

General Motors Increases Its Production Throughput

Author(s): Jeffrey M. Alden, Lawrence D. Burns, Theodore Costy, Richard D. Hutton, Craig A. Jackson, David S. Kim, Kevin A. Kohls, Jonathan H. Owen, Mark A. Turnquist and David J. Vander Veen

Source: *Interfaces*, Jan. - Feb., 2006, Vol. 36, No. 1, 2005 Franz Edelman Award for Achievement in Operations Research and the Management Sciences (Jan. - Feb., 2006), pp. 6-25

Published by: INFORMS

Stable URL: <https://www.jstor.org/stable/20141357>

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/20141357?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Interfaces*

General Motors Increases Its Production Throughput

Jeffrey M. Alden

General Motors Corporation, Mail Code 480-106-359, 30500 Mound Road, Warren, Michigan 48090, jeff.alden@gm.com

Lawrence D. Burns

General Motors Corporation, Mail Code 480-106-EX2, 30500 Mound Road, Warren, Michigan 48090, larry.burns@gm.com

Theodore Costy

General Motors Corporation, Mail Code 480-206-325, 30009 Van Dyke Avenue, Warren, Michigan 48090, ted.costy@gm.com

Richard D. Hutton

General Motors Europe, International Technical Development Center, IPC 30-06, D-65423 Rüsselsheim, Germany, richard.d.hutton@gm.com

Craig A. Jackson

General Motors Corporation, Mail Code 480-106-359, 30500 Mound Road, Warren, Michigan 48090, craig.jackson@gm.com

David S. Kim

Department of Industrial and Manufacturing Engineering, Oregon State University, 121 Covell Hall, Corvallis, Oregon 97331, david.kim@orst.edu

Kevin A. Kohls

Soar Technology, Inc., 3600 Green Court, Suite 600, Ann Arbor, Michigan 48105, kohls@soartech.com

Jonathan H. Owen

General Motors Corporation, Mail Code 480-106-359, 30500 Mound Road, Warren, Michigan 48090, jonathan.owen@gm.com

Mark A. Turnquist

School of Civil and Environmental Engineering, Cornell University, 309 Hollister Hall, Ithaca, New York 14853, mat14@cornell.edu

David J. Vander Veen

General Motors Corporation, Mail Code 480-734-214, 30003 Van Dyke Road, Warren, Michigan 48093, david.vanderveen@gm.com

In the late 1980s, General Motors Corporation (GM) initiated a long-term project to predict and improve the throughput performance of its production lines to increase productivity throughout its manufacturing operations and provide GM with a strategic competitive advantage. GM quantified throughput performance and focused improvement efforts in the design and operations of its manufacturing systems through coordinated activities in three areas: (1) it developed algorithms for estimating throughput performance, identifying bottlenecks, and optimizing buffer allocation, (2) it installed real-time plant-floor data-collection systems to support the algorithms, and (3) it established common processes for identifying opportunities and implementing performance improvements. Through these activities, GM has increased revenue and saved over \$2.1 billion in over 30 vehicle plants and 10 countries.

Key words: manufacturing; performance/productivity; production/scheduling; applications.

Founded in 1908, General Motors Corporation (GM) is the world's largest automotive manufacturer. It has manufacturing operations in 32 countries, and its vehicles are sold in nearly 200 nations (General Motors Corporation 2005b). GM's automotive brands include Buick, Cadillac, Chevrolet, GMC, Holden, HUMMER, Opel, Pontiac, Saab, Saturn, and Vauxhall. In 2004, the company employed about 324,000 people world-

wide and sold over 8.2 million cars and trucks—about 15 percent of the global vehicle markets—earning a reported net income of \$2.8 billion on over \$193 billion in revenues (General Motors Corporation 2005a, b).

GM operates several hundred production lines throughout the world. In North America alone, GM operates about 100 lines in 30 vehicle-assembly plants (composed of body shops, paint shops, and gen-

eral assembly lines), about 100 major press lines in 17 metal fabrication plants, and over 120 lines in 17 engine and transmission plants. Over 400 first-tier suppliers provide about 170 major categories of parts to these plants, representing hundreds of additional production lines.

The performance of these production lines is critical to GM's profitability and success. Even though the overall production capacity of the industry is above demand (by about 25 percent globally), demand for certain popular vehicles often exceeds planned plant capacities. In such cases, increasing the throughput of production lines increases profits, either by adding sales revenue or by reducing the labor costs associated with unscheduled overtime.

The Need to Increase Throughput

In the late 1980s, competition intensified in the automotive industry. In North America, this competition was fueled by an increasing number of imports from foreign manufacturers, customers' growing expectations for quality, and slow growth of the overall market, which led to intense pricing pressures that limited GM's opportunities to increase its prices and revenues. In comparison with its foreign competitors, GM was seen as wasteful and unproductive, seemingly unable to improve, and ineffectively copying Japanese methods without understanding the real production problems. Indeed, in competitive comparisons, *The Harbour Report*, the leading industry scorecard of automotive manufacturing productivity, ranked GM near the bottom in terms of production performance (Harbour and Associates, Inc. 1992). Many plants were missing production targets, working unscheduled overtime, experiencing high scrap costs, and executing throughput-improvement initiatives with disappointing results, while their managers argued about how to best meet production targets and become more productive and cost effective. GM was losing money, even with products in high demand, and was either cutting costs to the penny or opening the checkbook to solve throughput problems. Managers were extremely frustrated and had no logical plan to guide them in improving plant throughput and controlling production costs.

Vehicle launches (that is, the plant modifications required to produce a new vehicle) also suffered

greatly: it took plants years, not months, to achieve throughput targets for new model lines, typically only after major, costly changes in equipment and layout. Long launches meant lost sales and expensive unplanned capital investment in machines, buffers, and space to increase throughput rates. Several factors contributed to inadequate line designs:

- Production data was unavailable or practically useless for throughput analysis, crippling engineers' ability to verify the expected performance of new lines prior to the start of production.

- Inadequate tools for analyzing manufacturing throughput, mostly discrete event simulation (DES), took days or weeks to produce results, during which engineers may have changed the line design.

- Intense corporate pressure to reduce the costs for line investment led to overly optimistic estimates of equipment reliability, scrap rates, rework rates, space requirements, material-flow capability, operator performance, and maintenance performance. These optimistic estimates were promoted by vendors who rated equipment performance based on ideal conditions.

Under these conditions, many plants had great difficulty achieving the optimistic throughput targets during launch and into regular production. GM had two basic responses: (1) increase its pressure on managers and plants to improve the throughput rates with the existing investment, or (2) invest additional capital in equipment and labor to increase throughput rates—often to levels even higher than the original target rates to pay for these unplanned investments.

The confluence of these factors severely hindered GM's ability to compete; as a result, GM closed plants and posted a 1991 operating loss of \$4.5 billion (General Motors Corporation 1992)—a record at the time for a single-year loss by any US company.

Responding to the Need

Recognizing the need and the opportunity to improve plant productivity, GM's research and development (R&D) organization began a long-term project to improve the throughput performance of existing and new manufacturing systems through coordinated efforts in three areas:

- Modeling and algorithms,
- Data collection, and

—Throughput-improvement processes.

The resulting technical advances, implementation activities, and organizational changes spanned a period of almost 20 years. We list some of the milestones in this long-term effort:

- 1986–1987 GM R&D developed its first analytic models and software (C-MORE) for estimating throughput and work-in-process (WIP) inventory for basic serial production lines and began collecting data and defining a standard throughput-improvement process (TIP).
- 1988 GM's Linden body fabrication plant and its Detroit-Hamtramck general assembly plant implemented pilots of C-MORE and TIP.
- 1989–1990 GM R&D developed a new analytic decomposition model for serial lines, extending C-MORE to accurately analyze systems with unequal workstation speeds and to provide approximate performance estimates for parallel lanes.
- 1991 A GM R&D implementation team deployed C-MORE at 29 North American GM plants and car programs, yielding documented savings of \$90 million in a single year.
- 1994–1995 GM North America Vehicle Operations and GM Powertrain formed central staffs to coordinate data collection and consolidate throughput-improvement activities previously performed by divisional groups.
- 1995 GM R&D developed an activity-based network-flow simulation, extending C-MORE to analyze closed carrier systems and lines with synchronous transfer mechanisms.
- 1997 GM expanded the scope of implementation beyond improving operations to include designing production systems.
- 1999 GM Global Purchasing and Supply Chain began using C-MORE tools for supplier-development activities.
- 2000 GM R&D developed a new hybrid simulation system, extending C-MORE to

analyze highly complex manufacturing systems within GM.

- 2003 GM Manufacturing deployed C-MORE tools and throughput-improvement processes globally.
- 2004 GM R&D developed an extensible callable library architecture for modeling and analyzing manufacturing systems to support embedded throughput analysis within various decision-support systems.

Modeling and Algorithms

In its simplest form, a production line is a series of stations separated by buffers (Figure 1), through which parts move sequentially until they exit the system as completed jobs; such serial lines are commonly the backbones of many main production lines in automotive manufacturing. Many extensions to this simple serial-line model capture more complex production-system features within GM. Example features include job routing (multiple passes, rework, bypass, and scrap), different types of processing (assembly, disassembly, and multiple parts per cycle), multiple part types, parallel lines, special buffering (such as resequencing areas), and conveyance systems (chains, belts, transfer bars, return loops, and automated guided vehicles (AGVs)). The usual measure of the throughput of a line is the average number of parts (jobs) completed by the line per hour (JPH).

