Fixing planning in the VUCA world

Received (in revised form): 19th August, 2024

Carol Ptak

CAROL PTAK
Partner, Demand Driven Institute, USA

Carol Ptak is currently a Partner with the Demand Driven Institute (DDI) and was most recently at Pacific Lutheran University as Visiting Professor and Distinguished Executive in Residence. Previously, she was Vice President and Global Industry Executive for Manufacturing and Distribution Industries at PeopleSoft, where she developed the concept of demand driven manufacturing (DDM). Carol spent four years at IBM Corporation, culminating in the position of Global SMB Segment Executive. A leading authority in the use of enterprise resource planning (ERP) and supply chain tools to drive improved bottom line performance, Carol's expertise is well grounded in four decades of practical experience as a successful practitioner, consultant and educator in manufacturing operations. Her pragmatic approach to complex issues and dynamic presentation style has her in high demand worldwide on the subject of how to leverage these tools and achieve sustainable success. She holds an MBA from Rochester Institute of Technology and completed the EMPO programme at Stanford University. Carol is a frequent educator at university level and presents at many key technical conferences around the world, including South Africa, France, Israel, Australia, Ireland, the Netherlands and 11 APICS International conferences. In addition to her collaborative works with Chad Smith, she is the author and coauthor of the books *MRP and Beyond* (Irwin Professional, 1996), *ERP, Tools, Techniques and Applications for Integrating the Supply Chain* (Ptak and Schragenheim, CRC Press, 1999), *Necessary but Not Sufficient* (Goldratt, Schragenheim, Ptak, North River Press, 2005) and *Theory H.O.W. – How Organizations Could Work* (Ptak and Cavallaro, Specialty Publishing Company, 2006). Carol is certified through APICS/ASCM at the fellow level (CFPIM) and was certified in Integrated Resource Management (CIRM) with the first group internationally. Carol was the President and CEO of APICS, The Educational Society for Resource Management for the year 2000. Prior to her election as APICS President, she served on the society in a variety of positions.

Demand Driven Institute, PO Box 1432, Globe, AZ 85501, USA
Tel: +1 253 279 3291; E-mail: cptak@demanddriveninstitute.com

Chad Smith

CHAD SMITH
Co-founder and Partner, Demand Driven Institute, USA

Chad Smith is Co-founder and Partner of the Demand Driven Institute (DDI). Prior to DDI, Chad co-founded Constraints Management Group (CMG) in 1997, a services and technology company specialising in manufacturing, materials and project management solutions for mid-range and large manufacturers. He served as Managing Partner of CMG from 1998 to 2015. Clients past and present include Unilever, LeTourneau Technologies, Boeing, Intel, Erickson Air-Crane, Siemens, IBM, The Charles Machine Works (Ditch Witch) and Oregon Freeze Dry. In addition to the collaborative works with Carol Ptak he also coauthored *Demand Driven Performance: Using Smart Metrics* (Smith and Smith, Industrial Press, 2014). Chad is also a certified expert in all disciplines of the Theory of Constraints studying directly under the tutelage of the late Dr Eli Goldratt.

Demand Driven Institute, 1418 Gabriella Lane, Wenatchee, WA 98801, USA
Tel: +1 253 709 7330; E-mail: csmith@demanddriveninstitute.com

Abstract

The essence of any business is flow. The flow of materials and/or services from suppliers — perhaps through multiple manufacturing plants and then through delivery channels to customers. The flow of information to all parties about what is planned and required, what is happening, what has happened and what should happen. The flow of cash returns from the market to the organisation and through to the suppliers. This paper makes a critical connection between the concept of *flow and the material requirements planning (MRP) inputs, processing logic and assumptions and introduces a necessary change in processing logic to address the challenges companies experience today. The true purpose of planning — to promote and protect the flow of relevant information and materials — was the driving purpose behind the development of MRP. Throughout this planning and information system evolution, MRP's basic requirements, assumptions and processing logic have remained unchanged. Despite newer and more powerful planning, and control applications powered by cloud-based infrastructure, the basic approaches to planning manufactured and purchased items have remained largely a constant since the 1960s. This paper proposes an innovative pragmatic proven methodology that enables a company to successfully sense and adapt to changes in the volatility, uncertainty, complexity and ambiguity (VUCA) world.*

Keywords

flow, demand driven, DDMRP, planning, VUCA world, decoupling

DOI: 10.69554/GWUD5741

INTRODUCTION

What is the true business purpose of planning in general? In many cases today, that purpose tends to get lost in the proverbial fog of supply chain management technology. The purpose is to orchestrate, coordinate and synchronise an organisation's assets to the sale of an item or service. What do we make? When do we make it? What do we buy? When do we buy it? What do we deliver? When do we deliver it? What do we move? Where do we move it? The more complex the product, service and supply chain scenario, the greater the apparent need for effective orchestration, coordination and synchronisation. Material requirements planning (MRP) and the subsequent systems built around it were developed to facilitate this purpose.

But are we missing something fundamentally important about planning? The required orchestration, coordination and synchronisation is simply a means to an end. That much is quite easy to grasp. What is more difficult for many organisations to grasp is what fundamental principle should underlie the orchestration, coordination and synchronisation. Most supply chain personnel get distracted by the immediate quantities and tasks involved with orchestration, coordination and synchronisation rather than paying attention to what those quantities and the activities associated with them are intended to accomplish.

The essence of any business is flow. The flow of materials and/or services from suppliers — perhaps through multiple manufacturing plants and then through delivery channels to customers. The flow of information to all parties about what is planned and required, what is happening, what has happened and what should happen. The flow of cash returns from the market to the organisation and through to the suppliers.

Is this some sort of inspired revelation? In reality, flow has *always* been the primary purpose of most services and supply chains. Simply put, you must take things or concepts, convert or assemble them into different things or offerings, and then get these new things or offerings to a point that someone is willing to pay you for the new thing or offering. The faster you can make, move and deliver all things and offerings, the better the organisational performance by almost any metric of relevance. And it should be noted that this concept is not above the pay grade of planners and buyers — it is at the very core of what their job description should be. The concept of flow, known as Plossl's Law, is endemic to planning.

This paper makes a critical connection between these concepts and the MRP inputs, processing logic and assumptions and introduces a necessary change in processing logic to address the challenges companies experience today. The true purpose of planning — to promote and protect the flow of relevant information and materials — was the driving purpose behind the development of MRP. Throughout this planning and information system evolution, MRP's basic requirements, assumptions and processing logic have remained unchanged. Despite newer and more powerful planning, and control applications powered by cloud-based infrastructure, the basic approaches to planning manufactured and purchased items have remained largely a constant since the 1960s.

WELCOME TO THE VUCA WORLD

Table 1 introduces some of the changing circumstances in supply chain management since the codification of MRP in 1965. The table explains the need for a certain classification of inventory called fluctuation inventory; however, it does not fully encompass the nature of the changing business circumstances in supply chain management. The world is a much different place today than it was when the framework of conventional operational rules and systems was developed and codified. The circumstances under which this conventional framework was developed have dramatically changed. The world is a more volatile, uncertain, complex and ambiguous place, seemingly unimaginable 60 years ago. VUCA is an acronym to describe the world we live in today.

In the VUCA world, any supply chain professional must always keep two critical conclusions in mind:

- Frequent and severe disruptions *are* going to happen.
- Supply chain planning and execution systems *cannot* be based on strategies or rules that ignore the above conclusion.

These two conclusions require a fundamental rethinking of the way that most companies and industries plan, schedule and operate. Planning must take into account the VUCA world which results in extreme demand and supply continuity variability. MRP was developed with the idea that computational process and power was the solution for planning and execution.

MRP plans are based on the assumption that all its inputs are 100 per cent accurate at the time of planning: demand, inventory and product structure. These plans must follow basic rules as defined by those inputs when performing their calculations. Thus, every carefully synchronised precise schedule that MRP produces, regardless of how frequently it produces that schedule, is based on the following:

- There is sufficient time to accomplish all required activities to given demand.
- Demand is known and will not change.
- Execution will occur as planned.

