Performance Engineering for Enterprise Software Systems in Next Generation Data Centres

Jerome Rolia, Ludmila Cherkasova, Richard Friedrich
Enterprise Systems and Software Laboratory
HP Laboratories Palo Alto
HPL-2006-172
November 21, 2006*

software performance engineering, enterprise resource planning, model-based management

Software Performance Engineering (SPE) methods have been in use for over two decades as an approach to manage the risks of developing systems that fail to satisfy their performance requirements. In general, SPE advocates the use of performance oriented design principles to guide design decisions and predictive performance models to assess the performance impact of design alternatives. SPE methods have been used successfully to identify and overcome system design blunders early in the Information Technology (IT) project lifecycle before the blunders are built into a system and become expensive and time consuming to correct. While the methods have been used successfully in some IT project domains, they are not widely applied in the important domain of Enterprise Application (EA) systems. This experience paper considers the reasons for this and explores the role of SPE as new EA platform and data centre technologies become available.

We find that many risks traditionally addressed by SPE have been mitigated by the nature of existing EA platforms, the nature of today's IT projects for EA, and an attention to business process modeling. Furthermore, the design and implementation of future EA systems will see some performance risks reduced even further by new EA and IT system management platforms for Next Generation Data Centres. However, we expect that the nature of EA systems to be built is becoming more complex. As a result some familiar performance risks will re-emerge along with new runtime risks. We believe that SPE methods can help to mitigate such risks and describe research challenges that must be addressed to make this a reality.

* Internal Accession Date Only
A shorter version of this paper will appear in Proceedings of the Sixth ACN Workshop on Software and Performance (WOSP'07), 5-8 February 2007, Buenos Aires, Argentina
Approved for External Publication

Performance Engineering for Enterprise Software Systems in Next Generation Data Centres

Jerome Rolia
Hewlett-Packard Labs
1501 Page Mill Road,
Palo Alto, CA 94304, USA
Jerry.Rolia@hp.com

Ludmila Cherkasova
Hewlett-Packard Labs
1501 Page Mill Road,
Palo Alto, CA 94304, USA
Lucy.Cherkasova@hp.com

Richard Friedrich
Hewlett-Packard Labs
1501 Page Mill Road,
Palo Alto, CA 94304, USA
Rich.Friedrich@hp.com

ABSTRACT

Software Performance Engineering (SPE) methods have been in use for over two decades as an approach to manage the risks of developing systems that fail to satisfy their performance requirements. In general, SPE advocates the use of performance oriented design principles to guide design decisions and predictive performance models to assess the performance impact of design alternatives. SPE methods have been used successfully to identify and overcome system design blunders early in the Information Technology (IT) project lifecycle before the blunders are built into a system and become expensive and time consuming to correct. While the methods have been used successfully in some IT project domains, they are not widely applied in the important domain of Enterprise Application (EA) systems. This experience paper considers the reasons for this and explores the role of SPE as new EA platform and data centre technologies become available.

We find that many risks traditionally addressed by SPE have been mitigated by the nature of existing EA platforms, the nature of today's IT projects for EA, and an attention to business process modeling. Furthermore, the design and implementation of future EA systems will see some performance risks reduced even further by new EA and IT system management platforms for Next Generation Data Centres. However, we expect that the nature of EA systems to be built is becoming more complex. As a result some familiar performance risks will re-emerge along with new runtime risks. We believe that SPE methods can help to mitigate such risks and describe research challenges that must be addressed to make this a reality.

A shorter version of this paper will appear in Proceedings of the Sixth ACN Workshop on Software and Performance (WOSP'07), February 5-8, 2007, Buenos Aires, Argentina.

1. INTRODUCTION

Enterprise Application (EA) systems are a very important class of Information Technology (IT) systems. They are responsible for implementing business processes such as supply chain management, sales and distribution, customer relationship management, and human relations. These systems are critical to the world's economy and account for a large portion of annual IT spending. Performance is almost always a concern for these systems because poor performance may severely disrupt the ability of an enterprise to conduct its business. This paper focuses on Software Performance Engineering (SPE) as it relates to EA systems. We consider why SPE methods are not widely used in IT projects for EA systems today and why we expect the methods to be more widely used in the future. We identify research challenges that must be overcome to enable the general use of SPE methods to manage the risks.

