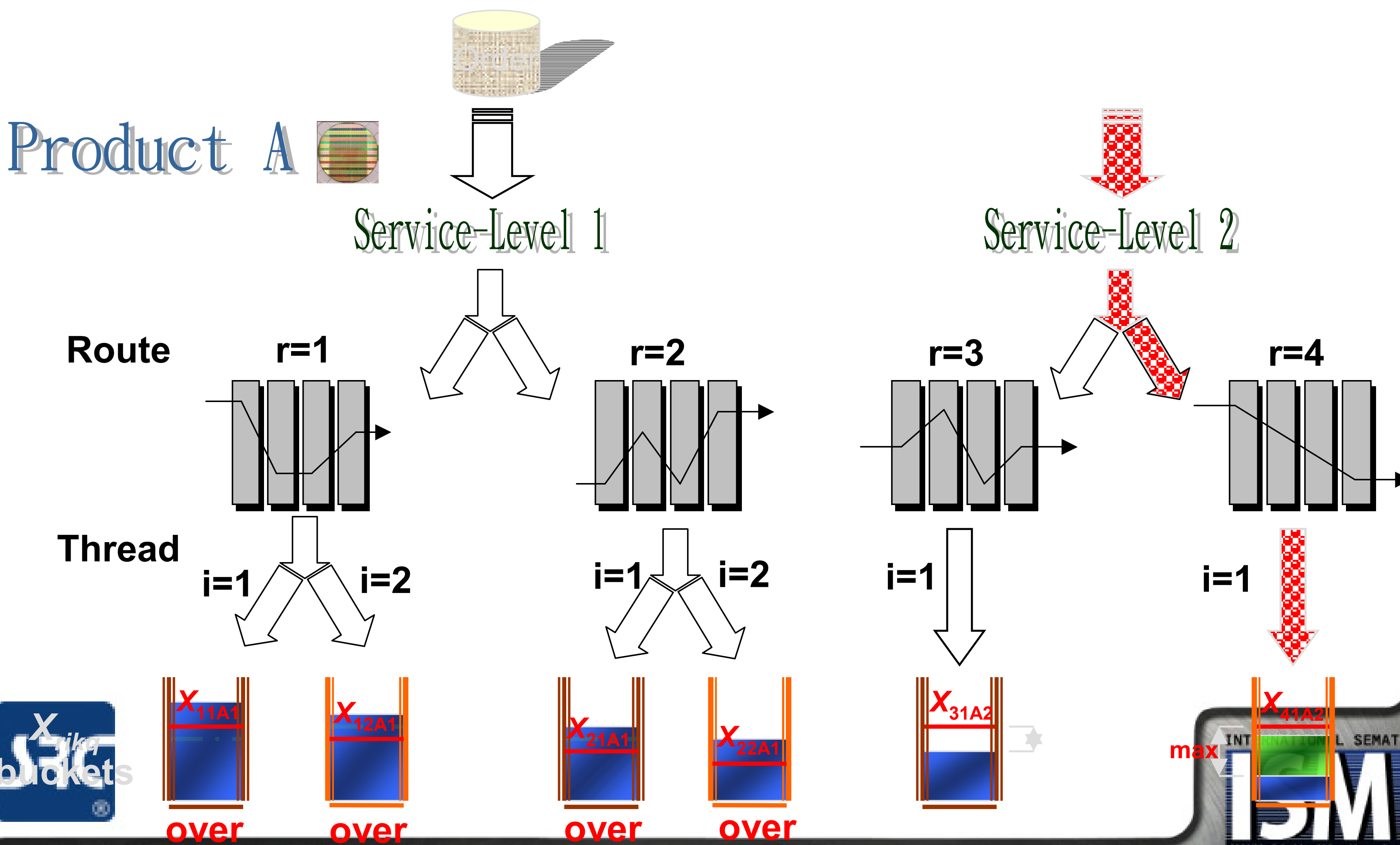
Implementation

Case 1: Order fulfilled based on X_{rikq}



Implementation

Case 2: Order fulfilled by a lower priority



Deliverables – Task 2

- Supply chain quadratic goal programming model and solution (Model, Methodology, Report) (July-05)
  - **Supply chain simulation model (March-05)**
  - **Supply chain planning goals (April-05)**
  - **Supply chain goal programs (May-05)**
  - **Baseline supply chain allocation model and solution (July-05)**
- Supply chain sensitivity analysis and monitoring methodology (Model, Methodology, Report) (July-06)

