ABSTRACT

Tactical and operational planning for manufacturing enterprises are more important today than ever before as their supply chains span the globe. Two state-of-the-art technologies that are critical to success are Discrete Event Simulation and Advanced Planning and Scheduling. They are commonly applied in designing and executing operations at each site within the supply chain. However, as supply chains become leaner and more responsive, operational constraints and stochastic influences within the manufacturing sites and the logistics network require a combination of both technologies applied to the entire supply chain. This paper describes a novel framework for advanced distributed simulation with integrated APS procedures for collaborative supply chain optimization. The framework can be used for fast optimization of both planning procedures and execution policies and also provides a base for easy implementation of simulation results. A prototype of a distributed semiconductor supply chain simulation has been developed and is currently being refined.

1 MOTIVATION

Supply chain management (SCM) involves managing the flow of material and information through multiple stages of manufacturing, transportation and distribution with the objective of minimizing costs and maintaining low inventories without compromising customer service level. The effective practice of SCM is critical to every company whose business environment involves global competition. The ability to navigate complex supply networks, crossing multiple enterprise boundaries with good visibility to material and information, fast information flow and collaborative supply chain optimization are all essential to achieving increased responsiveness to higher customer expectations while creating and sustaining cost competitiveness.

Two state-of-the-art technologies to tackle these challenges have found wide application: Discrete Event Simulation (DES) and Advanced Planning and Scheduling (APS). In both cases the main area of application is on performing what-if analysis, by varying different aspects of the supply chain. Whether one or the other technique is more appropriate depends on the nature of the business to be supported.

Discrete event simulation techniques have been used for high-level or aggregate analysis where the results of detailed local operational decisions can be represented by an appropriate random variable. For example, the details of shop floor control decisions are aggregated into a random variable representing total flow time. Or the details of transportation system operation are aggregated into a random variable representing total transport time. The power and appeal of DES is that it can represent system flow patterns with good fidelity. The difficulty of applying DES stems from the need to achieve reasonable fidelity between the representation of outcomes using random variables and the actual decision processes that drive outcomes in the real system. Particularly when sophisticated SCM systems are in use, it can be quite challenging to represent their impact in a DES model with reasonable fidelity. The typical question to be answered by DES is addressing the system performance (i.e. on a macro/aggregate level), and the typical result of a simulation run is an estimate of “how the real system would behave”.

Commercial simulation tools for analyzing supply chains have been released in recent years, for example the Supply Chain Analyzer by IBM (Archibald, Karabakal and Karlsson, 1999).

Advanced planning and scheduling techniques support operational decision-making by creating “high-fidelity” execution plans directly generated from customer demand (either sales forecast or orders), bills of materials (BOM), standard operating procedures (SOP), capacities, customer data, current inventory, and work-in-process (WIP). The
power and appeal of APS is the ability to utilize very detailed information about system status. With constraint-based planning algorithms, bottlenecks can be detected in advance and “feasible” execution plans generated. Excess demand is either rejected or dealt with through increased lead-time, increased inventory, increased capacity, or outsourcing. The challenge in using APS is incorporating models of system behavior (essentially, temporal input/output models) that are sufficiently accurate. The typical question to be answered by APS is addressing the feasibility of the (overall) demand or a particular customer order (i.e. on a micro level), and the typical result of a planning run is a specification of “how the real system should behave”.

A large number of integrated APS planning and optimization tools have been released into the market in recent years. Some of them such as the RHYTHM suite by i2 (Padmos et al., 1999) also incorporate simulation-based scheduling techniques for local optimizations. To draw accurate conclusions from a supply chain analysis, it has been shown that a detailed model needs to be built (Jain et al., 1999). By applying either discrete event simulation or advanced planning and scheduling, most of the challenges contained within a single supply chain node can be tackled. However, in many scenarios across a supply chain, a combination of both technologies is required to obtain high-fidelity results. With increasing re-planning capabilities across entire supply chains this issue becomes even more crucial.