EA systems are rarely built from scratch. Enterprises typically exploit EA platforms offered by EA vendors for at least part of their EA needs. This helps to manage the costs and quality of such processes. EA platforms are middleware that includes support for web, application, and database interactions while providing some transparency with respect to vendor specific hardware. Packaged software for business processes gets deployed to the middleware to implement the desired business processes.

Though enterprises rely heavily on packaged software for business processes, they must also differentiate themselves to their customers via business process innovation that enables more valuable or cost effective services. Innovation typically requires customized business processes. Customized and packaged processes are often tightly coupled via software development and configuration efforts. This makes the on-going maintenance of business processes and the IT systems that support them very challenging. This problem limits the ability of enterprises to extract value from their EA systems and is a major problem for EA today. Enterprises desire greater flexibility with respect to their EA and IT systems so that they can extract more business value.

There are several reasons why SPE methods are rarely used in such environments. SPE's focus on design principles is of little interest when adopting a packaged business process from an EA vendor. It is assumed that the vendor has already overcome performance blunders. Furthermore a packaged business process typically behaves in a similar manner for different enterprises.

Packaged processes typically dominate system behaviour.

Benchmarks are established by EA vendors to help estimate the hardware needed for systems that support a process based on simple characteristics such as numbers of users and process throughputs. The results of such exercises are typically good enough to size an IT system for an enterprise. However, this coarse approach often causes more IT resources to be provisioned than necessary. These additional resources then incur recurring management costs for the enterprise that are often greater than the costs to purchase the systems themselves.

Business process modeling tools are sometimes used to model customized processes [1, 2]. Analogies to SPE design principles apply in this domain. However, the models are typically at an abstraction that focuses on the relationship between the process and business value. The relationship between business processes and underlying IT systems, and hence IT system performance has been difficult to explore in detail. The cost of developing and maintaining SPE models that reflect such relationships by hand is prohibitive because there is so much information required for the models, the information needed to is rarely formally documented and must come from many sources, and is simply too difficult to obtain.

Fortunately EA and IT systems management platforms are evolving to better enable enterprises to create business value from their EA systems. EA platforms are being refactored to become more service oriented and more dynamically configurable. Next Generation Data Centers (NGDC) will create and operate virtual IT environments allowing resources to be automatically reassigned in response to changing business and IT requirements. To achieve this, management systems for the platforms are becoming more model-driven with repositories of models that maintain configuration information and tools that rely on the models to automate management tasks. The advances permit greater customization and flexibility for EA systems, a greater synergy between EA system management and IT systems management, and more self-management for IT systems.

The new platforms will lead to the development of more agile but more complex EA systems with re-emerging and new performance risks. We believe SPE methods have an important role to play in mitigating the performance risks. The repositories of models will enable greater automation for SPE model construction thereby reducing one of obstacles to the use of SPE methods.

The rest of the paper is organized as follows. Section 2 gives an overview of EA platforms and software. This is followed in Section 3 by a description of advances in EA and IT system platforms. Section 4 discusses model-driven management and Section 5 anticipates the impact of the advances on EA systems. Section 6 considers performance risks old and new and how they relate to SPE. Finally a summary and conclusions are offered in Section 7.

2. OVERVIEW OF EA PLATFORMS AND SOFTWARE

EA systems are typically based on business process architectures that describe abstract business process sets common to most enterprises. Figure 1 illustrates such an architecture. Business process sets cover related groups of processes such as financial management or customer relationship management. Each process in the set elaborates on intra and inter-enterprise steps needed to satisfy typical business and regulatory requirements. The steps may be implemented by people and/or IT systems. Each enterprise must decide how to do this in a manner that best meets their needs.

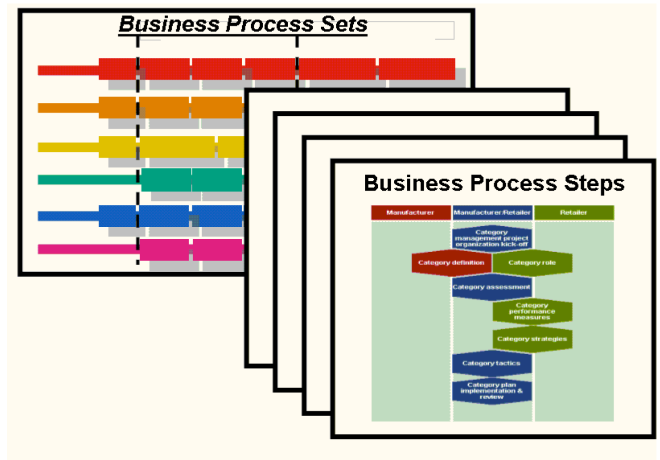

Figure 1: Business process sets and detailed steps.

