



Synchronous Flow Manufacturing: la sesta rivoluzione

Come e perché è possibile abbassare le scorte, ridurre i lead-time, aumentare la produttività, avere maggiore affidabilità e rapidità nelle consegne simultaneamente: non è un sogno, ma il risultato dell'ultima evoluzione del Manufacturing nata negli States

Il Manufacturing nasce con l'avvento dell'homo sapiens che sviluppa l'abilità di forgiare gli utensili per modificare l'ambiente circostante e acquisire la supremazia sulle altre specie. Grandi civiltà si succedono nell'arricchire la storia dell'Umanità, ma dal punto di vista della tecnologia la seconda grande discontinuità si manifesta solamente alcune migliaia di anni dopo, quando l'energia animale viene sostituita dalla macchina a vapore prima e dell'energia elettrica poi. Inizia così l'era moderna. L'Uomo non impiega molto a sfruttare questa nuova grande opportunità; nascono le grandi fabbriche che rendono disponibili ingenti quantità di beni. Migliora sensibilmente la qualità della vita di larghi strati delle popolazioni. Si sviluppano nuovi modelli di organizzazione sociale e del lavoro. Siamo nei primi del '900: inizia il taylorismo. Il lavoro, che per secoli ha avuto connotazione artigianale, diventa *scientifico*. Si deve produrre quanto più possibile, in modo ripetitivo e replicabile, perché grandi masse di un nuovo soggetto sociale, il consumatore, sono in grado di assorbire e contendersi tutto quanto verrà fabbricato. Si è in un regime in cui la domanda supera l'offerta, ovvero la capacità produttiva disponibile non riesce a soddisfare la domanda. Per descrivere quell'età dell'oro Henry Ford conia un famoso aforisma secondo cui “i clienti possono ordinare qualsiasi modello di automobile, purchè sia una Ford T di colore nero”. In questa fase storica si formano metodologie, modelli culturali e mentali ancora oggi presenti nel mondo industriale; naturalmente dopo un secolo alcuni di essi sono un po' “datati”. Una nuova rivoluzione avviene negli anni '60 e '70¹ con l'introduzione del MRP, che promuove una nuova attenzione alla pianificazione dei materiali. I progressi della tecnologia hanno cambiato il rapporto tra i costi della manodopera e dei materiali. Nei decenni successivi l'informatica entra con sempre maggiore forza nella gestione d'impresa, il MRP evolve in MRPII² e questo successivamente è inglobato negli ERP.

¹ MRP (Material Requirements Planning) vede la prima applicazione operativa nel 1960 a J.I Case, Racine – Wisconsin. La diffusione nel mondo industriale si manifesta negli anni '60 e '70.

² Anche Manufacturing Resources Planning



Agli inizi degli anni '80 l'industria automobilistica americana sperimenta una drammatica perdita di competitività a vantaggio di quella giapponese; la Toyota ha messo a punto il suo straordinario Sistema di Produzione³, che verrà successivamente conosciuto con l'espressione sintetica Just-in-Time. Un cocktail di metodologie operative di fabbrica, frutto della dedizione e dell'impegno di migliaia tra operai e tecnici Toyota, e della genialità di alcuni dei suoi ingegneri, rivoluziona nuovamente il Manufacturing. In Occidente la nuova filosofia evolve nella Lean Production e nel Lean Thinking, il cui senso profondo è il focus sulla creazione di valore per il cliente.

Negli anni '80 e '90 il fisico israeliano Ely Goldratt elabora la Theory of Constraints (TOC); l'impresa costituisce un sistema composto da processi interdipendenti la cui prestazione complessiva è determinata da un numero limitato di fattori, definiti constraint (vincoli). La focalizzazione sulla gestione dei constraint consente di finalizzare le limitate energie dell'impresa verso il conseguimento di alti livelli di prestazione, in antitesi alla consuetudine diffusa di disperdere le forze dell'azienda attraverso la indiscriminata ricerca di ottimi locali, la cui somma non produce l'ottimo globale.

La TOC non avuto il successo e la diffusione che avrebbe meritato, anche a causa del settarismo di una parte dei suoi adepti, ma rappresenta un passaggio chiave nell'evoluzione del Manufacturing. TOC e Lean Manufacturing si sono ibridati e potenziati vicendevolmente in una nuova metodologia nota come Synchronous Flow Manufacturing, oggetto di questo articolo.