As any of these items become less valid, the schedule becomes less valid, driving exception flags, order revisions and the need for another MRP run. The same fate awaits the new run — it is just a

TABLE 1 Changes in supply chain management 1965–2024

Circumstance	*1965*	*2024*
Supply chain complexity	Low. Supply chains looked like chains — they were more linear. Vertically integrated and domestic supply chains dominated the landscape.	High. Supply chains look more like 'supply webs' and are fragmented and extended across the globe.
Product life cycles	Long. Often measured in years and/or decades (eg rotary phones).	Short. Often measured in months (particularly in technology).
Customer tolerance times	Long. Often measured in weeks and months	Short. Often measured in days with many situations dictating less than 24-hour turns.
Product complexity	Low.	High. Most products now have relatively complex mechanical and electrical systems and micro-systems.
Product customisation	Low. Few options or custom feature available.	High. Lots of configuration and customisation to a particular customer or customer type.
Product variety	Low. Example — toothpaste. In 1965 Colgate and Crest each made one type of toothpaste.	High. In 2012 Colgate made 17 types of toothpaste and Crest made 42.
Long lead time parts	Few. Here the word 'long' is in relation to the time the market is willing to wait. If customer tolerance times were longer, it stands to reason that there were less long lead time parts. More so, however, is that fact that supply chains looked different. Most parts were domestically sourced and thus, often much 'closer' in time.	Many. Today's extended and fragmented supply chains have resulted in not only more purchased items overall, but more purchased items coming from more remote locations.
Forecast accuracy	High. With less variety, longer life cycles and high customer tolerance times forecast accuracy was almost a non-issue. 'If you build it, they will buy it.'	Low. The combined complexity of the above items is making the idea of improving forecast accuracy a losing battle.
Pressure for leaner inventories	Low. With less variety and longer cycles, the penalties of building inventory positions was minimised.	High. At the same time operations are asked to support a much more complex demand-and-supply scenario (as defined above), they are required to do so with less working capital.
Transactional friction	High. Finding suppliers and customers took exhaustive and expensive efforts. Choices were limited. People's first experience with a manufacturer was often through a salesperson sitting in front of them.	Low. Information is readily available at the click of the mouse. Choices are almost overwhelming. People's first experience with a manufacturer is often through a screen sitting in front of them.

matter of time. Let us summarise why these three points will not remain valid in today's environment.

Cumulative lead time as a constraint

MRP cannot bend the space–time continuum: it cannot make time when there is no time. MRP knows that and so do its users. MRP must back schedule through the entire cumulative horizon to create a complete and properly covered plan for an end item. Because of the VUCA world, it means that frequently orders are already late for release when the schedule is established. The exception to this is when there is sufficient on-hand inventory at a position to cover the required dependent demand.

MRP was not designed to leave on-hand stock at certain defined positions. Remember that the placement of safety stock does not figure into the netting equation, it is consumed during execution as required and then triggers a reorder and expedite. It is also common practice to not define safety stock at intermediate product structure levels.

What this means is that if MRP is truly utilised to its fullest potential, there will be no residual inventory at the lower levels of the product structures to stop the requirements explosion. Ordering and manufacturing activity cannot be initiated in the past. This then requires users to define a planning horizon at least as long as cumulative lead time to have a chance to meet the top-level demand numbers.

Demand will change

These longer planning horizons guarantee that the demand numbers being used to drive purchasing and low-level component schedules will change, as shown in Figure 1.

When forecast error is realised and order revisions are issued at the top level of a product structure (whether quantity, timing or both), there is a ripple effect through the lower levels in the replanning process. This is called *system nervousness*. This challenge of system nervousness has been known since the earliest days of MRP. It was the driving force behind the development of the demand time fence, firm planned orders and the master production schedule. All were created to mitigate system nervousness. None of these innovations, however, really solved the root of the problem. They were able to partially mask the issue at the time because system nervousness was more manageable due to a combination of factors:

- The gap between cumulative lead times and customer tolerance times was much smaller, resulting in less remote and consequently more accurate forecasts.
- MRP recalculations were less frequent (typically once per month).
- There was a significantly fewer number of end items to plan.
- Market volatility was significantly less than in the current VUCA world.

The complex and volatile environment characterised by more complex supply chain circumstances in combination with real-time computing capability makes the issue a much bigger challenge today. But given the hard-coded nature of the MRP calculation, the only real way to stop nervousness is to never replan. This would represent an enormous dilemma for companies, as it would mean significant service challenges, because the internally generated forecasted orders will increasingly vary from what the market really desires. Thus, MRP users are forced into compromises to slow down the rate of changes. These compromises come with penalties.

Safety stock at intermediate levels?

Some MRP users believe that they can attempt to combat nervousness by placing

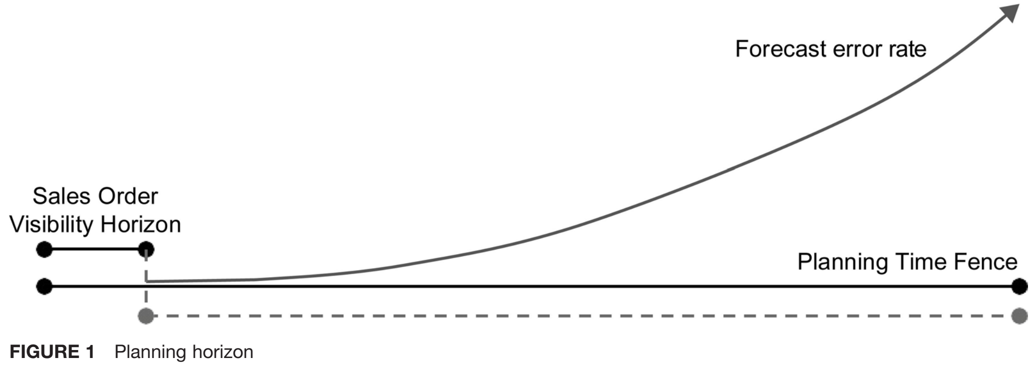

FIGURE 1 Planning horizon

safety stock at intermediate levels. This strategy could be extremely counterproductive for the following reasons:

- The netting equation will not change. Safety stock is not involved in the planned order coverage equation. Thus, any changes made in demand will still result in the same level of exception messages. The user may choose, however, to disregard some of these messages, knowing that they have safety stock in sufficient quantity to deal with the issue at their peril.
- Additional nervousness will most likely occur. As safety stock is consumed, MRP will attempt to restore the safety stock by launching and expediting orders.
- An increase in slow moving working capital will increase. Safety stock is always planned to be at full quantity — there is no allowed use or penetration below the level without a reactionary order to restore the position. This essentially relegates the inventory dedicated to safety stock as a pool of frozen working capital. The more items with safety stock, the larger the amount of dead inventory.

Less frequent replanning

As mentioned previously, the computing power has long since existed to recalculate MRP in real time. The mid-1990s saw the introduction of systems that could do just that. This meant a constant recalculation as any changes occur, which in turn meant constant nervousness. There is a direct relationship between the frequency of MRP recalculation and the amount of nervousness. Real-time recalculation was short-lived because most users found the level of nervousness to be nothing short of paralysing and led most companies to a more latent approach to recalculating requirements: once per week. Software companies added a configuration switch to display the planning information only at the desired interval and continued to calculate in real time.

Flattening the bill of material (BOM)

The flattened structure eliminates intermediate positions. This reduces the number of changes to intermediates (since there are none) and eliminates all the action flags and messages related to them. But does it produce an environment with more relevant information, or does it actually further distort the picture? The key to more relevant information is not to simply ignore dependencies. When we ignore critical dependencies, we risk oversimplification.