EA vendors provide platforms and packaged software to implement a portion of the processes described by process architectures while leaving the rest to people or customized IT systems. Fully implementing EA systems is a big challenge for enterprises. EA vendors typically offer a limited number of process variants to choose from. Typically enterprises must adapt to the processes when adopting them. As part of this onerous effort, the roles of people are likely to change and interactions with other processes must be taken into account. Ultimately IT systems must be made to support a consistent set of business processes for the enterprise.

While adopting packaged EA processes can be a difficult and expensive undertaking, developing and maintaining in-house processes can also be cost prohibitive. The development of a best practice business process may represent hundreds if not thousands of person years of effort. EA software vendors have an advantage in that they can amortize development and maintenance costs of packaged software over the many enterprises that adopt the software.

However, it can be a disadvantage for an enterprise to rely too heavily on an EA vendor's business process framework since it can severely limit business process innovation and related competitive advantage. For example, one enterprise may offer its customers the ability to purchase a fully customized product rather than one of several pre-defined configurations. This ability may be of great value to customers but may impact the supply chain processes used to manufacture the product. Other competing enterprises that want to offer the same customization capability may not be able to adapt unless their EA vendor provides the same supply chain support to all enterprises. Relying too much on packaged software may therefore have a hidden cost in terms of lost business opportunity. Enterprises typically adopt business processes from EA vendors where innovation is less likely. They then build upon them with new and/or customized processes to differentiate themselves.

3. ADVANCES IN EA AND IT SYSTEM PLATFORMS

To better support business process innovation for individual enterprises, EA software vendors are now offering integrated support for business process execution engines and (web) service-oriented architectures. Business processes are described in terms of control flow models with process steps [3]. These are executable models with steps that may be associated with software components. For these, a process execution engine invokes a step via web service or proprietary protocols. This avoids the direct coupling of business process logic with the software components that implement business process steps and makes the process visible. The software components act as services that can be re-used by many processes.

Executable business models make it easier to understand and modify the business process control flow, the relationship between process steps and software components, and the software components themselves. While it may seem obvious that this is desirable, legacy EA systems often implemented process logic within their application component code making it difficult to maintain and modify. An example of a business process control flow model is illustrated in Figure 2 (the text in the boxes is obscured as it is not relevant for our purpose). It shows branches, loops and process steps that may be implemented by people or IT systems.

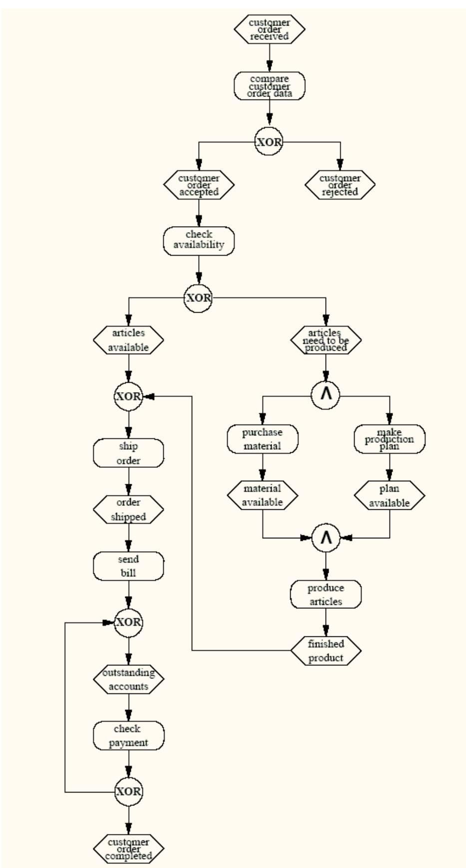

Figure 2: Business process control flow diagram.

