Estudo do tempo de ciclo de uma fábrica de semicondutores

Francisco Vieira da Silva Pires

Dissertação de Mestrado

Orientador na FEUP: Eng. Paulo Luís Cardoso Osswald

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Mecânica

2018-06-15

Resumo

Esta dissertação surgiu no âmbito de um projeto europeu que envolve empresas de semicondutores, investigação e de *software* e tem o objetivo de criar uma ferramenta online de planeamento que permita um controlo automático sobre decisões de planeamento. O trabalho efetuado abrange o estudo dos tempos de ciclo numa fábrica de semicondutores, matéria particularmente interessante nesta indústria devido localização geográfica das fábricas de semicondutores e dos seus clientes.

A indústria apresenta um grau de complexidade de produção invulgar, com ciclos de produção onde se repetem processos, um número peculiarmente alto de processos, produtos distintos, heurísticas de produção individuais para cada processo (sequenciamento, *batching*, regras de priorização), entre outros. Estes fatores originam uma variabilidade considerável nos tempos de espera o que dificulta a estimativa do tempo de ciclo e obriga a um planeamento de produção por objetivos.

Neste projeto foi importante começar por definir os fatores mais relevantes para o estudo do tempo de ciclo. Este tempo foi classificado de acordo com duas causas – tempos de processo e tempos entre processos. Os primeiros já estavam bem documentados e eram geralmente mais baixos e constantes que os tempos entre processo como tal o estudo focou-se nestes últimos. Verificou-se também que os tempos entre processo se deviam praticamente na sua totalidade aos lotes ficarem em fila de espera para serem processados. Somando os tempos de processo e tempos entre processos médios, obtém-se o tempo de ciclo previsto médio. Esta abordagem permitiu ainda obter o tempo de ciclo médio por processo e produto, assim como o tempo de ciclo médio até qualquer processo para qualquer produto.

Estudaram-se os processos com os tempos de espera mais elevados onde se seguiram lotes individualmente e se registaram as razões pelas quais ficaram em fila de espera. Observou-se que algumas heurísticas de processamento e planeamento da produção, entre as quais o sequenciamento, *batching* e gestão de WIP eram as causas principais destes tempos de espera elevados e que estavam a ser aplicadas de forma menos eficaz devido a vários fatores como a falta de sincronização dos produtos na fábrica (produtos da mesma família não estão próximos na cadeia de produção), tempos de *setup* elevados, relação inversa entre custos de produção e tempo de ciclo, entre outros.

Estes dados serviram para caracterizar o processo de produção da empresa, alimentar a ferramenta do projeto europeu e auxiliar o planeamento de produção. Os dados utilizados anteriormente pela empresa apoiavam-se apenas no *flow factor* do produto, ao contrário do método proposto por este projeto que se baseia em dados históricos e permite fazer um planeamento de produção mais preciso e correto. Sugere-se ainda que, de futuro, se estude a relação entre custo de produção e melhoria no tempo de ciclo, bem como a influência da diminuição do tamanho do *batch* no tempo de ciclo.

Cycle time study on a semiconductor factory

Abstract

This dissertation arised from an European project which comprises of semiconductor companies as well as other investigation and software companies. Its objective is to create an online planning tool which allows for an automatic control over factory floor planning decisions. This work encompasses the study of cycle times in the factory, a particularly interesting subject in this industry given the geographical localization of the semicondutor factories and their clients.

This industry features an unusually high production complexity, with production loops where processes get repeated, a peculiarly high number of processes, different products, individual production heuristics for each process (serial batching, parallel batching, prioritization rules), among others. These factors create considerable variability on process wait times which hinders the cycle time estimation and forces an objective-based production planning.

In this project it was important to start by defining the most relevant factors for the study of the cycle time. This time can be classified by two causes – process times and times between processes. The first are already well documented and are generally lower and more stable than the latter so the study focused heavily on the times between processes. It was also found that the times between processes were mostly due to the lots getting piled up in waiting lines for processing. Adding the average time between processes with the process times we get the predicted average cycle time. This method also allows for the calculation of the average cycle time by process and product, as well as the average time any product takes to get to a certain process.

Secondly, the processes with the highest wait times were studied and the lots were individually followed and the reasons behind their wait times were registered and timed. It was observed that specific production and planning heuristics including serial batching, parallel batching and WIP management were the main causes behind these high wait times and are being used sub-optimally due to several factors such as lack of product synchronization on the factory (products from the same family aren't close together on the production chain), high setup times, an inverse relationship between production costs and cycle time, among others.

These data provided a better production process description, some inputs for the European planning tool project and it improved the company's production planning. The data previously used by the company was based only on the product's flow factor, unlike the proposed method in this project which is based on historical data and allows for a more accurate production planning. The suggested future studies include the relationship between production cost and cycle time as well as the influence of reducing batch size on cycle time.

Agradecimentos

Em primeiro lugar agradeço à minha família pelo apoio, carinho e sacrifício que permitiram que atingisse este objetivo.

Agradeço também a todos os meus colegas de trabalho que tornaram a elaboração desta dissertação um verdadeiro prazer, em particular ao meu orientador na empresa, Pedro Vasconcelos.

Sem ficarem atrás, agradeço também aos meus amigos pelas noites de estudo que por vezes se transformaram em madrugadas, pela sua paciência e apoio em todos os momentos da escrita da dissertação e acima de tudo pelos finos partilhados que tornaram todo este caminho mais agradável.

Agradeço ainda ao meu orientador da FEUP – Paulo Osswald – pela sapiência e disponibilidade mostradas ao longo do semestre.

Finalmente deixo uma palavra de agradecimento a todos aqueles que direta ou indiretamentamente me auxiliaram no deccorrer desta dissertação.

Índice de Conteúdos

Siglas

5W – Five “whys” (cinco “porquês”)

ANN – Artificial neural network

C/O – Tempo de conversão

CR – Critical Ratio

CT – Cycle Time (tempo de ciclo)

FF – Flow factor

FIFO – First in first out

MLR – Multiple linear regression

PT – Process time (tempo de processamento)

RPT – Raw Process Time (Tempo de processamento de um processo)

RTT – Raw Tool Time (Tempo de processamento de um equipamento)

SB – Stand-by

SMED – Single minute Exchange of die

TH – Throughput time (cadência de um produto)

TRPT – Total remaining process time

UD – Unschedule down

UP – Uptime

VSM – Value stream mapping

WIP – Work In Progress/Work in Process (materiais em linha)

WP – Work packages/Wafer preparation

WT – Waiting Time (tempo de espera)

Índice de figuras

Índice de tabelas

1 Introdução

A dependência crescente de dispositivos eletrónicos no mundo atual, aliado à forte concorrência de mercado, obriga a indústria de semicondutores a estar em permanente evolução tecnológica. Devido à rápida globalização de ferramentas, materiais e bens de consumo genéricos que requerem um crescente poder de processamento, esta indústria sofre uma pressão constante para a miniaturização dos seus produtos, redução de custos e desenvolvimento tecnológico.

A ATEP – Amkor Technology Portugal S.A. é uma filial da Amkor que se especializa na produção de *packaging* de semicondutores no formato de *wafer* para aplicação na área das comunicações e automóvel. Este *packaging* é essencial para criar uma ponte elétrica entre o *chip* e o dispositivo onde vai ser aplicado, assim como para dissipar calor e proteger o *chip* de riscos ambientais, tais como a humidade. Este processo faz parte da produção de semicondutores o que efetivamente coloca a ATEP no centro de uma cadeia de processos, classificando-se como um prestador de serviços na produção de semicondutores sendo que os seus fornecedores e clientes são a mesma entidade.