The execution challenge

MRP creates a precisely synchronised plan based on its required inputs and assumptions. From an execution perspective, two critical assumptions affect how achievable and realistic this plan will be: item lead times are fixed, and full allocation will be achieved as planned. This means that any MRP plan assumes that all components in full quantity will always be available precisely when required by the parent order release. Thus, MRP plans are achievable and realistic only if everything in the entire dependent network goes precisely according to plan. In almost every modern environment, this is an impossibility.

Common cause variation

Any process, even one deemed to be statistically under control, still exhibits variation. In other words, under control does not mean without variation. Normal or random operational variability results in a process that may be statistically within calculated control limits but still varying between those limits. The Association for Supply Chain Management (ASCM) dictionary defines control limit as:

> A statistically determined line on a control chart (upper control limit or lower control limit). If a value occurs outside of this limit, the process is deemed to be out of control.[1]

Reducing the gap between the limits is a worthy goal, but the elimination of the gap is an impossibility; it would require the process to be absolutely perfect every time. A process that has a six sigma quality level still experiences 3.4 defects per million opportunities — not perfection. Unfortunately, perfection is exactly what MRP plans for and assumes will happen with purchase and manufacture of each and every item.

Complex delay accumulation

After establishing that single events or resources will always experience variation, we will turn to a much more common and problematic environment, particularly when considering the viability of plans generated by MRP systems. MRP is commonly utilised in environments that have relatively complex product structures. Typically, that complexity means that product structures can be both broad and deep. Broad means that the product structure fans out to multiple component legs. Deep pertains to the number of levels in the product structure.

Since, however, many components have extremely high variability and/or arrive early, the parent order release is still at the mercy of any one missing component. So, a simple rule is evident: delays accumulate, while gains do not. This means that in every conceivable case of even mild complexity, the plan can never be achieved. This delay accumulation can also be referred to as *supply continuity variability*.

Batching policies — an amplifier of variability

Batching policies can further exacerbate the effect of both nervousness and delay accumulation.

Many ways exist to determine the lot-sizing policies that the MRP calculation must obey. Lot-for-lot environments are less common due to the proliferation of cost-based lot-sizing techniques throughout the product structure and resource base.

Batching practices can also dramatically affect the way that material moves in a supply chain, contributing to or amplifying the accumulation of delays. Delay accumulation will occur while parts wait in a bin or queue in a larger transfer batch, or while an order waits on a trailer for other orders to fill up the truck if a transportation batching policy dictates that only full trucks are allowed.

THE BIMODAL INVENTORY DISTRIBUTION

With the understanding of both nervousness and execution delay accumulation, we begin to see a broader picture affecting the validity of any MRP-derived schedule across a product structure and enterprise. The broader challenge is of a bidirectional nature. Changes ripple and are amplified down the product structure, while delay accumulations, labelled as supply continuity variability, ripple up and are amplified in the product structure.

To appreciate the effect this has on an organisation, consider the value of inventory in facilitating flow through an enterprise or supply chain where an optimal zone can be depicted between points that represented an interruption to flow (stock out) or a drag on flow (extreme excess), as shown in Figure 2.

When the aggregate inventory position is considered in an environment using conventional MRP, there is typically a bimodal distribution noted. A bimodal distribution exhibits two distinct lumps or distribution groups. A bimodal distribution can occur at the single-part level over a period, as a part will oscillate back and forth between excess and shortage positions. In each position, flow is threatened or directly inhibited. What makes it bimodal is a clear separation between the two groups: the lack of any significant number of occurrences in the optimal range.

A bimodal distribution can also occur across a group of parts: at any one time, many parts will be in excess while other parts are in a shortage position. Shortages of any parts are particularly devastating in environments with assemblies and shared components because the lack of one part can block the delivery of many parent parts.

The bimodal distribution shown in Figure 2 depicts many parts that are in the 'too little' range, while still another large number of parts are in the 'too much' range. The solid line represents the number of parts at any particular point on the loss function spectrum. Not only is the smallest population in the optimal zone, but the time that any individual part spends in the optimal zone tends to be short-lived. Many items tend to oscillate between the two extremes every time a new MRP run occurs. The oscillation is driven by the transference and amplification of variability that builds up between MRP runs. The oscillation is depicted by a dashed curved line connecting the two disparate distributions. At any one time, any planner or buyer can have many parts in both extremes simultaneously. This bimodal distribution is rampant throughout industry. It can be very simply described as 'too much of the wrong and too little of the right' at any point in time and 'too much in total' over time.

There are three primary effects of the bimodal distribution evident in most companies.

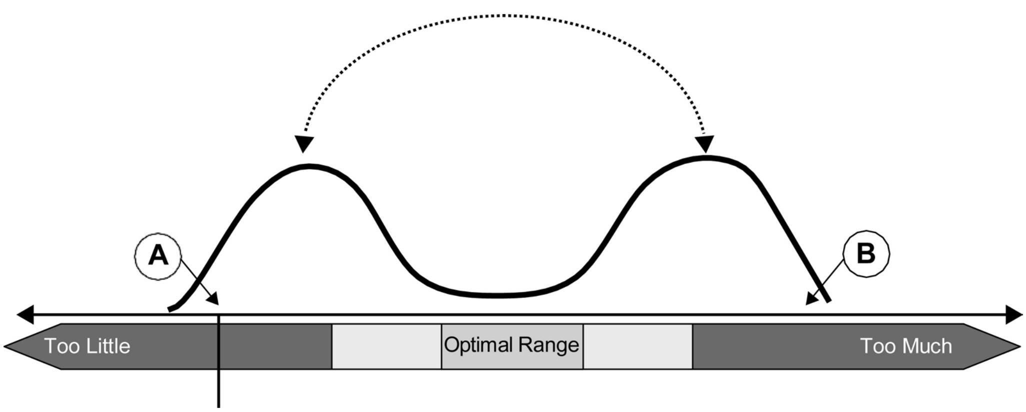

FIGURE 2 The bimodal inventory distribution

- *High inventories*: The distribution can be disproportionate on the excess side, as many planners and buyers will tend to err on the side of too much. This results in slow-moving or obsolete inventory, additional space requirements, squandered capacity and materials and even lower margin performance, as discounts are frequently required to clear out the obsolete and slow-moving items.
- *Chronic and frequent shortages*: The lack of availability of just a few parts can be devastating to many manufacturing environments, especially those that have assembly operations and common material/components. The lack of any one part will block an assembly. The lack of common material/components will block the manufacture of all parent items calling for that common item. This means an accumulation of delays in manufacturing, late deliveries and missed sales.
- *High bimodal-related expenses*: This effect tends to be under-measured and under-appreciated. It is the additional amount of money that an organisation must spend to compensate for the bimodal distribution. When inventory is too high, third-party storage space may be required. When inventory is too low, premium freight (LTL) and fast freight are frequently used to expedite material. Overtime is then needed to push late orders through the plant. Partial shipments are made to get the customer some of what they ordered, but with significantly increased freight expenses.

These three effects are indicative of major flow problems in most organisations' supply chains, but the bimodal distribution is only a localised measure. The transference and amplification of variability is not limited to the single enterprises of a supply chain — it propagates throughout the interdependent relationships and is known as the *bullwhip effect*.

THE BULLWHIP EFFECT

The bullwhip effect is a phenomenon that dominates most supply chains. The Association for Supply Chain Management (ASCM) dictionary defines it thus:

> An extreme change in the supply position upstream in a supply chain generated by a small change in demand downstream in the supply chain. Inventory can quickly move from being backordered to being excess. This is caused by the serial nature of communicating orders up the chain with the inherent transportation delays of moving product down the chain. The bullwhip can be eliminated by synchronizing the supply chain.[2]

This definition clearly deals with important points discussed earlier in this paper. 'Inventory can quickly move from being backordered to being excess' is descriptive of the oscillation effect with the bimodal distribution. Additionally, this definition deals with both information and materials. 'Communicating orders up the chain' is the information component, while 'moving product down the chain' is the materials component. The bullwhip effect is really the systematic and bidirectional breakdown of the flow of relevant information and materials in a supply chain. Figure 3 is a graphical depiction.