Le dimensioni della competizione

Una metodologia gestionale non è buona o cattiva, migliore o peggiore di altre in assoluto: lo diventa anche in ragione delle condizioni di contesto in cui viene applicata. Negli ultimi decenni si sono sommati gli effetti del miglioramento delle tecnologie produttive e della globalizzazione, che hanno ampliato ed esteso a tutto il mondo la base produttiva. Nella maggior parte dei settori industriali ciò ha portato ad un eccesso di capacità produttiva rispetto alla domanda riscontrabile, anche se, per contro, negli ultimi anni si sono aperti nuovi mercati di sbocco per le imprese. La selettività nell'acquisto da parte dei consumatori, l'accesso a fonti di approvvigionamento globali tramite internet, e la crescita della forza delle catene di distribuzione hanno sbilanciato ulteriormente i rapporti tra offerta e domanda nella direzione di quest'ultima.

Contrariamente a quanto avveniva nell'età dell'oro di Henry Ford, le risorse scarse sono sempre più i clienti, e sempre meno i fattori produttivi. In regime di eccesso di offerta, un sistema logistico-

³ TPS – Toyota Production System messo a punto da manager e ingegneri tra cui sono ben noti Taichi Ohno e Shigeo Shingo



produttivo dovrebbe essere fortemente orientato alla cattura e ritenzione del maggior numero possibile di clienti, e non solo alla ottimizzazione della propria efficienza. Di ritorno, avere più clienti induce comunque anche una maggiore saturazione delle risorse produttive.

I manager più accorti sono interessati a comprendere come il loro sistema logistico-produttivo si relazioni con il mercato, e come possa contribuire ad accrescere il vantaggio competitivo delle catene del valore dei clienti, e dei clienti dei clienti.

Vi sono delle “dimensioni” rispetto alle quali è possibile valutare l’attitudine di un sistema logistico-produttivo nel contribuire alla creazione di valore delle catene che operano a valle. Per confrontarsi positivamente con il mercato occorre riflettere su come il servizio erogato entri nelle catene del valore dei clienti che nutrono aspettative – implicite o esplicite - nei nostri confronti. Nella maggior parte dei casi le attese sono ragionevoli, ma condizionate dalla natura competitiva della concorrenza. Ciò significa che il livello della nostra prestazione non può essere valutato in assoluto, ma per confronto con quello dei competitori.

L’*affidabilità*, intesa come capacità di rispettare gli impegni, innanzitutto sulle tempistiche di consegna, sembra essere la dimensione più apprezzata; è essenziale per i piani di produzione e di vendita dei clienti potere poggiare su certezze rispetto alle forniture di materiali e servizi.

L’estensione del concetto di affidabilità ci porta a quello ben diffuso di *perfect order fulfillment*, con il quale intendiamo la capacità di far avere il prodotto giusto, al posto giusto, al momento giusto, nelle giuste condizioni e con l’imballo corretto, nella giusta quantità, con la giusta documentazione, al cliente giusto. Sembra poco?

La seconda dimensione è la *rapidità* con la quale si reagisce all’ordine del cliente. Come il precedente, questo è in generale un problema di supply-chain, nel senso che questo tipo di prestazione dipende dall’organizzazione dell’intera filiera e non solo dell’ultimo anello della catena. Si badi a non confondere il tempo di risposta – oggetto di prestazione sensibile al cliente - con quello totale di attraversamento, che hanno significati ben distinti e di cui il primo rappresenta una porzione del secondo.

La terza dimensione è la *flessibilità*. Questo tema è stato ampiamente dibattuto sotto una molteplicità di punti di vista. La flessibilità ha tre nature: al volume, al mix e al prodotto. Gli orientamenti più moderni tendono ad evidenziare come si dovrebbero considerare due livelli di profondità nella flessibilità. Si parla di *agility* e di *adaptability* per indicare l’attitudine delle supply-chain a reagire rispettivamente a variazioni di condizioni nel breve periodo, con modifiche tattiche, e nel medio periodo, con cambiamenti più strutturali. I sistemi logistico-produttivi che hanno ben sviluppato queste caratteristiche sembrano essere quelli che sopravvivono meglio nel tempo.

Infine non si possono trascurare *i costi*, che incidono anch'essi sulla capacità di sistema di realizzare performance di successo. Questo tema è probabilmente il più discusso in letteratura; ci preme in questo contesto sottolineare come la leva dei costi possa essere usata per conseguire un maggior utile a sostegno della capability presente e futura dell'impresa, o come strumento per guadagnare quote di mercato ove necessario attraverso politiche di prezzo mirate.

La value proposition, ovvero il carattere distintivo nella creazione di valore per i clienti, è definita dalla modalità con cui le aziende o le supply-chain articolano, in rapporto ai competitori, le proprie performance nelle dimensioni descritte.

Il manufacturing, come gli altri processi aziendali e in quanto parte di quello più ampio di supply-chain, deve dare il proprio contributo positivo al miglioramento di tali prestazioni. A tale scopo è bene conoscerne innanzitutto le dinamiche; sovente accade di imbattersi in contesti nei quali dei pur bravi manager di produzione operano trascurando alcune delle regole che governano i sistemi logistico-produttivi. Nel seguito di questo articolo presentiamo un'analisi su una serie di problemi gestionali endemici del manufacturing, e su una metodologia innovativa che porta all'eccellenza delle performance.