O prazo de entrega dos produtos é geralmente fixo e depende apenas dos requisitos do cliente. As ordens de produção dos clientes não têm uma data de entrega específica, mas existe uma data específica para serem entregues certas quantidades de produtos, independentemente da capacidade da fábrica, ou seja, mesmo que os produtos estejam completamente processados apenas devem ser entregues nas datas e quantidades estipuladas pelos clientes.

Não obstante, a fábrica é a líder europeia em produção de uma classe de produtos denominada por *fanout* que muitos consideram ser o futuro da indústria de semicondutores tendo aplicações principalmente no mercado dos *smartphones* devido ao seu tamanho inferior, custos reduzidos e melhor desempenho (Tracy e Tseng 2016). Ainda assim, a fábrica também produz uma quantidade considerável de *wafers fanin*.

1.1 Enquadramento e motivação do projeto

Este trabalho surgiu no âmbito de um projeto europeu multi-organizacional da indústria 4.0 denominado SEMI40, que envolve a ATEP e empresas de investigação e *software* com o objetivo de criar uma ferramenta online de planeamento para a indústria de semicondutores que permita um controlo automático sobre decisões de fábrica. A ferramenta é desenvolvida e testada recorrendo a dados reais da ATEP.

A fábrica foi adquirida pela multinacional Amkor durante o ano de 2017 e está no processo de integração no grupo, no entanto este projeto iniciou-se anteriormente, quando a empresa era independente e designada por NANIUM, S.A e os seus objetivos não foram afetados pela alteração de estratégias.

Devido à localização geográfica da empresa, dos seus clientes e da concorrência e devido ao facto de os clientes pedirem os produtos para um certo espaço temporal, conseguir atingir um *lead time* curto sem afetar a qualidade dos produtos torna-se fulcral para manter a vantagem competitiva da empresa. Há pouco interesse em reduzir o *lead time* para valores inferiores ao requerido pelo cliente se isso implicar um aumento dos custos de produção e armazenamento, assim o que se pretende é um controlo mais preciso e eficaz do tempo de processamento dos produtos.

Devido a situações particulares desta indústria, como repetições de processos, filas de espera elevadas, máquinas em paralelo para o mesmo processo, tempos de *setup* elevados, entre outros, o tempo de ciclo tende a apresentar uma alta variabilidade. Este trabalho pretende então estudar este tempo de ciclo dos produtos a um nível mais preciso com especial foco nos tempos de espera dos produtos. O estudo pretende encontrar as razões para os tempos de espera

elevados, mapeá-los e quantificá-los, de modo a conseguir perceber os fatores que mais afetam o tempo de ciclo. Neste momento estes tempos de espera são pouco claros e não estão quantificados de forma detalhada, apenas são conhecidos os seus efeitos no *lead time*.

O trabalho foi realizado no departamento de planeamento e logística da ATEP e permitirá obter uma melhor visualização e controlo dos produtos na linha de produção, o que por sua vez permite fazer uma melhor estimativa sobre o tempo de ciclo restante de cada lote em qualquer processo da linha de produção. Vai ser ainda possível perceber o que está a afetar mais o tempo de ciclo e em que processos, o que vai permitir melhorá-los de modo a reduzir custos com equipamentos e melhorar a fiabilidade de entrega de produtos.

1.2 Estudo do tempo de ciclo na ATEP

O projeto SEMI40 é composto por 7 atividades, denominadas *work packages* (ou WP), e está neste momento no WP2 que consiste no uso de algoritmos informáticos para fazer análises preditivas e de causas, para serem usados na automatização de decisões de fábrica. Para esta automatização ser possível, é necessário conhecer as heurísticas usadas na fábrica e a sua razão, assim como quantificar o seu efeito no tempo de ciclo. A dissertação insere-se neste projeto na medida em que o "alimenta" com estes dados e com uma referência do cálculo do tempo de ciclo para confirmar os resultados dos seus algoritmos.

A dissertação vai então incidir no estudo do tempo de ciclo e mapeamento dos tempos de espera dos produtos. Para estudar os tempos de ciclo decidiu-se abordar o projeto por duas frentes: tempos de processo e tempos entre processo. Os primeiros estão bem definidos e atualizados em sistema, no entanto os segundos não são conhecidos de todo, pelo que o estudo focar-se-á nesses tempos.

Este trabalho é fundamental não só para o projeto SEMI40, mas também para a empresa visto que, devido à imposição dos clientes de entregas em datas específicas, é de grande interesse conseguir perceber os fatores que afetam o tempo de ciclo. Neste momento as heurísticas usadas para priorização dos produtos não estão bem definidas e podem variar de área para área sem ter em atenção o tempo de ciclo global dos produtos. Por esta razão, o tempo de espera dos produtos antes dos processos varia consideravelmente e sem razão aparente. Este fator tem de ser estudado e quantificado para se poder ter mais controlo sobre o tempo de processamento dos produtos.

1.3 Objetivos

Com este trabalho pretende-se fazer uma estimativa eficaz do tempo de ciclo de qualquer produto em qualquer ponto da produção através da quantificação dos tempos entre processo e tempos de processo, assim como implementar estes resultados no projeto do SEMI40 e conhecer os fatores e heurísticas relevantes que causam o tempo de espera dos produtos e a sua quantificação. Com estes dados será possível identificar e corrigir eventuais problemas encontrados que promovam um tempo de ciclo superior ao esperado. Tal informação será vital para o conhecimento da linha e para futuras ações de melhoria tanto do planeamento como da produção.

1.4 Métodos seguidos

Durante a elaboração deste projeto foram seguidos vários ensinamentos e teoremas que serviram de apoio às análises realizadas e ao trabalho efetuado.

Entre os métodos seguidos destaca-se o teorema na lei de Little para a definição e cálculo de tempo de ciclo bem como a sua relação com a cadência de um processo e a quantidade de materiais em processo.

Foi também utilizada a teoria das filas de espera que explica que a espera dos produtos numa linha de produção pode ser prevista e quantificada através da equação de Kingman, desde que cumpra certas suposições. Apesar desta teoria ser robusta o suficiente para incluir a variabilidade de tempos de chegada e saída dos produtos no cálculo do tempo em fila de espera, mais uma vez, a variação de recursos na fábrica ou o uso de heurísticas de priorização mal definidas causa sérios problemas à aplicação desta teoria.

Foi também usada a metodologia dos 5W (5 "porquês") para identificar a origem dos altos tempos entre processo encontrados. Embora simples, esta metodologia demonstra ser bastante eficaz e permite sugerir mudanças para redução destes tempos.

Foi ainda tida em conta a teoria das restrições enquanto filosofia de gestão de fábrica que é útil para identificar as limitações da empresa a nível de produção e tentar melhorá-las.

1.5 Estrutura da dissertação

A dissertação divide-se em 5 capítulos, compostos pela introdução, revisão do estado de arte, descrição do processo produtivo e das heurísticas de planeamento da empresa, análises das heurísticas e os seus efeitos no tempo de ciclo e conclusão.

O capítulo 2 abrange a revisão dos conceitos, metodologias e heurísticas utilizadas na indústria de semicondutores, assim como as definições e termos específicos desta indústria.

Seguidamente, no capítulo 3, apresenta-se a empresa de forma mais detalhada com ênfase na sua linha de produção. A produção de circuitos integrados é bastante própria e conta com algumas particularidades únicas, pelo que é necessário perceber de um modo geral o processo de produção para conseguir entender algumas dificuldades acrescidas desta indústria. São ainda apresentados os casos particulares presentes na linha de produção desta fábrica, assim como a forma como o planeamento e logística da produção são efetuados. As observações mais relevantes, tais como a discriminação do *bottleneck* da fábrica e dos tempos de ciclo dos produtos, são também expostas e analisadas neste capítulo.