Analyzing even simple serial production lines is complicated because stations experience random failures and have unequal speeds. When a single station fails, its (finite) input buffer may fill and block upstream stations from producing output, and its output buffer may empty and starve downstream stations of input. Speed differences between stations can also cause blocking and starving. This blocking and starving creates interdependencies among stations, which make predicting throughputs and identifying bottlenecks very difficult. Much research has been done on the throughput analysis of production lines (Altiok 1997, Buzacott and Shantikumar 1993, Chen and Chen 1990, Dallery and Gershwin 1992, Gershwin 1994, Govil and Fu 1999, Li et al. 2003, Papadopoulos and Harvey 1996, Viswandham and Narahari 1992).

Although analysts naturally wish to construct very detailed discrete-event simulation models to ensure

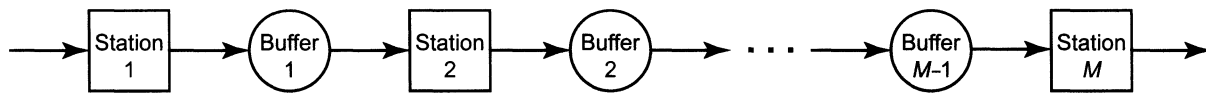


Figure 1: A simple serial (or tandem) production line is a series of stations separated by buffers. Each unprocessed job enters Station 1 for processing during its work cycle. When the work is complete, each job goes into Buffer 1 to wait for Station 2, and it continues moving downstream until it exits Station M . The many descriptors of such lines include station speed, station reliability, buffer capacity, and scrap rates.

realistic depiction of all operations in a production system, such emulations are difficult to create, validate, and transport between locations. The number of analyses and the time they take make such simulation models impractical. In analyzing production systems, a major challenge is to create tools that are fast, accurate, and easy to use. The trade-offs between these basic requirements motivated us in developing decomposition-based analytic methods and custom discrete-event-simulation solvers that we integrated with advanced interfaces into a suite of throughput-analysis tools called C-MORE.

While the specific assumptions we made in modeling and analysis depended on the underlying algorithm used for estimating performance, we generally assumed that workstations are unreliable, with deterministic operating speeds and exponentially distributed times between successive failures and times to repair. We also assumed that all buffers have finite capacity and that the overall system is never starved (that is, jobs are always available to enter) or blocked (that is, completed jobs can always be unloaded). Among the performance measures C-MORE provides are the following outputs:

- The throughput rate averaged over scheduled production hours, which we use to predict performance, to plan overtime, and to validate the model based on observed throughput rates;
- System-time and work-in-process averages that provide lead times and material levels useful for scheduling and planning;
- The average state of the system, which includes blocking, starving, processing, and downtimes for each station and average contents for each buffer;
- Bottleneck identification and analysis, which helps us to focus improvement efforts in areas that will indeed increase throughput and to assess opportunities to improve throughput;

—Sensitivity analysis to estimate the impact of varying selected input parameters over specified ranges of values, to account for uncertainty in input data, to identify key performance drivers, and to evaluate suggestions for system improvements; and

—Throughput distribution (per hour, shift, or week) to help managers to plan overtime and to assess system changes from the perspective of throughput variation (less variation is desirable).

The basic trade-off in the C-MORE tools and analysis capabilities is between the complexity of the system that can be modeled and the execution speed of the analysis (Figure 2). Because we always need the fastest possible, accurate analysis, what performance-estimation method we choose depends on the complexity of the manufacturing system of interest. C-MORE's basic analysis methods are analytic decomposition and discrete simulation.

Analytic-Decomposition Methods

The analytic solvers we developed represent the underlying system as a continuous flow model and rely on a convergent iterative solution of a series of nonlinear equations to generate performance estimates. These methods are very fast, which enables automatic or user-guided exploration of many alternatives for system design and improvement. Furthermore, tractable analytic models require minimal input data (for example, workstation speed, buffer size, and reliability information).

Except in the special case of identical workstations, no exact analytic expressions are known for throughput analysis of serial production lines with more than two workstations. However, heuristics exist; they are typically based on decomposition approaches that use the analysis of the station-buffer-station case iteratively as the building block to analyze longer lines (Gershwin 1987). The fundamental challenge is

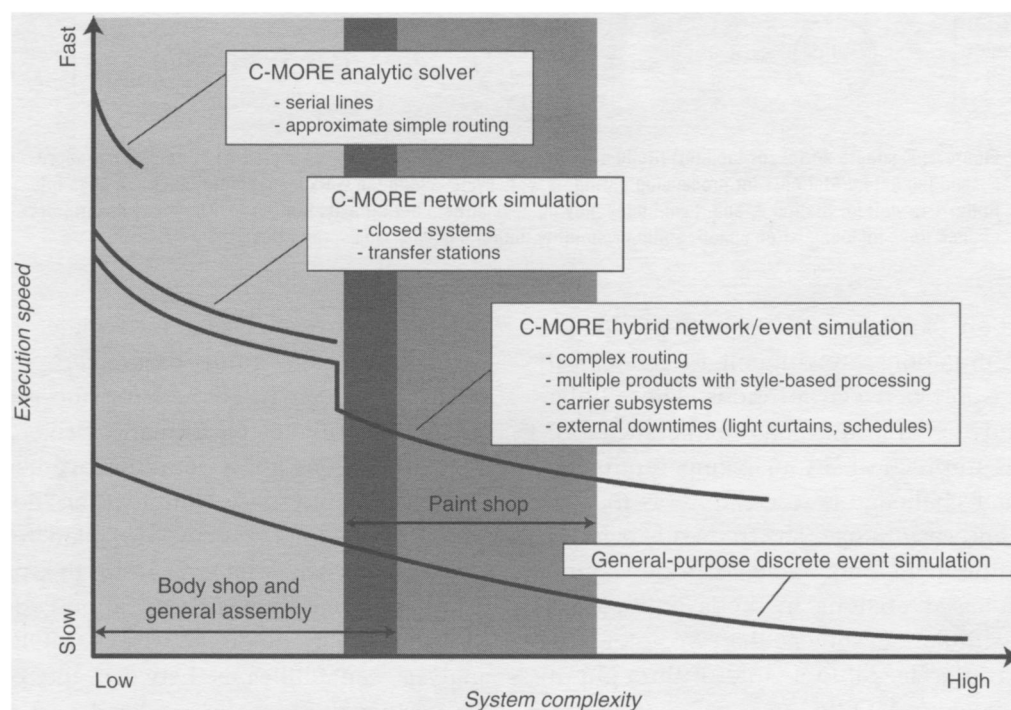


Figure 2: The basic trade-off among the C-MORE tools and general-purpose discrete-event simulation (DES) is between the complexity of the system that can be modeled and the execution speed of analysis. For typical systems, the C-MORE analytic solver produces results almost instantaneously, while C-MORE simulations usually execute between 50–200 times faster than comparable DESs. Because execution speed is important, throughput engineers typically use the fastest analysis tool that is capable of modeling the system of interest. For example, in the context of GM's vehicle-assembly operations, the analytic solver and network simulation tools are most suitable for the production lines found in body shops and general assembly, while the complexity of paint shop operations requires use of the hybrid network/event simulation or general-purpose DES.

the analysis of the two-station problem. Our analysis approach to the two-station problem for our situation gives a closed-form solution (Alden 2002). We used this approach in the analytic throughput solver of C-MORE.

We based our analysis on a work-flow model approximation of the two-station system in which stations act like gates on the flow of work and buffers act like reservoirs holding work. This paradigm allows us to use differential equations instead of the more difficult difference equations. We assumed that station speeds are constant when processing under ideal conditions (no blocking, starving, or failures). We also assumed that each station's operating time between failures and its repair time are both distributed as negative exponential random variables. We used operating time instead of elapsed time, because most

stations rarely fail while idle. We assumed that all variables are independent and generally unequal. We made one special simplifying assumption to handle mutual station downtime: when one station fails and is not working, the remaining station does not fail but processes at a reduced speed equal to its stand-alone throughput rate (its normal operating speed multiplied by its stand-alone availability). This assumption simplifies the analytic analysis by allowing convenient renewal epochs (for example, at station repairs) while still producing accurate results. We believe that we can relax this last assumption and still produce a closed-form solution—a possible topic of future work.

Our analysis approach focuses on the steady-state probability distribution of buffer contents at a randomly selected time of station repair. Under the (correct) assumption that this distribution is the weighted

sum of two exponential terms plus two impulses at zero and full buffer contents, we can derive and solve a recursive expression by considering all possible buffer-content histories between two consecutive times of station repair (Appendix). Given this distribution, we can derive closed-form expressions for throughput rate, work in process, and system time. These expressions reduce to published results for special cases, for example, equal station speeds (Li et al. 2003). Extensive empirical observation and simulation validations have shown this approach to be highly accurate, with errors typically within two percent of observed throughput rates for actual systems at GM that have many stations. Large prediction errors in practice are usually caused by poor quality data inputs; in these cases, we often use C-MORE to help us to identify the stations likely to have data-collection problems by using sensitivity analysis and comparisons with historical results. We obtain performance estimates for typical GM systems almost instantaneously and bottleneck and sensitivity analyses in a few seconds on a desktop PC.