For example, it might be possible to simulate realistic outbound logistics scenarios of a factory towards different distribution centers, only if customer demands are checked against a master supply chain plan generated by an APS procedure integrated into the simulation, taking into account the factory’s sequencing constraints. At the same time the logistics network might be subject to significant stochastic influences.

Another common obstacle for high-fidelity supply chain simulation is the lack of availability of consistent and correct, possibly real-time, input data from multiple, distributed ERP and legacy systems.

Last but not least, it often seems difficult if not impossible to actually benefit from simulation results because they cannot be implemented sufficiently fast and without losing release-capability in an IT landscape which is typically quite inflexible.

These shortcomings are all to be addressed in this paper. A novel framework for collaborative supply chain optimization across globally distributed locations, including the possibility of shielding sensitive company data from other supply chain nodes, is described. It is based on ultra-fast distributed simulation with integrated APS procedures. Eventually the framework can also be used to optimize such APS procedures. The major stochastic uncertainties within the entire supply chain are taken into account. Easy implementation of simulation results into an existing IT landscape can be accomplished with zero downtime.

The paper is organized as follows. Section 2 gives a conceptual and functional outline of the framework and also describes the general architecture. This is followed by an overview of the technologies required and some issues regarding input data and synchronization in Section 3. In Section 4 the industries for which this framework will be feasible and the potential benefits are discussed. Finally, Section 5 concludes with open issues and challenges to be addressed in future research.

2 FRAMEWORK

2.1 General Outline

The distributed simulation framework presented in this paper comprises a number of nodes (manufacturers, suppliers, logistics service providers) forming a supply chain as depicted in Figure 1. A subset of them might share a planning domain with the task of generating a common overall supply chain execution plan (production and/or procurement and/or distribution) for mutual benefit.

Figure 1: Subset of a Supply Chain Sharing a Planning Domain

Within the simulation framework each participating corporation/company will be able to run their own simulation model of manufacturing and/or logistics operations at their own site where users interact with the system. This is shown in the example in Figure 2. For the simulation, planning and scheduling systems (eventually APS procedures) are logically separated from the execution systems. The simulation models interact and exchange data with the planning and scheduling systems in the same way as the real manufacturing or logistics nodes.

Detailed model information (application codes and data) is encapsulated within each (either planning or operational) model itself. Typically this is required for the supply chain simulation because corporations would not want
to share sensitive data such as certain execution policies (Gan et al., 2000). This is even more critical in the case when members of the planning domain are competitors (see Figure 1). Also, suppliers of planning and scheduling software are not willing to disclose their algorithms.

The participating corporations only need to define essential data flows from one supply chain node to another. In the background the system initiates a remote model. Data representing the simulated material and information flow is then exchanged as messages during the simulation run. These messages can be transmitted through a network connecting the participating corporations. The network is also needed for the transfer of the relevant results to all partners for animation-based analysis.

In models describing a single supply chain node, combination of simulation and planning is already commonly applied, because in many cases this might be the only way to generate appropriate input data for the simulation of the actual situation. For example, in a scenario with many new products to be taken into account it might be impossible to calculate realistic release rates of input materials without generating a production plan.

However, across entire supply chains that are connected by several simulation models, the approach of combining distributed simulation and APS is rather novel.

### 2.3 High-Fidelity Simulation

Results of a simulation run that comprises several supply chain nodes naturally should be of high fidelity, therefore the external behavior of the individual models should be as realistic as possible. This means that all systematic internal effects, constraints and allocations that have a significant influence on both the relevant key performance indicators and the behavior of other nodes need to be visible. This is the main reason why in many cases it is actually indispensable to incorporate the logic of APS procedures into the simulation. This is illustrated in the following examples.

![Figure 2: Distributed Simulation Framework](image)

Two critical issues for implementing such a distributed supply chain simulation are: (a) the specification of the interfaces between models, and (b) the mechanism for supporting inter-model communication.