O desenvolvimento do projeto é apresentado no capítulo 4, onde se utilizam as metodologias já referidas nos capítulos anteriores. São também comparados os tempos de ciclo por camada da empresa com os da indústria com o intuito de demonstrar o desempenho da empresa. É posteriormente demonstrado o raciocínio utilizado nas diversas análises com o objetivo de encontrar a melhor forma de prever e estimar os tempos entre processos para cada produto e implementá-los na ferramenta de planeamento do projeto SEMI40. Também são discriminadas e quantificadas as razões pelas quais estes tempos são tão elevados.

Por fim, o capítulo 5 refere-se às conclusões sobre a abordagem a este projeto onde se examina criteriosamente os resultados obtidos e se propõem trabalhos futuros complementares.

2 Estado de arte

2.1 Revisão bibliográfica

No presente capítulo são apresentados os conceitos teóricos e metodologias usados na abordagem ao problema. O estudo bibliográfico sobre o cálculo de tempo de ciclo na indústria de semicondutores, apesar de vasto, reconhece as dificuldades inerentes à complexidade da indústria, sendo comum encontrar soluções pouco eficientes (Shanthikumar et al. 2007; Li et al. 2014) ou sensíveis ao estado da linha de produção (Schelasin 2011) . É, portanto, crucial encontrar uma forma eficaz e mais robusta de calcular e prever o tempo de ciclo nesta indústria.

A indústria de semicondutores tem um processo de produção particularmente complexo em que os tempos de ciclo de cada produto variam entre cada produto por terem diferentes tipos de processo com ciclos diferentes consoante uma vasta quantidade de variáveis (tipo de *wafer*, especificação do cliente, entre outros), processos de *batching*, tempos de espera obrigatórios e tempos de transporte variáveis, entre outras. No fundo, o problema principal tem origem na variação extrema que se observa nos tempos de espera antes dos processos, principalmente nos processos que requerem sequenciamento ou *batching*.

2.2 Definições

Em primeiro lugar, é útil definir os termos usados na indústria e nesta dissertação, visto que certos termos podem ter significados mais genéricos, podem variar consoante o autor e contexto ou são específicas a certas indústrias. No âmbito desta dissertação, seguir-se-á as definições usadas na indústria de semicondutores.

***Raw tool time* (RTT):** ou tempo de processamento de um equipamento, é o tempo bruto que uma máquina demora a processar um produto, expresso geralmente em horas. O RTT engloba apenas o tempo do processo e é obtido através dos dados históricos do equipamento, desde que um produto começa a ser processado pelo equipamento, até o equipamento completar o processo. Não considera tempos inativos do equipamento, tempos de processo realizados pelo operador, etc.

***Raw process time* (RPT):** é o tempo total de processo de um produto incluindo tempos de movimentação, de *setup*, de ajustes, documentação, entre outros. Apenas não são considerados para este tempo os tempos inativos do equipamento, tempos de espera e tempos de transporte. A soma de todos os RPT numa linha de produção de um produto corresponde a todo o tempo "útil" e ativo do produto, é todo aquele tempo que não pode ser reduzido com ações de planeamento.

WIP (*Work in process*): corresponde ao número de materiais em curso de fabrico ou à espera de serem processados por um equipamento ou numa linha de produção.

***Starving* ou *stand-by* (SB):** é o tempo em que um equipamento não está ativo por falta de material.

***Unschedule down* (UD):** é o tempo ponderado de avaria de um equipamento – tempo médio por processo em que o equipamento fica parado devido a uma avaria e a sua respetiva resolução. Por definição este tempo não pode ser facilmente previsto, mas usando dados históricos pode-se assumir uma média por equipamento para um dado tempo. Juntamente com o SB forma o tempo inativo do equipamento.

***Uptime*:** é a relação do tempo produtivo do equipamento com o tempo total, isto é, o tempo médio histórico em que o equipamento está a funcionar e disponível para produzir em função do tempo total.

Tempo de espera (*queueing*/WT): também denominado tempo em fila de espera, é definido como o tempo que o produto fica a aguardar para ser processado antes de uma célula de trabalho ou equipamento. Pode ter várias origens, incluindo excesso de WIP, ordens de produção superiores à cadência de produção, heurísticas de processo, entre outros.

Tempo de transporte: é o tempo que o produto demora a ser transportado de um processo para a fila de espera do processo seguinte. Não deve ser confundido com tempo de movimentação, que representa o tempo de o produto ser movimentado da fila de espera para o equipamento.

Operações administrativas de *setup*: é o somatório dos tempos necessários para fazer a seleção das unidades, movimentação das mesmas para os equipamentos, documentação e qualquer outra tarefa que o operador precise de fazer para a preparação das unidades a serem processadas. Estes tempos são geralmente baixos, mas podem ser uma origem de desperdícios considerável.

Tempo de ciclo (CT): vários autores têm várias definições para tempos de ciclo. Hopp e Spearman (2001) definem tempo de ciclo como o tempo médio entre a entrada no início da linha de produção até entrar em inventário final. Neste projeto, é mais útil definir tempo de ciclo como o tempo que um processo leva a ser realizado, sendo contado desde que o produto sai do processo anterior até sair deste processo, contemplando o RPT, o tempo em fila de espera, tempos de transporte e os tempos ponderados de avaria, como ilustra a figura 1. O somatório dos CT de todos os processos de um produto é denominado como tempo de ciclo total (CTt) e corresponde ao *lead time* médio.

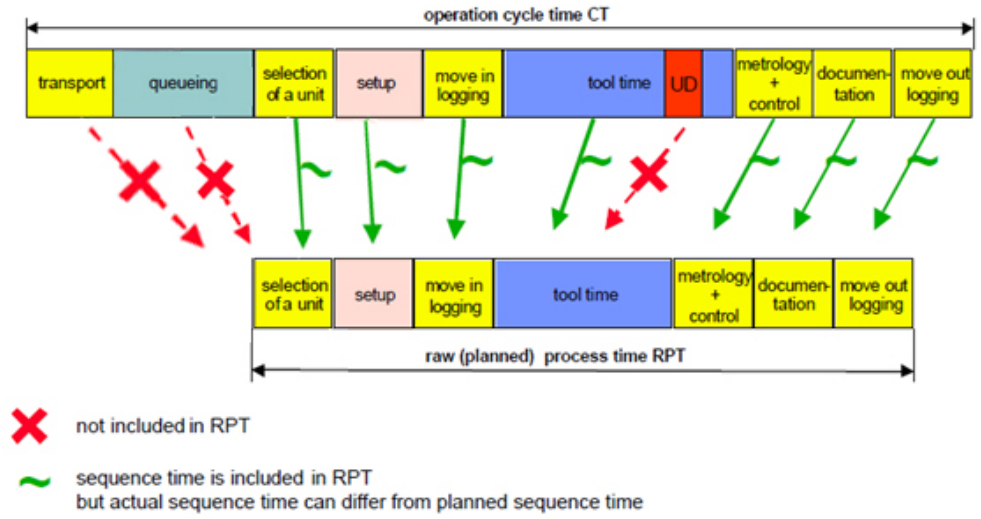

Figura 1 – representação do CT e RPT, adaptado de um documento fornecido pela empresa

Cadência (TH): frequência de saída de um produto por processo ou por linha de produção, podendo ser definido em produtos por unidade de tempo ou inverso. Alguns autores (Hopp e Spearman 2001; Rother e Shook 1999) usam o termo *takt time* para o ritmo de produção de um processo ou linha e *throughput* (TH) (Hopp e Spearman 2001) para o inverso – quantidade de produtos produzidos por unidade de tempo.

