Tales of Time Scales

Ward Whitt

AT&T Labs – Research

Florham Park, NJ

New Book

Stochastic-Process Limits

An Introduction to Stochastic-Process Limits and Their Application to Queues

Springer 2001

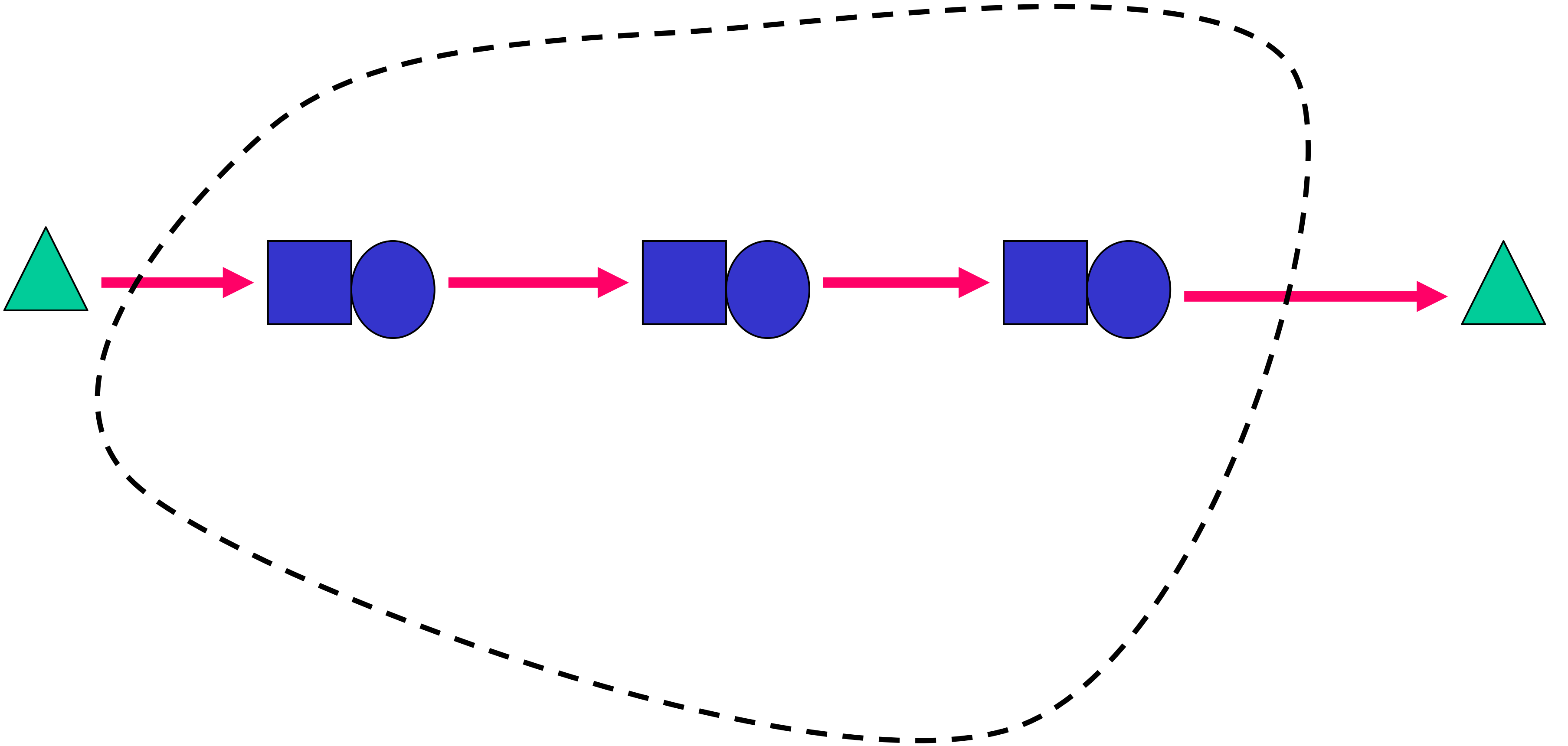
“I won’t waste a minute to read that book.”