Ultimately EA systems are implemented on IT system platforms that include servers, networking, and storage. These IT system platforms are evolving into Next Generation Data Centres. Rather than associating specific EA platform middleware servers with explicit hardware resources, hardware is becoming virtualized. This permits many applications to be associated with each hardware resource. Sharing hardware can significantly reduce the overall cost of ownership for enterprise IT because it reduces management costs.

In Next Generation Data Centres, hardware and virtualization vendors will programmatically offer server, networking and storage infrastructure on demand to EA systems. Figure 3 illustrates a Next Generation Data Center with a resource pool of servers that supports virtualized resources. Within such a pool it is possible to dynamically create and delete resource containers that act as virtual hardware servers, to increase or decrease the performance capacity associated with them and, with some choices of virtualization technologies, to migrate resource containers from one physical resource to another without interrupting the functional behaviour of any EA systems they support.

Many enterprises are beginning to exploit large shared resource pools in data center environments to offer shared access to computing capacity. The goal is to lower their infrastructure and management costs. These environments may have tens, hundreds, or even thousands of server resources. Workload and capacity management frameworks will exist to automate management tasks in these environments.

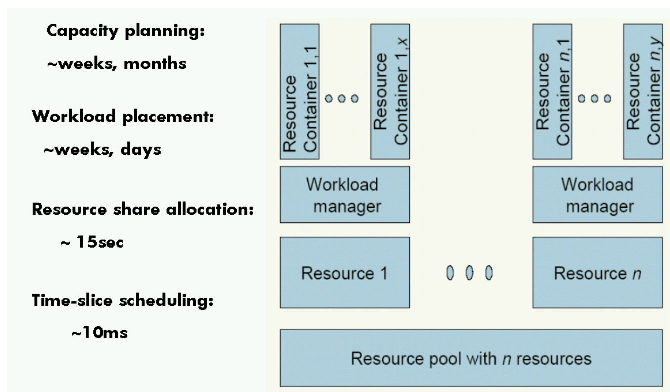

Figure 3: Resource pool environment with Illustrative Management Timescales.

Figure 3 illustrates several aspects of capacity management and their management timescales. Long term management corresponds to capacity planning; the goal here is to decide when additional capacity is needed for a pool so that a procurement process can be initiated. Over a medium timescale (e.g., weeks to days), groups of resource containers are chosen that are expected to share resources well [4]. Each group is then assigned to corresponding resources. Assignments may be adjusted periodically as service levels are evaluated. Once resource containers are assigned to a resource, a workload manager for the resource [5, 6] adjusts workload capacity allocations over short timescales based on time-varying workload demand. Finally, resource schedulers operate at the time-slice granularity (sub-second) according to these allocations.

Adjustments to allocations in response to changing workloads can greatly increase the efficiency of the resource pool while providing a degree of performance isolation for the containers.

These advances are being exploited by data centre operators to make more efficient use of the hardware resources thereby decreasing overall IT management costs. IT systems management platforms aim for greater resource pool self-management where EA system performance requirements are specified and used to automatically guide the run-time behaviour of capacity management systems in resource pools [7, 8].

4. MODEL-DRIVEN MANAGEMENT

To better manage the additional dynamism and complexity that comes with new EA and IT system platforms, management systems are becoming more model-driven and service oriented. They maintain repositories of model information that describe the configuration of the system. Management tools rely on the model information to manage the lifecycle of EA [9] and IT systems. The management tools are responsible for keeping the model information up to date and consistent. Ultimately management tools for EA system design, development, and deployment and IT system platforms for deployment and on-going management will interact to enable more cost effective management and more adaptive enterprises. Figure 4 illustrates aspects of IT service management that can be expected to be more integrated over time [10].

Figure 4: IT Service Management Reference Model.

The model-driven approach aims to support cost-effective management services by capturing configuration information in models in a way that can be reused. For example, suppose an enterprise will introduce a new product to be sold during the back-to-school season. Support for the new product is expected to cause a business process change for the system. The CEO asks a capacity planner whether or not the IT systems will be able to support the additional workloads. To answer the question a capacity planner may browse models to discover which business process are being used, which resources they use, references to historical information about process throughputs and response times, and to business process resource usage. The capacity planner may then bind such information as inputs to a capacity planning tool that is used to predict the impact of future changes to a business process. A similar approach may be used to support the re-design of business processes and/or their relationships with IT systems.