***Bottleneck* (gargalo):** conjugando o tempo de ciclo com o número de equipamentos disponíveis para cada processo, podemos calcular a capacidade da linha e identificar o *bottleneck* (gargalo) da produção. O gargalo de uma linha de produção é o equipamento ou conjunto de equipamentos com a cadência mais lenta, causando efetivamente o processo mais lento da produção. A cadência de produtos à saída, por definição, em média, nunca pode ser superior à cadência do gargalo e, portanto, para evitar falhas de eficiência, este processo não deve parar em ocasião alguma, pois é o processo crucial que define a frequência máxima de

saída de produtos na fábrica. No entanto, esta definição torna-se mais complicada de aplicar quando há diferentes produtos em linha que partilham alguns processos, mas têm RPTs diferentes, como é o caso na indústria de semicondutores.

Zhou e Rose (2009) propõe uma abordagem diferente, considerando o *bottleneck* como o equipamento com a taxa de utilização mais elevada. De facto, a principal restrição a nível de produção deverá encontrar-se nos equipamentos com menos folga de tempo (por dia ou por semana) para processarem os produtos que têm de processar. Assim, consoante a mistura de produtos diferentes na linha e o plano de produção, o *bottleneck* pode também variar diariamente, no entanto a sua identificação fica bastante simplificada.

Para Zhou:

$$Bottleneck = Machine_j = \max_{1 \le j \le m} \frac{WT_j(T) + OT_j(T)}{T} \tag{2.1}$$

Em que *j* é o número de máquinas do processo, *m* é o número de máquinas total da fábrica, $WT_j(T)$ é o tempo produtivo da máquina *j* nas últimas *T* horas e $OT_j(T)$ é o *downtime* da máquina *j* nas últimas *T* horas (tempo perdido com avarias, manutenção, *setup*, etc). De acordo com esta definição, o *bottleneck* será a máquina com menor tempo livre (em *stand-by*).

***Batching* em paralelo:** há equipamentos que permitem o processamento de vários produtos ao mesmo tempo. Quando isto acontece, diz-se que o equipamento faz *batching* em paralelo (ou apenas *batching*), pois agrega vários produtos que são processados em simultâneo.

***Batching* em série (sequenciamento):** Hopp e Spearman (2001) utilizam a expressão *batching* em série para definir processos onde se agrupam os produtos de modo a evitar fazer mudança de *setup* de equipamento. Desta forma, os equipamentos processam apenas um certo número de produtos do mesmo tipo ou durante um certo tempo, fazendo depois a conversão para o *setup* de outro tipo de produto. Diferencia-se do *batching* em paralelo pois o processamento dos produtos é feito individualmente e em sequência, ao contrário do *batching* em paralelo onde o processamento dos produtos é feito simultaneamente. Os produtos são processados em série, como num equipamento normal, apenas são agrupados e processados por família.

2.3 *Flow factor*

Este fator é o termo utilizado na indústria de semicondutores para relacionar o tempo de ciclo médio com o RPT de um produto. É aplicado tanto a processos individuais, à linha de produção de um produto ou de todos os produtos, desde que as unidades se mantenham consistentes. Por definição, o *flow factor* serve como um indicador das perdas por tempos de espera, transporte e UD. É dado pela equação 2.1:

$$FF = \frac{CT}{RPT} \tag{2.2}$$

2.4 Regras de priorização

Quando existem vários produtos em WIP, alguém tem de tomar a decisão sobre que produto processar primeiro. Em cada fábrica são utilizados diferentes métodos de priorização de produtos, sendo que talvez o mais comum seja o FIFO (*first in first out*) onde o primeiro produto a chegar ao posto de trabalho é o produto que começa a ser processado. Algumas fábricas preferem outras regras de priorização, entre as quais o CR (*Critical ratio*) que considera o tempo até à data de entrega e o RPT restante (TRPT) de modo a obter-se a margem relativa de tempo livre restante do lote. Aquele produto com o CR menor, ou seja, menor margem de tempo para finalização, tem prioridade (Rose 2003). Usa-se o seguinte método para calcular CR:

$$\begin{matrix} \frac{1 + Due - Now}{(1 + TRPT)}, se\ Due > Now, e \\ \frac{1}{(1 + Now - Due) * (1 + TRPT)}, se\ Now \geq Due \end{matrix} \quad (2.3)$$

2.5 Métodos de cálculo de tempo de ciclo

Segundo Chung e Huang (2002) há quatro tipos de métodos gerais para o cálculo do tempo de ciclo: métodos analíticos, de simulação, de análise estatística e métodos híbridos. Os primeiros dois são os que tiveram historicamente mais sucesso, auxiliados pelo aumento do poder de processamento dos computadores (Schelasin 2011).

Os métodos de simulação tomam partido de algoritmos informáticos, apoiando-se na rápida expansão do poder de processamento dos computadores, de modo a replicarem fiavelmente os processos de uma fábrica. Este método tem sido extensivamente estudado e a sua aplicação na indústria de semicondutores tem sido cada vez mais relevante, no entanto, para garantir a sua precisão, este método requer um alto nível de recursos, tanto para o criar como para o manter com o nível de detalhe desejado e o seu tempo de processo pode ser impraticavelmente longo. Por esta razão, as comunidades de simulação e modelação parecem concordar que, de modo a obter resultados válidos com este método, é necessário um nível de esforço pouco prático (Schelasin 2011).

Os métodos analíticos baseiam-se na estimativa matemática de tempo de ciclo usando heurísticas desenvolvidas com o propósito de serem aplicadas em indústrias de diversos tipos e são detalhadamente descritos no livro *Factory Physics* de Hopp e Spearman (2001). Apesar da fragilidade dos métodos analíticos em sistemas complexos, podemos considerar um sistema complexo como vários sistemas mais simples, para os quais este método consegue ser mais robusto e flexível (Hopp e Spearman 2001).

Métodos de análise estatística têm como objetivo criar modelos que estabeleçam uma relação entre dados históricos e previsões de tempo de ciclo. Essencialmente, o método resolve problemas complexos usando algoritmos simples com uma grande quantidade de dados. De acordo com Wang et al. (2018), o método estatístico que ultimamente tem atraído mais atenção é o de ANN (*artificial neural network*), superando métodos de MLR (*multiple linear regression*) e de árvores de decisões em estudos anteriores (Sha e Hsu 2004; Tirkel 2013) na previsão de tempo de ciclo de lotes de *wafers*.

Métodos híbridos, como o nome indica, são métodos que usam heurísticas de outras metodologias em paralelo com o objetivo de obterem dados mais precisos.

2.5.1 Teorema de Little em filas de espera

O teorema de Little é uma das leis fundamentais de gestão da produção que relaciona o CT com o WIP de forma simples para um dado TH:

$$WIP = TH * CT \tag{2.4}$$

Esta equação é aplicável a qualquer sistema, seja equipamento, célula de trabalho ou linha de produção, desde que seja feita para uma amostra temporal infinita e assumindo uma taxa de chegada de produtos estável e constante. Na realidade, os sistemas não são infinitos, ainda assim a lei de Little continua a ser uma excelente aproximação para quaisquer fábricas, com a exceção de períodos de transição como mudanças de produção ou fábricas novas.

Devido à sua funcionalidade, esta lei pode ser usada para calcular não só o tamanho das filas de espera num processo como também o tempo de ciclo previsto ou atual, desde que se mantenham as mesmas unidades. A figura 2 demonstra como esta lei se verifica numa fábrica simples de apenas uma linha com 4 equipamentos com o mesmo TH.