#### 2.2 Features

When users interact with the system to start a simulation run within the framework, they typically want to tackle and resolve various operational and tactical challenges within their site, possibly comparing different sales forecasts or demand scenarios. Typical questions to be addressed are the optimal safety stock to reach desired customer service level or the optimal lot size strategy to minimize costs, or more general, the effect of a certain execution policy on the relevant key performance indicators.

Since APS procedures are integrated into the simulation, the framework can eventually be used to optimize these planning procedures as well. Moreover, it is possible to demonstrate the benefits in terms of industrial and logistics costs as well as customer service compared to a scenario without such APS procedures.
In a simulation scenario which does not incorporate the generation of this execution plan, the deadline \( D_{n+1} \) for customer order \( C_{n+1} \) would be missed by \( \Delta_{n+1} \) as shown in Figure 3. However, if an APS procedure is applied as in a real scenario where customer order \( C_{n+1} \) would actually have been rejected because of the known capacity constraints, the simulation could be much more realistic.

Figure 4 illustrates the circumstances under which the interdependencies between local scheduling procedures and optimization criteria have to be taken into account in a supply chain simulation. In the example customer A orders product 2 and customer B orders products 2 and 5. A planning procedure that takes into account all the constraints would plan the required quantity of product 2 for customer order A in the first block of product 2, and the quantity for customer order B in the second block as depicted. The reason is that customer order B would not be available for outbound logistics until product 5 has been made in the last block, which cannot be advanced due to sequencing constraints. Unless there are other constraints such as procurement lead-times for additional customer-specific components which would not allow product 2 for customer order A to be made in the first block, this scenario would be optimal: Customer order A would be available for outbound logistics as early as possible, and work-in-process for customer order B could be minimized.

![Production Scenario Under Sequencing Constraints](image)

However, a simulation scenario in which the required quantity of product 2 for customer order B is made before the quantity required for customer order A, simply for the reason that it comes in first, would not be realistic.

### 2.4 Required Input Data

The input data required for the type of simulation as described basically has to serve two tasks: (a) describe structures and execution rules in each operational node, and (b) feed the integrated planning and scheduling procedures.

The input data comprises bills of materials or recipes, standard operating procedures and capacities, customer data, work-in-process and inventory at the beginning of the simulation run, as well as demand data such as sales forecast and/or customer orders.

Only the data required by the integrated APS procedures needs to be uploaded to the relevant simulation models, it can still be shielded from nodes which do not belong to the planning domain. The other data is required for the respective local simulation models only and does not have to be disclosed to other nodes.

To avoid data redundancy, input data should be uploaded from ERP and legacy systems wherever available, and also maintained in these systems. One way is to upload BOM, SOP, capacities, customer data, and WIP and inventory data directly from the ERP/legacy system(s) for each simulation run. This would be of advantage in an environment where the product structure changes frequently. Moreover, with “near-real-time” WIP and inventory data the framework can be used to tackle even operational issues. On the other hand, this approach raises questions regarding the repeatability of simulation results.

Repeatability is not much of an issue in the alternative approach where data is input from intermediate files, which are updated from the ERP/legacy systems only upon request. Such a procedure is naturally more cumbersome, and it might make the framework impracticable to address tactical or operational issues. In the end, both approaches need to be supported by the framework to ensure maximum flexibility.

Demand data will typically have to be generated from historical customer orders and needs to be supported with sales forecast data in the case that future demand is significantly different from past demand and/or new products are to be phased in.

### 2.5 Interface Modeling

In order to encapsulate the operation of each individual element of the supply chain (or its model), and yet have the models interact, an interface specification is required. Analogous to an application program interface (API) the specification should be complete yet concise. At present, we know of no industry standard describing this type of specification.

The approach we have taken is to apply UML (Unified Modeling Language) to identify the key interactions and then to specify the objects and messages to be shared between nodes in the supply chain simulation. An example of a shared object is an order, which contains information about the items being ordered and subsequently shipped from a supplier and transported by a transportation node. An example of a message is an inventory status enquiry from a business process to a source for a product ordered by a customer.