Supports Needed – Task II

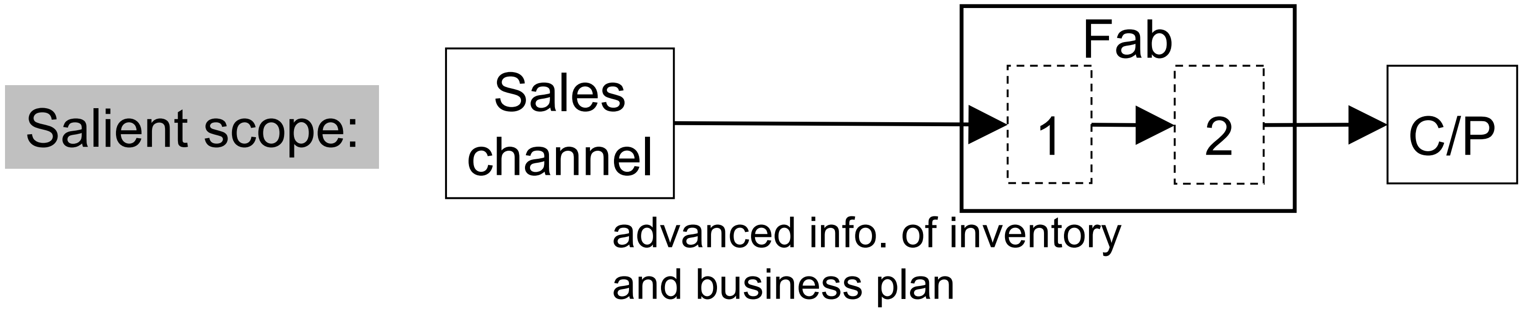
- Supply chain network data
  - Number of supply chain levels
  - Number of facilities at each level
  - Capacity and capability of each facility
  - Locations of facilities, etc.
- Supply chain operations data
  - Facility reliability data
  - Cycle time
  - Dispatching policies
  - Control policies
  - Order fulfillment policies, etc.
- Supply chain allocation practice
- Supply chain performance data

Task 3: Dynamic Control

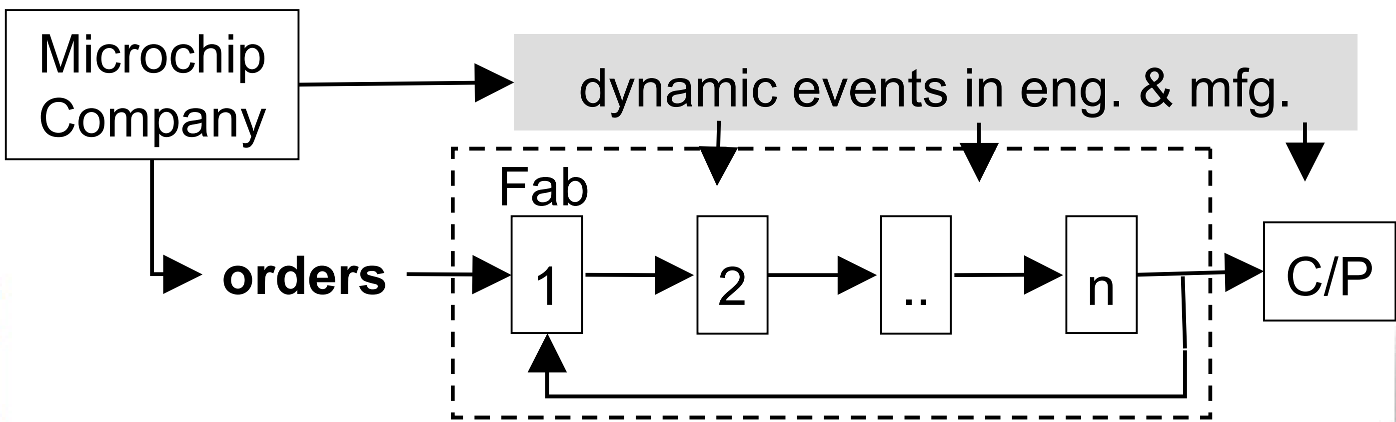
Will develop:

PI: Yon Chou
Co-PI: Shi-Chung Chang

1. A control model for demand support



2. A control method to enhance delivery, speed and service

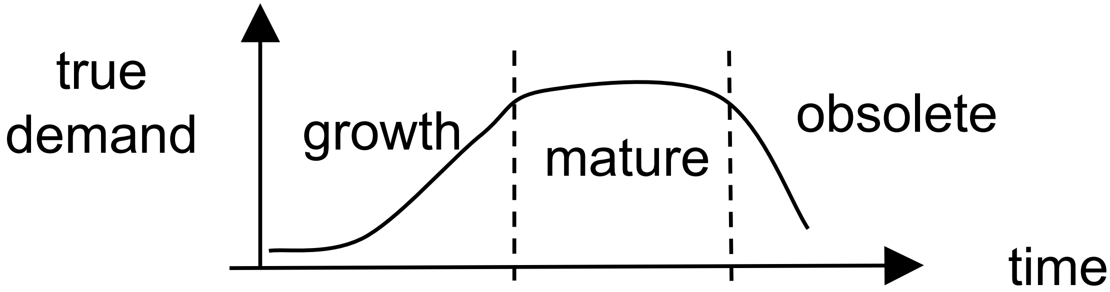


Mid-year Progress – Task 3

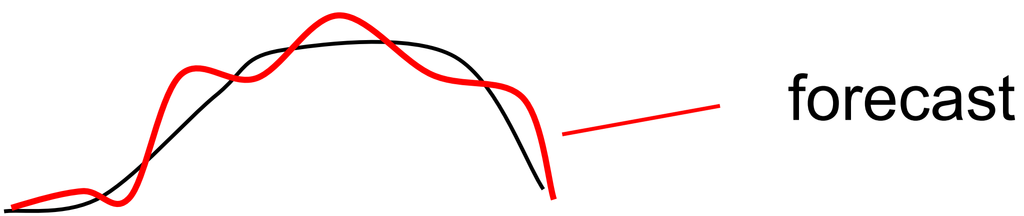
1. A control model for demand support
   - Have defined the problem scope
   - Have outlined the model
2. A control method to enhance delivery, speed and service
   - Have developed a workload flow model
   - Have developed an integer program (with preliminary implementation)

3.1 Salient feature

- Demand (technology, product, etc.) has a life cycle



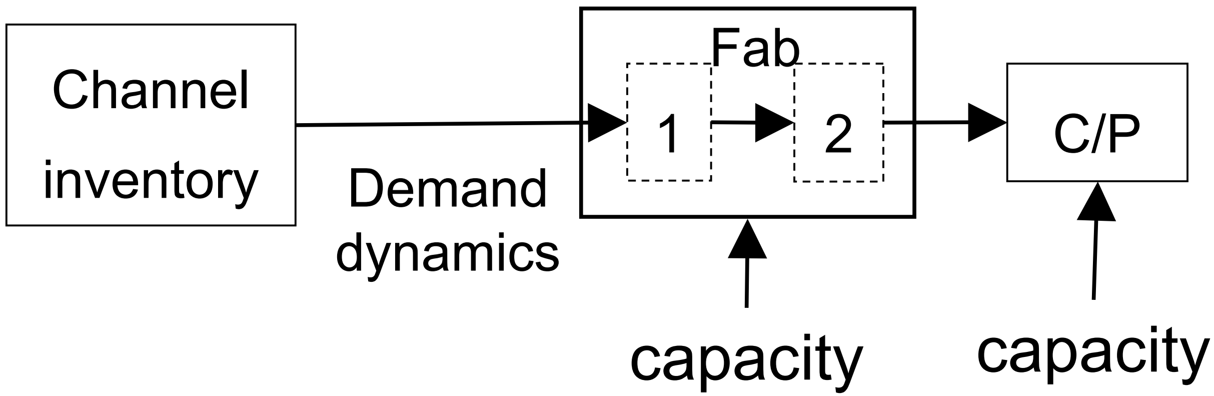
- Demand forecasts and channel inventory are signals. The total demand is a more reliable estimate.



Mean-reverting model

3.1 A control model for demand support

- Objectives
  - To monitor demand-capacity mismatch in medium-long term
  - To support the demand-capacity synchronization by capacity decisions (expansion, reservation, prioritization)
- Model scope
  - Relationship between capacity, cycle time, WIP and throughput
  - Integrating channel inventory and demand dynamics with supply capability