The wavy arrow moving from right to left is the distortion to relevant information in the supply chain. The arrow wave grows in amplitude to depict that the farther up the chain you go, the more

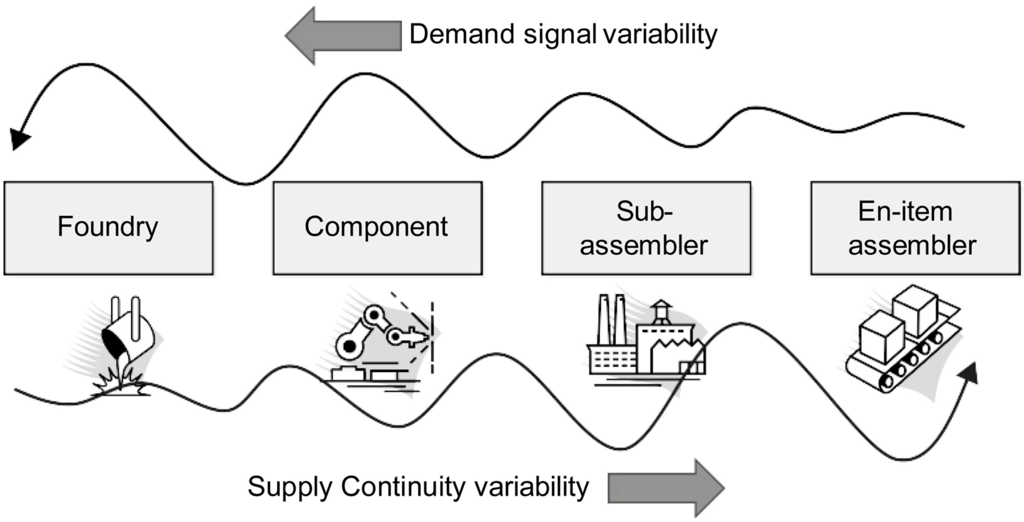

FIGURE 3 The bidirectional bullwhip effect

disconnected the information becomes from the origin of the signal, as signal distortion is transferred and amplified at each connection point.

A massive amount of research and literature has been devoted to the phenomenon known as the bullwhip effect; however, very little, if any, of that body of knowledge has been devoted specifically to its bidirectional nature. Most of the research has been dedicated to understanding how and why demand signal distortion occurs and how to potentially fix it by synchronising the supply chain around a better forecast. Yet because the issue is bidirectional, it cannot be solved by only addressing one direction. Even a perfect forecast still leaves any MRP plan subject to the inevitable supply continuity variability that will occur. A bidirectional problem requires a bidirectional solution.

The core problem underlying the bullwhip effect

We have summarised the challenges in using MRP systems in the VUCA world. Those challenges connect directly with the bimodal effect seen in most manufacturers as well as the bullwhip effect occurring throughout supply chains. If we want to identify a bidirectional solution, we must trace these issues to a root cause or core problem endemic in MRP based on what has been established so far. Figure 4 is a logical construct to trace that core issue.

At the bottom of the construct, we see two things in shaded boxes that we know to be true:

1. Cumulative lead times exceed customer tolerance times. This is well established in the VUCA world.
2. MRP calculates a schedule based on all defined dependencies. Of key significance is the impact of the defined product structure, its fixed lead times, and the lot-sizing policies for items.

The reader can follow the construct from bottom to top, culminating in the erosion of the protection and promotion of the flow of relevant information and

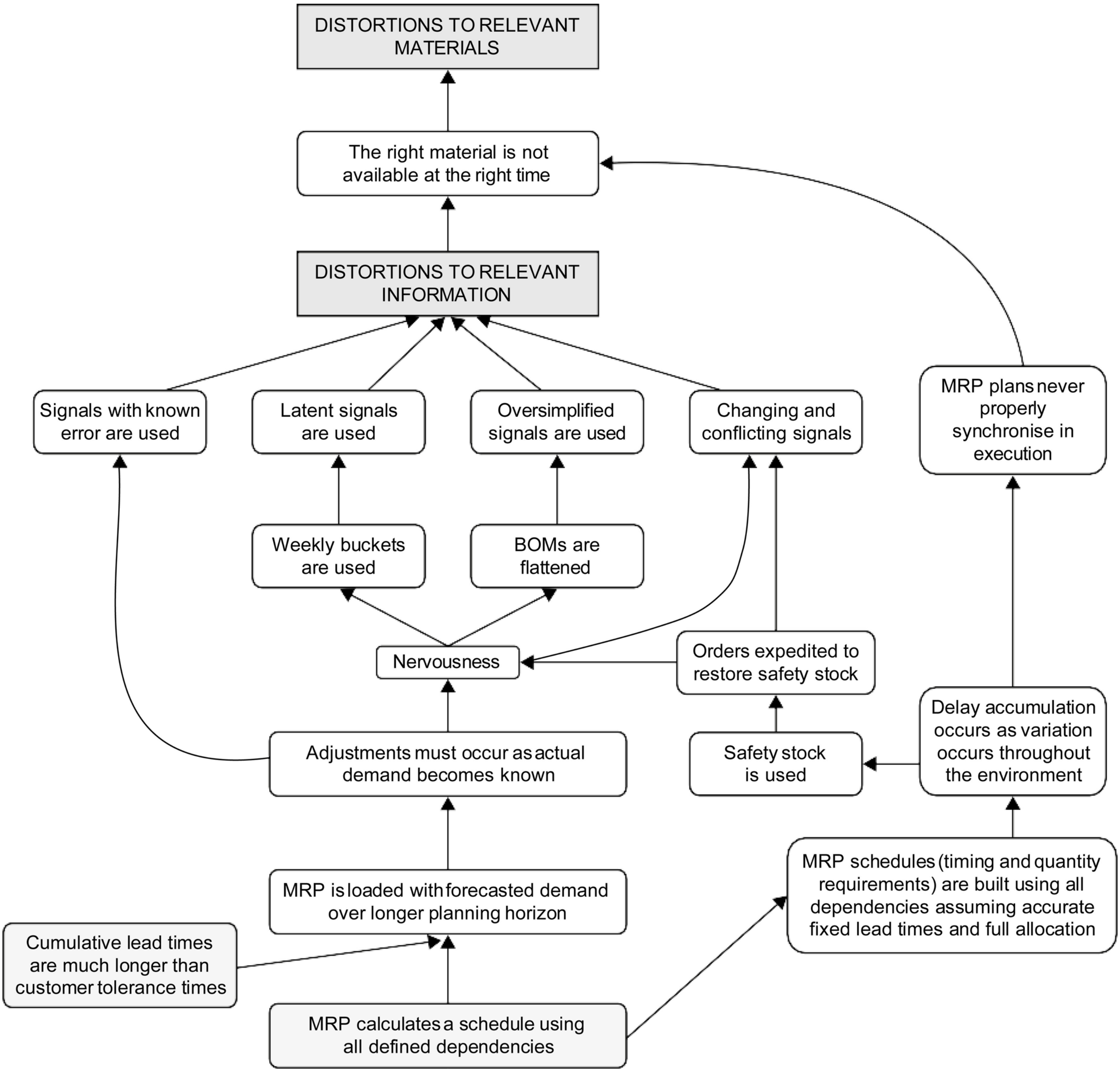

FIGURE 4 Finding the core problem behind the bimodal distribution and the bullwhip effect

materials. The boxes at the tips of the arrows are effects of the boxes at the tail of the arrow. The two shaded entry points on this construct are both candidates to address in a proposal for a solution direction. In the VUCA world, one is feasible and realistic, while the other is a near impossibility.