Problemi del manufacturing: effetti del sistema tradizionale

Il modello proposto dal MRP ha fortemente condizionato la gestione dei processi produttivi negli ultimi decenni; senza dubbio l'adozione del MRP è stata favorita anche dalla crescita esponenziale della capacità di calcolo offerta dalla tecnologia informatica. Nonostante l'evoluzione degli anni '80, il sistema MRP è comunque sempre basato sugli stessi concetti di fondo che ne hanno permesso la creazione e la divulgazione. Il MRP opera con la logica "batch and queue", che spinge il materiale all'interno del processo produttivo senza verificare la disponibilità di capacità per lavorare il materiale stesso.

Vediamo l'esempio in figura 1 per semplificare il concetto.

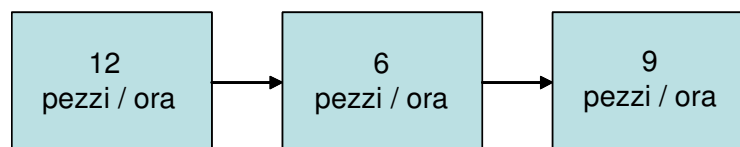


Figura 1

Il processo produttivo è composto da tre centri di lavoro in sequenza che lavorano pezzi di un solo item a differenti velocità: il primo centro opera a 12 pezzi/ora, il secondo a 6 pezzi/ora, il terzo a 9



pezzi/ora. Ipotizziamo che sia un sistema perfetto, entro il quale i tempi di processo siano sempre rispettati e non vi siano condizioni che perturbano il funzionamento del sistema stesso. La gestione push prevede di pompare materiale all'interno del sistema per soddisfare previsioni e ordini. La capacità produttiva effettiva è determinata dalla risorsa più lenta, e pari a 6 pezzi/ora. Se i materiali sono immessi via via che il primo centro di lavoro li completa al ritmo di 12 pezzi/ora, in breve tempo il processo produttivo collassa, perché si forma una coda di materiale in attesa a monte del secondo centro di lavoro.

Vediamo come l'aumento di materiale in corso lavorazione – work in progress – ha impatto anche sul tempo di attraversamento complessivo. Ipotizziamo di evidenziare un pezzo con colore differente prima di introdurlo nel processo in modo da avere una sorta di tracciante nel flusso. Questo pezzo sarà lavorato immediatamente dal primo centro di lavoro, resterà in attesa al secondo centro fintantoché non sarà smaltito il cumulo dei pezzi in coda, sarà quindi processato senza nessuna attesa dal terzo centro di lavoro. Da questo esempio appare chiaro come rilasciare il materiale in quantità indiscriminata, cioè senza tener conto della capacità produttiva del sistema, ovvero della velocità con cui lo smaltisce, influisca negativamente sul tempo di attraversamento complessivo a causa delle code. Questo effetto è espresso da un indicatore numerico noto come VAR – Value Added Ratio – che esprime il rapporto tra le attività a valore aggiunto e il tempo di attraversamento. Nell'esempio appena descritto il valore aggiunto è dato dalla somma dei tempi di lavorazione necessari al completamento di un pezzo sui tre centri di lavoro. Il tempo di attraversamento è composto dalla somma dei tempi di lavorazione e del tempo di coda in attesa davanti al secondo centro di lavoro: ne consegue che il lead time – a denominatore – tende a crescere quanto maggiore è il tempo trascorso in coda. L'effetto complessivo evidenzia un sistema collassato dall'eccessivo afflusso di materiale, espresso in termini numerici da un valore di VAR molto basso.

Introduciamo ora una regola che governa il funzionamento dei processi produttivi.

Esiste una relazione di proporzionalità, nota come Little's law, tra il livello di materiale presente nel processo (I), e il tempo di attraversamento (T), dove R è la velocità del processo.

$$I = R \times T$$

Per comprendere, si pensi di bere una bottiglia di acqua minerale al giorno (R = 1 bottiglia / giorno). Se si fa la spesa al supermercato una volta alla settimana, si acquisteranno sette bottiglie, onde avere una scorta fino al successivo ripristino. Se si acquista l'acqua nel negozio sotto casa ogni giorno, si compra una bottiglia per volta. Nel primo caso la scorta iniziale sarà di sette



bottiglie, nel secondo caso di una bottiglia. A beneficio dei pignoli, nell'ipotesi di consumo regolare, la scorta media vale la metà.