Discrete Simulation Methods

The analytic method we described and other similar methods provide very accurate performance estimates for simple serial lines. Unfortunately, they are not applicable to GM's most complex production systems (for example, its paint shops). These production systems employ splits and merges with job-level routing policies (for example, parallel lanes and rework systems), use closed-loop conveyances (such as pallets or AGVs), or are subject to external sources of downtime that affect several disjoint workstations simultaneously (for example, light-curtain safety devices for nonadjacent operations). To analyze such features with the available analytic-decomposition methods, one must use various modeling tricks that yield approximate and sometimes misleading results.

General-purpose discrete-event-simulation (DES) software can analyze complex production systems with a high degree of accuracy. It takes too long, however, to develop and analyze these models for extensive what-if scenario analysis. GM needs such extensive scenario analysis to evaluate large sets of design alternatives and to explore continuous improvement opportunities. To overcome the limitations of the available DES software and the modeling

deficiencies of existing analytic methods, we developed and implemented new simulation algorithms for analyzing GM's production systems.

In these simulation algorithms, we used activity-network representations of job flow and efficient data structures (Appendix). We used these representations to accurately simulate the interaction of machines and the movement of jobs through the system, without requiring all events to be coordinated through a global time-sorted queue of pending actions. We implemented two simulation algorithms as part of the C-MORE tool set (Figure 2); one is based entirely on an activity-network representation, while the other augments such a representation with the limited use of an event queue to support analysis of the most complex manufacturing systems. Our internal benchmarking of these methods demonstrated that their analysis times are typically 50 to 200 times faster than those of comparable DES methods and their results are identical. Chen and Chen (1990) reported dramatic savings in computational effort for similar simulation approaches.

With the C-MORE simulation tools, we provide modeling constructs designed specifically for representing GM's various types of production systems. Examples include "synchronous line" objects to represent a series of adjacent workstations that share a single transfer mechanism to move jobs between the workstations simultaneously, convenient mechanisms to maintain the sequence order of jobs within parallel lanes, and routing policies to enforce restrictions at splits and merges. With such modeling constructs, the C-MORE simulation tools are much easier to use than general-purpose DES software for modeling GM's production systems. Although C-MORE's usability comes at the expense of flexibility, it provides two important benefits: First, end users spend far less time on model development than they would with commercial DES software (minutes or hours versus days or weeks); GM Powertrain Manufacturing, for example, reported that simulation-analyst time per line-design project has dropped by 50 percent since it began using these tools in 2002. Second, the use of common modeling constructs facilitates communication and transportability of models among different user communities within the company; a benefit that is especially important in overcoming organizational and geographic boundaries between global regions.

Delivery Mechanism

To facilitate use of the C-MORE analysis tools (both analytic and simulation approaches), we designed and developed an extensible, multilayer software architecture, which serves as a framework for deploying the C-MORE tools (Figure 3). We implemented it as a set of callable libraries to provide embedded access to identical modeling and analysis capabilities for any number of domain-specific end-user software applications. This approach promotes consistency across the organization and allows disparate user groups with divergent interests to access identical analysis capabilities while sharing and reusing system models. As a concrete example, the throughput predictions obtained by engineers designing a future manufacturing line are consistent with the results obtained later by plant personnel using the line's actual performance data to conduct continuous-improvement activities, even though the two user groups access the C-MORE tools through separate user interfaces.

The C-MORE software architecture provides a modeling isolation layer, which decouples system mod-

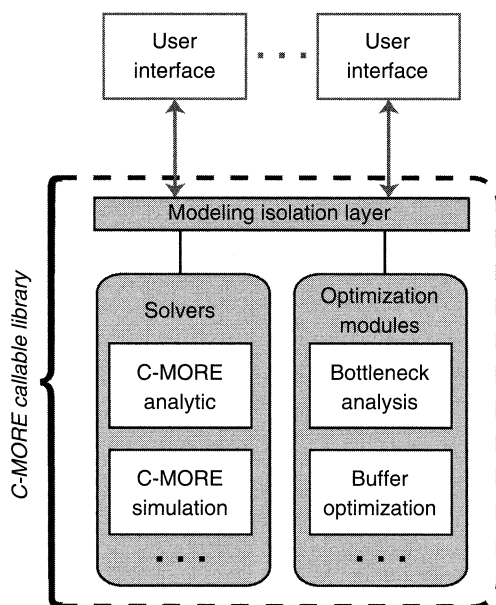


Figure 3: The C-MORE callable library decouples system modeling from analysis and provides embedded throughput-analysis capabilities for multiple, domain-specific user interfaces. The software architecture design supports future extensibility as new analysis techniques are developed and modeling scope is broadened.

eling from analysis capabilities (much like AMPL decouples mathematical modeling from such solvers as CPLEX or OSL). The C-MORE software architecture provides analysis capabilities through “solvers” and “optimization modules.” We use the term *solver* to refer to a software implementation of an algorithm that can analyze a model of a manufacturing system to produce estimates of system performance (for example, throughput, WIP levels, and station-level blocking and starving). Solvers may employ different methods to generate results, including analytic and simulation methods. We use the term *optimization module* to refer to a higher-level analysis that addresses a specific problem of interest to GM, typically by executing a solver as a subroutine within an iterative algorithmic framework. For example, C-MORE includes a bottleneck-analysis module that identifies the system components that have the greatest impact on system throughput. Other examples of optimization modules include carrier analysis, which determines the ideal number of pallets or carriers to use in a closed-loop system, and a buffer-optimization module, which heuristically determines the distribution of buffer capacity throughout the system according to a given objective, such as, maximum throughput. Each of these optimization modules comprises an abstract modeling interface and one or more algorithm implementations. Generally, these optimization modules can use any available solver to estimate throughput performance.

Because its modeling and analysis capabilities are decoupled, the C-MORE software architecture allows users to reuse system models with different analysis approaches. In addition, the extensible architecture design facilitates rapid deployment of new analysis capabilities and provides a convenient mechanism for developing and testing new approaches.

Data Collection

Collecting production-line data is a difficult and massive task. A line may have hundreds of workstations, each with its own processing times and operating parameters. Furthermore, models with many data inputs tend to be complex and time consuming to analyze. Hence, for practical use, it is important to develop models and algorithms with modest data

requirements that still produce meaningful results. To determine the appropriate data inputs, we conducted repeated trials and validation efforts that eventually defined the minimal data requirements as a function of production-line characteristics (routing, scrapping, failure modes, automation, conveyance systems, and so on).

In our first pilot deployments, we had almost no useful data to support throughput analyses. The data available was often overwhelming in quantity, inaccurate, stale, or lacking in internal consistency. GM had not established unambiguous standards that defined common measures and units. This lack of useful data presented a tremendous obstacle. Good data collection required ongoing tracking of changes in workstation speed, scrap counts, and causes of workstation stoppages (equipment failures, quality stops, safety stops, part jams, feeder-line starving, and so forth). We gradually instituted standardized data-collection practices to address these issues. Early on, we collected and analyzed well-defined data manually at several plants to prove its value in supporting throughput analysis. This success helped us to justify manual data-collection efforts in other plants for the purpose of increasing throughput rates. It also helped us to understand what data we needed and obtain corporate support for implementing automatic data collection in all GM plants.

Automatic data collection presented further challenges. GM's heritage as a federation of disparate, independently managed businesses meant that its plants employed diverse equipment and technologies. In particular, plant-floor control technologies varied widely from plant to plant, and standard interfaces did not exist. We could not deploy data-collection tools and techniques we developed at one plant to other plants without modifying them extensively. Likewise, we could not readily apply the lessons we learned at one plant to other plants. We needed common equipment, control hardware and software, and standard interfaces. GM formed a special implementation team to initiate data collection and begin implementation of throughput analysis in GM's North American plants. This activity continues today, with established corporate groups that develop and support global Web-accessible automatic data collection. These groups also support throughput analysis in all

plants by training employees, analyzing data, and providing on-site expertise.

Automatic Data Collection, Validation, and Filtering

The automated machines and processes in automotive manufacturing facilities include robotic welders, stamping presses, numerically controlled (NC) machining centers, and part-conveyance equipment. Industrial computers called programmable logic controllers (PLCs) provide operating instructions and monitor the status of production equipment. To automate data collection, we implemented PLC code (software) that counts production events and records their duration; these events include machine faults and blocking and starving events. Software summarizes event data locally and transfers them from the PLC into a centralized relational database. We use separate data-manipulation routines to aggregate the production-event data into workstation-performance characteristics, such as failure and repair rates, and operating speeds. We then use the aggregated data as input to C-MORE analyses to validate models, to identify bottlenecks, and to improve throughput.

We validate and filter data at several stages, which helps us to detect missing and out-of-specification data elements. In some cases, we can correct these data anomalies manually before we analyze the plant-floor data with C-MORE. We compare the C-MORE throughput estimates to actual observed throughput to detect errors in the model or data. We pass all data through the various automatic filters before using them to improve throughput. GM collects plant-floor data continually, creating huge data sets (typically from 15,000 to 25,000 data points per shift, depending on plant size and the scope of deployment). Therefore, we must regularly archive or purge historical performance data.

