INT. J. PROD. RES., 2001, VOL. 39, NO. 14, 3183–3193

Peak management

B. RONEN†*, A. COMAN‡ and E. SCHRAGENHEIM§

Peaks occur when a firm accustomed to operating under market constraint conditions occasionally encounters peaks in market demand that temporarily exceed the firm's supply capacity. This paper defines the demand peak, classifies peak management (PM) conditions and prescribes a framework for PM. The Theory-of-Constraints (TOC) is expanded to handle the PM concept, introducing the two-policy concept: one policy for peak periods and another for non-peak periods. The paper outlines techniques to elevate the firm's peak performance by loading during the peak period. Distinct pricing policies for peak and off-peak periods are presented and a comprehensive methodology is suggested to deal with the PM problem, including labour policy, outsourcing policy and pricing policy. The paper demonstrates the methodology and techniques in a broad range of real-life management situations. Under TOC principles, the present analysis shows the different decision rules for the two periods and their substantial impact on the firm's overall strategy. Finally, the paper discusses the ramifications of operating under two distinct sets of rules, as well as the rules for making the transitions between an off-peak and a peak period.

1. Introduction

A peak period is when demand increases temporarily, returning later to the normal level. It is neither a ripple or noise, nor is it permanent. Peak management (PM), one of those important issues that are seldom discussed in the literature, is analysed here through the tools and philosophies of the Theory-of-Constraints (TOC) (Goldratt 1997), and a PM methodology is prescribed relating to the philosophies, concepts and techniques on which it based.

Capacity planning and management is a major component of the operations management literature, in particular when demand fluctuations are significant. Cox and Spencer (1998) analysed key problem areas at Trane Company and remarked: 'The first area of concern is the seasonality in demand for the final product. Building construction is seasonal in many parts of the country. The Master Production Scheduling is developed with the expectation that a leveled production rate can be produced over a relatively long period of time.... The peak tends to occur at the end of the summer and during the fall time period, while the valley tends to be in the winter and early spring.'

The APICS Dictionary (Cox and Blackstone 1998) defines idle capacity as 'The capacity generally not used in a system of linked resources. Idle capacity consists of protective capacity and excess capacity.' Protective capacity is defined as 'A given amount of extra capacity at non-constraints above the system constraint's capacity,

Revision received January 2001.

† Faculty of Management, Tel Aviv University, Tel Aviv 69978, Israel.
‡ Holon Academic Institute of Technology, Holon, Israel.
§ Elyakim Management Systems Ltd., Ra'anana, Israel.
* To whom correspondence should be addressed. e-mail: boaz_r@netvision.net.il

International Journal of Production Research ISSN 0020–7543 print/ISSN 1366–588X online © 2001 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00207540110054588

used to protect against statistical fluctuation.' Excess capacity is defined as 'A situation where the output capabilities at a non-constraint resource exceed the amount of productive and protective capacity required to achieve a given level of throughput at the constraint.'

This analysis of PM is based on an expansion of the TOC methodology. The essence of the TOC view is to simplify the entire system, which consists of the organization as a whole, its clients and suppliers. This is accomplished by focusing on exploitation of a very small number of critical factors called constraints, demanding subordination to these constraints from the rest of the system and elevating their throughput (Goldratt 1986, Ronen and Starr 1990). We extend the TOC view by distinguishing between two different sets of constraints that prevail in the same system at different times. Most of the time, during the so-called off-peak period, the system constraints lie with market demand. Occasionally, market demand peaks, the system enters the peak period and resource constraints emerge. The implications of recognizing the regular interchange between the two sets of constraints are so great that both global strategy and tactics across all facets of the organization's activity must be adapted accordingly (figure 1).

PM is relevant to any type of organization, as every organization has customers whose needs it endeavours to supply. It is assumed that the market demand, imposed by the customers, is not flat: there are times when demand is relatively high and times when it is relatively low. Though most organizations operate under conditions of excess capacity (market constraints), even they face peak periods when demand exceeds capacity.

According to the Pareto principle, for many organizations 80% of the throughput is generated in 20% of the time, namely the peak period. Consider the following examples.

- A takeaway pizzeria in downtown Manhattan. While oven and counter capacities exceed demand most of the time, the reverse is true during the lunch hour, when crowds of people from neighbouring offices need to be fed. Though clearly most of the revenues are created then, the pizzeria becomes a bottleneck at this time and some customers are lost due to the waiting time ('actual lost sales'). In addition, owing to time pressure, sales personnel miss the opportunity to suggest drinks or desert, resulting in the loss of potential revenues from served customers ('potential lost sales').