Questo implica che tutte le volte che noi rileviamo un lungo tempo di attraversamento, ci dobbiamo attendere elevate scorte di materiale, e viceversa, quando generiamo livelli di scorta lungo il processo, faremo i conti con lunghi tempi di attraversamento. Si badi bene che il tempo di attraversamento va considerato nella sua interezza. Nella maggior parte dei casi la quota più grande è generata dalle code di attesa davanti alle risorse, che è come dire che si opera in condizioni di basso Value Added Ratio.

La legge empirica appena enunciata fu formulata e poi dimostrata matematicamente da John Little. Preghiamo di non trasformare lo shop floor in un laboratorio newtoniano per cercare di confutarla o dimostrarla sperimentalmente: si faccia tesoro del concetto, la cui comprensione o non comprensione, è però nell'esperienza alla base di molti successi ed insuccessi.

Quanto affermato ha valenza generale ed è applicabile ad ogni tipo di processo che contenga un flusso, sia esso fisico o informativo. Per esempio l'order-entry degli ordini di vendita segue la medesima logica; se si accumulano i fax dei clienti questi generano un backlog di ordini che hanno anch'essi un tempo di smaltimento inversamente proporzionale al tempo medio di imputazione di un ordine nel sistema informativo. In questo contesto si assume come unità di flusso la riga d'ordine. Quando si effettuano analisi sulle supply-chain, o su porzioni di esse, si devono considerare flussi fisici ed informativi che concorrono a determinare il tempo di attraversamento totale del processo.

Sulla base del ragionamento esposto, consideriamo ora l'effetto del tempo di attraversamento e della quantità di materiale sul sistema di misura delle performance, introdotto nella parte iniziale dell'articolo. In un sistema Make To Stock, l'affidabilità è strettamente connessa alla disponibilità del prodotto finito a magazzino: lunghi tempi di attraversamento determinano elevati livelli di scorta e, conseguentemente, considerevole impegno finanziario e alti costi per l'impresa. In un sistema Make To Order lungo lead time vuol dire lunghi tempi di risposta al cliente, o, in altre parole, bassa prestazione in termini di rapidità, a cui spesso si associa anche bassa affidabilità quando i tempi di attraversamento, oltre ad essere lunghi, finiscono per essere sempre meno certi. Tutto ciò ci porta a comprendere meglio le ragioni per le quali è desiderabile operare con un sistema logistico-produttivo leggero, ovvero con bassi livelli di materiale al suo interno.

Un siffatto sistema è rapido e può trarre le proprie promesse di consegna ad un tempo non molto lontano; in tal modo cresce ovviamente la probabilità di essere credibili, perché chiunque comprende come sia più facile stimare cosa sarà realizzato tra una settimana piuttosto che tra un



mese. Inoltre poco materiale consente anche migliore flessibilità: diventa un'impresa ardua cambiare rotta quando la produzione e i fornitori annegano negli ordini inevasi e lo shop floor è fisicamente vessato da tanto di quel materiale che non si riesce nemmeno a camminare.

Molti ordini rilasciati in fabbrica innescano poi il multi-tasking, quella pratica per cui sono in lavoro molte commesse, e si passa da una all'altra per soddisfare le nuove priorità dettate dai ritardi. Ma ciò dura fin tanto che quel set di priorità ha validità perché, in un gioco che sembra non aver fine, le commesse sospese hanno acquisito un ritardo tale per cui si rende necessario farle correre nuovamente a danno di quelle che le avevano soppiantate. In queste condizioni l'expediting diventa la regola per gestire la produzione e non le eccezioni: si pagano costi per straordinari, riattrezzaggi e trasporti speciali, si ha la sensazione che il tempo non basti mai, e come beffa la frustrazione di non vedere uscire un chiodo dalla fabbrica!

Alti livelli di materiale nella pipeline creano anche problemi all'introduzione di nuovi prodotti, perché al momento del loro lancio sul mercato si deve decidere cosa fare di quello in corso di lavorazione o nei magazzini di prodotto finito. Le scelte possibili sono l'esaurimento delle scorte con conseguente ritardo del lancio e perdita di opportunità, oppure la vendita a prezzi più bassi per esaurire rapidamente le giacenze, con impatto negativo sulle strategie commerciali dell'azienda. Nei casi estremi si arriva alla rottamazione del prodotto obsoleto. In ogni caso un danno.

Queste considerazioni sono derivate dall'analisi del funzionamento di un sistema semplice, definito perfetto in quanto privo di qualsiasi forma di perturbazione; vediamo nel prosieguo l'effetto di un altro fenomeno che affligge i processi produttivi.

Variabilità

Consideriamo ora un altro elemento, la variabilità, che influisce sulle prestazioni dei sistemi produttivi con effetti potenzialmente devastanti.