In general, PLCs can detect a multitude of event types; we had to identify which data are meaningful in estimating throughput. The C-MORE analysis tools helped GM to determine which data to collect and standard ways to aggregate the data. GM's standardization of PLC hardware and data-collection techniques has helped it to implement C-MORE and the throughput-improvement process (TIP) throughout North America, and it makes its ongoing globalization efforts possible.

Historical Data for Line Design

As we collected data, we developed a database of actual historical station-level performance data. This database, which is organized by equipment type (make and model) and purpose, has become an integral part of GM's processes for designing new production lines. Previously, GM's best sources of performance estimates for new equipment were the machine specifications equipment vendors provided. Because vendors typically base their ratings on assumptions of ideal operating conditions, the ratings are inherently biased and often unsuitable as inputs to throughput-analysis models. GM engineers now have a source of performance estimates that are based on observations of similar equipment in use within actual production environments. Using this historical data, they can better predict future performance characteristics and analyze proposed production-line designs realistically, setting better performance targets and improving GM's ability to meet launch timetables and throughput targets.

Throughput-Improvement Processes

In our pilot implementations of throughput analysis in GM during the late 1980s, we faced several cultural challenges. First, we had difficulty assembling all the right people and getting them to focus on a single throughput issue unless it was an obvious crisis. Because those who solved large crises became heroes, people sought out and tackled the big problems quickly, rather than working on the little persistent problems. These little problems often proved to be the real throughput constraints, and identifying and quantifying them requires careful analysis. Second, our earliest implementation efforts appeared to be yet another set of not-invented-here corporate program-of-the-month projects, to be tolerated until they fizzled out. As a result, we had difficulty getting the support and commitment we needed for successful implementation. Finally, plant personnel lacked confidence that a black-box computer program could really understand the complexities of a production line.

We responded to these problems in several ways. We prepared solid case studies and excellent training materials to prove the value of throughput analysis,

and we developed a well-defined process the plant would own to support and reward ongoing throughput improvements. To this last point, we developed and taught a standard TIP to promote the active and effective use of labor, management, data collection, and throughput analysis to find and eliminate throughput bottlenecks. The TIP consisted of the following steps:

- Identify the problem: Typically, it's the need to improve throughput rates.

- Collect data and analyze the problem: Use throughput analysis to find and analyze bottlenecks.

- Create plans for action: The planners are usually multidisciplinary teams of engineers, line operators, maintenance personnel, suppliers, managers, and finance representatives.

- Implement the solution: Assign responsibilities and obtain commitment to provide implementation resources.

- Evaluate that the solution: Ensure that the solution is correctly implemented and actually works.

- Repeat the process.

To further cultural transformation, we distributed free copies of *The Goal* (Goldratt and Cox 1986) to personnel at all plants. We conducted classes in the basic theory of constraints (TOC) to help plant personnel understand, from a very fundamental perspective, the definition and importance of production bottlenecks, why they can be difficult to identify, and why we needed C-MORE to find them.

With the TIP in place and the support of plant managers and personnel, plants soon improved throughput and discovered new heroes—the teams that quickly solved persistent throughput problems and helped the plants meet their production targets. To further reward and promote the use of throughput analysis and TIP, GM gave a popular "Bottleneck Buster" award to plants that made remarkable improvements in throughput.

After plants improved throughput of existing production lines, we applied the process in a similar fashion to designing new production lines. We established formal workshops, especially in powertrain (engine and transmission) systems, to iteratively find and resolve bottlenecks using the C-MORE tools and a cross-functional team of experts.

An Early Case Study—The Detroit-Hamtramck Assembly Plant

In 1988, GM conducted a pilot implementation of C-MORE and TIP at the Detroit-Hamtramck assembly plant, quickly achieving substantial throughput gains and cost savings. This experience serves as a good example of improvements that can be realized through the use of these tools.

In the late 1980s, GM considered the Detroit-Hamtramck plant a model for its future, with hundreds of training hours per employee and vast amounts of new (and untested) advanced manufacturing technology throughout the facility. It was designed to produce 63 JPH with no overtime; yet the plant could barely achieve 56 JPH even with overtime often exceeding seven person-hours per vehicle. The associated costs of lost sales and overtime exceeded \$100,000 per shift. Rather than increase efficiency, the plant's advanced manufacturing technology actually hindered improvement efforts because the added operational complexity made it difficult for plant workers and managers to determine where to focus improvement efforts. Under intense pressure to achieve its production targets, the plant worked unscheduled overtime and conducted many projects to increase throughput rates. Unfortunately, most of these initiatives produced meager results because they focused on local improvements that did not affect overall system throughput. The lack of improvement caused tremendous frustration and further strained the relationships among the labor union, plant management, and corporate management.

In 1988, GM researchers presented their new throughput-analysis tools (an early version of C-MORE) to the Detroit-Hamtramck plant manager, who in turn asked a plant-floor systems engineer to visit the GM Research Labs to learn more. Through interactions with GM researchers, the engineer learned about bottlenecks, throughput improvement, and C-MORE's analysis capabilities. He also read the book they recommended, *The Goal* (Goldratt and Cox 1986), to learn more about continuous improvement.

Back at the plant, he started collecting workstation performance data. After several days, he had gathered enough data to model and analyze the system

using C-MORE, and he found the top bottleneck operation: the station installing the bulky headliner in the ceiling of each car. The data showed this station failing every five cycles for about one minute. Although every vehicle had to pass through this single operation, managers had never considered this station a significant problem area because its downtime was small and no equipment actually failed. The engineer investigated the operation and found that the operator worked for five cycles, stopped the line, went to get five headliners, and then restarted the line. The downtime was measured at slightly less than one minute, and it occurred every five cycles. Further investigation revealed that the underlying problem was lack of floor space. The location of a supervisor's office prevented the delivery and placement of the headliners next to the line. Ironically, the plant had previously considered relocating the office but had delayed the move because managers did not consider it a high priority. Based on this first C-MORE analysis, the engineer convinced the planner to move the office, despite its low priority. On the Monday after the move, the headliners were delivered close to the line and the line stops ended. Not surprisingly, the throughput of the whole plant increased.

Up to this point, the prevailing opinion had been that plant operations were so complex that one could not predict cause-and-effect relationships, and the best one could do was to fix as many things as possible and hope for the best. Now GM could hope to systematically improve plant throughput.

We repeatedly collected data and addressed bottlenecks until the plant met its throughput targets consistently and cut overtime in half (Figure 4). The plant's achievement was recognized and documented in *The Harbour Report*, which placed the Detroit-Hamtramck plant in the top 10 most-improved vehicle-assembly plants with a 20 percent improvement in productivity (Harbour and Associates, Inc. 1992).

Because of this success, plant operations changed. First, the language changed; such terms as *bottlenecks*, *stand-alone availability* (SAA), *mean cycles between failures* (MCBF), and *mean time to repair* (MTTR) became part of the vernacular. Armed with the C-MORE analysis, the plant could prioritize repairs and changes

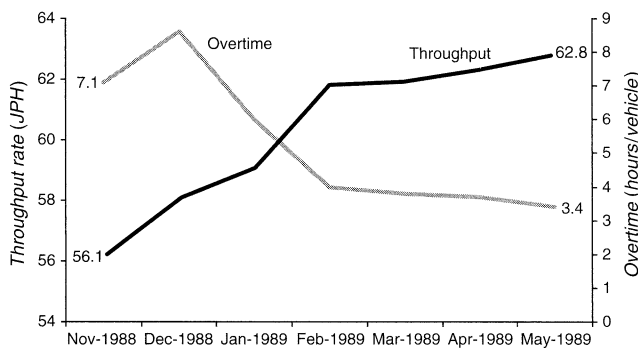


Figure 4: Within six months of using C-MORE in the Detroit-Hamtramck assembly plant in November 1988, we found and removed bottlenecks, increased throughput by over 12 percent, attained the 63 jobs-per-hour (JPH) production target, and cut overtime hours per vehicle in half.

more effectively than ever before. Accurate data collection exposed the true performance characteristics of workstations, and the finger pointing that caused distrust in the plant gave way to teamwork, with one area of the plant volunteering resources to another to help it to improve the bottleneck station.

Building on this successful pilot, the Detroit-Hamtramck plant continued its productivity improvement efforts in the ensuing years and achieved results that were described as “phenomenal” in *The Harbour Report North America 2001* (Harbour and Associates, Inc. 2001).

Implementation and the Evolution of the Organization

Even after early pilot successes, it took a special effort and top management support to implement throughput analysis widely within GM. In 1991, with a vice president as eager champion, we formed a special implementation-project team. This team consisted of 10 implementation members and 14 supporting members to help with management, consultation, training, programming, and administrative work. Our stated goal was to “enable GM’s North American plants to save \$100 million in 1991 through the implementation of C-MORE.” We trained team members and stationed them in 15 of the most critical plants. They collected data, modeled and analyzed systems using C-MORE to find the top throughput bottlenecks, ran TIP meetings, tracked performance, and provided

overall training. The results were remarkable: plants achieved dramatic throughput increases (over 20 percent was common), reduced their overtime hours, and often met their production targets for the first time. Our most conservative accounting of dollar impact for this effort was over \$90 million in a single year (1991) through C-MORE and TIP implementations for 29 GM plants and car programs. This work proved the value and feasibility of throughput analysis.