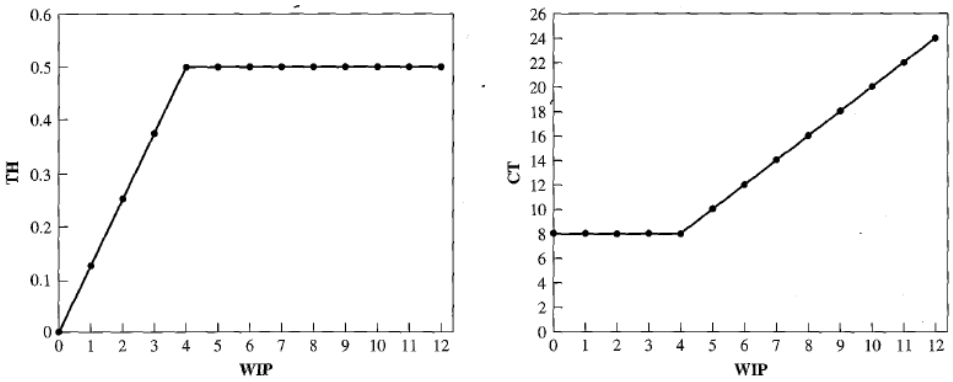

Figura 2 – evolução do TH e CT em relação ao WIP, imagem retirada de Hopp e Spearman (2001)

Como se pode verificar, o TH vai aumentando com o WIP até atingir o número de equipamentos na linha, mantendo-se depois estável. O contrário acontece com o tempo de ciclo, mantendo-se estável até WIP=4 e aumentando linearmente após esse valor. Isto acontece, porque, até a linha "encher", os equipamentos demoram sempre o mesmo tempo a fazer menos produtos do que a sua capacidade permite (TH de cada equipamento é de 1 produto por unidade de tempo), por outro lado, para WIP>4, há mais produtos na linha do que os equipamentos, obrigando assim alguns produtos a ficarem à espera de ser processados, aumentando o tempo de ciclo, mas mantendo a mesma cadência.

O ponto ideal de produção, neste caso seria para WIP=4, em que a cadência dos produtos é máxima sem afetar o seu tempo de ciclo.

2.5.2 Tempo de ciclo médio

Calcular o tempo de ciclo usando a média histórica do tempo de ciclo de vários lotes num certo horizonte temporal é a abordagem mais direta, mas requer uma base de dados histórica bem preenchida e é vulnerável a mudanças de fatores na linha como tempos de *setup*, número de máquinas, entre outros. Esta abordagem é fortemente baseada na definição de tempo de ciclo total de Hopp e Spearman (2001) – "*The average cycle time in a line is equal to the sum of the cycle times at the individual stations, less any time that overlaps two or more stations*". Desta forma podemos calcular o tempo de ciclo total como o somatório do tempo de ciclo de cada processo, diminuindo a complexidade do sistema para vários sistemas mais simples. Mesmo tratando-se de uma produção complexa, a sua dimensão deverá equilibrar os eventuais desvios, o que permitirá obter resultados perto da realidade.

É usada a média dos tempos de espera de cada processo obtidos historicamente e o respetivo desvio-padrão, assim como o RPT de cada processo para calcular o tempo de ciclo de cada processo. Neste caso assume-se que o tempo parado não programado da célula é nulo, pois o presente projeto não pretende abordar temáticas de manutenção corretiva. São, portanto, apenas considerados os tempos de funcionamento normal da célula.

Esta abordagem permite, de uma forma simplista, não só obter o tempo de ciclo total esperado para cada produto conhecido como também o tempo de ciclo esperado até cada processo, permitindo uma verificação sobre o estado do produto em qualquer ponto da linha, pois permite comparar o processo onde o produto se encontra com o processo onde seria esperado encontrar-se.

2.5.3 Teoria das filas de espera (*Queueing theory*)

A teoria das filas de espera é uma ferramenta geralmente eficaz para analisar o comportamento de filas de espera, pois consegue integrar de forma concreta dados que seriam normalmente complexos de quantificar tais como a variabilidade de chegada de produtos. Naturalmente, este teorema é limitado a sistemas com filas de espera para os quais estas fórmulas são adaptáveis. Esta teoria requer alguns pressupostos, entre os quais, uma produção estável em que a variabilidade não mude consideravelmente (tempos de chegada e saída de produtos têm de seguir sempre o mesmo tipo de distribuição), uma fábrica com alta utilização e que a produção seja gerida com heurísticas bem definidas e automáticas e sem alterações humanas na linha, em particular na priorização dos produtos.

Para os casos em que a teoria das filas de espera não consegue proporcionar respostas válidas, os métodos de simulação são geralmente a melhor opção (Killian 2010). Estes casos incluem sistemas em que o tempo de chegada ou de processos não pode ser aproximado a distribuições probabilísticas ou sistemas em que a sequência de operações segue heurísticas baseadas em decisões humanas como, por exemplo, evitar conversões de *setup*, no entanto, estas heurísticas podem ser acopladas a esta teoria desde que sirvam como limite superior para o tempo em fila de espera, isto é, o valor que obtivermos com este método não pode ultrapassar o das heurísticas definidas. Na indústria de semicondutores as fábricas tendem a ser demasiado complexas (como vai ser explicado em maior detalhe no capítulo 3) para serem bem representadas com apenas um sistema de filas de espera (FabTime 2018). De acordo com Schelasin (2011) podemos considerar fazer o cálculo individual para cada processo e obter resultados positivos, porém estes resultados dependem de uma base de dados bem desenvolvida assim como fórmulas precisas que permitam obter taxas de utilização esperadas e produção alta e estável.

As vantagens da teoria das filas de espera comparada com os métodos de simulação são a obtenção de resultados exatos e invariáveis, a perceção direta dos fatores relevantes para o tempo de espera e o cálculo direto sem longos períodos de processo. A principal desvantagem é a sua limitada aplicabilidade a sistemas mais complexos (Killian 2010).

A equação fundamental usada é a equação de Kingman com os pressupostos subjacentes de que tanto os tempos de chegada como os tempos de processo seguem uma distribuição normal, que os equipamentos funcionam em paralelo e que o número máximo de lotes no sistema é infinito. Os sistemas de filas de espera podem ser caracterizados pela notação de Kendall como A / B / m, sendo *A* a distribuição dos tempos de chegada, *B* a distribuição dos tempos de processo e *m* o número de máquinas paralelas. Usando esta notação, este tipo de sistema também é conhecido como o sistema de fila de espera G / G / m (Hopp e Spearman 2001).

A equação para o sistema de fila de espera é então definida como:

$$CT_q = \left(\frac{c_a^2 + c_e^2}{2}\right)\left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)}\right)t_e \quad (2.5)$$

Em que:

CT_q = tempo em fila de espera

c_a = coeficiente de variação dos tempos de chegada ao grupo de equipamentos

c_e = coeficiente de variação dos tempos efetivos do processo de um grupo de equipamentos

m = o número de equipamentos disponíveis

u = taxa de utilização dos equipamentos disponíveis

t_e = tempo efetivo do processo.

Na sua forma simplificada, esta equação é também conhecida por equação de VUT, onde:

$$CT_q = VUt_e \quad (2.6)$$

A utilização dos equipamentos, assim como o número de equipamentos disponíveis, pode ser calculada a partir da quantidade de produtos em linha e o tempo efetivo de processo é o conhecido RPT, no entanto calcular os coeficientes de variação dos tempos torna-se complicado na indústria de semicondutores. Por esta razão, Schelasin (2011) propõe encontrar os valores de variabilidade calculando-os através da mesma equação , mas utilizando dados históricos. Simplificando e resolvendo a equação em função de V, obtém-se a seguinte equação:

$$CT_q = VUt_e <=> V_b = \frac{CT_q}{\left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)}\right)t_e} \quad (2.7)$$

Em que os valores se mantêm com a mesma definição, mas são referentes a dados históricos representativos do estado da linha para o qual se pretende fazer o cálculo:

CT_q = tempo médio de espera;

u = taxa de utilização real dos equipamentos no tempo para o qual se está a calcular a variabilidade;

m = número de equipamentos disponíveis no mesmo período de tempo;

te = tempo efetivo de processo desse período;

V_b = variabilidade calculada baseada nos valores históricos (*backward variability*)

t_e tende a não variar muito, a não ser que haja uma diferença significativa na mistura de produtos em linha.