Le attività produttive sono soggette a perturbazioni che derivano da "eventi casuali" quali mancate consegne dei fornitori, guasti alle macchine, annullamento di ordini cliente, assenteismo. Inoltre i piani produttivi sono costruiti assumendo la natura deterministica di alcune grandezze, quali i tempi delle lavorazioni, di set-up, ecc. Nella realtà queste grandezze appartengono a popolazioni statistiche, per cui il loro valore reale non può essere determinato a priori con esattezza. Ci riferiamo all'insieme di questi fenomeni con l'accezione *variabilità*. Vediamo rapidamente come si manifesta: se un lotto di produzione subisce un ritardo in una lavorazione, se lo porta nella lavorazione successiva, alterando il piano di lavoro della risorsa corrispondente. In questo modo le perturbazioni si propagano rapidamente in punti del processo lontani da quello di origine. Il



risultato è intuitivo: gli appuntamenti previsti non vengono rispettati e il sistema tende ad andare fuori controllo.

Molte organizzazioni operano con la logica del pompiere, impegnate a spegnere gli incendi che divampano improvvisamente nei punti meno prevedibili, in un continuo mutamento dei piani di lavoro per assecondare l'insorgere di cambiamenti interni ed esterni. La variabilità mina il funzionamento del processo produttivo, con impatti negativi facilmente immaginabili sia sul livello di servizio offerto al mercato, sia sull'efficienza complessiva del sistema manifatturiero. Inoltre la variabilità aumenta l'incertezza e con essa l'insicurezza, per cui si innesca un meccanismo di protezione mediante ulteriori scorte le cui conseguenze sono state illustrate precedentemente. La causa prima di tutto questo è il non controllo della variabilità⁴.

Alla necessità di gestire la complessità, l'IT ha cercato di porre rimedio sfruttando l'evoluzione tecnologica e lo sviluppo del software: di fatto ha messo a punto soluzioni complesse per risolvere un problema complesso. Gli algoritmi contemplan situazioni intricate – frutto anche dell'evoluzione generata dall'esperienza di precedenti implementazioni del software stesso – che richiedono grandi quantità di dati, onerosi sia nella raccolta che nella manutenzione. La velocità di calcolo permette di ripianificare sempre più frequentemente, ma i sistemi IT non sono in grado di fornire soluzioni che incarnano il completo funzionamento di un sistema manifatturiero! Gli algoritmi sono perfetti, ma i cambiamenti del mondo reale sono sempre più rapidi: i piani non sono mai abbastanza aggiornati, e non sono validi che per poche ore, a volte per pochi minuti.

La corsa a inseguimento tra problemi complessi e soluzioni complesse non sembra essere la via migliore per controllare il processo manifatturiero: il Synchronuos Flow Manufacturing ci guida verso una riprogettazione del sistema logistico-produttivo in grado di migliorare le prestazioni semplificandone la gestione e il controllo.

Synchronous Flow Manufacturing

In sintesi, un sistema leggero è intrinsecamente anche rapido, flessibile e affidabile; ciò è quanto abbiamo visto essere richiesto al manufacturing e alle supply-chain per ottenere un premio dal mercato.

Le aziende che abbiamo aiutato ad intraprendere questa strada hanno sperimentato quanto sia premiante questa filosofia. Naturalmente occorrono degli strumenti che traducano queste idee in fatti concreti. Il Synchronous Flow Manufacturing, d'ora in poi SFM, è una disciplina che fonde

⁴ La scuola di Deming ha compiuto molti studi rispetto a queste problematiche.



Lean Production e Theory of Constraints, ponendo finalmente rimedio ai problemi precedentemente evidenziati.

SFM è una metodologia di tipo *demand-driven* o - se preferite - *pull*. Come anche indicato dal guru della Lean Production Jim Womack, il senso profondo del pull consiste nel portare il mercato all'interno del sistema produttivo. In termini operativi ciò significa fare tirare la produzione direttamente dagli ordini clienti quando possibile. Questi risalgono la filiera attivando le risorse. Quando i tempi di attraversamento non sono compatibili con quanto richiede il mercato – interno o esterno che sia -, la domanda viene soddisfatta con scorte precedentemente predisposte. Le scorte però non sono spinte nel processo produttivo su base previsionale; la previsione sul consumo, unitamente alla velocità di ripristino, viene utilizzata per definire il livello di scorta necessario. Quando si consuma il prodotto o il semilavorato, si genera un ordine che risale la filiera per ripristinare al livello prefissato la scorta consumata. Si parla in questo caso di *replenishment*. Il kanban è un'applicazione di questa logica.

È riconosciuto da tutta la letteratura e da chiunque lo abbia sperimentato che i sistemi pull operano con livelli di scorta più bassi dei sistemi push, e per quanto enunciato dalla Little's Law, anche con tempi di attraversamento più brevi.

