

CS5239 Computer System Performance Evaluation

2016/17 – Semester I

www.comp.nus.edu.sg/~teoym/cs5239-16



Teo Yong Meng

Room: COM2, #04-39

Department of Computer Science

National University of Singapore

E-mail: teoym@comp.nus.edu.sg

Tel: 6516 2830

What I do

Teach

- Parallel Computing
- Cloud Computing
- Systems Modeling & Simulation
- Distributed Systems
- Applied Parallel Computing (joint teaching with MIT)
- Computer Systems Engineering (joint teaching with MIT)
-

Research - parallel & distributed computing

- Performance evaluation
- Heterogeneous, Cloud Computing
- Modeling and Simulation

Faster is better

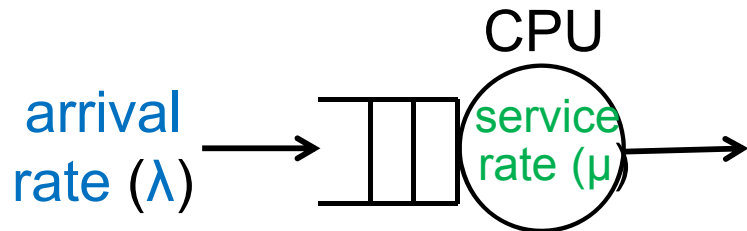
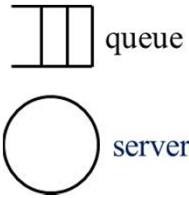
Outline

- What, why and how
- Learning objective
- What we cover
- Module Assessment
- Course Schedule & Webpage
- Resources

What is Performance

- Concept of work
 - Latency (time)
 - Bandwidth (rate)
- How well a computer system performs a given job or activity?
- Why do we care about computer performance?
- What is hard?
 - Performance of a computer system is multidimensional
 - Complex component interactions
 - Hard to predict how it will scale
 - ...

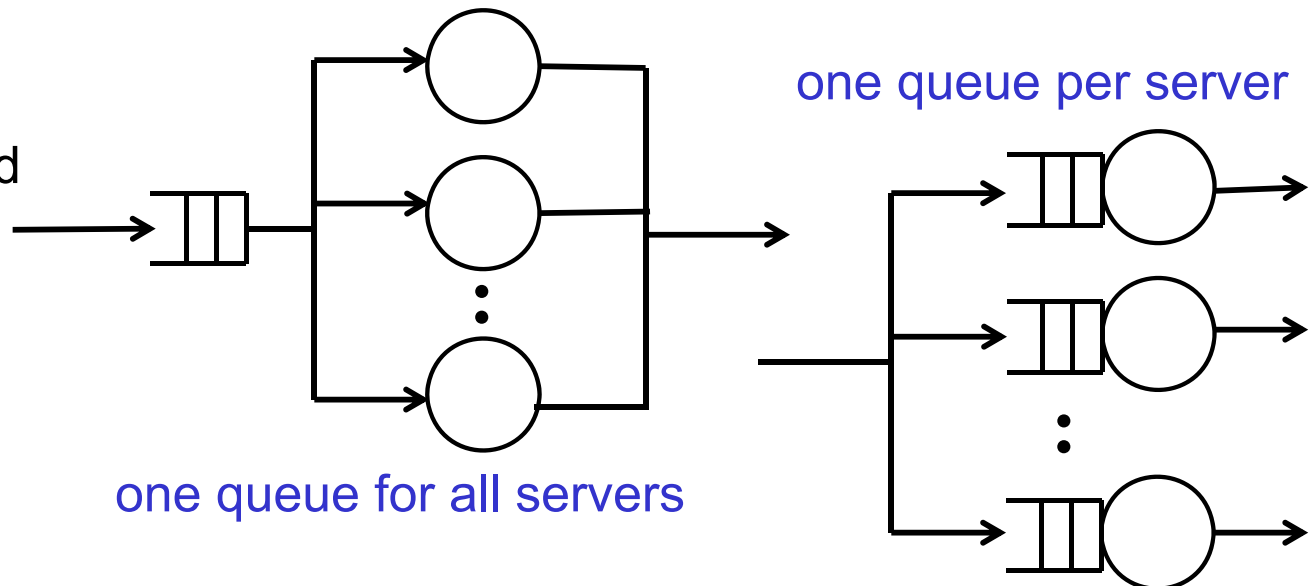
Performance



1. What is the average **time** it takes a job to complete service?
2. What is the **throughput** of the system (number of jobs completed per unit time)?
3. If arrival rate is doubled ($\lambda \rightarrow 2\lambda$), how much should μ increase? Do we do nothing or do we need another CPU?