In addition to the specification of the interfaces, there must be a method by which the interface specification is communicated to each participating model and enforced in the operation of the distributed simulation. As outlined be-
low, this requirement is satisfied by using the High Level Architecture infrastructure.

3 TECHNOLOGY

3.1 Parallel and Distributed Simulation
Using High Level Architecture

The integration of a set of corporations’ simulation models and APS procedures to form a high-fidelity supply chain simulation can be accomplished by adopting the standards of the High Level Architecture (HLA).

The High Level Architecture is an architecture for reuse and interoperability of simulations (Kuhl, Weatherly, and Dahmann, 1999). In HLA terms, each simulation model (which in our case represents either an operational node or an APS procedure within the supply chain) is referred to as a federate, while a collection of such federates makes up a federation. The HLA supports the possibility of distributed collaborative development of a complex simulation application as well as the reuse of capabilities available in different simulations. Thus, a set of simulation and planning models, possibly developed independently and implemented using different languages and hardware platforms, can be put together to form a large federation of simulations. At the same time this can help corporations to avoid recreating simulation models for different supply chains they might be part of. The HLA has been adopted by Object Management Group (OMG) and IEEE as a standard to realize distributed simulation.

Several challenges arise due to this integration and can be resolved or partly resolved by the natural characteristics and the implementation of this architecture. These challenges are (a) simulation time synchronization of distributed corporations’ models, (b) protecting and sharing of sensitive data, (c) lengthy execution time of large scale supply chain simulation, and (d) online visualization of a distributed simulation.

3.1.1 Synchronization between Federates

Synchronizing simulation time advancement is crucial for guaranteeing the correctness of a distributed simulation. Future events triggered by one federate/model must not affect the past of another. In the supply chain context, each federate is simulating one corporation and the federates are running at geographically distributed sites. The problem of synchronization arises because of the exchange of timestamp messages between the geographically distributed federates.

The time management service of the HLA implementation helps to tackle this challenge by enforcing several rules when external (timestamp) communications between federates are necessary (Fujimoto, 1998). These rules guarantee that a federate will only be allowed to advance to a simulation time when no events with timestamp less than or equal to this simulation time will be received in the future.

The criticality of synchronization in a distributed simulation scenario that incorporates APS and available-to-promise features is illustrated in the following example.

The generation of a new supply chain execution plan can be triggered by any kind of event that changes demand or capacity within the supply chain. Typically this is accomplished regularly (for example once a month when a new sales forecast comes in). For such decision-making, “real-(simulation) time” execution data about inventory, WIP, backlog, etc. from all relevant locations is needed, thus requiring synchronization of the corresponding federates before the information can be exchanged between them. This is illustrated in Figure 5.

![Figure 5: Synchronization in a Distributed Supply Chain](image)

In the case that a customer order is checked for “availability-to-promise”, the relevant business process federate does not necessarily have to be synchronized with the other federates because the business process would check the customer order only against the existing supply chain plan of which all the information is contained in the business process federate itself. However, if the ATP check also involves “real-(simulation) time” inventory data from finished good warehouses which might be the case in a make-to-stock environment, real-(simulation) time data would be required.

3.1.2 Information Shielding

Using the HLA, each participating simulation in the federation can define the information that it likes to share with others, but its internal behavior (and other sensitive data) is completely invisible to the outside world (i.e. the other federates). Even though the HLA can hide information that a corporation does not want to share, it lacks the capability to share a subset of the sensitive data with a subset of corporations that make up the supply chain.

This limitation can be resolved by a technique called hierarchical HLA (Cai, Turner, and Gan, 2001). This approach shows significant potential of being further developed to resolve other technical challenges of distributed supply chain simulation such as improving the scalability of the simulation and relaxing the synchronization requirements between federates.
3.1.3 Execution Time

Lengthy execution time is another major concern when it comes to large scale supply chain simulation that involves more than one corporation. Any one federate that runs slowly (typically because of the complexity of its model) will hinder the progress of the whole supply chain simulation.