First, let us address the impossibility. The ability to compress cumulative lead times to approximate customer tolerance times would require an extreme reconfiguration in a company's supply chain, as characterised by purchased and manufactured component lead times. While the COVID-19 pandemic made many companies reconsider the length and vulnerabilities of their cumulative lead times, the ability to dramatically reduce the fixed lead times of all components so that the cumulative lead times match customer tolerance times would require massive amounts of innovation and capital. To address the other shaded entry point would simply require a change in the basic MRP calculation regarding which dependencies are considered and calculated in a product structure at any one point in time.

The answer is obvious. Changing the basic MRP calculation is several orders of magnitude simpler and cheaper and can be consistently applied across industries. Yet this will cause experienced MRP practitioners to pause. The ability to calculate all dependencies across product structures was billed as the revolutionary promise of MRP — 60 years ago. How can the answer be the problem?

And if all dependencies are not calculated together, how can the entire environment be planned and synchronised? The key word in the question is *together*. The solution will calculate all dependencies to plan and synchronise an environment, but not all at the same time.

DECOUPLING

The role of MRP in the modern supply chain is significant, powerful, and should not be understated. MRP is the conductor of the supply chain symphony. Each node in the supply chain has an MRP system supporting each different manufacturing operation.

The core problem in MRP has remained in existence because calculating all dependencies together was the real power behind the MRP tool. The more dependencies and the faster the calculation, the more powerful MRP was thought to be. As a result, conventional planning grew more complex but exhibited more system nervousness. Supply chain professionals treated the symptoms. Firm planned orders were created. The demand time fence and planning time fence were created. The MPS was invented. DRP evolved with all these tools. Advanced planning and scheduling (APS) was invented. Workarounds proliferated. Forecasting software became more and more sophisticated. The effort to maintain these complex systems and workarounds grew as the world became more variable, volatile, uncertain and ambiguous.

Investments in these sophisticated planning systems failed to provide a return on that investment; billions of dollars were squandered, and no one could specifically answer why. Instead, conventional planning advocates pointed to data accuracy, the failure to follow the rules in the system, overall discipline, and the need for still better forecasts. Thus, industry has continued to focus on and treat symptoms regarding the breakdown of the MRP equation.

Figure 5 highlights the symptom level that most companies attack regarding planning and scheduling problems.

Companies put inordinate amounts of time and money into generating better forecasts without real benefit. Planners use spreadsheets to gain better visibility and separate 'real priorities' from the long list of exception messages. Capacity and inventory are added to compensate for delays in both the short and long term.

This is not just a recent phenomenon. These issues have been growing in

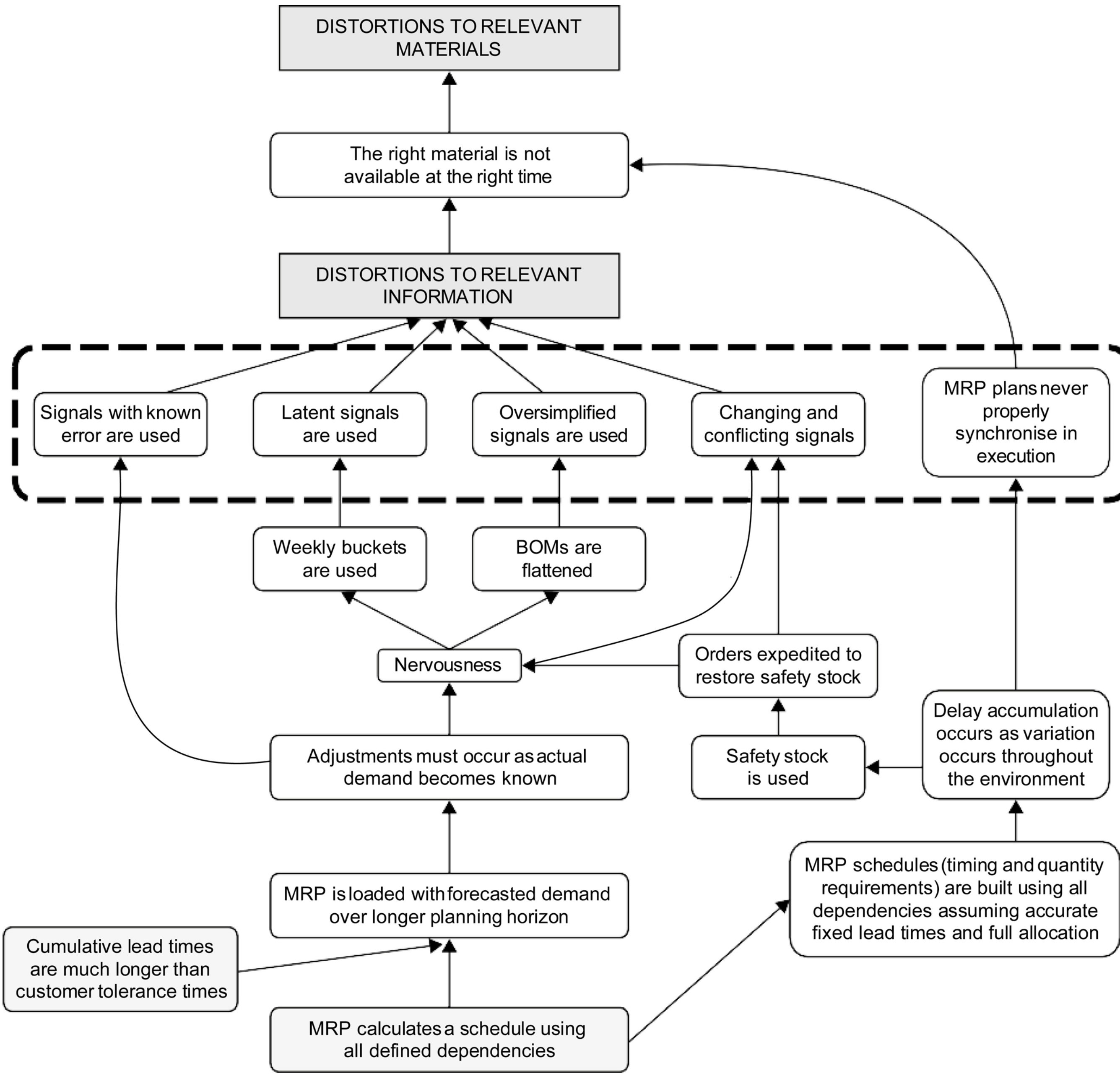

FIGURE 5 The symptom level

magnitude as the world became more volatile, uncertain, complex and ambiguous. But throughout all these decades, MRP's basic attribute (the core problem) remained essentially untouched: it was deemed inviolate. To remove that attribute would seem to render it useless and return planners to the 1940s with order point. But rather than completely remove the attribute, we must understand how to apply it better and more selectively. This relatively simple solution may unlock the key to achieve the purpose of planning. The solution is called *decoupling*.

What is decoupling?

Allowing a system to decouple is a way to alter the MRP logic in a profound and impactful way. This is the innovation behind the next evolution of planning, called demand driven material requirements planning (DDMRP). The ASCM dictionary defines decoupling as:

> Creating independence between supply and use of material. Commonly denotes providing inventory between operations so that fluctuations in the production rate of the supplying operation do not constrain production or use rates of the next operation.[3]

Decoupling disconnects one entity from another. This isolates events that happen in one entity or portion of a system from affecting other entities or other portions of the system. But this disconnection does not change the actual product structure. What is being disconnected is the dependent demand requirements on the component from the parent item generated by an MRP run, in addition to disconnecting the execution variability of delay in component supply on scheduled parent order release. This disconnect can only work if there is something between these two items that can simultaneously absorb both directions of variability. This is called *decoupling inventory*.

Decoupling inventory is a level of inventory placed between entities that allows for decoupling (the bidirectional disconnect) to be consistently maintained. Think of decoupling inventory as a firewall isolating the events or environment on one side from the other side. Obviously, the extent and/or thickness of any firewall is relative to the level of potential threat (variability) that is encountered on either side.