Nel semplice sistema produttivo utilizzato nell'esempio di figura 1, si è dimostrato come un flusso produttivo non bilanciato sulla capacità produttiva reale tenda ad accumulare materiale prima della risorsa più lenta. Questo è proprio ciò che avviene con la produzione governata dal MRP: il materiale è rilasciato e spinto in produzione da monte verso valle senza il controllo preventivo della effettiva capacità produttiva.

Nel SFM invece il flusso dei materiali è regolato sulla base della effettiva capacità produttiva dei colli di bottiglia, come insegnato dalla Theory of Constraints. Ciò avviene ribaltando il piano delle consegne sul collo di bottiglia, schedato a capacità finita.

Il passaggio successivo è il controllo del rilascio degli ordini alla produzione, per evitare che questa sia ingolfata da materiali che non possono essere smaltiti: l'ingresso dei materiali nel flusso avviene alla velocità con cui il collo di bottiglia li processa. Abbiamo oramai imparato che l'unica conseguenza di un comportamento diverso sarebbe quella di produrre accumuli di materiale, allungamento dei lead-time e inefficienze produttive. Già la Lean Production ci ha insegnato ad aggettivare come *muda* (spreco) la pratica di *overproduction*. Nel SFM la fase di controllo dell'introduzione dei materiali nel processo è detta *gating* (in inglese gate = cancello). Di fatto è un ulteriore segnale che, in logica pull, risale la corrente fino alla sorgente del processo, autorizzando il rilascio dei materiali nella stessa sequenza con cui li dovrà processare il collo di bottiglia.



Un sistema siffatto ha dunque due punti di controllo: il collo di bottiglia, che viene schedulato sulla base delle priorità di consegna recepite dal programma di produzione, e a ritroso, in conseguenza di questa schedulazione, l'operazione di inizio lavorazione degli ordini di produzione, precedentemente definita gating.

Questa metodologia genera un controllo costante sul livello di materiale, perché in prima approssimazione un ordine di produzione può entrare solo quando ne è uscito un altro: ciò significa anche controllare i tempi di attraversamento. Le risorse non collo di bottiglia hanno l'ordine di lavorare il materiale appena possibile quando è disponibile, al fine di mantenere il flusso costante e regolare. Negli insegnamenti Lean questo costituisce un altro punto cardine, noto come *Flow*. A questo effetto collabora anche la trovata più geniale della Theory of Constraints, il buffer management. La Theory of Constraints propone una soluzione brillante per attutire l'impatto della variabilità, che introduciamo con una metafora. Se abbiamo un treno a mezzogiorno non pianificheremo certo di essere in stazione alle ore 11.58, perché questo comporterebbe un'altissima probabilità di perderlo. Punteremo invece ad essere in stazione con un certo anticipo temporale rispetto all'evento atteso (in questo caso la partenza del treno); chiamiamo questo anticipo *time buffer*. Il time buffer è dunque l'anticipo che prevediamo sull'inizio dell'evento. Quanto deve essere grande il time buffer? Tanto più grande quanto più è elevata la variabilità sul percorso fino alla stazione. Se siamo in un giorno ferialo con molto traffico, l'anticipo dovrà essere grande perché grande è l'incertezza sul tempo necessario a percorrere il tragitto fino alla stazione; di domenica il time buffer potrà essere sicuramente ridotto.

Uscendo dalla metafora del treno, nel mondo produttivo realizzare un time buffer significa fare in modo che la disponibilità del materiale sia pianificata in anticipo rispetto all'utilizzo previsto. Di quanto tempo? Di un time buffer!

Il time buffer si concretizza attraverso i materiali che si avvicinano secondo la schedulazione davanti alla risorsa che intendiamo proteggere dalla variabilità dei processi a monte. Si usano i time buffer per proteggere il collo di bottiglia e la spedizione: il primo – denominato constraint buffer – salvaguarda la quantità di materiale che esce dal processo e l'efficienza di sistema, il secondo – definito shipping buffer – il rispetto delle date di consegna. I nomi constraint buffer e shipping buffer identificano il medesimo concetto di time buffer precedentemente spiegato: la differente denominazione è dovuta alla diversa posizione in cui sono applicati all'interno del processo produttivo.

Il buffer diventa il termometro della fabbrica. Controllando il modo in cui questo viene alimentato con la logica del semaforo verde-giallo-rosso, si realizza il monitoraggio del flusso e del processo

produttivo. Il *buffer management* diventa uno degli strumenti operativi fondamentali per la gestione della fabbrica e contraddistingue in maniera sostanziale la Theory of Constraints e come derivazione il SFM, come potenti metodologie sistemiche⁵ in grado di gestire e assorbire gli effetti della variabilità nel manufacturing.