To tackle this problem, the internal parallelism of the bottleneck federates can be exploited using a parallel federate architecture (Ji, Gan, Turner, and Cai, 2001). This parallel federate architecture partitions the bottleneck federate to form logical processes (LPs) that are simulated in parallel on shared memory multiprocessor system. It integrates a parallel discrete event simulation (PDES) protocol (Gan and Turner, 2000) and HLA-based distributed simulation and facilitates the formation of a hybrid distributed simulation that consists of both sequential and parallel federates. With this parallel federate architecture, the performance of the overall supply chain simulation can thus be improved.

3.1.4 Visualization

The ability to see the simulation activities while a simulation is running offers several advantages. Users can better understand the simulation, and observe trends that cannot be captured using average statistics (that are typically available only at the end of the simulation run). Further, visualization allows user to take immediate corrective measures on the model, instead of waiting until the simulation ends, if a modeling problem is observed. This latter advantage is particularly crucial for a distributed supply chain simulation since simulation of this scale takes relatively longer (execution time) to complete.

An approach of realizing the visualization support of a distributed supply chain simulation was proposed by Gan, Cai, Turner, and Xavier (2001). The visualization was implemented as a separate federate (one per corporation) that observes the activities of the simulation. Techniques such as dynamic information subscription and data distribution management were employed to control the amount of information being transmitted across the network to ensure the applicability of the visualization support.

3.2 Integration of Simulation Results into Real Systems

Flexible and seamless integration of new functionality suggested by simulation results into existing applications without losing release-capability can be achieved with novel component-based, scalable enterprise software systems that represent business processes in an object-oriented framework. One of the most advanced systems has been developed by SKYVA International (SKYVA, 2000).

Entirely Java-based technology ensures platform independence and allows application distribution. As changes have to be made only to the relevant business objects, it is not required that the entire connection to the external systems are changed or taken down. The new application code is generated automatically from the modeling framework. Any constraints can be defined by "rules". This allows to create individual “tailor-made” solutions without losing release-capability.

Extendable data models can be generated dynamically based on the requirements of the business process, using agent technology. This makes availability of “near-real-time” input data possible and allows seamless connection of both the simulation framework and execution applications to the underlying ERP or legacy systems. One of the leaders is agent technology is living systems (IDC, 2000).

4 INDUSTRY FEASIBILITY AND BENEFITS

The framework is feasible in industries which are subject to favorable characteristics as follows:

- A mass production environment is subject to both make/deliver-to-order characteristics (that require “available-to-promise” features) and make-to-stock characteristics (with stochastic uncertainties) in the supply chain.
- The bills of materials are not too complex and easy to configure.
- The logistics content in the value-added is significant.
- The non-repetitive labor content in the value added is low.
- Manufacturing activities are subject to little variance only, their parameterization in master data might be difficult but not impossible, therefore participation of the shopfloor at planning and scheduling is rather low.
- The number of customer orders to be handled is large.
- The need for optimization of sequence and capacity utilization in manufacturing is high.
- The flexibility regarding capacity adaptations (for example because of high capital costs) is low.

A good example for this kind of environment is the semiconductors/electronics industry. Most of the other mass production environments are subject to these characteristics as well.

The tremendous potential benefits of the application of this kind of framework across supply chains can be summarized as follows:

- More realistic optimization results can be obtained because APS algorithms are incorpo-
rated into the simulation, the dynamic behavior of the supply chain and stochastic uncertainties are taken into account, and (near-)real-time input data is used.

- **Collaborative supply chain optimization** becomes possible across globally distributed locations without having to disclose sensitive company data.
- **Fast optimizations**, and especially re-optimizations of supply chains can be obtained when they are needed rather than when they are already obsolete.
- **High flexibility** accounts for today’s frequent changes of business requirements and market places: Simulation models and applications can be changed very easily and the framework is not hampered by growth limitations, i.e. it is scalable.
- **Easy integration** of the framework into an existing landscape of multiple, diverse IT systems is possible with zero downtime and without the necessity of any reprogramming.