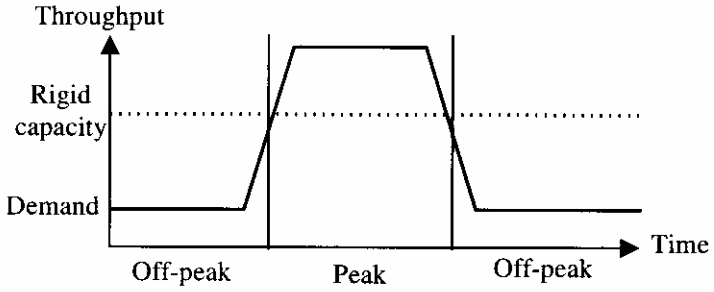

Figure 1. Peak and off-peak periods.

- The transportation system exhibits the same phenomenon: while subway, taxi and limousine services are in a state of over-capacity most of the time, customers vie for service during rush hours.
- *Fortune* (1999), discussing securities exchange peaks, describes the bathtub curve of how volume typically flows over the course of a day: 'heavy in the morning (everyone gets in); flat during the day (lunch); heavy at the close'.
- A soft drink can manufacturer faces peak demand during the summer. Demand during peak months exceeds the maximum capacity of the company by 30%.
- Demand for electricity in residential areas during the evening follows the same pattern.
- Broadcasting networks face heavy demand for advertising during prime time and significantly lower demand during the rest of the day.
- In supermarkets and banks, cashiers become bottlenecks for <20% of the business hours.
- Some businesses operate only for short periods: cinema buffets, refreshment sellers at sports events, etc. During their business hours they become throughput constraining bottlenecks.
- The demand for emergency services varies significantly. Ambulances and fire engines are idle most of the time and action peaks during emergencies. Such systems are idle by definition.

Figure 2 describes the conflict facing firms operating in peak conditions. For the firm to reach its objective of maximizing value to shareholders (A), it must at the same time (B) streamline a structured process and (C) take advantage of opportunities. In order to effect a streamlined structured process, the company must define one clear uniform policy (D) and in order to take advantage of opportunities the firm must adopt improvized *ad-hoc* solutions (D′).

The underlying assumptions are as follows.

- B to A: clear definition helps the company define, measure, manage and continually improve its business processes.
- C to A: business opportunities should be taken advantage of.

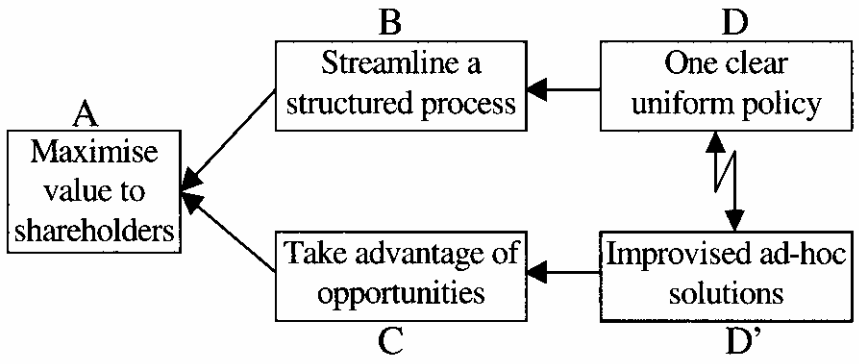

Figure 2. Peak management conflict.

- D to B: one clear uniform policy is best for implementation throughout the company.
- D′ to B: improvized *ad-hoc* solutions enable fast response to opportunities and crises.

The conflict is resolved through differentiation (mainly breaking the D to B assumption). In many cases there are two well-defined and expected situations: the peak and non-peak situations. We suggest a methodology to manage the two periods using two different policies to maximize performance, thus maximizing the value to shareholders. The literature usually deals with similar problems by determining the capacity needed for certain situations.

Section 2 classifies the types of peaks. Section 3 expands the TOC concept to the PM problem. Section 4 defines rules for the off-peak period. Section 5 establishes a methodology to manage the peak period. Section 6 prescribes how to manage the transition between off-peak and peak periods. Section 7 examines the impact of PM on global strategy and different pricing policies. Section 8 concludes and suggests further research.

2. Classifying peak conditions

We now classify peak conditions into four categories by the predictability of their occurrence and the organization's preparedness for them.