In the planning world, decoupling inventory is also commonly referred to as *stock buffers*. Stock buffers are amounts of inventory that provide reliable availability to the consumers of the stock while at the same time allowing for the aggregation of demand orders, creating a more stable and efficient supply signal to suppliers of that stock. This implies two basic requirements for decoupling inventory: an appropriately defined level of inventory to keep that disconnection, and an appropriate management mechanism to maintain and adjust that level over time. If these requirements can be addressed, then decoupling promises a fundamental break from the hard-coded net requirements calculations of conventional MRP.

Next, the question becomes, where to decouple? Decoupling at all product structure connections seems an extreme opposite, one that would serve to dramatically expand the inventory footprint of most companies with extensive BOMs and large end item product lines. Decoupling nowhere is simply a restatement of the core problem of MRP by forcing dependency throughout the product structures. Thus, the only answer is to decouple in the places that are meaningful and impactful for the given environment. The points at which we choose to decouple are appropriately called *decoupling points*. The ASCM dictionary defines decoupling points as:

> The locations in the product structure or distribution network where strategic inventory is placed to create independence between processes or

> entities. Selection of decoupling points is a strategic decision that determines customer lead times and inventory investment.[4]

The transference and amplification of variability is dampened or stopped at the point where decoupling is imposed. There are two key observations to be made:

- At a single point, a bidirectional solution is created. Stopping the demand signal distortion and supply continuity variability simultaneously does not require two independent or different solutions; decoupling is just the opposite. Decoupling implies a relative simplicity that should be welcome in most supply chain and planning environments.
- The decoupling point does not eliminate variability at a discrete level — that is not its intention; rather, its intention is to mitigate the effect of variability on system flow at critical points. The bidirectional variability does begin to transfer and accumulate again in both directions after the decoupling point, but the point is chosen to minimise the impact on total systemic flow.

While the concept of decoupling has been defined for decades, it remained elusive in practical implementation at scale until the development of DDMRP. The concept is simply incongruent with the primary feature and calculating design of MRP. If that were to change, there are some immediate implications or impacts.

Decoupling implications for MRP

Before we speak of what change decoupling brings for MRP, we must first understand what does not change. The use of decoupling has no impact on the three required inputs for MRP. A product structure and its attributes are still required. Inventory policies and status are still required. Finally, there must also be a source of demand. With decoupling in place, however, the way these inputs are considered and processed will be altered.

Netting in a decoupled environment

With decoupling employed, a critical change to the basic netting logic must occur. Decoupling points should never be netted to zero. Netting a decoupling point to zero means that it ceases to be a decoupling point. Thus, a level of decoupling point inventory must be maintained to guarantee a decoupling point's effectiveness. This means that an effective process or method to set and adjust this level is required as well as amended netting logic to determine requirements against the position.

LEAD TIME COMPRESSION

Decoupling point buffer placement has huge implications for planning lead times. By decoupling supplying lead times from that certain point in a supply chain, lead times are instantly compressed to the customer. This lead-time compression has immediate service and inventory implications. Market opportunities can be exploited, while working capital required in the stock buffers placed at higher levels in the product structure (end item or closer to complete) can be minimised.

Figure 6 shows decoupling points placed in a sample environment. Each large 'X' represents a decoupling point.

Their placement has created a series of independent or decoupled planning horizons visible at the bottom of the figure. The length of these planning horizons is determined by the longest time sequence between the decoupling points. This span of time is called *decoupled lead time* (more on that in the next section).

Figure 6 also reveals an additional benefit due to decoupling: its impact on relevant information. MRP, due to the core problem identified in this paper, forces users into using planning horizons that are at least the length of the cumulative lead time. The longer the planning horizon, the more error associated with the demand signal.

Signal accuracy (and relevance) increases with shorter decoupled planning horizons. Furthermore, when the decoupling points are placed inside the sales order visibility horizon, it will allow for the system to exclusively use actual demand — the most accurate and relevant form of demand available for planning. This means that there are no planned orders derived from forecast and no nervousness associated with their reconciliation. By using and maintaining decoupling points, we have essentially found the time we lacked that forced the use of forecasted orders in the first place.

Decoupled lead time

Decoupling point placement has significant implications for the lead times recognised by the planning system for the purposes of calculating supply order receipts and releases as well as the necessary decoupling point protection level. The use of decoupling points in a planning system led to a key innovation — a new form of lead time called decoupled lead time (DLT).

DLT is the longest unprotected sequence in a BOM or distribution network. DLT is essentially a qualified cumulative lead time for a decoupled item. This sequence is determined solely by decoupling point placement. It is the time between either

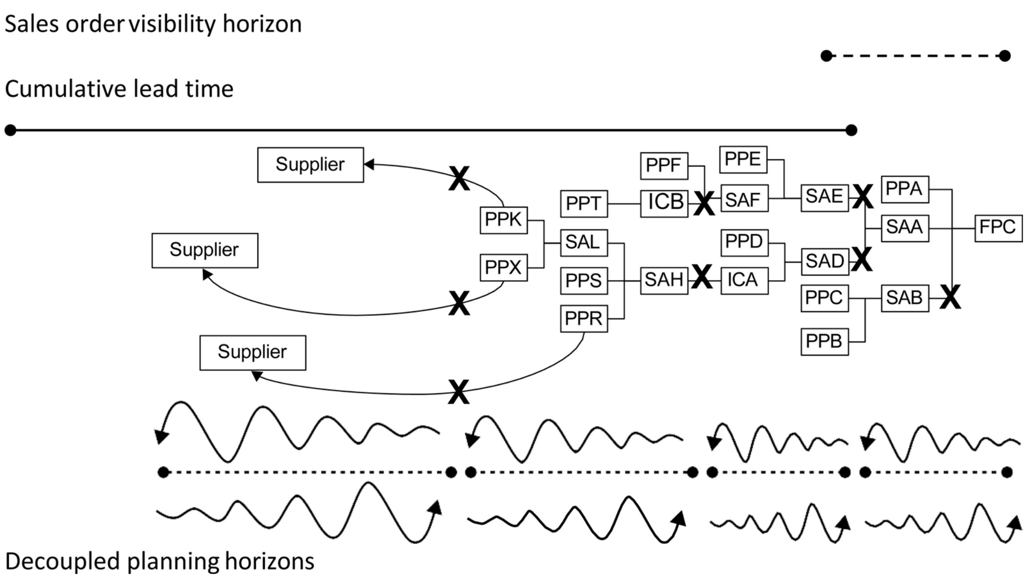

FIGURE 6 Decoupling points and their impact

two decoupling points, the last decoupling point and a customer, or the first decoupling point and a supplier. Note that the cumulative lead time of the item does not change; it is simply no longer relevant for planning purposes when there are decoupling points placed in the item's planning structure.

In Figure 7, decoupling points have been placed and the DLT chains have been bolded. The number inside the circle represents the manufacturing or purchasing lead time of the item. With the decoupling point placement depicted in the figure we can see two distinct DLT chains, one for Finished Product A (FPA) and one for Sub Assembly A (SAA). While the cumulative lead time of FPA remains unchanged at 30 days, the DLT is five days. It is the sum of the manufacturing lead times of FPA, Intermediate Component A (ICA) and Sub Assembly B (SAB). In addition, we can see that SAA has a decoupled lead time of four days — the sum of the manufacturing lead times of SAA, Intermediate Component B (ICB) and Sub Assembly G (SAG). In this example, all purchased components have been decoupled. They have their own DLT that is the same as their purchasing lead times.

Decoupled lead times play a key role in the following:

- Compressing response times to market required ranges.
- Determining realistic due dates when needed.
- Setting decoupling point buffer levels properly.
- Finding high-value inventory leverage points for decoupling point placement. DLT is combined with the matrix BOM to find the relevant key shared components.