To build on the success of the implementation project team, central management established a dedicated staff to spread the use of C-MORE and TIP to other plants. Thus, in 1994, it formed a central organization within GM North America Vehicle Operations to continually help plants with all aspects of throughput analysis, education, and implementation. Over time, this group, which grew to almost 60 members, developed common C-MORE data-collection standards as part of all new equipment going into plants and produced Web-enabled bottleneck reports for use across the corporation. It developed training courses and custom games to teach employees and motivate them to use throughput-improvement tools and concepts. Improvement teams identified and attacked launch problems before new equipment started running production parts. Integral to this activity, C-MORE is now a standard corporate tool for improving plant throughput and designing production lines. To support its widespread use, GM undertook several activities:

- GM integrated C-MORE into multiple desktop applications to provide various users easy access to throughput-analysis capability.

- GM provided formal training and software to over 400 global users.

- GM formed a C-MORE user’s group to share developments, lessons, and case studies across the organization and summarize results in a widely distributed newsletter.

- GM established common methods and standards for data collection to support throughput analysis and its rapid implementation worldwide.

- GM established a “Bottleneck Buster” award to recognize plants that make remarkable improvements in throughput.

We are continuing to expand the use of the C-MORE tools and TIP in GM’s global operations.

The C-MORE tools and data provide a common platform and language that engineers, managers, and analysts across GM's worldwide operations use to develop and promote common best-manufacturing practices. Standard data definitions and performance measures facilitate communication among GM's geographically dispersed operations and enable the fast transfer of lessons learned among manufacturing facilities, no matter where they are located. C-MORE underpins GM's new global manufacturing system (GMS), which it is currently deploying throughout its worldwide operations.

Because of the success GM realized internally using C-MORE and TIP, it formed a separate corporate organization to conduct supplier-development initiatives. Through its activities, GM can improve the manufacturing capabilities of external part suppliers by helping them to design and improve their production systems. Previously, GM plants had to react to unexpected constraints on part supplies (often during critical new vehicle launches); now GM engineers collaborate with suppliers to proactively develop and analyze models of suppliers' manufacturing systems. With GM's assistance, key suppliers are better able to design systems that will meet contractual throughput requirements. As a result, GM's suppliers become more competitive, and GM avoids the costs and negative effects of part shortages. In one recent example, 12 (out of 23) key suppliers for a new vehicle program had initially designed systems that were incapable of meeting throughput-performance requirements; GM discovered this situation and helped the suppliers to resolve it prior to the vehicle launch period, avoiding costly delays. Although their impact is difficult to quantify, these supplier-development initiatives become increasingly important as GM relies more heavily on partnerships with external suppliers.

Impact

The use of the C-MORE tools, data, and the throughput-improvement process pervades GM, with applications to improve manufacturing-system operations, production-system designs, supplier development, and production-systems research. The impact of work in these areas is multifaceted. It has yielded measured productivity gains that reduced costs and increased

revenue, and it has changed GM's processes, organization, and culture.

Profits

Use of the C-MORE tools and the TIP for improving production throughput alone yielded over \$2.1 billion in documented savings and increased revenue, reflecting reduced overtime and increased sales for high-demand vehicles. Although it is impossible to measure cost avoidance, internal users estimate the additional value of using the C-MORE tools in designing manufacturing systems for future vehicles to be several times this amount; they have improved decisions about equipment purchases and GM's ability to meet production targets and reduce costly unplanned investments. We have not carefully tracked the impact on profits of using the C-MORE tools in developing suppliers.

Competitive Productivity

The use of C-MORE and the TIP has improved GM's ability to compete with other automotive manufacturers. The remarkable improvements GM has achieved have been recognized by *The Harbour Report*, the leading industry scorecard of automotive manufacturing productivity. When GM first began widespread implementation of C-MORE in the early 1990s, not one GM plant appeared on the Harbour top-10 list of most productive plants. In fact, of the 15 least productive plants measured by Harbour from 1989 through 1992, 14 were GM plants. During this same time period, *The Harbour Report* estimate for GM's North American manufacturing capacity utilization was 60 percent, and its estimate of the average number of workers GM needed to assemble a vehicle was more than 50 percent higher than its largest domestic competitor, Ford Motor Company (Harbour and Associates, Inc. 1992). This unbiased external assessment of GM's core manufacturing productivity clearly attested to its need to improve throughput.

In the ensuing years, GM focused continually on increasing throughput, improving vehicle launches, and reducing unscheduled overtime. C-MORE enabled these efforts, helping GM to regain a competitive position within the North American automotive industry. GM has made steady, year-after-year improvements in manufacturing productivity

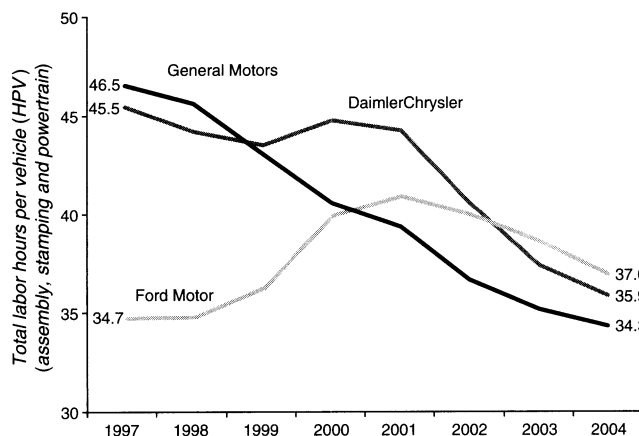


Figure 5: Productivity improvements in General Motors' North American manufacturing operations have outpaced those of its two major domestic competitors, DaimlerChrysler and Ford Motor Company. From 1997 through 2004, GM's total labor hours per vehicle improved by over 26 percent (Harbour and Associates, Inc. 1999–2003, Harbour Consulting 2004).

since 1997 (Figure 5). Its earlier use of C-MORE also led to measurable improvements, but direct comparison with prior years is problematic because Harbour Consulting changed its measurement method in 1996.

According to the 2004 *Harbour Report*, GM's total labor hours per vehicle dropped by over 26 percent since 1997. GM now has four of the five most productive assembly plants in North America and sets the benchmark in eight of 14 vehicle segments (Harbour Consulting 2004). The estimate of overall vehicle-assembly capacity utilization is now 90 percent. In addition, GM's stamping operations lead the industry in pieces per hour (PPH), and GM's engine operations lead the US-based automakers. C-MORE and TIP, combined with GM's new global manufacturing system (GMS) and other initiatives, have enabled GM to make these outstanding improvements.

Standard Data Collection

Common standards for data definitions, plant-floor data-collection hardware and software, and data-management tools and practices have made the otherwise massive effort of implementing and supporting throughput analysis by GM engineers and in GM plants a well-defined, manageable activity. Plant-floor data collection and management is now a global effort, with particular emphasis in Europe and the Asia-Pacific region.

Line Design

GM engineers often design new production lines in cross-functional workshops whose participants try to minimize investment costs while achieving quality and throughput goals. These workshops use C-MORE routinely because they can quickly and easily develop, analyze, and comprehend models. Using C-MORE, they can evaluate hundreds of line designs for each area of a plant, whereas in the past they considered fewer than 10 designs because of limited data and analysis capability. Also, GM now has good historical data on which to base models. Although the impact of line designs on profit is more difficult to measure than the impact of plant improvements, GM users estimate that line designs that meet launch performance goals have two to three (sometimes up to 10) times the profit impact of throughput improvements made after launch. This impact is largely because of better market timing for product launches and less unplanned investment during launch to achieve throughput targets. By using C-MORE, GM has obtained such savings in several recent product launches, achieving record ramp-up times for the two newest engine programs in GM's Powertrain Manufacturing organization.

Plant Operations and Culture

GM's establishment of throughput analysis has fostered important changes in its manufacturing plants. First, they use throughput sensitivity analysis and optimization to explore, assess, and promote continuous-improvement initiatives. They can use it to resolve questions (and avoid speculation) about buffer sizes and carrier counts.

Throughput analysis forces plants to consider all stations of a line as integral components of a single system. This perspective, together with the cross-functional participation in TIP meetings, reinforces a focus on system-level performance. With many local improvement opportunities competing for limited resources (for example, engineers, maintenance, and financial support), C-MORE and TIP help the plants concentrate their efforts on those initiatives that improve bottleneck performance and have the greatest impact on overall system throughput. This change greatly reduces the frustration of working to get project resources, only to find the improvements

did little to improve throughput; with throughput analysis, such efforts find quick support and provide satisfying results.

Tailored throughput reports improve scheduling of reactive and preventive maintenance by ranking areas according to their importance to throughput. These reports also enhance plant suggestion programs; managers can better evaluate and reward suggestions for improving throughput, acting on the better-targeted suggestions and obtaining higher overall improvements. They thus motivate employees to seek and prevent problems rather than reacting only to the problem of the moment.