2.1. *Peak occurrence*

Peaks can be either expected or unpredictable.

2.1.1. *Expected peaks*

Expected peaks, whose occurrence can be predicted quite reliably, can be the result of a business initiative, e.g. an advertised promotional campaign, or the result of temporal cycles. Daily peaks occur in traffic rush hours. Variations in restaurant, cafe and bar occupancy occur during the morning, lunch and evening. Banks receive trading instructions in the morning. Theatre and cinema patronage peak in the evening as does television prime time. Typical weekend peaks are pressure to deposit pay cheques, which clogs banks, Saturday Night Fever, and weekend demand on hotel rooms. Quarterly peaks occur when firms are anxious to deliver goods in order to improve their quarterly reports. Annual peaks occur as a result of weather: beaches, ski resorts, annual tax reporting, summer school break, etc. The jewellery and gift industries face peaks during the holidays. The fashion industry predicts its seasons reasonably well. Greater cycles occur as a result of presidential elections, etc.

2.1.2. *Unpredictable peaks*

Unpredictable peaks may be the result of competitor campaign initiatives, stochastic variation in the frequency of calls originating from a telephone switchboard or customers arriving at a restaurant. They can result from unpredictable natural causes such as earthquakes or tornadoes, whose exact timing and direction cannot be precisely determined in advance. Greater economic cycles such as bull or bear markets, technological breakthroughs, wars or major accidents such as the *Exxon Valdez* are likewise unpredictable.

2.2. *Firm's condition*

We identify two firm conditions based upon their preparedness. Firms are either *prepared* or *unprepared* to handle the peak when it occurs. Prepared firms have two sets of policy measures and procedures for peak and off-peak conditions. Thus, prepared firms may have inventories created in pre-peak conditions, agile and versatile operations, and options for abruptly elevating capacity to the temporary peak level.

PM situations are classified into four categories (figure 3):

- Classical PM: occurs when the firm is prepared to handle a peak whose occurrence is predictable (typically due to seasonality). This is the case with rush hour food and transportation services. Tools for handling this condition include the definition of a pre-peak period during which the firm subordinates off-peak time to elevate its peak performance and the establishment of two sets of performance measures—one for peak and one for off-peak performance.
- Rapid response: the situation of an organization prepared to handle a peak whose occurrence is unpredictable. It is typical of emergency services throughout the world: snow removal during the winter, police SWAT team handling of crisis and hostage situations, etc. A pre-peak period is less practical when peak occurrence is unpredictable. Tools prescribed for this state include the creation of reserve resources. Agility—a quick and versatile response is the key to containing the costs of the firm's preparedness to exploit unexpected business opportunities.
- 'Unexpected Christmas': occurs when the peak is predictable but finds the firm unprepared. Such situations are encountered as a result of planning failure. Retail sales outlets occasionally running out of inventory during the holiday seasons is an example of this phenomenon. It is argued that there is no justification for a firm to be unprepared for a predictable peak. It is management's responsibility to transfer the unexpected Christmas state to the classical PM state.
- Opportunity/crisis management: occurs when unpredictable business opportunities or catastrophes occur such as competitor bankruptcy, unprecedented success rate in business tenders and accidents disabling production or service facilities. Firms cannot be prepared for every possibility and will therefore always be exposed to this situation. Tools appropriate for this condition include rapid outsourcing, robust, rapid decision-making processes and alliances with competitors or partners.

		Peak Occurrence	
		Expected	Unexpected
Firm's Condition	Prepared	Classical PM	Rapid Response
	Unprepared	'Unexpected Christmas'	Opportunity/Crisis Management

Figure 3. Peak classification.

This paper has two objectives: (1) to develop a methodology for preparing the organization to benefit from expected peaks, and (2) to consider how organizations can build fast response capabilities.

3. Expansion of TOC to handle peak situations

TOC is a managerial philosophy based on two main principles (for further details on TOC, see Goldratt 1997 and Schragenheim and Ronen 1991).

- The organization is a complex system of dependent variables striving to attain a specific goal that can be measured. Ideally, according to this holistic view, every member of the organization strives to maximize the organizational goal according to the global measures.
- The complex system can be simplified because, in a certain state, only a very small number of variables limit the organization's performance relative to its goal. These are called the organizational constraints. The attention of management should be focus on the few organizational constraints. Goldratt (1990) outlined five focusing steps as the kernel of TOC. Ronen and Spector (1992) enhanced and modified this kernel into the following seven steps.