The 'decoupled explosion'

How can we understand the practical implementation and impact of decoupling on an order planning and generation mechanism? As discussed earlier, decoupling points are strategic in nature and carefully selected. It is rare that with

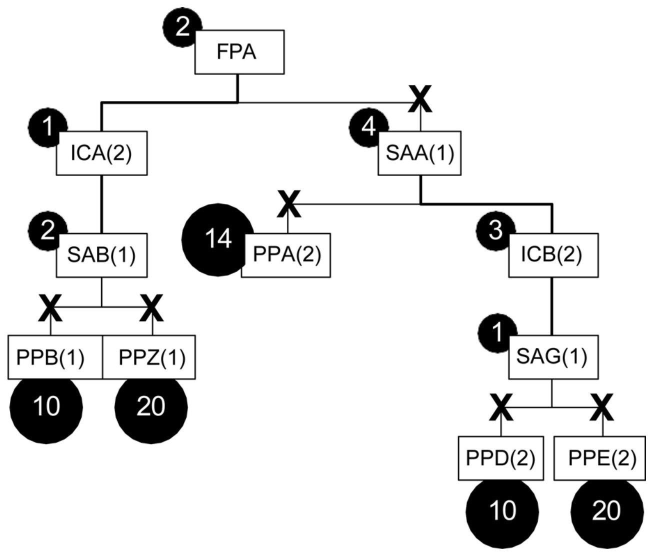

FIGURE 7 Calculating DLT

even moderately extensive product structures there will be decoupling points at all connections. If this is true, it implies that there are still dependent points and/or sequences that must be considered and calculated.

In planning, blending the use of decoupling with the notion of managing and accounting for dependencies is accomplished using a *decoupled explosion*. A decoupled explosion is the cessation of a dependent requirements explosion at any decoupling point. The term itself seems to be an oxymoron, since it literally means independent dependence. Yet that is exactly what is occurring with decoupled explosion.

When a supply order is generated at a higher level, decoupling stops the explosion of the bill of material at decoupling points placed at lower levels. The explosion can be stopped without risk because that decoupling point is buffered and carefully managed with decoupling inventory. The explosion then restarts only when the decoupled position (through an independent calculation) determines that it needs resupply. This independent calculation is called the *net flow equation*.

The explosion then independently restarts (at the appropriate time and for the appropriate quantity, according to its net flow position). These decoupled purchased items will then independently call for resupply at the appropriate time and for the appropriate quantity (based on their independently determined net flow positions).

This type of explosion is obviously different from conventional MRP, where high-level demand over a planning horizon is typically driven all the way through to the purchased component/material level. There are some exceptions to this rule in MRP, but they are simply that—exceptions.

While there are obvious differences, there are also similarities between a conventional MRP explosion and a decoupled explosion. There is independence at the decoupling points, but between decoupling points there is dependence. That dependence between decoupling points is no different from conventional MRP (see Figure 8).

Can conventional MRP decouple?

Decoupling is not a new idea; the concept has been around for many years, but with no practical and impactful way to implement it in MRP. MRP was

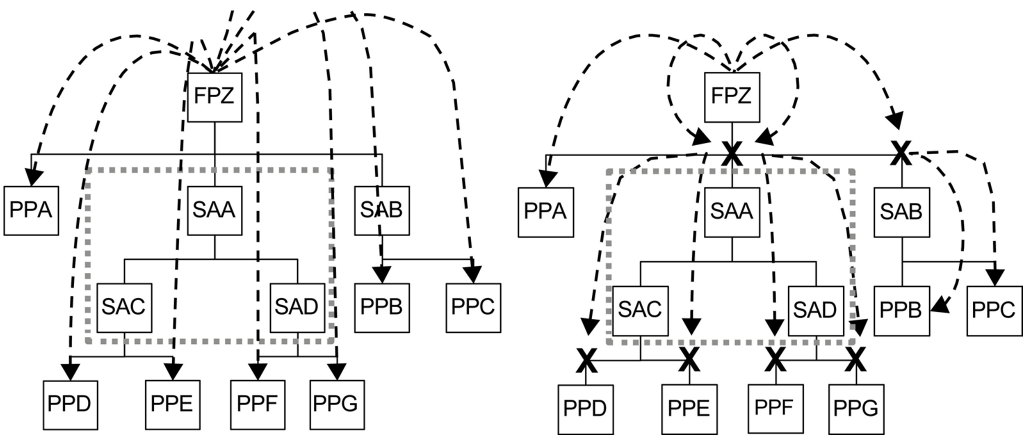

FIGURE 8 Conventional MRP explosion versus a decoupled explosion

designed with the explicit intention of tightly coupling everything so that precise equations could be performed to synchronise the environment. Limited and sporadic amounts of decoupling can occur in MRP but typically it is happenstance, comes with dramatic complications, or may prove to be completely unsustainable. There are at least five tactics that are commonly explored to accomplish decoupling in MRP; none are very effective.

Using safety stock

Safety stock could be added at desired decoupling points in a sufficient quantity to attempt to guarantee on-hand quantities, but as noted earlier, safety stock is not accounted for in the planned order coverage in MRP. This will negate the ability to perform a decoupled explosion based on the safety stock level. Also, when safety stock is consumed, it will generate additional orders and expedites. Safety stock was intended to be a supplementary inventory mechanism to be used in conjunction with conventional demand input in execution; it was not intended to be the primary order-generation mechanism at a decoupling point. At best, the functionality of safety stock is more akin to a fire extinguisher than the firewall of a true decoupling point.

Using order points

Order points could be added at desired decoupling points. Order point was, in fact, designed to be a primary order-generation mechanism, but this approach also incorporates the use of safety stock. This means that the problems that come with safety stock are also realised with order points. Additionally, order points do not recognise or qualify actual future demand. This means they are blind to potential spikes related to actual orders. To cover this blind spot, additional levels of inventory must be deployed.

Over-planning

Over-planning is a term used to describe intentionally ordering more than is deemed required in the hope that there will be an abundance of on-hand inventory throughout the system to stop the effects of nervousness by absorbing new or adjusted dependent demand requirements. To say this is an inefficient approach is an understatement. The working capital load this requires would be enormous and unaffordable.

Using a stop explosion flag

Many conventional MRP systems allow the deployment of a *stop explosion flag* or an *externally planned* setting for specifically designated parts. In this case, the explosion can be stopped at that position, but there are two subsequent challenges. First, the buffer level must be constructed at that position, and secondly, a mechanism must be employed to restart the explosion when needed. Conventional MRP systems cannot automatically restart the explosion when a stop explosion flag or externally planned manual setting is used. The question becomes, when do you restart the explosion? This means that it will fall to planners to manually restart the explosion for each decoupled intermediate and purchased position. This would most commonly be accomplished using an assortment of spreadsheets.

Using a multilevel MPS

A multilevel MPS process can be used. Under this case, an explosion can be stopped and then restarted, but at what cost? Setting up and effectively maintaining multiple master production schedules and their connections proves extremely difficult for most planning teams. Furthermore, the decoupling buffer levels must have a mechanism that alerts planning when to reorder and there is no mechanism in conventional MRP systems to effectively do that (see safety stock and order point discussion).

Mitigating the bullwhip effect

Decoupling becomes crucial in preventing nervousness and ensuring supply continuity. This has huge implications not just for one company but for entire supply chains that embrace the concept (see Figure 9).

Historically it has been thought that connecting MRP systems more tightly together and with faster calculations would result in better synchronisation. Yet what we have learned is that without forcing some points of independence, variability transfers, amplifies and accumulates within and across companies with increasing distortion. Flow breaks down inside and across entities, resulting in unacceptable overall performance.

Decoupling in the 'real world'

The best theory may sound terrific — but does it work in the real world? Table 2 lists just a few of the varied industries that have applied these concepts.