The plants have integrated throughput analysis in their organizations by designating throughput engineers to oversee data collection, to conduct throughput analyses, to help with problem solving, and to support TIP meetings. They have made this commitment even though they are under pressure to reduce indirect labor costs.

The C-MORE users group (active for several years) and the central staff that support it collect and communicate best practices for throughput analysis and data collection among the plants. They thus further support the continuous improvement of plant operations and capture lessons learned for use by many plants.

Organizational Learning

The organization has learned many lessons during its effort to institute throughput analysis. While some of these lessons may seem straightforward from a purely academic perspective—or to an operations research (OR) professional experienced in manufacturing-systems analysis—they were not apparent to the organization's managers and plant workers. Developing an understanding of these concepts at an organizational level was challenging and required paradigm shifts in many parts of the company; in some cases, the lessons contradicted active initiatives. We conducted training sessions at all levels of the organization, from management to the plant floor. This training took many forms, from introductory games that illustrate basic concepts of variability and bottlenecks while demonstrating the need for data-driven analysis, to formal training courses in the tools and processes. Here are some lesson topics:

- Understanding what data is needed and (more important) what data is not needed.

- Understanding our data-collection capabilities and identifying the gaps.

- The importance of accurate and validated plant-floor data for analyzing operations improvements and for designing new systems.

- The importance of basing decisions on objective data and analysis, not on beliefs based on isolated experiences.

- The impact of variability on system performance and the importance of managing buffers to protect system performance.

- Reducing repair times to improve throughput and decrease variability.

- Understanding that extreme applications of the lean mantra (zero buffers) are harmful.

- Understanding that bottleneck operations aren't simply those with the longest downtimes or lowest stand-alone availability.

- The need for robust techniques to find true bottlenecks.

- Understanding that misuse of simulation analysis can foster poor decisions.

- The importance of consistency in modeling and in the transportability of models and data.

- The use of sensitivity analysis to understand key inputs and modeling assumptions that work well in practice.

- The impact of early part-quality assessment on cost and throughput.

The organizational changes made during the course of establishing throughput analysis included the formation of central groups to serve as "homerooms" for throughput knowledge and to transfer best practices and lessons learned among plants. Ultimately, GM's inclusion of throughput analysis in its standard work and validation requirements testify to its success in learning these lessons.

Stimulation of Manufacturing-Related Operations Research Within GM

The success of throughput analysis in GM has exposed plant and corporate staffs to OR methods and analysis. General Motors University, GM's internal training organization, has had over 4,500 enrollments in its eight related courses on TOC, the C-MORE analysis tools, data collection, and TIP.

Employees' growing knowledge has greatly facilitated the support and acceptance of further manufacturing-related OR work in GM. Since this project began, GM has produced over 160 internal research publications concerning improving plant performance. GM researchers have developed and implemented analysis tools in related areas, including maintenance operations, stamping operations, plant-capacity planning, vehicle-to-plant allocation, component make-buy decisions, and plant layout decisions. These tools generally piggyback on C-MORE's reputation, data collection, and implementations to gain entry into plants and line-design workshops. Many of these tools actually use embedded throughput calculations from C-MORE through its callable library interface. The cumulative impact on profit of the additional OR work enabled by our throughput-analysis work is an indirect benefit of the C-MORE effort.

Concluding Remarks

Improving GM's production throughput has required a sustained effort for over 15 years and has yielded tangible results that affect its core business—manufacturing cars and trucks. Today, we continue this effort on a global scale in both plant operations and line design with strong organizational support. The benefits include reduced costs and increased revenues. Our work has established standardized, automatic collection of data, fast and easy-to-use throughput-analysis tools, and a proven and supported throughput-improvement process. Success in this work results from a focused early implementation effort by a special task team followed by establishment of a central support staff, plant ownership of throughput-analysis capability with designated throughput engineers, strong management support and recognition of successes, and an ongoing collaboration between Research and Development, central support staffs, and plant personnel.

GM's need for fast and accurate throughput-analysis tools gave us exciting opportunities to apply and extend state-of-the-art results in analytic and simulation approaches to throughput analysis and optimization and to conduct research on many topics related to the performance of production systems. GM recognizes that much of this work provides it with an important competitive advantage in the auto-

motive industry. In addition, our work has fostered an increased use of OR methods to support decision making in many other business activities within GM.

From a broader perspective, our project represents a success story of the application of OR and highlights the multifaceted approach often required to achieve such success in a large, complex organization. The impact of such projects goes beyond profit. At GM, this effort has advanced organizational processes, employee education, plant culture, production systems research and related initiatives, and it has greatly increased the company's exposure to OR.

Appendix

An Analytic Solution for the Two-Station Problem

The solution of the two-station problem can be used as a building block to analyze longer lines using a decomposition approach that GM developed in the late 1980s and recently disclosed (Alden 2002). We first describe the solution of the two-station problem with unequal speeds and then discuss the solution of longer lines. We begin with a job-flow approximation of the two-station problem under the following assumptions and notation:

ASSUMPTION 1. *Jobs flow through the system so that for each fraction of a work cycle completed in processing a job, the same fraction of the job moves through the station and enters its downstream buffer.*

ASSUMPTION 2. *The intermediate buffer, on its own, does not delay job flow, that is, it does not fail and has zero transit time. It has a finite and fixed maximum job capacity of size B . (We increase buffer size by one to account for the additional buffering provided by in-station job storage in the representation of the job flow.)*

ASSUMPTION 3. *Station i 's ($i = 1, 2$) processing speed S_i is constant under ideal processing conditions, that is, when not failed, blocked, or starved.*

ASSUMPTION 4. *While one station is down, the other station does not fail, but its speed drops to its normal speed times its stand-alone availability to roughly account for possible simultaneous downtime. We thus simplify the analysis with little loss of accuracy.*

ASSUMPTION 5. *Station i 's processing time between failures (not the elapsed time between failures) is an exponentially distributed random variable with a fixed failure*

rate λ_i . This means a station does not fail while idle, that is, when it is off, failed, under repair, blocked, or starved.

ASSUMPTION 6. Station i 's repair time is an exponentially distributed random variable with a fixed repair rate μ_i .

ASSUMPTION 7. The first station is never starved and the second station is never blocked.

ASSUMPTION 8. All parameters (station speed, buffer size, failure rates, and repair rates) are positive, independent, and generally unequal.

ASSUMPTION 9. The system is in steady state, that is, the probability distribution of the system state at a randomly selected time is independent of the initial conditions and of the time.

For now, we also assume that the processing speed of Station 1 is faster than that of Station 2, that is, $S_1 > S_2$. While both stations are operational, the finite-capacity buffer will eventually fill because $S_1 > S_2$. While the buffer is full, Station 1 will be idle at the end of each cycle until Station 2 finishes its job and loads its next job from the buffer. This phenomenon is called *speed blocking* of Station 1. (A similar *speed starving* of Station 2 occurs when $S_1 < S_2$.) Speed blocking briefly idles Station 1; as stations typically do not fail during idle time, we reduce the failure rate λ_1 by the ratio S_2/S_1 in the flow-model representation during periods of speed blocking.

We focus on the state of the system at times of a station repair, that is, when both stations become operational. At such times, the state of the system is completely described by the buffer contents (ranging from 0 to B), and the associated state probability distribution is given by

P_0 = probability the buffer content is zero (empty) when a station is repaired,

P_B = probability the buffer content is B (full) when a station is repaired, and

$p(x)$ = probability density that the buffer content is x for $0 \leq x \leq B$ when a station is repaired, excluding the finite probabilities (impulses) at $x = 0$ and $x = B$.

Let $\hat{p}(x)$ denote the entire distribution given by $(P_0, p(x), P_B)$ so that $\hat{p}(x)$ is the mixed probability distribution composed of $p(x)$ plus the impulses of areas P_0 and P_B located at $x = 0$ and $x = B$, respectively. In a

similar manner, let $\hat{q}(x)$ or $(q(x), Q_B)$ be the full probability state distribution at a station failure (note for $S_1 > S_2$, $Q_0 = 0$).

Under our assumptions, the system has convenient renewal times at station repairs, that is, when the probability distributions of the system's state are identical. We break the evolution of the distribution of buffer contents between subsequent times of station repairs into parts: (a) evolution from station repair to a station failure, and (b) evolution from station failure to its repair (Figure 6).

For $S_1 > S_2$, there are five general cases of buffer-content evolution from station repair to subsequent failure (Figure 7) and seven general cases of buffer-content evolution from station failure to subsequent repair (Figure 8). Assumption 4 greatly simplifies this case analysis by reducing the infinite number of possible station failure-and-repair sequences before both stations are operational to a manageable finite number.