(1) State the *goal* of the organization.
(2) Develop *global performance measurements* of the goal.
(3) *Identify* the system constraints.
(4) Decide how to *exploit* the system constraints.
(5) *Subordinate* everything else to the above decisions.
(6) *Elevate* the system's constraints.
(7) If a constraint has been broken, go back to Step 3. Warning: do not let *inertia* become the system's constraint.

4. Management during the off-peak period

An off-peak period is defined as a period during which demand does not exceed the system's capacity, meaning that market demand is the only constraint. In other words, the organization can serve more customers without additional resources and there is excess capacity in the system. This is certainly not an infrequent state of affairs.

Let us apply the seven steps to the off-peak situation. Step 1, the goal-definition, does not change. Step 2, the setting of performance measures, does not change. In Step 3, identification, market demand is the only constraint.

In Step 4—*exploitation*, the market is the constraint and every business opportunity should be pursued. For example, no minimum order size should be required; products can be customized to customer needs; special changes can be made in products. If the pricing policy constrains the system from selling more, it should be changed (see Section 4). A buffer of finished goods should be positioned in front of the market.

In Step 5—*subordination*, all system decisions should be subordinated to market demand. The quality, performance, and features of the product and services should be subordinated to the preferences of the various markets.

In Step 6—*elevation of the constraint*, the market constraint should be broken through differentiation, OEM, private labels, broadening of the product base and similar techniques.

5. Management during the peak period

A peak period is defined as a period during which demand exceeds the system's capacity and cannot be fully met. Let us apply the seven steps to the off-peak situation.

- In Step 1, the *goal's* definition does not change.
- In Step 2, the setting of *performance measures* does not change.
- In Step 3, *identification*, the constraining resource should be specifically pinpointed. For example, the pizzeria suffers from a shortage of counters. Other places encounter shortages of cashiers or sitting space. Call centres and web sites encounter shortage of bandwidth during peak periods. In R&D, when an unexpected project arrives, team leaders, senior programmers or microwave experts constrain the system, causing delays and unsatisfied specifications. Trade exhibitions such as CeBit create a shortage in integrators.
- In Step 4—exploitation is conventionally achieved through efficiency and effectiveness.
  - Efficiency: constraint should be utilized 100% of the time. Usually it is doing irrelevant jobs part of the time, e.g. sales people assist their clients in logistic tasks (providing a supply cable to a customer), R&D experts assist peers in installing upgrades of standard software packages.
  - Effectiveness: prioritize the jobs that give the maximum contribution per hour of constraint resource (see below).
- Step 5—*subordination* is carried out in two modes. First, as in conventional TOC, during the peak period all resources are subordinated to the constraint, assisting it to maximize throughput. However, we introduce the concept of *temporal*-subordination, i.e. the subordination of operations during off-peak periods to the needs of peak periods. Temporal subordination is carried out in two modes: (1) temporal subordination during off-peak periods, and (2) temporal subordination during pre-peak periods. Periods of off-peak time immediately preceding the peak period are defined as *pre-peak* (figure 4). During the pre-peak period resources are subordinated to assist the constrained resource in the peak. All activities that do not have to be carried out during the peak are performed during the off-peak period (e.g. preventive maintenance, training and recruiting). During the pre-peak period one can take the following measures.
  - Building up protective inventory (Cox and Blackstone 1998) of finished goods that will be sold with certitude.
  - Recruiting and training on-call temporary manpower.
  - Signing contracts with subcontractors and outsourcers and establishing a working relationship.
  - Preparing subassemblies and subcomponents.
- Step 6—*elevation of the constraint*: traditional elevation can be done in two ways;

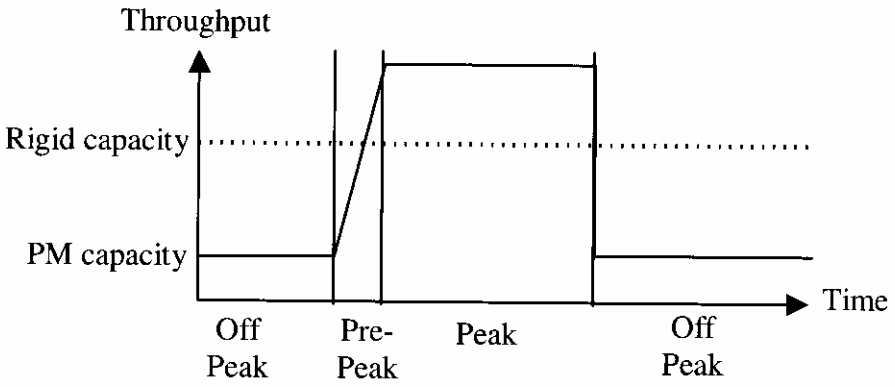

Figure 4. pre-peak period.