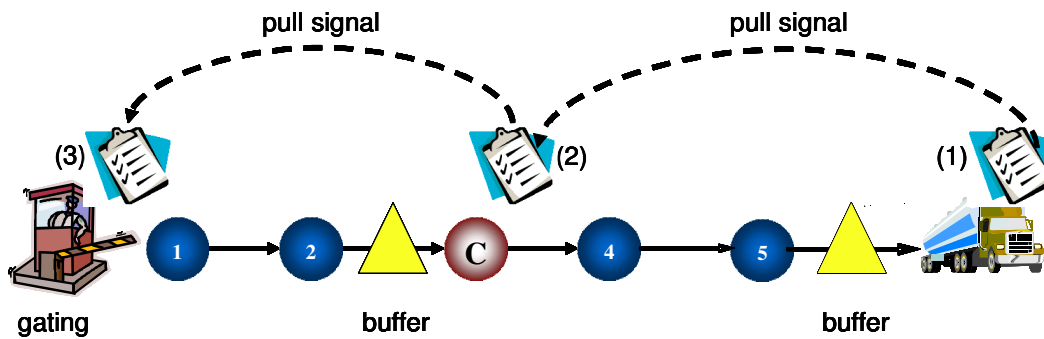


Figura 2

In figura 2 è riportato lo schema di funzionamento del SFM applicato ad un processo manifatturiero, che in logica pull è da leggersi da destra verso sinistra risalendo il flusso dalla spedizione verso il rilascio dei materiali alla produzione. Il prodotto è realizzato con cinque operazioni in sequenza lineare rappresentate dai pallini blu, di cui la terza avviene su una risorsa collo di bottiglia (C). Il mercato richiede determinate consegne, che si traducono in un piano di spedizioni (1). Queste risalgono la corrente con un segnale *pull* fino al collo di bottiglia, schedato – lui solo! - a capacità finita per rispettare il piano consegne. La schedulazione del collo di bottiglia (2) diventa il pacemaker⁶ di tutto il processo; infatti da qui si genera un nuovo segnale *pull* che risale il flusso fino al punto di rilascio dei materiali (3). Qui avviene il gating, cioè si regola l'ingresso dei materiali per assecondare la schedulazione (2) del collo di bottiglia, aprendo il cancello in modo che sia mantenuto costante il livello di WIP all'interno della filiera. I rilasci sono calcolati in modo che i materiali abbiano il tempo di ricevere le operazioni di trasformazione e vi sia un margine di sicurezza – il time buffer, eccolo qua finalmente! – che protegga la schedulazione del collo di bottiglia e la spedizione dalla variabilità che si potrebbe manifestare lungo il percorso.⁷

⁵ Six Sigma offre strumenti operativi per l'analisi e la riduzione della variabilità nei processi. Questo ben si sposa con l'approccio sistemico proposto da TOC e SFM

⁶ Gli appassionati della Theory of Constraints chiamano questa fase "drum", che in inglese significa tamburo, ad indicare che qui si scandisce il ritmo al quale tutto il processo deve muoversi, in analogia a quanto avveniva nei tempi passati quando i plotoni militari marciavano appunto al tempo scandito da un tamburino

⁷ Il modello funziona anche quando le operazioni sono collegate tra loro da kanban, perché i colli di bottiglia quando sono reali assumono questo ruolo con stabilità nel tempo. Anche in un sistema a kanban il collo di bottiglia è il pacemaker del processo



Il livello della variabilità di sistema cresce con la complessità dei flussi: quanto più una fabbrica ha necessità di uno schedatore a capacità finita per organizzarne gli appuntamenti, tanto meno questo può funzionare. Il motivo è da attribuire al fatto che gli schedatori a capacità finita, e gli APS che eventualmente li contengono, non considerano la variabilità, e in un vano tentativo di ottimizzare l'efficienza delle singole risorse produttive rilasciano quantità indiscriminate di materiale, generando inefficienza di sistema e perdita di controllo della fabbrica. Che cosa si fa allora? Si sfrutta la potenza di calcolo per schedare e rischedare continuamente *tutta* la fabbrica, facendo diventare matti gli operativi che in una situazione simile hanno mille buone ragioni per fare di testa propria. L'expediting elettronico, magari ben supportato da coloratissimi grafici drag and drop, è dunque il frutto del progresso!

Viceversa il Synchronous Flow Manufacturing, che contempla il controllo dei flussi e degli effetti della variabilità (attraverso il buffer management), consente il salto di qualità nel manufacturing, risolvendo la diatriba tra le spinte verso la Lean Production e i vincoli imposti dai sistemi informativi che i manager di produzione ben conoscono.