5 CONCLUSIONS AND OUTLOOK

A novel framework for ultra-fast distributed simulation with integrated APS procedures has been described. Clearly, this is a challenging undertaking. While the basic material flows are relatively simple, the web of decision-making, particularly the interactions between supply chain elements, can be quite complex. In this attempt to model supply chains as a federation of independent, interacting models, there are a number of open challenges to be addressed. The following list illustrates a few of these issues:

- What level of detail in each federate is appropriate for a given supply chain issue and environment?
- How can the interactions be designed to permit federate models with different levels of granularity to be used?
- How is the federation simulation process to be initiated, monitored, and regulated?
- What additional assumptions and infrastructure are required to support supply chains consisting of federates representing individual companies, who may be both collaborators and competitors?
- If a very large set of federates is created, is it possible to create federation instances with subset of federates or are all federates required always?

Each of the issues identified, as well as others, requires careful design and assessment of the federation framework and the specification of federate interactions.

The work described in this paper is ongoing in the collaboration between Gintic and Georgia Tech to develop the basic methodology and computational tools. A laboratory prototype federation representing a semiconductor supply chain has been developed and is currently being tested. It uses HLA to integrate independent models of waferfabs, an assembly and test facility as well as a number of distribution nodes. Other computational models such as stochastic production planning are also being addressed. The prototype federation incorporates federates located both in Atlanta and in Singapore.

Synchronization and time-management issues are being investigated in order to optimize federation simulation performance. Alternative visualization mechanisms and media, so that the federation simulation can be observed as it proceeds, are also being explored.

As the current work demonstrates a proof of concept, the nature of our collaborative efforts will change focus to address the potential impact in industry. In particular, the future work will engage one or more industrial partners to develop industrial prototypes and extend the business operations aspect of the framework to allow seamless integration of manufacturing and inbound/outbound logistics on a scheduling level.

ACKNOWLEDGMENTS

This work was funded by the Singapore National Science and Technology Board, and by the W. M. Keck Foundation through a grant to The Georgia Institute of Technology.

REFERENCES


**AUTHOR BIOGRAPHIES**

**PETER LENDERMANN** is a Research Fellow with the Manufacturing Planning & Scheduling Group at Gintic Institute of Manufacturing Technology in Singapore. Previously he was a Managing Consultant with agiConsult in Germany where his focus was on the areas of supply chain management and production planning. He also worked as a Research Associate at the European Laboratory for Particle Physics CERN in Geneva (Switzerland) and Nagoya University (Japan). He obtained a Diploma in Physics from the University of Munich (Germany), a Doctorate in Applied Physics from Humboldt-University in Berlin (Germany) and a Master in International Economics and Management from Bocconi-University in Milan (Italy). His research interests include advanced methods for supply chain planning as well as parallel and distributed simulation. His email address is <peterl@gintic.gov.sg>.

**BOON PING GAN** is an Associate Research Fellow with the Manufacturing Planning & Scheduling Group at Gintic Institute of Manufacturing Technology in Singapore. He received a Bachelor of Applied Science in Computer Engineering and Master of Applied Science from Nanyang Technological University of Singapore in 1995 and 1998 respectively. His research interests are parallel and distributed simulation, parallel programs scheduling, and application of genetic algorithms. His email address is <bpooi@gintic.gov.sg>.

**LEON McGINNIS** is Eugene C. Gwaltney Professor of Manufacturing Systems at Georgia Tech, where he also serves as Associate Director of the Manufacturing Research Center and founding Director of the Keck Virtual Factory Lab. His research focuses on the application of operations research and computer science to solve decision problems arising in the design and operation of industrial logistics systems. His email address is <leon.mcginnis@isye.gatech.edu>.