- Felix Pollaczek

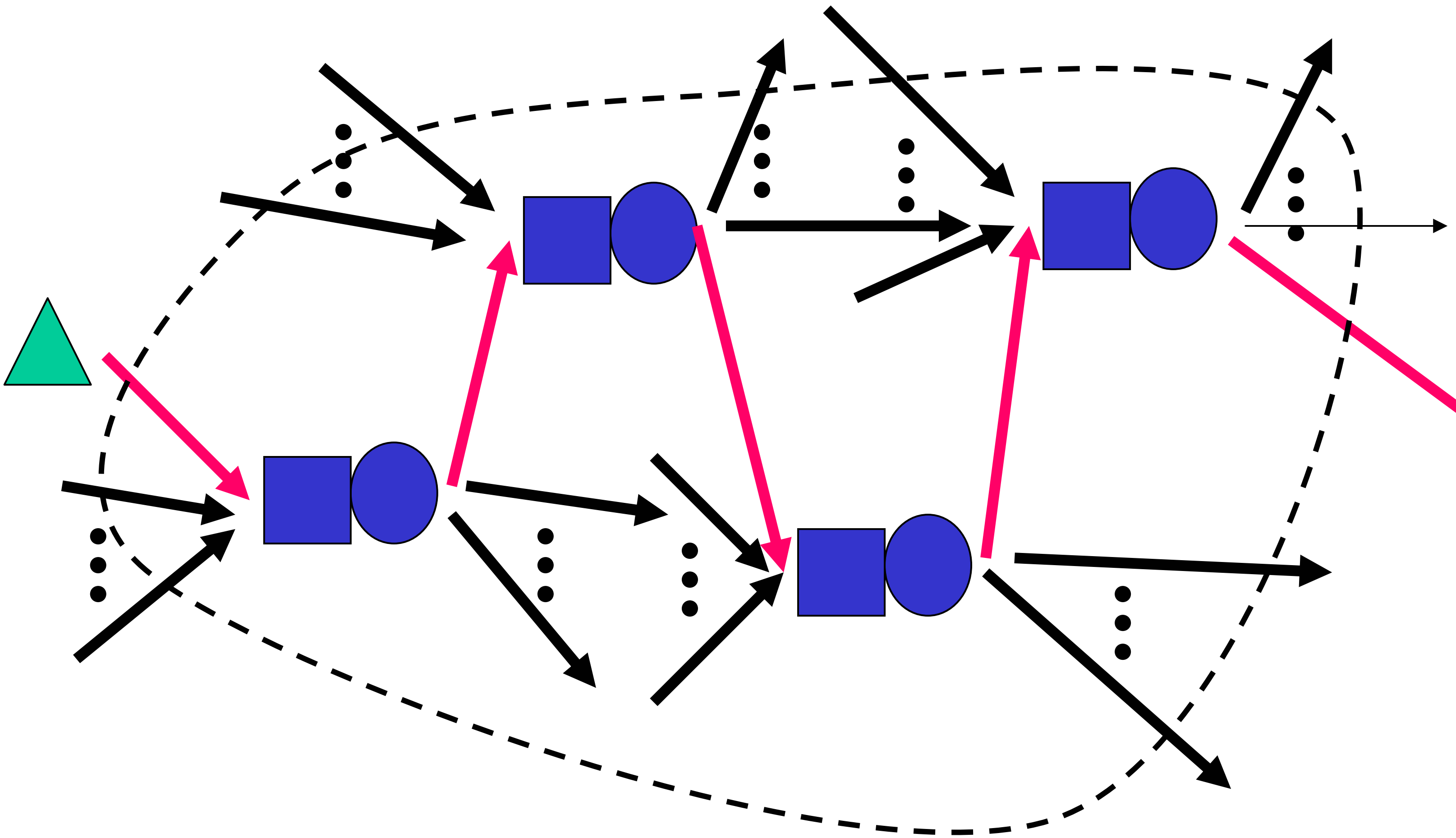
“Ideal for anyone working on their third Ph.D. in queueing theory”