By considering all five cases of buffer-content evolution from repair to failure, we can express $\hat{q}(x)$ in terms of $\hat{p}(x)$. By considering all seven cases of buffer-content evolution from failure to repair and the renewal property of station repairs, we can derive $\hat{p}(x)$ in terms of $\hat{q}(x)$. Alden (2002) showed that $\hat{q}(x)$ can be removed by substitution to develop a recursive equation for $p(x)$ that suggests it may be the weighted sum of three exponential terms (actually only two are required—the first term drops out). Assuming this is the case, we can solve the recursive equations to obtain

$$p(x) = a_2 P_0 e^{\alpha_2 x} + a_3 P_0 e^{\alpha_3 x}, \quad \text{where}$$

$$a_2 = \frac{(\alpha_2 + r_2)(\alpha_2 + r)}{\alpha_2 - \alpha_3}, \quad a_3 = \frac{(\alpha_3 + r_2)(\alpha_3 + r)}{\alpha_3 - \alpha_2},$$

$$r = \frac{\lambda_1 + \lambda_2}{S_1 - S_2}, \quad r_1 = \frac{\mu_1(\mu_2 + \lambda_2)}{S_2 \mu_2}, \quad r_2 = \frac{\mu_2(\mu_1 + \lambda_1)}{S_1 \mu_1},$$

$$\alpha_2 = -\frac{r - r_1 + r_2}{2}$$

$$+ \sqrt{\left(\frac{r - r_1 + r_2}{2}\right)^2 + \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)r(r_1 + r_2)}, \quad \text{and}$$

$$\alpha_3 = -\frac{r - r_1 + r_2}{2}$$

$$- \sqrt{\left(\frac{r - r_1 + r_2}{2}\right)^2 + \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)r(r_1 + r_2)}.$$

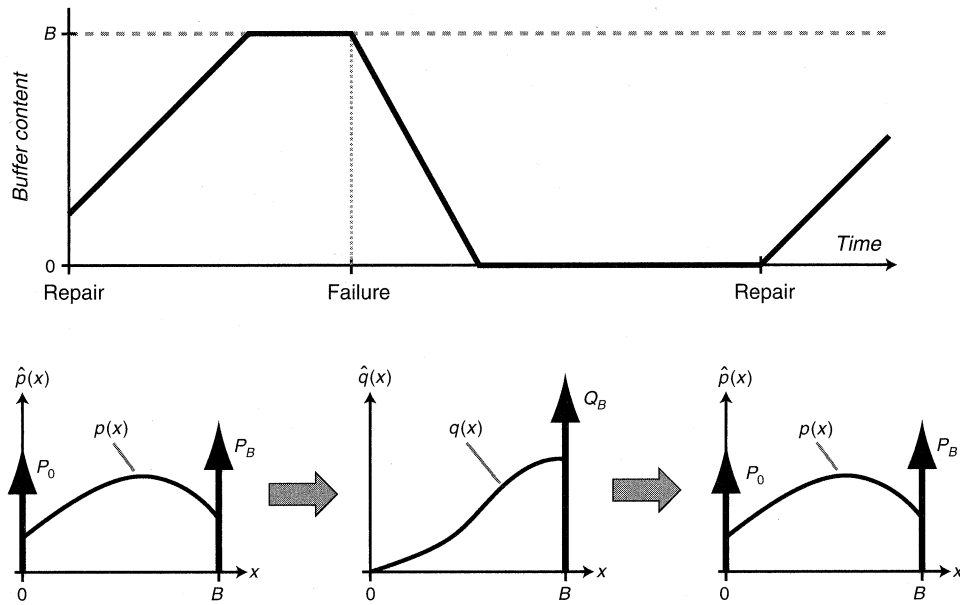


Figure 6: The probability distribution of buffer contents at (any) station repair $\hat{p}(x)$ evolves to become the probability distribution of buffer contents at the subsequent failure of any station $\hat{q}(x)$ and continues to evolve into the same distribution of buffer contents $\hat{p}(x)$ when the station is repaired—a renewal. In this particular example, Station 1 was repaired, and then it failed, and subsequently it was repaired again.

Given $p(x)$, we can derive the equations

$$P_B = \frac{(\hat{\theta}_2(\alpha_2 + r_2) + \theta_2 r)e^{\alpha_2 B} - (\hat{\theta}_2(\alpha_3 + r_2) + \theta_2 r)e^{\alpha_3 B}}{\hat{\theta}_1(\alpha_2 - \alpha_3)} P_0$$

and

$$P_0 = (\hat{\theta}_1 \alpha_2 \alpha_3 (\alpha_2 - \alpha_3)) \cdot (\hat{\theta}_1 r r_2 (\alpha_2 - \alpha_3) + \alpha_3 ((\alpha_2 + r_2)(\alpha_2 + \hat{\theta}_1 r) + \theta_2 \alpha_2 r) e^{\alpha_2 B} - \alpha_2 ((\alpha_3 + r_2)(\alpha_3 + \hat{\theta}_1 r) + \theta_2 \alpha_3 r) e^{\alpha_3 B})^{-1}, \text{ where}$$

$$\theta_2 = \frac{\lambda_2}{\lambda_1 + \lambda_2}, \quad \hat{\theta}_1 = \frac{S_2 \lambda_1}{S_2 \lambda_1 + S_1 \lambda_2}, \quad \text{and}$$

$$\hat{\theta}_2 = \frac{S_1 \lambda_2}{S_2 \lambda_1 + S_1 \lambda_2}.$$

Given $\hat{p}(x)$, we can derive closed-form expressions for the various system states, throughput rate, work-in-buffer (WIB) and work-in-process (WIP), and system time (Alden 2002).

We can easily derive results for the case $S_1 < S_2$ from the above results by analyzing the flow of spaces for jobs that move in the opposite direction of job flow (a duality concept). With this understanding, we can quickly derive the following procedure to analyze the

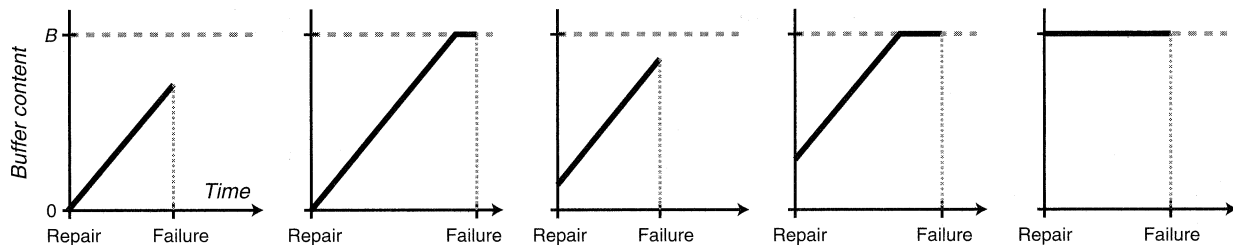


Figure 7: For $S_1 > S_2$, there are five cases of buffer-content evolution from a station repair to a subsequent failure of any station.

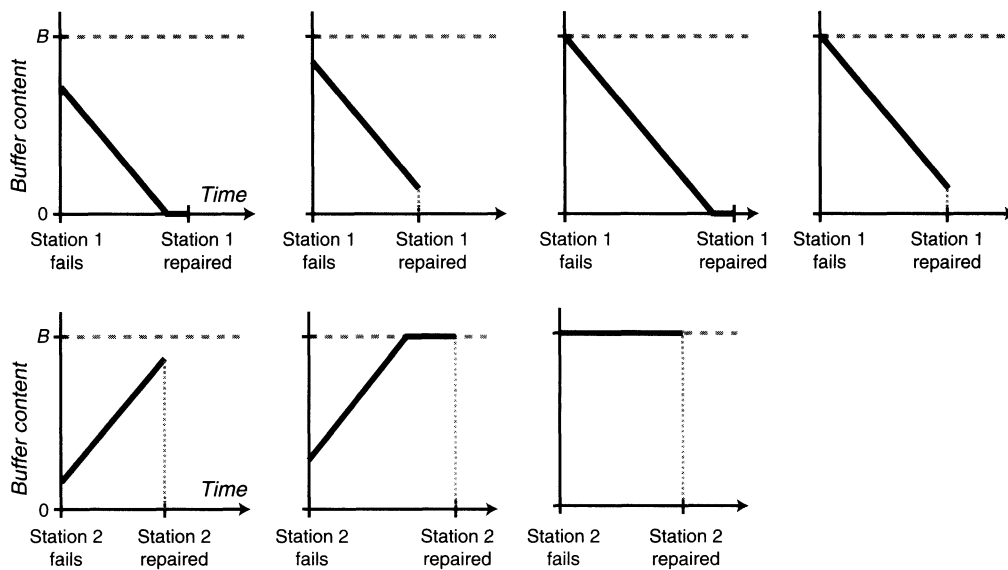


Figure 8: For $S_1 > S_2$, there are seven cases of buffer-content evolution from a station failure to its subsequent repair.

case $S_1 < S_2$:

- (1) Swap S_1 , λ_1 , and μ_1 for S_2 , λ_2 , and μ_2 , respectively.
- (2) Calculate the results using the equations for the case $S_1 > S_2$.
- (3) Swap buffer state probabilities as follows: blocking with starving, full with empty, and filling with emptying.
- (4) Swap the expected buffer contents of the flow model, $E[\text{WIB}]$, with $B - E[\text{WIB}]$.

We can use this analysis approach (or L'Hospital's rule) to derive results for the case of equal speeds, $S_1 = S_2$. Also, a search for possible zero divides reveals one case, when $\alpha_2 = 0$, for which we can use L'Hospital's rule to derive useful results (Alden 2002).