Obtendo o valor de variabilidade, pode-se então voltar à fórmula inicial para calcular CT_q:

$$CT_q = V_b\left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)}\right)t_e \quad (2.8)$$

Onde:

u = taxa de utilização prevista;

m = número de equipamentos disponíveis para o período de tempo para o qual se pretende encontrar os tempos de espera;

t_e = tempo efetivo de processo.

2.6 Sistema *pull* e *push*

Taiichi Ohno (1988) definiu a ideia de *pull* da seguinte forma:

> *Manufacturers and workplaces can no longer base production on desktop planning alone and then distribute, or push, them onto the market. It has become a matter of course for customers, or users, each with a different value system, to stand in the frontline of the marketplace and, so to speak, pull the goods they need, in the amount and at the time they need them.*

A distinção entre sistemas *pull* e sistemas *push* é mais visível se pensarmos no mecanismo que causa a produção numa fábrica. Enquanto que nos sistemas *push* os produtos são lançados em fabrico mesmo que não haja uma encomenda especifica por parte do cliente (sendo que o cliente tanto pode ser interno como externo), no sistema *pull* a produção apenas inicia quando há encomenda do cliente. Ambos os sistemas são válidos consoante o tipo de produção que se pretende.

No sistema *push*, acumula-se inventário de modo a garantir que não há ruturas. A produção é contínua e baseada na previsão de vendas, mesmo que as encomendas ainda não tenham sido realizadas. Esta filosofia, apesar de garantir entregas ao cliente e uma boa ocupação do equipamento, cria uma inércia bastante considerável o que torna o sistema mais lento a responder à variabilidade da procura. É, no entanto, um sistema válido e preferível em atividades sazonais que requerem a disponibilidade prévia imediata do produto como lojas de venda ao público.

No sistema *pull*, temos o inventário diminuído ao mínimo possível, procurando garantir sempre as entregas ao cliente num tempo reduzido, mantendo a agilidade de um tempo de ciclo reduzido. As vantagens são óbvias, menos inventário significa não só menos custo de armazenamento como também uma melhor velocidade de resposta ao cliente, com um tempo de ciclo o mais reduzido possível e menos variável.

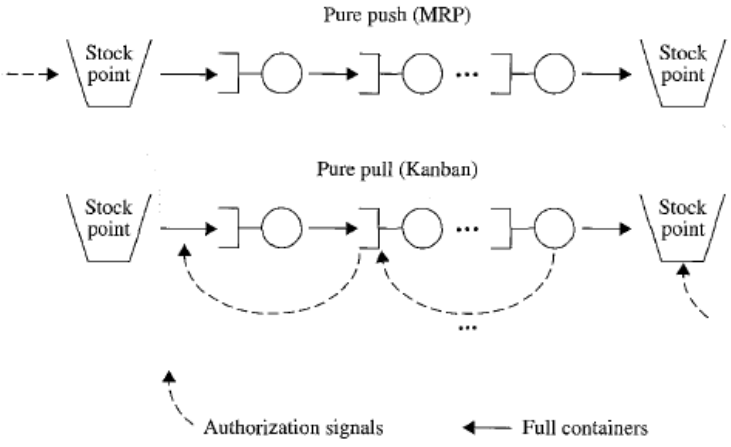

Figura 3 – representação do sistema *push* e *pull*, imagem adaptada de Hopp e Spearman (2001)

2.7 *Value Stream Mapping*

"Value Stream Mapping" é uma ferramenta de auxílio à gestão e planeamento da produção de uma fábrica que permite visualizar de forma rápida e eficaz o fluxo dos materiais e da informação pela linha de produção. Segundo Rother e Shook (1999), o objetivo desta ferramenta é analisar o estado atual da linha de produção, desde que o produto ou serviço são expedidos para a fábrica até serem entregues ao cliente; encontrar os pontos críticos onde se perde mais tempo ou recursos e desenhar um estado futuro da linha como sugestão de melhoria.

Como o nome indica, VSM engloba o mapeamento de todos os fluxos na linha de produção, no entanto seria impraticável fazê-lo para todos os produtos pois ficaria incompreensivelmente complicado, derrotando o propósito original da ferramenta. Por estas razões, o VSM deve ser feito apenas para uma família de produtos.

Num VSM existem dados de tempo de ciclo, utilização de equipamentos, fluxo de materiais e informação, tempo em inventário e todos os dados que forem úteis para compreender o fluxo dos produtos. Todas as figuras usadas na construção do VSM devem ser explícitas e intuitivas, de modo a requerer o mínimo de explicação a qualquer pessoa que não esteja familiarizada com a linha de produção da fábrica. É crucial que a pessoa responsável pelo VSM faça o mapeamento do fluxo do produto ao nível da linha de produção, com uma visão geral e imparcial da fábrica, de modo a projetar o mais fiavelmente possível o fluxo de materiais e informação da fábrica. Rother e Shook (1999) sugerem ainda que todos os tempos e dados sejam retirados manualmente pela pessoa mesmo que estes já existam em base de dados, para evitar usar dados desatualizados ou específicos de uma data onde haja produção fora do normal. Na figura 4 apresenta-se um exemplo de um VSM de uma fábrica de produção de peças estampadas e montadas.