- Agner Krarup Erlang

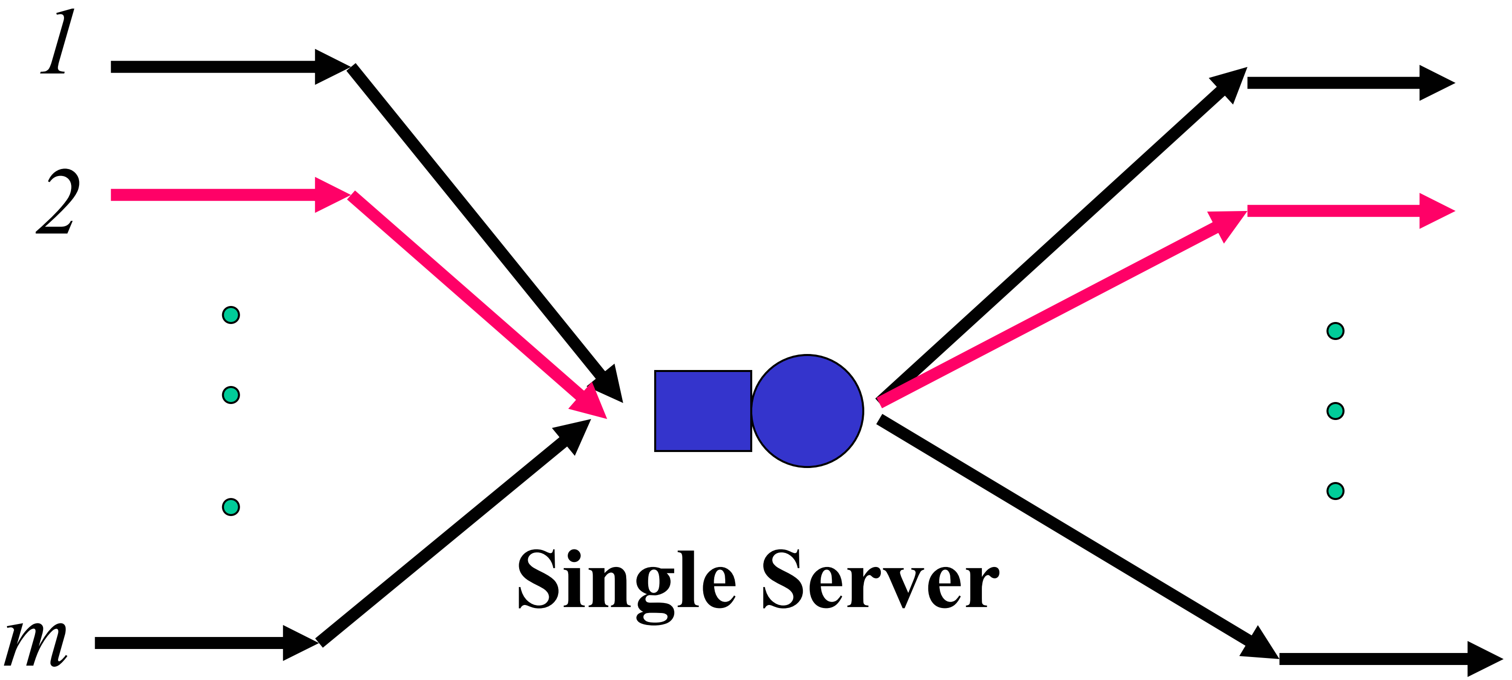
Flow Modification



Flow Modification



Multiple Classes



$m \longrightarrow \infty$

- **Whitt, W. (1988)** A light-traffic approximation for single-class departure processes from multi-class queues. ***Management Science*** **34**, 1333-1346.
- **Wischik, D. (1999)** The output of a switch, or, effective bandwidths for networks. ***Queueing Systems*** **32**, 383-396.