We also developed a method of using the two-station analysis as a building block to analyze longer lines: we analyzed each buffer in an iterative fashion while summarizing all stations upstream and all stations downstream of the buffer as two separate aggregate stations; this allowed us to use the results of the two-station analysis. We derived aggregate station parameters that satisfied conservation of flow, expected aggregate station-repair rate, and expected aggregate station-failure rate. These expected rates considered the actual station within an aggregate sta-

tion closest to a buffer and the smaller "nested" aggregate station two stations away from the same buffer. We calculated the resulting equations (in aggregate parameters), while iteratively passing through all buffers until we satisfied a convergence criterion. Convergence is not guaranteed and occasionally is not achieved, even with a small step size. However, it is very rare that convergence is too poor to provide useful results. Gershwin (1987) describes this general approach, called decomposition, to derive expressions for aggregate station parameters (repair rate, failure rate, and speed), which we extended to handle stations with unequal speeds. Due to the similarity of approaches, we omit the details of our particular solution.

Activity-Network Representation of Job Flow

An activity network is a graphical representation of events or activities (represented as nodes) and their precedence relationships (represented with directed arcs). Muth (1979) and Dallery and Towsley (1991) used a network representation of job flow through a production line to prove the reversibility and symmetry properties of production lines. These networks are sometimes referred to as sample path diagrams.

To analyze serial production lines, the C-MORE network simulation uses an activity network with a

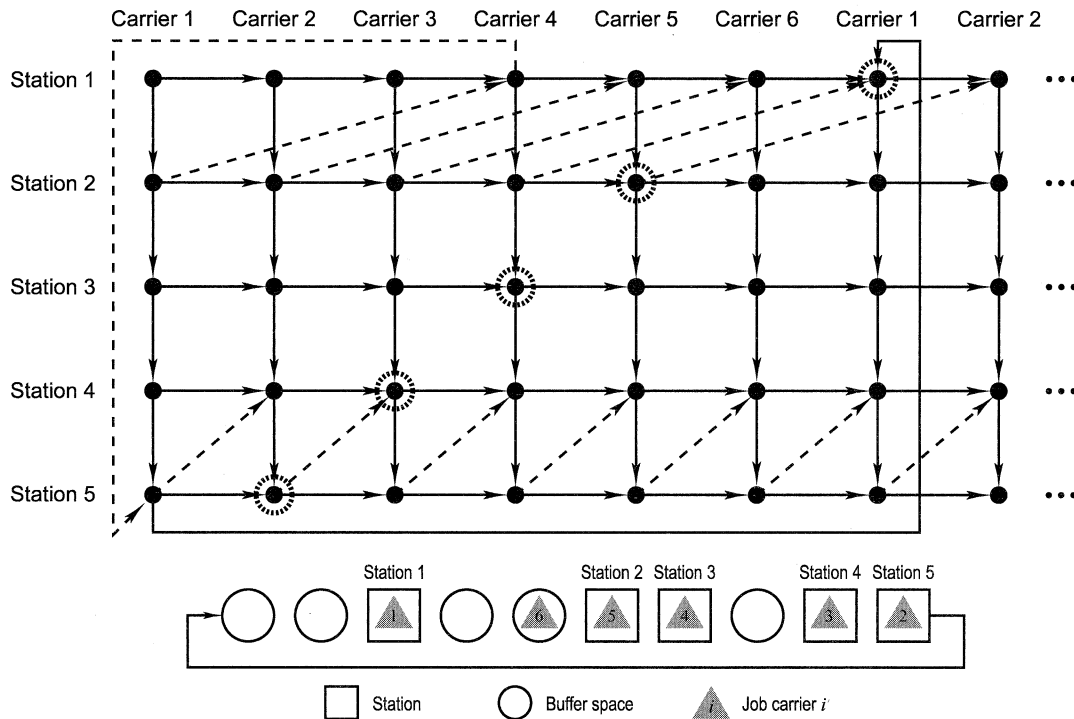


Figure 9: In the activity-network representation (top) of a closed-loop serial production line (bottom), each node represents a job being processed by a workstation. Each row of nodes corresponds to the sequence of jobs processed by a specific station, and each column of nodes represents a single job (moved by a specific carrier) as it progresses through all five stations. The arcs in the network indicate precedence relationships between adjacent nodes (for clarity, we have omitted some arcs from the diagram). Solid arcs represent the availability of jobs (vertical arcs) and the availability of workstations (horizontal arcs). Dashed arcs represent time dependency due to potential blocking between neighboring workstations, where the number of columns spanned by an arc indicates the number of intermediate buffer spaces between the workstations. We can calculate the time when a job enters a workstation as the maximum time among nodes with arcs leading into the specific node corresponding to that workstation. The update order for information in the activity network depends on the initial position of job carriers (dashed circles).

simple repeatable structure, where each node represents a station processing a job (Figure 9). Conceptually, the nodes in the network are arranged in a grid, so that each row of nodes corresponds to the sequence of jobs processed by a specific workstation and each column of nodes corresponds to a specific job as it progresses through the series of workstations. The arcs in the activity network represent the time-dependent precedence of events. Within a given row, an arc between nodes in adjacent columns represents the time dependency of a specific workstation's availability to process a new job upon that workstation's earlier completion of the previous job. Within a given column, an arc between nodes in adjacent rows represents the time dependency of a specific job's availabil-

ity to be processed by a new workstation upon that job's completion at the upstream workstation. A diagonal arc between nodes in adjacent rows and different columns represents the dependency of an upstream station's ability to release a job upon downstream blocking, where the number of columns spanned by the arc indicates the number of intermediate buffer spaces. We use the time dependencies represented by these arcs to analyze the dynamic flow of jobs through the production line.

For C-MORE, the activity-network representation provides the conceptual foundation for a very efficient simulation that avoids using global time-sorted event queues and other overhead functions found in traditional discrete-event simulation. Although we

can view the hybrid simulation algorithm in C-MORE as a generalization of the network-simulation method, in our implementation, we use very different representations of the internal data structures for the activity network, which improve computational performance when combined with a global event queue for managing the time-synchronized events.

References

- Alden, J. M. 2002. Estimating performance of two workstations in series with downtime and unequal speeds. GM technical report R&D-9434, Warren, MI.
- Altioik, T. 1997. *Performance Analysis of Manufacturing Systems*. Springer, New York.
- Buzacott, J. A., J. G. Shantikumar. 1993. *Stochastic Models of Manufacturing Systems*. Prentice Hall, Englewood Cliffs, NJ.
- Chen, J. C., L. Chen. 1990. A fast simulation approach for tandem queueing systems. O. Balci, R. P. Sadowski, R. E. Nance, eds. *Proc. 22nd Conf. Winter Simulation*. IEEE Press, Piscataway, NJ, 539–546.
- Dallery, Y., S. B. Gershwin. 1992. Manufacturing flow line systems: A review of models and analytical results. *Queueing Systems* 12(1–2) 3–94.
- Dallery, Y., D. Towsley. 1991. Symmetry property of the throughput in closed tandem queueing networks with finite buffers. *Oper. Res. Lett.* 10(9) 541–547.
- General Motors Corporation. 1992. *1991 Annual Report*. Detroit, MI.
- General Motors Corporation. 2005a. *2004 Annual Report*. Detroit, MI.
- General Motors Corporation. 2005b. Company profile. Retrieved February 25, 2005 from http://www.gm.com/company/corp_info/profiles/.
- Gershwin, S. B. 1987. An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking. *Oper. Res.* 35(2) 291–305.
- Gershwin, S. B. 1994. *Manufacturing Systems Engineering*. Prentice Hall, Englewood Cliffs, NJ.
- Goldratt, E. M., J. Cox. 1986. *The Goal: A Process of Ongoing Improvement*. North River Press, Croton-on-Hudson, NY.
- Govil, M. C., M. C. Fu. 1999. Queueing theory in manufacturing: A survey. *J. Manufacturing Systems* 18(3) 214–240.
- Harbour and Associates, Inc. 1992. *The Harbour Report: Competitive Assessment of the North American Automotive Industry 1989–1992*. Troy, MI.
- Harbour and Associates, Inc. 1999. *The Harbour Report 1999—North America*. Troy, MI.
- Harbour and Associates, Inc. 2000. *The Harbour Report North America 2000*. Troy, MI.
- Harbour and Associates, Inc. 2001. *The Harbour Report North America 2001*. Troy, MI.
- Harbour and Associates, Inc. 2002. *The Harbour Report North America 2002*. Troy, MI.
- Harbour and Associates, Inc. 2003. *The Harbour Report North America 2003*. Troy, MI.
- Harbour Consulting. 2004. *The Harbour Report North America 2004*. Troy, MI.
- Li, J., D. E. Blumenfeld, J. M. Alden. 2003. Throughput analysis of serial production lines with two machines: Summary and comparisons. GM technical report R&D-9532, Warren, MI.
- Muth, E. J. 1979. The reversibility property of production lines. *Management Sci.* 25(2) 152–158.
- Papadopoulos, H. T., C. Harvey. 1996. Queueing theory in manufacturing systems analysis and design: A classification of models for production and transfer lines. *Eur. J. Oper. Res.* 92(1) 1–27.
- Viswanadham, N., Y. Narahari. 1992. *Performance Modeling of Automated Manufacturing Systems*. Prentice Hall, Englewood Cliffs, NJ.