3.1 Elements of the model

- Channel inventory: an input, based on market intelligence data

  $I(t)$

- Demand dynamics
  - Demand lifecycle

    $X(\mathrm{t})$: cumulative demand by time t

  - Demand learning effect

    $\dot{X}(t) = k(X(t))$

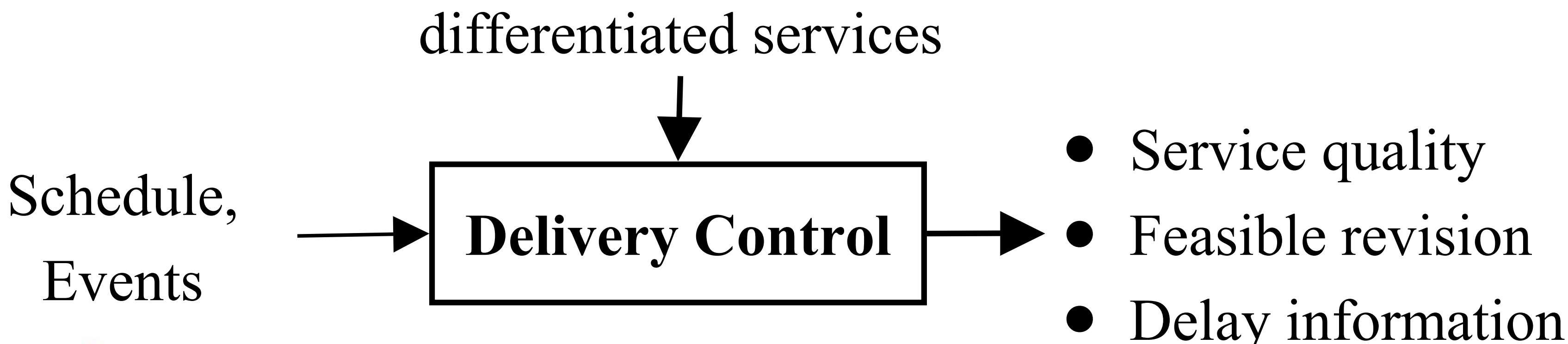
- Supply capability of the nodes
  - Cycle time, WIP, throughput

    $WIP = (TH)(CT)$

- Objective functional
  - Capacity allocation (to control shortage points)

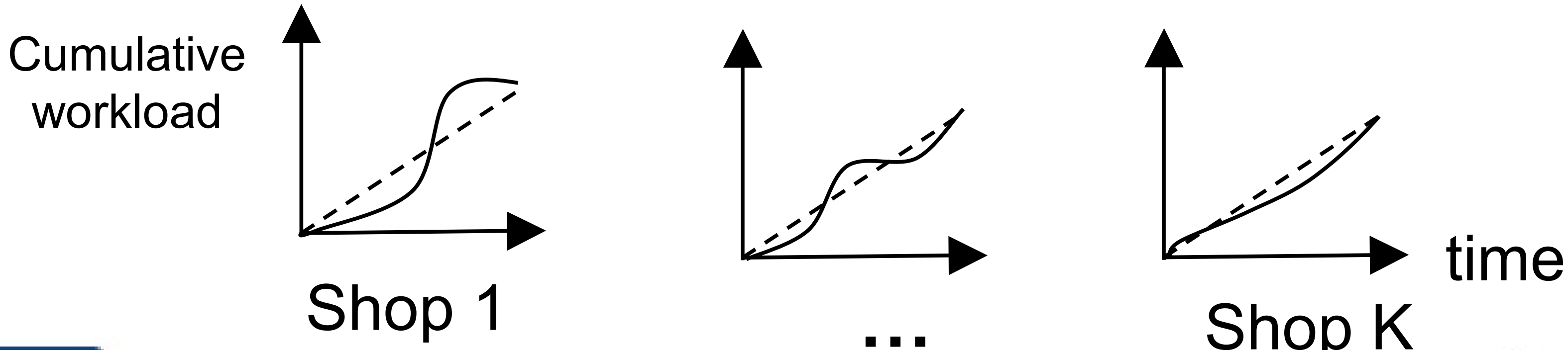
3.2 Delivery control

- Objectives:
  - To assess the impact of dynamic events on the performance under high-mix environment
  - To identify feasible revision, shortfall points and delay information in order to enhance delivery, speed and service