M/G/1 Queue in a Random Environment

Let the arrival rate be a stochastic process with two states.

MMPP/G/1 **is a special case.**

	environment state	
	1	2
mean holding time (in environment state)	**1**	**5**
arrival rate	**0.92**	**0.50**
mean service time	**1**	**1**

Overall Traffic Intensity ρ = **0.57**

	environment state 1	environment state 2
mean holding time (in environment state)	**1**	**5**
arrival rate	**2.00**	**0.50**
mean service time	**1**	**1**

Overall Traffic Intensity ρ = **0.75**

A Slowly Changing Environment

	environment state 1	environment state 2
mean holding time (in environment state)	**1*M***	**5*M***
arrival rate	**0.92**	**0.50**
mean service time	**1**	**1**

Overall Traffic Intensity ρ = **0.57**

What Matters?

- **Environment Process?**

- **Arrival and Service Processes?**

- ***M*?**

Nearly Completely Decomposable Markov Chains

P. J. Courtois,
***Decomposability*, 1977**

What if the queue is unstable in one of the environment states?

	environment state	
	1	2
mean holding time (in environment state)	**1*M***	**5*M***
arrival rate	**2.00**	**0.50**
mean service time	**1**	**1**

Overall Traffic Intensity ρ = **0.75**

What Matters?

- **Environment Process?**
- **Arrival and Service Processes?**
- ***M*?**

Change Time Units

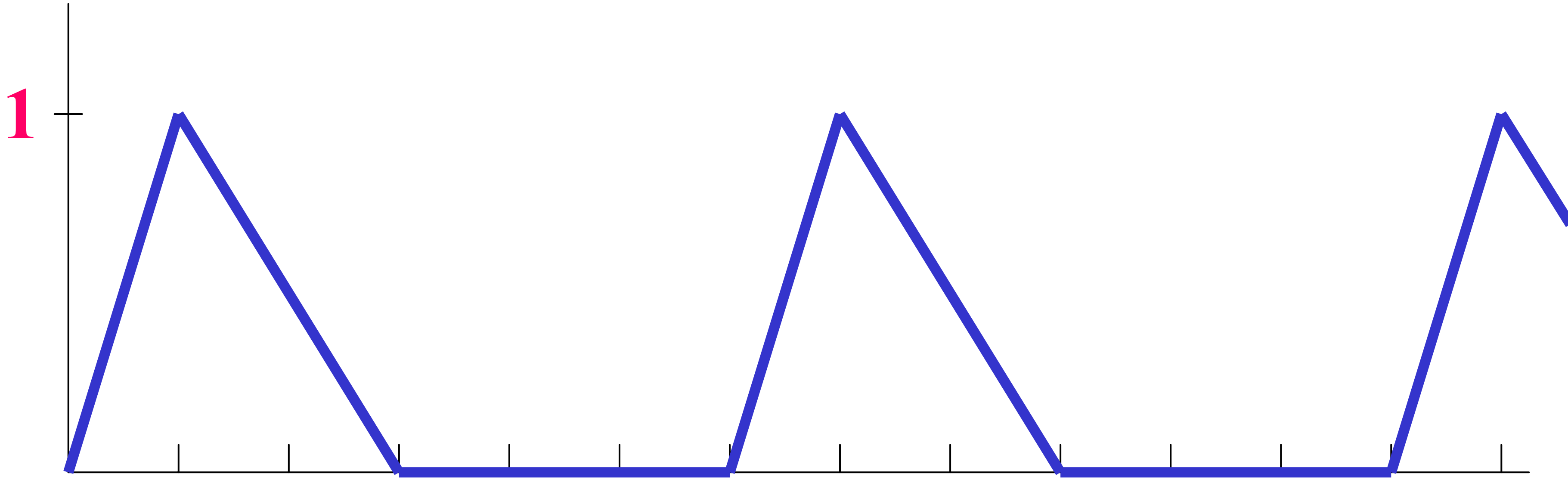
Measure time in units of ***M***

i.e., divide time by ***M***

	environment state	
	1	2
mean holding time (in environment state)	**1**	**5**
arrival rate	**2.00*M***	**0.50*M***
mean service time	**1/*M***	**1/*M***