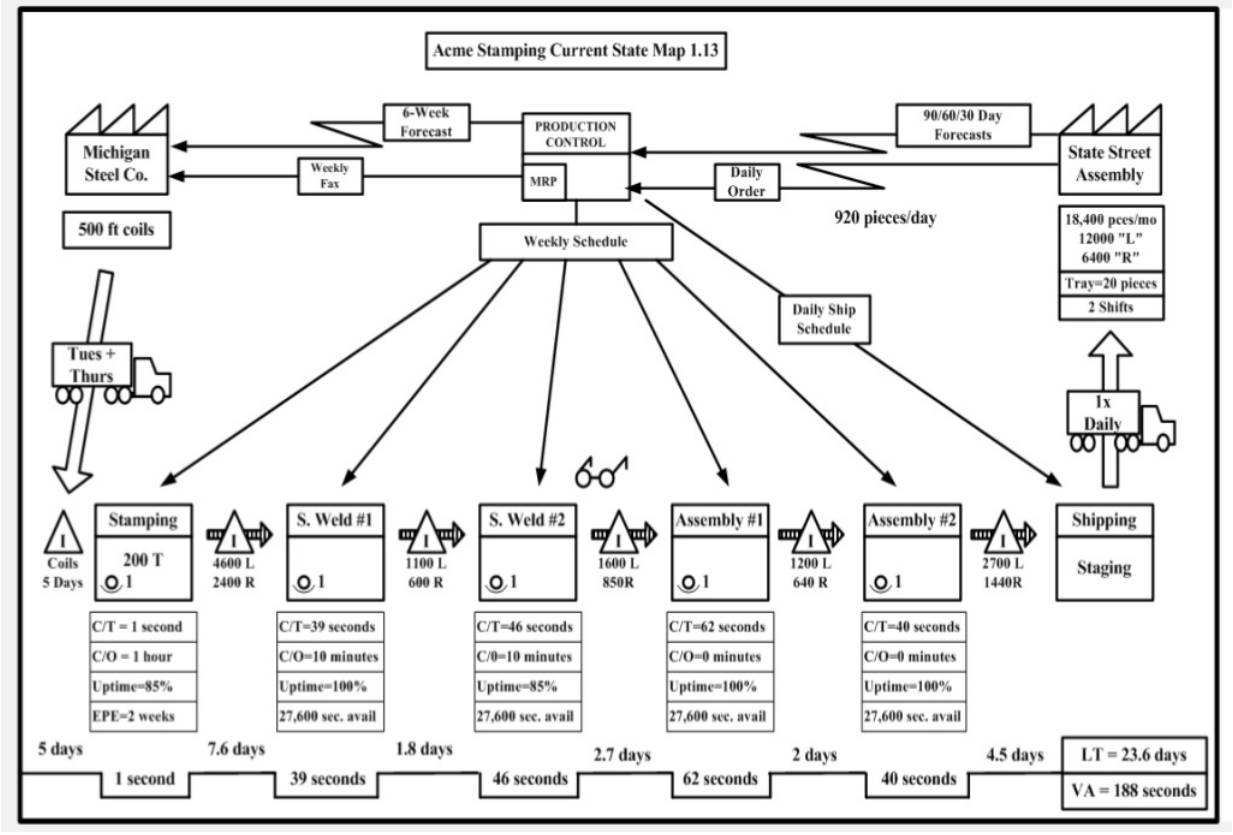

Figura 4 – exemplo de um VSM, imagem adaptada de Rother e Shook (1999)

Como se pode ver, neste caso particular o tempo de ciclo total do produto é de 23,6 dias, enquanto que o seu RPT total (VA no VSM) é de apenas 188 segundos, demonstrando uma

diferença surpreendentemente grande entre estes dois fatores. A partir deste diagrama podemos verificar imediatamente uma sobreprodução na área de *Stamping*, que é um processo extremamente rápido (apenas 1 segundo por produto) comparado com o seguinte, o que cria um inventário de 7,6 dias. Esta sobreprodução pode ser parcialmente explicada pelo alto tempo de conversão/*setup* (C/O), que no caso é de uma hora. Para evitar fazer demasiadas destas conversões morosas, possivelmente a estampagem produz uma quantidade superior àquela que a soldadura consegue processar de modo a evitar que a soldadura fique sem material para processar. Um melhor planeamento de produção entre as duas áreas poderia diminuir este tempo de espera, sendo o *Stamping* tão mais rápido que *S. Weld #1*, uma produção em formato *pull* seria aconselhável. Por outro lado, se o equipamento de estampagem tiver a capacidade de servir várias outras linhas, pode perder essa capacidade se for sincronizado com uma linha específica. É este tipo de visão global e análise rápida que o VSM permite obter.

Também podemos identificar o gargalo da produção como sendo o *Assembly #1*, com um tempo de ciclo de 62 segundos. Tendo um *uptime* de 100%, significa que a fábrica está a trabalhar a 100% da capacidade deste equipamento e, portanto, a cadência deste material na linha de produção é de uma peça por 62 segundos ou ~1peça/min. Com estes dados podemos controlar a entrada de material de modo a apontar para este valor, sabendo que se tivermos uma cadência de entrada superior a este valor, a fábrica vai obrigatoriamente acumular material em inventário.

2.8 *Mizusumashi*

Existem vários tipos de transporte de produtos de um equipamento para o próximo. A vasta maioria deles pode ser intuitivamente compreendida, como tapete de rolos ou transporte manual, no entanto existem alguns tipos mais complexos. De acordo com Coimbra (2013), o *mizusumashhi* (palavra japonesa para "aranha de água") é um operador logístico responsável pelo transporte interno de materiais, utilizando uma rota e tempos fixos, semelhante ao funcionamento de um autocarro.

Este operador transporta toda a informação relacionada com ordens de produção (*kanbans*) assim como todos os produtos entre áreas/células de trabalho, repetindo os mesmos movimentos num circuito fechado e fixo (geralmente 20 a 60 minutos). Neste ciclo, o operador para num certo número de estações ao longo da sua rota e verifica a necessidade de materiais nessa área. Este operador, geralmente usa um pequeno comboio para o transporte que deve ser capaz de fornecer material a todas as estações da sua rota. O seu tempo de ciclo é também conhecido como *pitch time*.

O potencial do *mizusumashi* para melhorar a eficácia do transporte é vasto. O objetivo principal de usar esta solução é transferir o trabalho sem valor agregado para um operador especializado de modo a otimizar o trabalho dos operadores dos equipamentos (Reis et al. 2016), evitar o tempo em vazio de retorno numa rota de ida e volta e eliminar causas de variabilidade na entrega de materiais.

2.9 Diagrama de causa-efeito

O diagrama de *Ishikawa*, também conhecido como diagrama de causa-efeito, foi desenvolvido por Kaoru Ishikawa na década de 60 com o objetivo de encontrar as causas raiz da variabilidade da qualidade dos produtos. Com o tempo, o diagrama foi adaptado a outros tipos de problemas e é agora uma das ferramentas mais utilizadas para encontrar causas de problemas em qualquer indústria. Originalmente, de acordo com Ishikawa (1976), na maior parte dos casos há 3 causas principais para esta dispersão (as 3 primeiras da lista), porém esta doutrina já foi atualizada para incluir um total de 6 causas:

1. Matéria prima
2. Equipamentos
3. Método de trabalho
4. Pessoas
5. Medição
6. Meio ambiente

É, portanto, aconselhável começar a desenhar o diagrama usando estas causas. O diagrama (figura 5) consiste numa linha central da qual emergem as causas aparentes do problema que se pretende resolver, identificado na caixa à direita. Para cada causa, aplica-se a mesma metodologia, efetivamente integrando o diagrama a cada uma dessas razões. Seguindo este conceito, podemos, eventualmente, encontrar as causas raiz do problema.

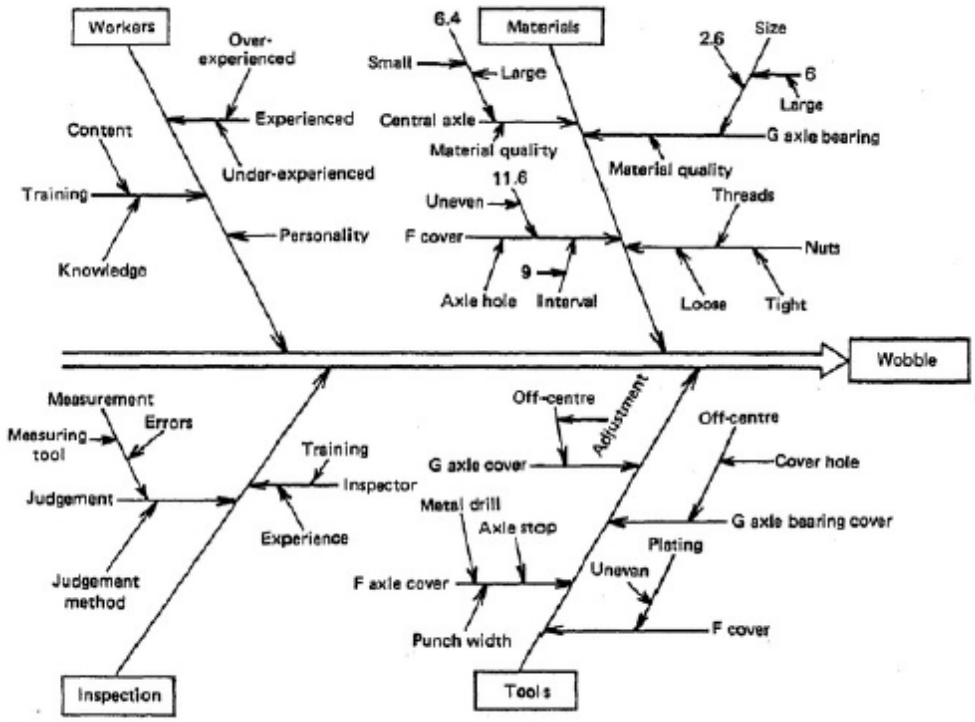

Figura 5 - diagrama de causa-efeito, imagem adaptada de Ishikawa (1976)