- elevating the effective capacity of the constraint by buying additional resources; and
- increasing the effective capacity of the constraint without additional resources.

Additional resources can be bought by hiring more people, working additional shifts, overtime, renting more equipment, hiring temporary manpower. Offloading the constrained resource can elevate capacity without additional resources. Some non-bottleneck resources can take simple structured tasks off the constrained resource. The difference between traditional and peak-management conditions is that under peak-management demand will return to off-peak levels and therefore elevation must be temporary or flexible. Flexibility and agility in resource allocation should enable resource mobility between tasks and smooth expansion through subcontractors during peak periods as well as reduced fixed costs during off-peak periods.

It should be noted that the pre-peak period applies for predictable peaks. Consider, for example, a canning company facing peak demand during the summer that applies the concept of PM by building up a buffer of finished goods during pre-peak time. After calculating the trade-off between building up inventory and working three shifts they came to the conclusion that it is much cheaper to create throughput by working a third shift. They reduced the pre-peak time from 3 months to 6 weeks by working full time during pre-peak and peak periods. Ten per cent of the orders that were not really needed during the peak time were offloaded to non-peak time. Since set-up times are 'free' during off-peak time, orders that were originally manufactured using a single batch to save on set-up time were split and the minimal quantity was delivered during peak time. An extra set-up delivered the rest during off-peak time. Company workers are encouraged and given incentives to take their annual vacations during off-peak time, and all preventive maintenance, training and recruiting are done then.

Another typical example is found in the insurance industry, which regularly faces demand peaks towards the end of the fiscal year when clients take advantage of tax shelters. For this reason no vacations are allowed during this period, and temporary employees are hired to off-load sales and administrative personnel.

6. Transition between off-peak and peak behaviour

The transition between peak and off-peak operational modes is complex. The existence of two distinct sets of internal rules that interchange unexpectedly creates problems. However good the human resource management is and however well the reasons for the periodic changes in policy are explained, implementation may still be difficult, especially when it concerns workers at the capacity constraint workstation.

Two features characterize prepared firms: predefined and rehearsed peak-time policies that differ from off-peak or routine policies; and *peak-time resource allocation* that differs from off-peak resource allocation. Peak-time resources belong to one of three categories.

- Permanent peak-dedicated resources, where long-term resource planning corresponds with peak conditions. Examples include extra teller booths in a supermarket that are vacant during off-peak periods.
- Pre-peak resource build-up in anticipation of predictable peaks. For example, food and equipment accumulated in anticipation of an imminent tornado.
- Reserve resources that are deployed only when the permanent peak-dedicated resources cannot meet the demand. For example, 'on call' doctors at the hospital are a reserve resource that constitutes a response option without incurring regular costs. Certain countries maintain a corps of trained and equipped reserve soldiers who are mobilized only when unexpected emergency situations occur.

The rules for pre-production should be based on two distinct considerations.

- Any forecast should note a 'minimum' demand for every product, such that the probability that demand will be higher than this minimum is >90%. This amount can be produced prior to the start of the peak. If the calculations show that the constrained resource can cope with the rest of the expected demand, with a certain safety margin, then this minimum should be the amount to produce prior to the peak.
- If larger quantities are needed before the peak actually starts, the products that consume more of the anticipated constraint's time are those for which higher levels of finished goods inventory should be considered. This will free more constrained capacity when it is really needed. Of course, materials cost should be taken into account—at higher levels of finished goods, these costs may be prohibitive. Notice that the time spent in this early production is not an investment. In off-peak periods the alternative price for regular working time is practically zero. This is also true for the anticipated constraint time—when we are still in off-peak period. One might argue that there is an alternative price because of the impact on timely response and quality. True, but it is negligible relative to the value of better exploiting the opportunities of the peak period.

The potential capacity appears to be that of a peak period, but it would be a mistake to treat it this way. The extra load (load for peak) should clearly be in the least-priority list. The marketing and pricing considerations should be as at off-peak times, in the manner discussed above. We suggest starting production for a coming peak somewhat earlier than necessary for this load in order to prevent deterioration of response time while still in an off-peak period.