Overall Traffic Intensity ρ = **0.75**

Workload in Remaining Service Time With Deterministic Holding Times



Steady-state workload tail probabilities in the MMPP/G/1 queue

size factor M	service-time distribution	tail probability $P(W > x)$			
		$x = 0.5$	$x = 2.5$	$x = 4.5$	$x = 6.5$
10	D	0.40260	0.13739	0.04695	0.01604
	E_4	0.41100	0.14350	0.05040	0.01770
	M	0.44246	0.16168	0.06119	0.02316
	$H_2^b, c^2 = 4$	0.52216	0.23002	0.01087	0.05168
10^2	D	0.37376	0.11418	0.03488	0.01066
	E_4	0.37383	0.11425	0.03492	0.01067
	M	0.37670	0.11669	0.03614	0.01120
	$H_2^b, c^2 = 4$	0.38466	0.12398	0.03997	0.01289
10^3	D	0.37075	0.11183	0.03373	0.01017
	E_4	0.37082	0.11189	0.03376	0.01019
	M	0.37105	0.11208	0.03385	0.01023
	$H_2^b, c^2 = 4$	0.37186	0.11281	0.03422	0.01038
10^4	D	0.37044	0.11159	0.03362	0.01013
	E_4	0.37045	0.11160	0.03362	0.01013
	M	0.37047	0.11164	0.03363	0.01013
	$H_2^b, c^2 = 4$	0.37055	0.11169	0.03366	0.01015