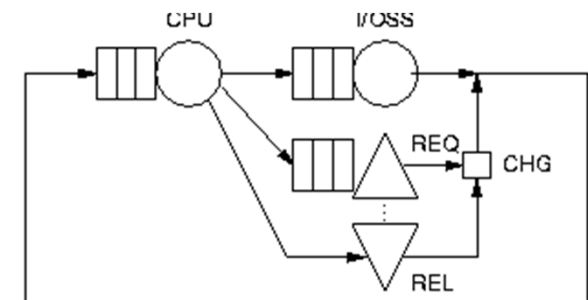
If we need more server capacity, what are our options?

- a. buy a new server with the needed capacity
- b. buy a few smaller servers that adds up to the required capacity
 - a. one queue for **all** servers
 - b. one queue for **each** server
 - c. does it matters?

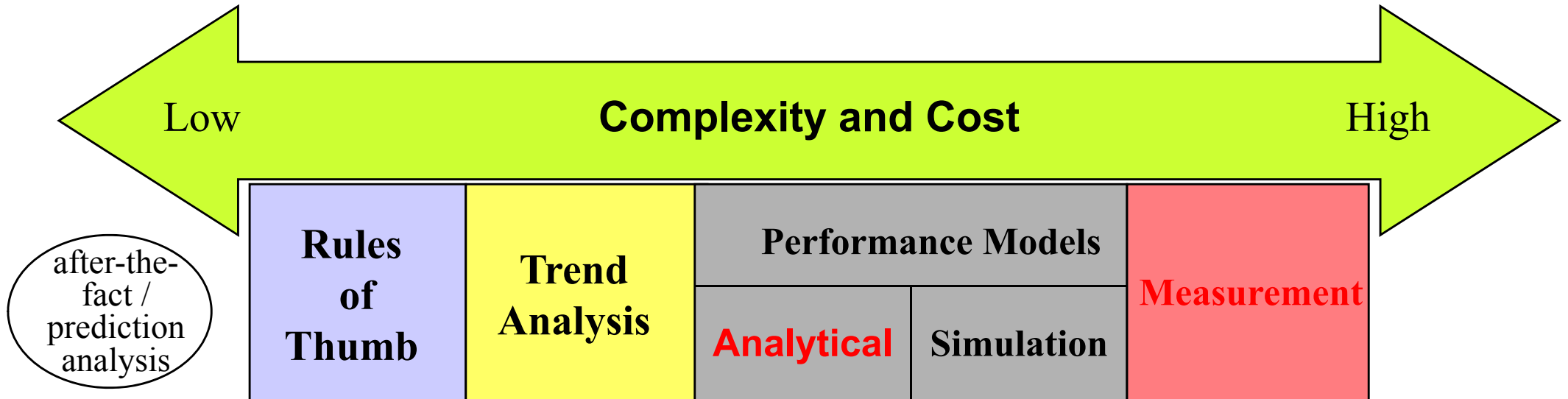


Performance Evaluation: How

- **Measurements** of actual systems
- **Simulations** using software models
- **Mathematical modeling** using techniques as queuing analysis



Performance Evaluation: How



CS5239 Computer System Performance Evaluation

CS5271 Performance Analysis of Embedded Systems

CS6211 Analytical Performance Modelling for Computer Systems

CS5233 Simulation and Modelling Techniques

CS6205 Advanced Modelling and Simulation

Course Catalogue

CS5239 Computer System Performance Analysis

Modular Credits: 4

Workload: 2-1-0-3-4

Prerequisite(s): CS1020 or its equivalent

The objective of this module is to provide students a working knowledge of computer performance evaluation and capacity planning. Students will be able to identify performance bottlenecks, to predict when performance limits of a system will be exceeded, and to characterise present and future workload to perform capacity planning activities. Topics include: performance analysis overview; measurement techniques and tools including workload characterisation, instrumentation, benchmarking, analytical modelling techniques including operational analysis, stochastic queuing network analysis; performance of client-server architectures; capacity planning; case studies.