Today's performance analysis and design services require time consuming human interview processes to discover relevant aspects of a system and then begin manual data collection processes to obtain further information. These exercises offer only restricted views into systems (as time permits), are rarely accurate, and leave ad-hoc model artifacts that become more and more inaccurate with respect to continuous change in the real system. Model-driven methods provide the opportunity to reduce costs by simplifying the process of finding information, automating data gathering and supporting the integration of tools. While virtually all of today's enterprises lack detailed models for any of their systems, we expect the use of models to become more prevalent as the new platforms are adopted.

EA vendor tool chains now automatically maintain models that document important aspects of their systems. These models support the vendors' own tool chains, are tightly integrated with change management processes, and can be expected to become more robust over time. The existence of these new models presents opportunities for providing advanced cost-effective model-driven performance analysis and performance management services.

5. ADVANCED EA SYSTEMS

Enterprise business needs will drive the use of advanced platforms to create systems that are more customized, that have more linkages with EA systems from other enterprises, and that will be more integrated with sensors from the physical world. These systems will be even more complex than today, and will necessarily rely on model-driven management tools from EA and IT system platform vendors to better support on-going management and change management.

As an example, consider a large international retailer with a complex supply chain that includes thousands of suppliers. The retailer may expose inventory level information via EA systems to the suppliers so that they can be responsible for ensuring that just enough inventory is always available at each of its locations. Each supplier may have its own customized business processes to best interact with the retailer and the supplier's own suppliers and transportation providers. Items to be sold, shipments, and transportation vehicles may all be tracked by sensor networks to better enable the just-on-time yet cost-effective delivery of product to store shelves and ultimately to the consumer.

It is not practical for all of the EA systems from all participants to be homogeneous. Yet they must all operate and evolve over time together. This presents a challenge for design, configuration, and change management with respect to the EA systems and IT systems that support them. A change in one business process or software component may have a significant impact on the behaviour of other processes and their use of IT resources. Any interruptions to IT systems that result from such changes could disrupt supply chains and ultimately affect the profitability of the retailer and/or its suppliers.

6. SPE RISKS, OLD AND NEW

The new flexibility for EA systems permits a more rapid design and modification of business processes for enterprises. These may result in significant changes in workloads presented to IT systems. Furthermore, IT systems may dynamically vary the capacity associated with parts of EA systems to better utilize resources. We classify the corresponding performance risks as due to:

- customization,
- agility,
- interdependence,
- resource sharing and late binding,
- scalability, and
- model composition.

Below, we discuss how they relate to SPE.

Customization: customized EA systems, with respect to processes and/or software components, may behave very differently from benchmarks reported for packaged business processes. Sizing methods previously used become less reliable when customization is exploited.

Agility: additional flexibility permits enterprises to make greater changes to their business processes and IT systems more often. These often change increases the likelihood of incurring a performance problem.

Interdependence: new business processes are typically layered upon existing business processes within an enterprise and may interact with EA systems of other enterprises. These EA systems are a web of application and IT infrastructure that supports them. The interdependencies among business processes may have a significant impact on resource usage for IT systems.

Resource sharing and late binding: to better manage IT costs, many enterprises are exploiting pools of shared resources. Workload management features in these environments may provide time varying capacity to various components of an EA system. Workloads may get access to more or less resource as their demands change. Resources may adjust service rates to better manage their use of power. Virtualization mechanisms may be chosen at runtime. The impact of such dynamism may affect the quality of service for interactive users or the performance qualities of the business processes for EA systems.

Scalability: development and performance testing is typically done in small testbed systems. Yet these smaller environments may not expose the performance risks incurred in a large scale environment.

Model composition: models require information from many sources. Open mechanisms are needed to obtain the information at all. SPE performance models must be built from this information and advanced methods are needed to validate that the behaviour of each model and the composite behaviour of derived models are as intended. For example, business process models may include sufficient information to guide control but additional annotations are needed to describe expected behaviours (e.g., loop counts, branching probabilities, performance requirements) and measurements are needed to validate them. Standard models will be needed for resources that take into account aspects such as virtualization, resource, and power management.

We now consider the role of SPE for managing these risks. We assume that repositories of management models will ultimately enable a more automated creation of SPE performance models to support SPE exercises.

Customization and interdependence present traditional performance risks that have been managed by SPE. SPE performance models need to be augmented to take into account business process control flows, their steps and relationships to IT systems and people. However the principles of SPE apply in the same way. Business process modeling tools have done this to some extent [1, 2] but more progress is needed to relate the processes to the IT systems that support them.