Figure 10 shows the results companies have achieved as reported by Camelot Consulting with research from the French Association for Supply Chain Management.

SUMMARY

Convention's emphasis on the necessary precision of both inputs and outputs directly undermines its credibility and usability. With convention, everything must go exactly according to plan to reach the expected outcome. We know that given the increasing volatility, uncertainty and complexity in today's environment this is unrealistic. Instead, decoupling allows for — even assumes

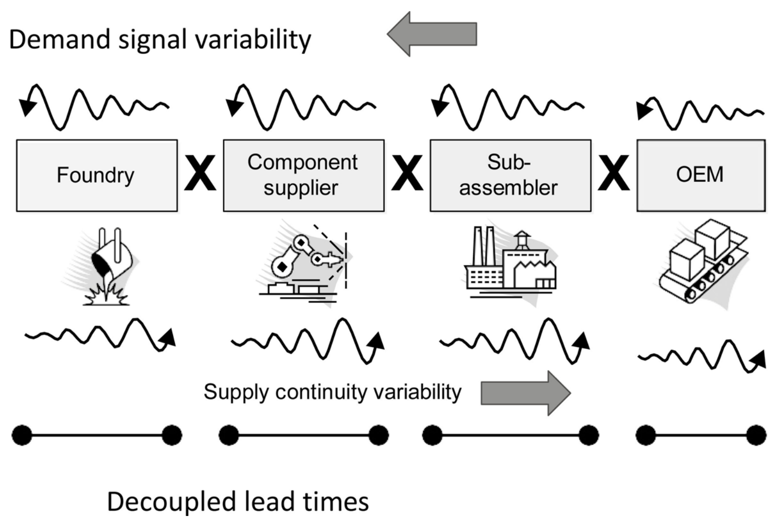

FIGURE 9 Decoupling at the supply chain level

TABLE 2 Real-world examples of decoupling

Macquila Internacional de Confeccion SA (MIC)	Consumer products	MIC designs, produces and sells children's garments under licence from companies such as Disney and Mattel. MIC also supplies direct sales channels for ladies' garments.
Oregon Freeze Dry (OFD)	Consumer products	OFD is the world's largest custom freeze-drying operation in the world with plants in North America, Europe and Asia. Craig Jolly, Director of IT, presents their journey implementing DDMRP methods since 1997. OFD is one of the first adopters of what would become DDMRP methodology in the world.
Tube Forgings of America, Inc. (TFA)	Forging, steel and machining	TFA has been manufacturing welding fittings since 1955 for industries ranging from oil refining to chemical and petrochemical processing, from gas transmission to power generation, including nuclear, and from shipbuilding to a broad assortment of commercial construction applications. On the eve of their largest capacity expansion ever, Wally shared how demand driven methods protected the company through the downturn and has positioned TFA to absorb huge growth over the last three years and into the future.
Productos Tubulares	Forging, steel and machining	Productos Tubulares won the Internationally recognised Ptak Prize for Supply Chain Excellence in 2016. Their presentation described the implementation of a demand driven operating model (DDOM) and the amazing results achieved in a relatively short period of time.
ABE Construction Chemicals	Petrochemical	ABE, a South African company and part of the global French company Chryso Group, presented their DDMRP results. ABE has over 17,500 stock keeping units (SKUs). They reported that many products have seen 200–300% growth in annual stock turns — some achieving 45–52 turns per year. Back-orders as a percentage of sales have dropped from 16.3% to 2.5% with a 54% inventory reduction.[5]

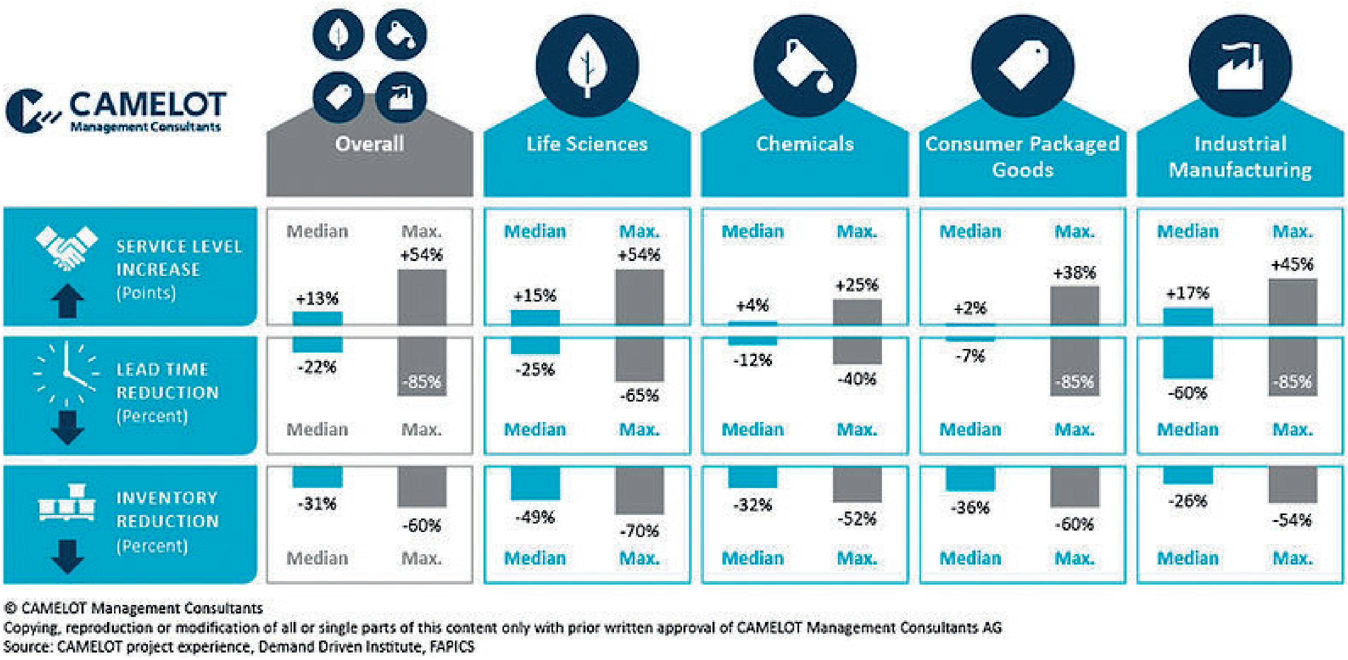

FIGURE 10 Composite benchmarking of DDMRP and DDOM implementations

— that most things will not go according to plan. It presupposes variability so that it can protect and promote flow by being approximately right rather than precisely wrong. This means that the true purpose of planning is more achievable using decoupling.

Given the increasing challenges faced by conventional MRP systems in the VUCA world, enabling decoupling as part of DDMRP makes sense. Companies have extended and integrated supply chains globally and have made them more complex and fragile at the same time. These longer and more complex supply chains are subject to much higher levels of variability and are much harder to plan. Breaking dependencies in key places dramatically simplifies the planning equation and provides shorter horizons with much more relevant information.

The concept of decoupling poses an ironic situation. To promote and protect the flow of relevant information and materials in a system, the flow must strategically and purposefully be slowed or temporarily interrupted at certain critical points (decoupling point buffers). The size of the decoupling point buffers defines the length of the slowdown or interruption at these caching points. This approach is pragmatic and proven across a variety of industries around the world.

REFERENCES

(1) Association for Supply Chain Management (ASCM), 'ASCM Dictionary', p. 41, available at https://learn.ascm.org/s/ascm-dictionary/view-dictionary?fId=069R3000006PlDHIA0 (accessed 19th August, 2024).
(2) *Ibid*., p. 25.
(3) *Ibid*., p. 52.
(4) *Ibid*.
(5) More case study details can be found at Demand Driven Institute, 'Demand Driven Case Studies', available at https://www.demanddriveninstitute.com/case-studies (accessed 19th August, 2024).