When a peak cannot be anticipated, there is not much we can do. The challenge is to anticipate a peak in good time. Buffer management (BM) (Schragenheim and Ronen 1991) extends the Drum–Buffer–Rope scheduling technique to control the production flow. The BM tool divides the buffer (the shipping or the constraint buffer) into three control zones: green, yellow and red. Whenever a batch fails to arrive at the planned time it creates a 'hole' in a certain buffer zone. If the hole occurs in the green zone no special action is required, but management should watch and wait. If the hole occurs in the yellow zone, corrective action is required. If it occurs in the red zone then expediting is required. Usually in a peak situation the main buffer is the constraint buffer while the shipping buffer is almost empty. During off-peak situations the shipping buffer is loaded while the constraint buffer may be partially empty. This can assist in identifying peaks as well as off-peaks.

7. Impact of PM on global strategy and pricing policy

The changing of internal rules should be part of the strategy for running the organization. Moreover, distinguishing between peak and off-peak periods has more to offer. Off-peaks create pressure on management due to the highly visible excess capacity, which creates the notion of 'waste'. Acknowledging that long off-peak periods exist, they can be used for bettering workers' conditions—letting them have more than one annual vacation. The actual cost is nil, and if this turns out to be a way to attract better qualified people, then all the better. Strategic offloading, from the peak periods to off-peaks, is another possibility. As already noted, the pricing considerations are very different in the two periods. See Goldratt (1990) for a detailed analysis of price/cost considerations. The fact that no internal constraint exists during off-peak periods is of paramount importance. Special arrangements can be worked out to shift part of the peak demand to the off-peak demand. If these deals attract new business, the bottom line will be enhanced. Candidates for offloading are those products that consume more time on the peak's constraint.

Analysing the performance at peak periods can lead management to consider elevating the internal constraint—or, better still, ways to elevate both the market demand constraint and the internal constraint. Assuming this elevation requires considerable investment, a careful analysis of the benefits, taking into account both peak and off-peak periods, should of course be carried out. For a cost/benefit analysis of elevating a capacity constraint, see Ronen and Spector (1992).

Therefore, two strict pricing policies should be implemented: during peaks pricing should be based on the exploitation of the resource constraint, i.e. it should be based on the contribution per constraint time. During off-peak hours, having excess capacity leads to marginal costing. Thus, the same product is to be sold at different prices during different time brackets. For example, business lunches offered at lunchtime are less expensive than during peak dinnertime. In the agricultural manufacturing industry it should be common for agricultural irrigation projects to be discounted during off-peak time. Thus, pricing policies should be related to resource utilization, and finance people should work together with those in marketing and operations.

8. Conclusions

Most organizations have to recognize the different operational tactics of peak and off-peak periods. What is quite evident is that market demand is always a constraint. What also emerges from the above discussion is that a capacity constraint

also exists in the vast majority of organizations. Though the capacity constraint is active as such only part of the time, it is still a constraint by definition, because it limits the performance of the whole organization relative to its goal. Its impact is substantial, and it requires different policies than those advocated by TQM and JIT, which do not recognize the possibility of an internal constraint. Failure to recognize these factors creates a 'policy constraint'—meaning that the policy in itself limits the success of the organization.

References

COX, J. F. and BLACKSTONE, J. H. (eds), 1998, *APICS Dictionary*, 9th edn (Falls Church: APICS).

COX, J. F. and SPENCER, M. S., 1998, *The Constraints Management Handbook*. APICS Series on Constraints Management (Boca Raton: Lucie Press).

FORTUNE, 1999, Buzzwords. Fortune, 29 March, 121.

GOLDRATT, E. M., 1986, *The Goal* (Croton-on-Hudson: North River).

GOLDRATT, E. M., 1990, *The Haystack Syndrome* (Croton-on-Hudson: North River).

GOLDRATT, E. M., 1997, *Critical Chain* (Croton-on-Hudson: North River).

RONEN, B., 1992, The complete kit concept. *International Journal of Production Research*, **30**, 2457–2466.

RONEN, B. and SPECTOR, Y., 1992, Managing system constraints: a cost/utilization approach. *International Journal of Production Research*, **30**, 2045–2061.

RONEN, B. and STARR, E. M., 1990, Synchronized manufacturing as in OPT: from practice to theory. *Computers and Industrial Engineering*, **18**, 585–600.

SCHRAGENHEIM, E. M. and RONEN, B., 1991, Buffer management: a diagnostic tool for production control. *Production and Inventory Management Journal*, Second Quarter, 74–79.