Support for performance oriented design requires a characterization of demands for fine grained application components [11]. Methods are needed to characterize EA vendor components that are expected to affect the performance of EA systems most in a way that the characterizations can be re-used in performance models. The characterizations should estimate capacity requirements for likely usage scenarios and for different IT system platforms (e.g., different physical and/or virtualized resource containers).

Agility will enable change to occur more frequently. It will be important to continuously maintain performance models for EA systems so that the impact of changes can be explored in a timely manner.

Resource sharing has diametric impacts on performance. On the one hand, in large shared resource pools the amount of capacity associated with an EA system can be dynamically scaled up or down. This reduces the risk of not having enough capacity for a particular EA system as the capacity can be changed quickly. Capacity planning becomes a responsibility for the shared IT system environment as a whole with risks amortized across many EA systems. On the other hand, dynamic capacity adjustments must be well understood. A local resource allocation optimization in one IT system environment may have a negative impact on the global behaviour of an EA system. SPE methods can be used to better understand the sensitivity of an EA system to dynamic resource sharing and late binding so that the use of workload managers in support of EA systems is planned in advance to reduce resource sharing risks for the EA system as a whole. In this way SPE models may also be well suited to support self-management methods for resource pools.

7. SUMMARY AND CONCLUSIONS

Enterprise Application systems are a very important domain of IT systems. Yet SPE methods have not been very successful in this domain. This is largely a result of the way in which EA software has been adopted and the difficulty in creating SPE performance models for these systems.

However, enterprises have a need for more advanced EA platforms. They require more flexibility from their EA and IT systems so that they can create more business value. EA and IT system platform vendors are responding with new advanced platforms that provide for more adaptive enterprises. The new platforms will enable greater customization, agility, interdependence among business processes, and IT system resource sharing, but will incur increased performance risks. The performance risks are similar to those traditionally addressed by SPE but also include new risks due to resource sharing and late binding. SPE methods may be necessary at runtime to better understand system behaviour in such complex environments.

Fortunately, the new platforms have model-driven management platforms. We expect the breadth and depth of model-driven management to increase over time. Ideally repositories from management system vendors will jointly support the automated generation of SPE models for EA systems so that SPE principles can be more easily applied to mitigate the performance risks. Thus we see an increasing role for SPE methods in the design and management of advanced EA environments.

8. ACKNOWLEDGEMENTS

The authors wish to than their many colleagues within Hewlett Packard that have helped to explain concepts of current and future EA and IT system platforms.

9. REFERENCES

[1] IDS Scheer, ARIS Platform, www.aris.com.

[2] Chun Ouyang, Eric Verbeek, Wil M. P. van der Aalst, Stephan Breutel, Marlon Dumas, Arthur H. M. ter Hofstede: WofBPEL: A Tool for Automated Analysis of BPEL Processes. ICSOC 2005: 484-489.

[3] WS-BPEL, www.oasis-open.org.

[4] J. Rolia, L. Cherkasova, M. Arlitt, A. Andrzejak: A Capacity Management Service for Resource Pools. Proc. of the Fifth International Workshop on Software and Performance (WOSP'2005), Spain, July 11-15, 2005.

[5] HP-UX Workload Manager, http://www.hp.com/products1/unix/operating/wlm/

[6] IBM Enterprise Workload Manager, http://www.ibm.com/developerworks/autonomic/ewlm/

[7] J. Rolia, L. Cherkasova, M. Arlitt, V. Machiraju. Commun. ACM 49(3): 55-60 (2006).

[8] L. Cherkasova and J. Rolia: R-Opus: A Composite Framework for Application Performability and QoS in Shared Resource Pools. To appear in the proceedings of the International Conference on Dependable Systems and Networks (DSN 2006), Philadelphia, PA, USA, June 2006.

[9] SAP Solution Manager http://www.sap.com/solutions/netweaver/solutionmanager /index.epx

[10] HP ITSM Reference Model, ftp://ftp.hp.com/pub/services/itsm/info/itsm_rmwp.pdf

[11] V. Vetland, Measurement-Based Composite Computational Work Modelling of Software. Ph.D. Thesis, University of Trondheim, Aug. 1993.