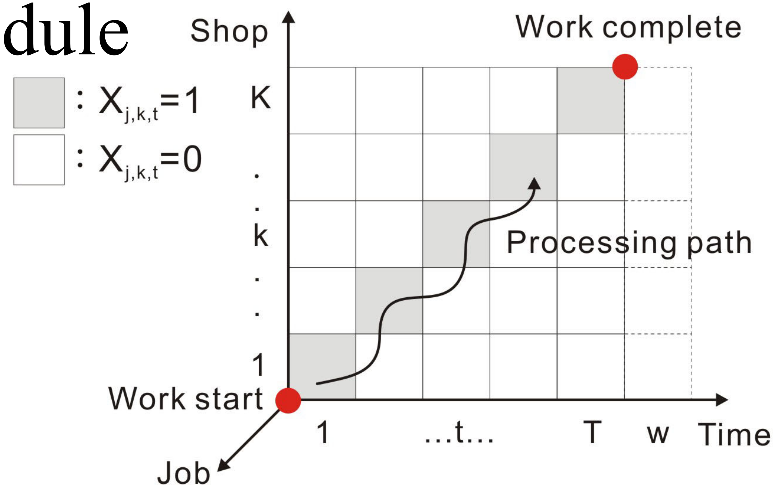
3.2 Workload variation propagation

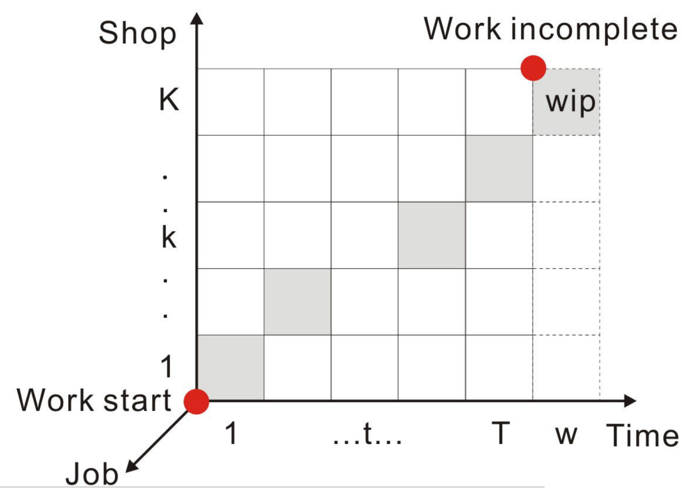
- Elements
  - Events: uncertain job arrivals, urgent orders, disrupting events, and material availability
  - Modeling of capacity loss due to variety, variation and dynamic events
  - Cumulative workload



3.2 Behavior modeling of re-allocation

- Entities of allocation schedule
  - T time periods (weeks)
  - K shops (nodes)
  - J orders
- Variables $X_{j,k,t} \in \{0,1\}$
- Dynamic events
  - Hold
  - Hold-release
  - Order insertion

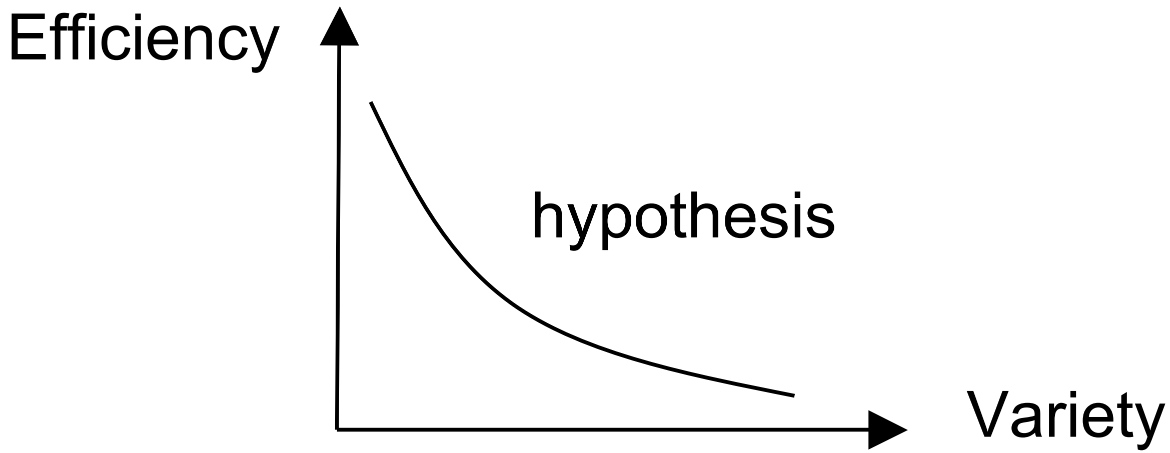




Is coding an integer program for studying the behavior

3.2 Variety-efficient relationship

- There are many parallel machine systems in semiconductor manufacturing.
- How to measure variety?
- How to characterize the relationship between variety and efficiency?



Deliverables – task 3

- A control model for demand support (Model, Report) (July-05)
- A delivery control method (Methodology, Report) (July-06)