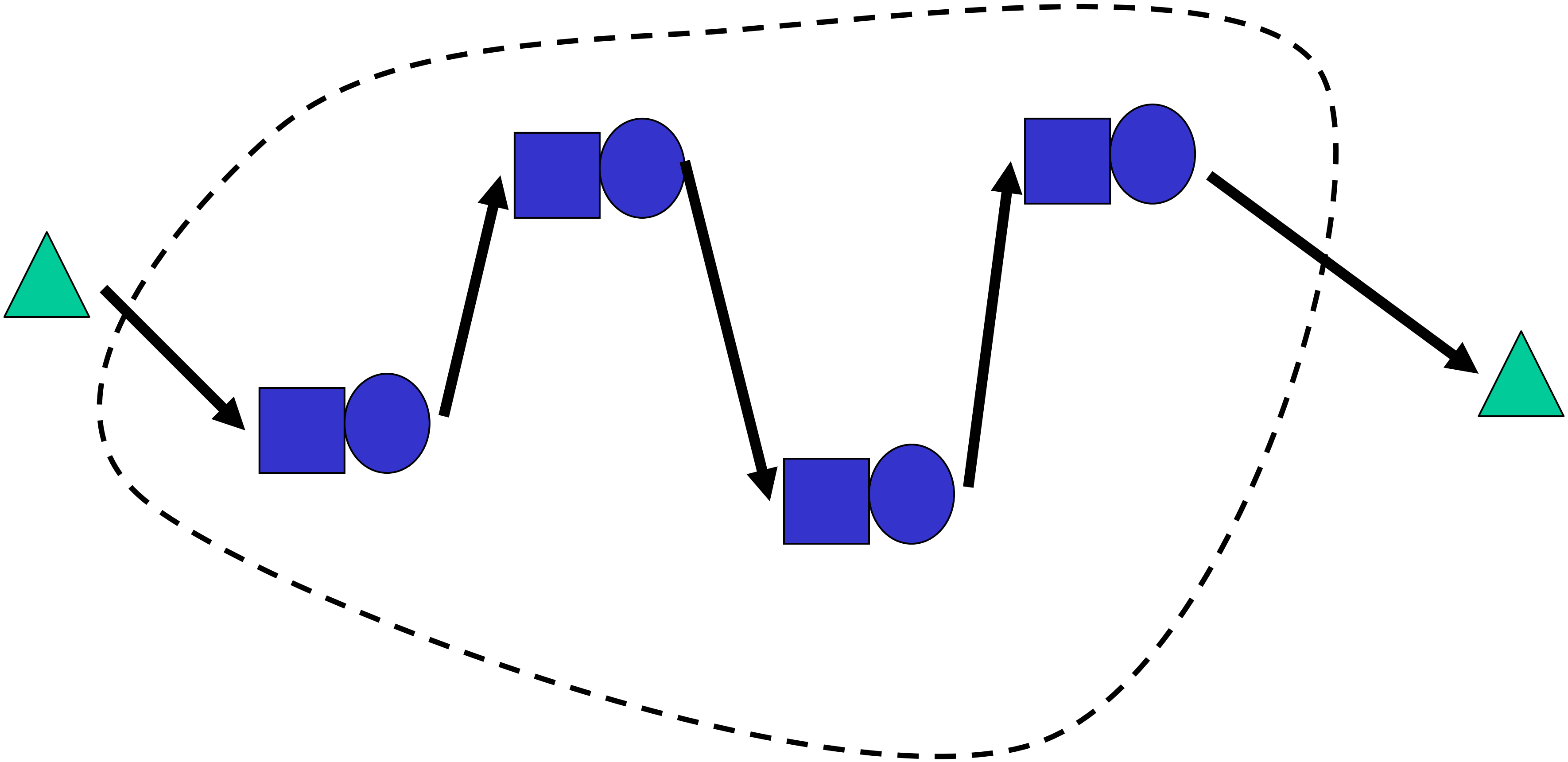
G. L. Choudhury, A. Mandelbaum, M. I. Reiman and W. Whitt, "Fluid and diffusion limits for queues in slowly changing environments." ***Stochastic Models*** **13 (1997) 121-146.**

Thesis:

Heavy-traffic limits for queues can help expose phenomena occurring at different time scales.

Asymptotically, there may be a separation of time scales.