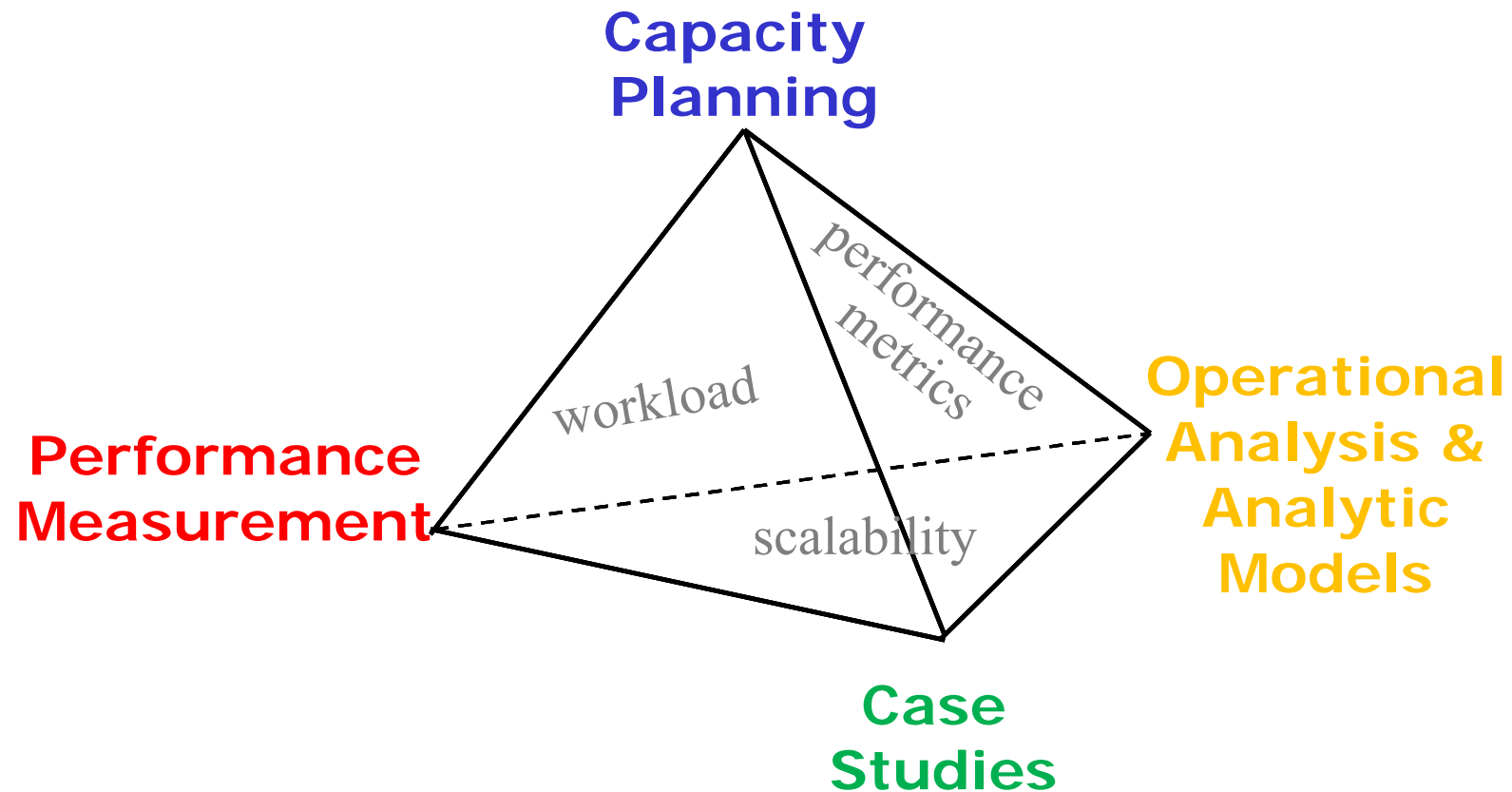
Prerequisites

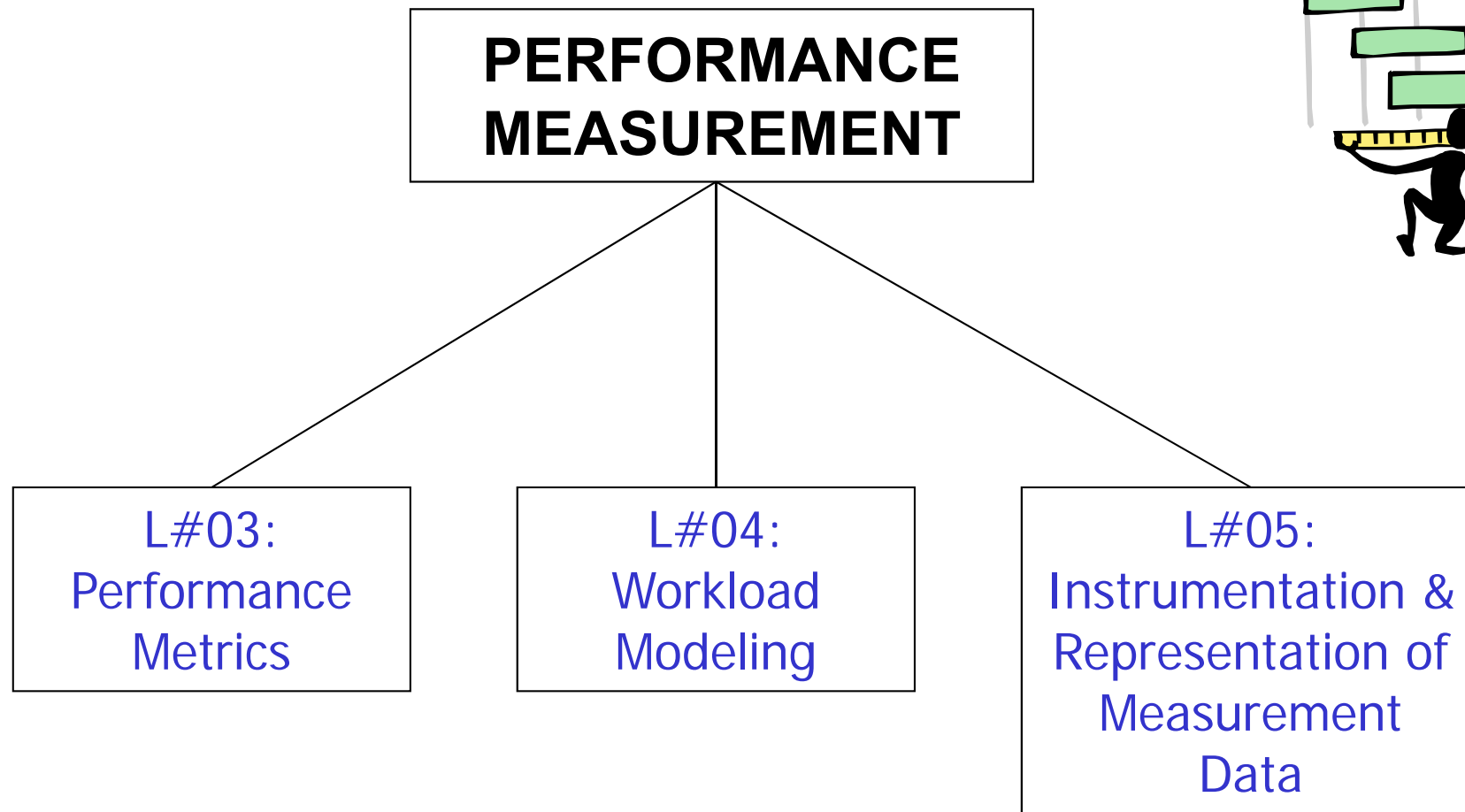
- CS1020 Data Structures and Algorithms I or its equivalent
- ST2334 Probability and Statistics (preferred)

Learning Objective

- performance analysis of computer systems
 - capacity planning
 - bottleneck and modification analyses
 - measurement and analytic model analyses
 - scalability analysis

What we cover





“Measurements are not to provide numbers but insights.”
Ingrid Bucher

OPERATIONAL ANALYSIS & ANALYTIC MODELS

L#06:
Introduction &
Notation

L#07-09:
Techniques

L#10:
Performance Laws
& Scalability

L#11-12
Case Studies

L#07:
Operational
Analysis

L#08:
Analysis of
Single Queue

L#09: Analysis of
Queuing Networks /
Multiple Classes

- bottleneck analysis
- performance bounds

System - open, closed, hybrid
Component - fixed capacity, delay,
load-dependent

Workload - single, multiple classes

Module Assessment

1. Continuous Assessment (60%)

- Test (20%)
- Assignment 1 (20%)
- Assignment 2 (20%)

2. Open Book Exam (40%)

- 2 hrs

Course Schedule & Webpage

- Lecture: Tue, 6.30-8.30pm, Com1, 02-02
- Tutor: Sunimal Rathnayake (Com2, #B1-01)
- Consultation: Wed, 10-12am
- Webpage:
 - IVLE for course announcement
 - www.comp.nus.edu.sg/~teoy/cs5239-16 for lecture slides, assignments, etc