Synchronous Flow Manufacturing⁸ permette di mantenere i materiali in flusso costante e regolare; abbiamo visto come il sistema descritto ottemperi a questa istanza, mantenendo costante il livello di WIP⁹ e prevenendo la formazione di code. Le risorse non collo di bottiglia ricevono i semilavorati e li trasformano prima possibile nella sequenza con cui arrivano; il meccanismo del gating (3) - tirato dagli ordini (1) attraverso il constraint (2) – garantisce la coerenza del piano di produzione con quello delle spedizioni. Inoltre, la gestione dello shop floor attraverso questa semplice regola sulle priorità risparmia la necessità di un oneroso lavoro di controllo e coordinamento.

Altro concetto caro alla scuola Lean è il presidio del *Value Stream*, ovvero l'orientamento all'eliminazione di quanto è spreco perché non aggiunge valore al prodotto (*muda*). La strategia di focalizzazione sui constraints impedisce l'*overproduction* e la nascita di scorte non necessarie (*unnecessary inventory*), perché il flusso dei materiali è bilanciato e continuamente regolato sulla esatta capacità (finita) della fabbrica, determinata da quella dei suoi colli di bottiglia. SFM realizza la riduzione del livello complessivo di scorta, perché i materiali sono pianificati, acquistati e rilasciati sulla base della reale capacità del processo di smaltirli. La schedulazione continuamente accoppiata alla domanda originale del mercato garantisce di produrre ciò che realmente serve. Infine, un altro precetto della Lean è racchiuso nella parola *perfection*, a significare la tensione verso il miglioramento continuo. SFM recepisce questa esigenza potenziandola con la gestione

⁸ La metodologia del SFM, per quanto concettualmente semplice, necessita il supporto di un software che ne incarni le logiche

⁹ Il funzionamento a WIP costante è noto come CONWIP (Constant WIP)



selettiva delle risorse acquisita dalla Theory of Constraints, che insegna a guardare i processi in ottica sistemica, così finalizzando gli sforzi di miglioramento ove questi producono il maggiore effetto in termini di risultato complessivo piuttosto che puntuale. Gli effetti sul flusso sono differenti quando si agisce sulle risorse collo di bottiglia, piuttosto che su quelle che stanno a valle o a monte: la Theory of Constraints ci aiuta a focalizzare i punti del processo su cui attivare azioni di miglioramento, sostenute, per esempio, da gruppi kaizen e da analisi Six Sigma.

Altri tecnicismi del SFM sul controllo dei flussi e sul processo di pianificazione non possono essere qui raccontati per necessità di brevità.

Conclusioni

Abbiamo visto come il contributo del Manufacturing alla prestazione della supply-chain e, in senso più ampio alla competitività, può essere riferito ad affidabilità, rapidità, flessibilità e costi.

Si è spiegato come e perché le fabbriche gestite con le logiche tradizionali quali push, MRP, ecc. tendano naturalmente a ingolfarsi di materiale, con conseguenze negative proprio sulle dimensioni della performance di cui sopra.

Abbiamo quindi presentato le linee generali del Synchronous Flow Manufacturing, sintesi di Lean Production e Theory of Constraints. SFM si focalizza sul controllo dei flussi e fa operare le fabbriche con livelli di WIP controllato e brevi lead time, previene la formazione di code e ingolfamenti, spinge verso le logiche della Lean Production riducendo i muda, assorbe gli effetti nefasti della variabilità. L'efficienza aumenta perché attraverso la politica di focalizzazione sui colli di bottiglia il processo ottiene il massimo delle proprie potenzialità, senza sforzi inefficaci quali expediting, straordinari non necessari e coordinamento affannoso. Il sistema snello così costituito è intrinsecamente flessibile, in quanto rapidamente riconfigurabile.

Synchronous Flow Manufacturing orienta realmente la fabbrica alle logiche demand-driven, ponendola continuamente in tiro con il mercato mediante il doppio segnale pull dalle spedizioni al collo di bottiglia, e da qui al gating.

L'implementazione di questa metodologia presuppone il cambio del sistema di pianificazione, delle modalità operative della fabbrica e naturalmente un salto culturale. Si consideri che rispetto ad un progetto di Lean Production è estremamente più leggero, rapido e abbordabile, perché agisce prevalentemente sulle modalità di coordinamento e controllo piuttosto che sulla struttura fisica e sugli assetti organizzativi della fabbrica.

La serie di problemi evidenziati dalla gestione tradizionale dei processi produttivi non è una crociata contro il MRP, ma aiuta a rendersi conto come uno strumento vecchio di 45 anni



probabilmente oggi non risponda più alle esigenze per cui era nato, determinate da condizioni di contesto non più attuali. Synchronous Flow Manufacturing costituisce la sesta rivoluzione del Manufacturing, ed è la metodologia che promuovendo affidabilità, rapidità, flessibilità e costi meglio si adatta alla corrente economia della domanda.