Network Status Probe



Heavy-Traffic Perspective

$$n^{-H}\, W_n(nt) \Rightarrow W(t)$$

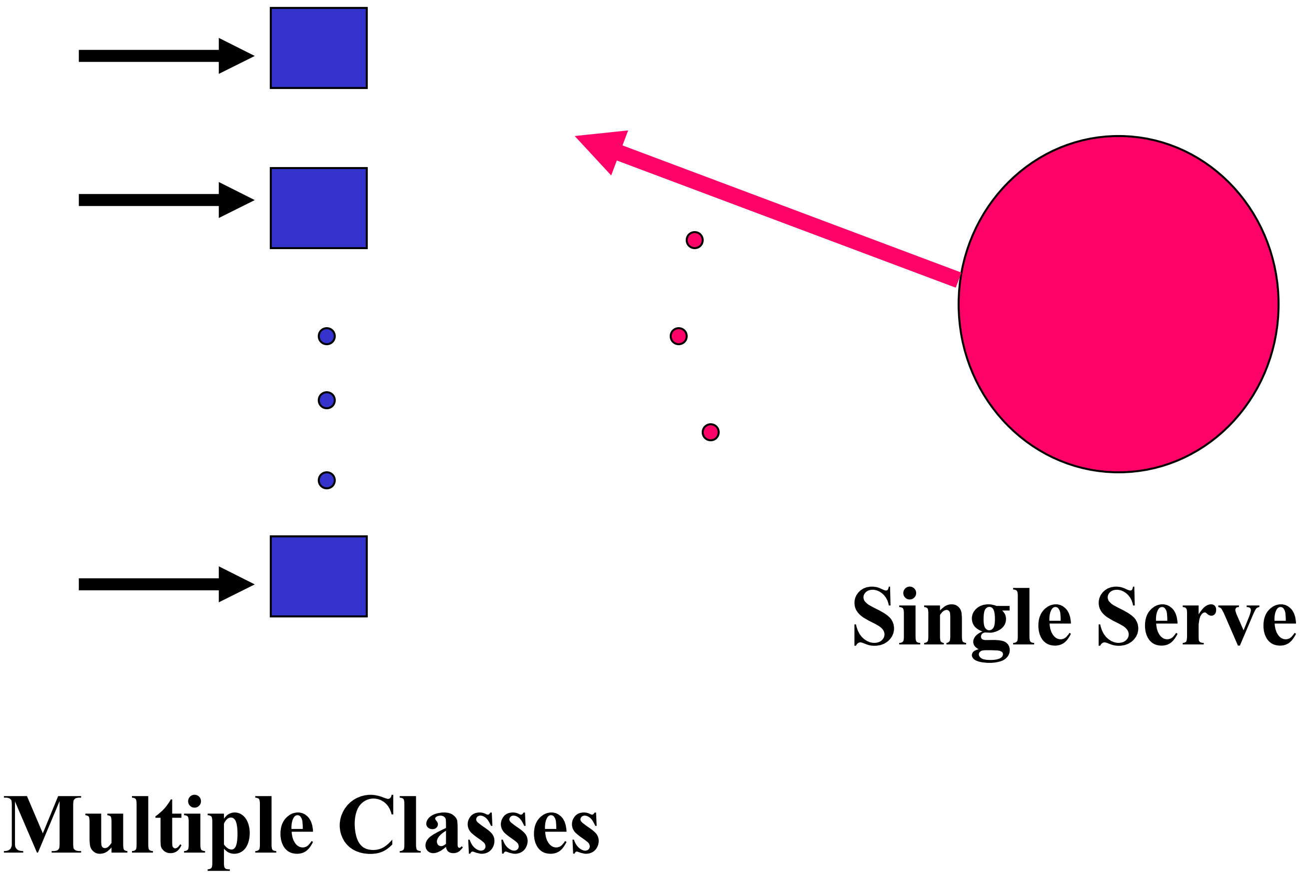
$$0 < H < 1$$

$$n = (1 - \rho)^{-1/(1-H)}$$

Snapshot Principle

Server Scheduling

With Delay and Switching Costs



Heavy-Traffic Limit for Workload

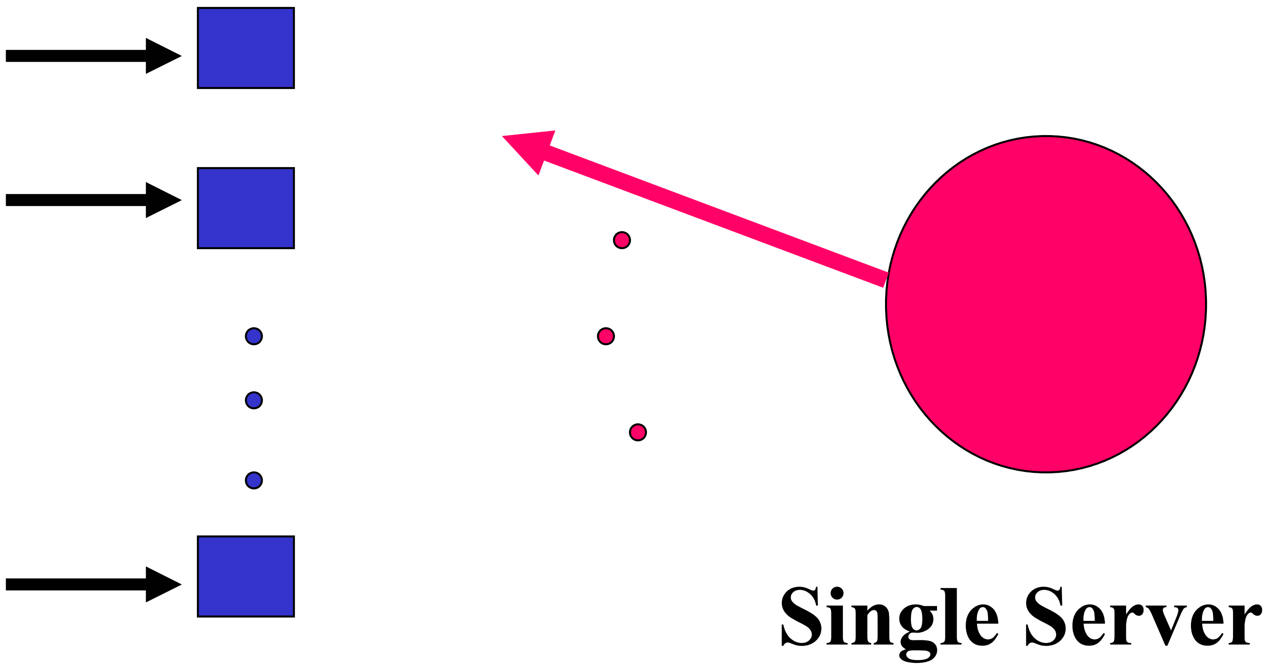
$$\boldsymbol{W_n} \Rightarrow \boldsymbol{W}$$

$$\boldsymbol{W_n}(t) = n^{-H}\, W_n(nt)$$

$$0 < H < 1$$

$$n = (1 - \rho)^{-1/(1-H)}$$

One Approach: Polling



Multiple Classes

Heavy-Traffic Averaging Principle

$$h^{-1} \int_{s}^{s+h} f(\boldsymbol{W}_{\boldsymbol{i,n}}(t))\, dt \Longrightarrow h^{-1} \int_{s}^{s+h} \left(\int_{0}^{1} f(a_i u \boldsymbol{W}(t))\, du \right) dt$$

$$\boldsymbol{W}_{\boldsymbol{i,n}}(t) = n^{-H}\, W_{i,n}(nt)$$

- **Coffman, E. G., Jr., Puhalskii, A. A. and Reiman, M. I. (1995)** Polling systems with zero switchover times: a heavy-traffic averaging principle. ***Ann. Appl. Prob.*** **5**, 681-719.
- **Markowitz, D. M. and Wein, L. M. (2001)** Heavy-traffic analysis of dynamic cyclic policies: a unified treatment of the single machine scheduling problem. ***Operations Res.*** **49**, 246-270.
- **Kushner, H. J. (2001)** ***Heavy Traffic Analysis of Controlled Queueing and Communication Networks,*** Springer, New York.

"My thesis has been that one path to the construction of a nontrivial theory of complex systems is by way of a theory of ***hierarchy****."*

- H. A. Simon

- Holt, Modigliani, Muth and Simon, **Planning Production, Inventories and Workforce**, 1960.
- Simon and Ando, *Aggregation of variables in dynamic systems*. **Econometrica**, 1961.
- Ando, Fisher and Simon, **Essays on the Structure of Social Science Models**, 1963.

Application to Manufacturing

MRP

mrp

material requirements planning

Management information system

Expand bill of materials

Orlicky (1975)

MRP

Manufacturing Resources Planning

Long-Range Planning
(Strategic)

Intermediate-Range Planning
(Tactical)

Short-Term Control
(Operational)

Hierarchical Decision Making in Stochastic Manufacturing Systems

- Suresh P. Sethi and Qing Zhang, 1994

$$\inf_{u \in A} E \int_0^\infty e^{-\rho t}[h(x(t)) + c(u(t))]dt$$

$$\dot{x}(t) = u(t) - z(t),$$

$$0 \le u(t) \le K(t)$$