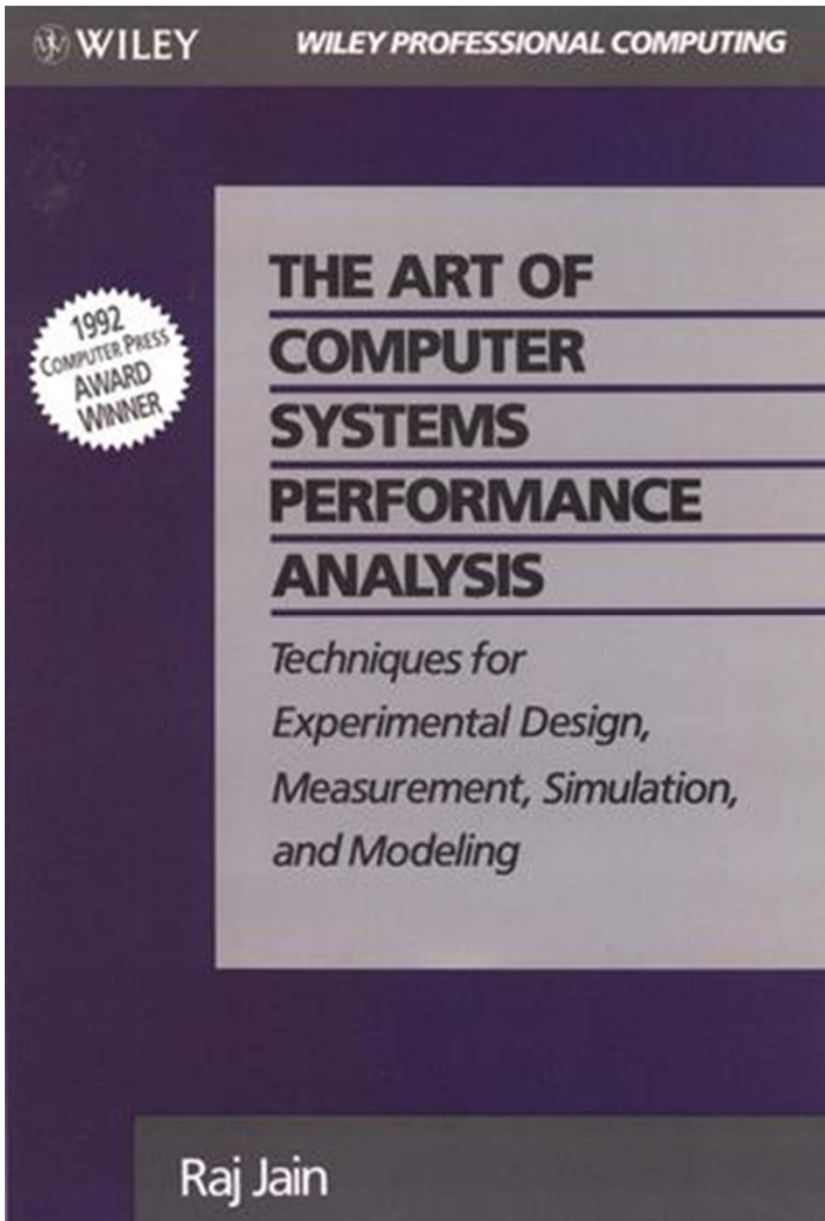
Resources

Main Textbooks

- ◆ **The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling**, R. Jain, John-Wiley, 1991.
- ◆ Quantitative System Performance, E.D. Lazowska et al., Prentice-Hall, 1984, <http://www.cs.washington.edu/homes/lazowska/qsp/>.
- ◆ Measuring Computer Performance - A Practitioner's Guide, D.J. Lilja, Cambridge University Press, 2000.

Reference Books

- ◆ Capacity Planning and Performance Modeling - From Mainframes to Client-Server Systems, Daniel A. Menasce, et al., Prentice-Hall, 1994.
- ◆ Capacity Planning for Web Performance – Metrics, Models and Methods, D.A. Menasce, et al., Prentice-Hall, 1998.
- ◆ Simulation Modeling and Analysis, A.M. Law and W.D. Kelton, McGraw Hill, 3rd edition, 2000.
- ◆ Introduction to Parallel Computing, A. Grama, et al., Addison-Wesley, 2nd Edition, 2003.



**The Art of Computer Systems
Performance Analysis: Techniques for
Experimental Design, Measurement,
Simulation, and Modeling**

Raj Jain

ISBN: 978-0-471-50336-1

720 pages

April 1991

Problems



If you're not sure,
don't guess...**ASK!**



Wrong guesses are **COSTLY!**



- consultation hours – Wed, 10-12, catch me after lectures, email