2.10 Os 5 porquês

A metodologia dos 5 "porquês" (5W) é frequentemente atribuída a Taiichi Ohno que, na década de 50, desenvolveu uma forma de abordar problemas na indústria automóvel que mais tarde foi expandida para as restantes indústrias. Ohno via os problemas "não como um fator negativo, mas antes como uma oportunidade para melhorar os processos" (Ohno 1988), aplicando na sua mais natural forma o conceito de *kaizen* (melhoria continua). A metodologia 5W, como o nome indica, consistia em fazer a pergunta "porquê" cinco vezes para determinar a causa raiz de um problema, no entanto o número de perguntas é flexível. Ele costumava usar o exemplo de um processo de soldadura robotizada que estava parada para provar a eficácia do seu método (Toyota Motor Corporation 2006):

1. **Porque é que o robot parou?**
   O circuito estava sobrecarregado o que causou a detonação de um fusível.
2. **Porque é que o circuito estava sobrecarregado?**
   Falta de lubrificação nos rolamentos, o que causou a sua paragem
3. **Porque é que havia falta de lubrificação nos rolamentos?**
   A bomba de óleo no robot não estava a bombear óleo suficiente.
4. **Porque é que a bomba não estava a bombear óleo suficiente?**
   A entrada da bomba entupiu com aparas de metal.
5. **Porque é que a bomba entupiu com aparas de metal?**
   Porque não havia nenhum filtro na bomba.

Chegamos então à conclusão que o *robot* parava porque a bomba de lubrificação dos seus rolamentos não conseguia filtrar as aparas de metal criadas pelo processo. Neste caso a sequência de perguntas não precisa de continuar pois já conseguimos solucionar o problema com estes dados. Deve-se, no entanto, ser cuidadoso na aplicação deste método de modo a evitar entrar em ciclos de causalidade e resistir à tentação de escolher uma resposta genérica que impeça o normal desenvolvimento da metodologia.

Apesar de esta metodologia ser atribuída a Taiichi Ohno, existe um provérbio inglês mais antigo, com origens germânicas, que também a explica bem esta ideia:

For want of a nail the shoe is lost;
For want of a shoe the horse is lost;
For want of a horse the rider is lost;
For want of a rider the battle is lost;
For want of a battle the kingdom is lost;
And all for the want of a horseshoe nail.

-George Herbert

2.11 Teoria das restrições

A teoria das restrições foi apresentada por Eliyahu M. Goldratt em 1984 com o seu livro intitulado *The Goal* e recorrentemente atualizada ao longo dos anos pelo mesmo autor sendo a última versão *Isn't It Obvious?* de 2009. Entenda-se uma restrição como qualquer fator que esteja a limitar as possibilidades de uma organização atingir os seus objetivos. Há vários tipos de restrições, que se podem dividir em 2 grupos: restrições internas e externas. As restrições internas podem ser facilmente reconhecidas quando a procura do mercado é superior à capacidade de uma fábrica e podem ter uma variedade de origens, desde equipamentos com capacidade insuficiente a falta de operários ou políticas empresariais que limitam a produção. As restrições externas identificam-se precisamente pela razão contrária, quando há mais capacidade de produção do que a procura do mercado, diz-se que a restrição é o próprio mercado, pois é o que limita a empresa de produzir mais.

Esta teoria apoia-se na premissa de as organizações poderem ser controladas pelo rendimento da empresa, custos operacionais e inventário. O rendimento da empresa é a frequência a que o dinheiro é gerado ao vender os seus produtos ou prestar os seus serviços, os custos operacionais, como o nome indica, é a frequência a que a empresa gasta o dinheiro ao

produzir os produtos ou a prestar os seus serviços e o inventário são todos os ativos da organização. Como se pode entender por estas definições, para Goldratt e Cox (1992) o objetivo final de um negócio é gerar lucro.

Outra premissa base desta filosofia de gestão é a de qualquer sistema está restringido de atingir os seus objetivos (sejam eles gerar dinheiro, aumentar a produção, etc) por um número reduzido de restrições, geralmente uma. Esta ideia é na realidade muito semelhante à definição de *bottleneck*, porém é mais genérica pois aplica-se não só a uma linha de produção, mas a qualquer sistema com um objetivo específico. Apenas libertando o sistema desta restrição, ou alargando a mesma, se consegue produzir mais. De acordo com (Goldratt e Cox 1992), há 5 passos (denominados *the five focusing steps*) que podemos seguir para melhorar o desempenho de um sistema:

1. Identificar a restrição do sistema
2. Decidir como melhorar a restrição
3. Submeter o resto do sistema à restrição
4. Elevar (quebrar) a restrição
5. Voltar ao passo 1 (melhoria contínua)

Depois de identificada a restrição e decidido como a melhorar, mas antes de o fazer, deve-se equilibrar o sistema em relação a essa restrição, pois, por definição, o sistema não consegue produzir mais do que a restrição permite. Este passo permite libertar alguma inércia desnecessária do sistema, permitindo agilizar os processos mantendo o mesmo rendimento sem ser necessário nenhum investimento monetário. Quando o passo 4 é concluído com sucesso, se o desempenho da restrição identificada for elevado de tal modo que já não está a restringir o sistema, diz-se que a restrição foi quebrada. Nestes casos, apesar de parecer que o sistema ficou otimizado, o passo 5 não deve ser ignorado. Quebrando uma restrição apenas assegura que existe outra restrição a limitar o desempenho do sistema, mesmo que esta seja uma limitação superior.

3 Situação atual

3.1 A empresa

A ATEP é uma empresa que se especializa na produção do *packaging* de semicondutores em formato *wafer* para aplicação na área das comunicações e automóvel, o que a localiza no centro de um processo geral para a produção de semicondutores. Devido ao facto de os clientes terem a maioria da sua produção localizada no Oriente, onde também têm empresas de produção de semicondutores, existe uma pressão adicional em produzir com mais celeridade para compensar o tempo despendido no transporte. Por esta razão, o estudo do tempo de ciclo torna-se particularmente interessante e o conhecimento aprofundado do mesmo e de onde o melhorar cria uma vantagem competitiva para a empresa.

Ainda assim, dois dos principais custos da empresa são a energia e os materiais consumíveis e é necessário gerir estes fatores, mesmo que afetem negativamente o tempo de ciclo de um produto. Este é um dos vários fatores que complicam o cálculo do tempo de ciclo e estudo dos tempos de espera, visto que estes tempos são manualmente afetados para minimizar outros custos e por essa razão as prioridades e os tempos em fila de espera não obedecem a uma distribuição normal e têm uma variabilidade elevada. É, portanto, necessário encontrar um equilíbrio entre um tempo de ciclo e o custo dos recursos consumidos.

A empresa opera com máquinas automáticas em praticamente todos os processos, apenas precisando de operadores para fazer tarefas administrativas e inspeção manuais. Estes processos já se encontram bem documentados, com *raw tool times* (RTTs) de reduzido valor e variabilidade quando comparados com os tempos de espera que tendem a ser 2 a 3 vezes superiores ao RTT. Por esta razão torna-se crucial focar o estudo do tempo de ciclo nos tempos de espera. A diferença entre *raw process time* (RPT) e RTT é também diminuta, o único fator que faria variar consideravelmente estes dois tempos são os tempos de *setup*, no entanto, nos equipamentos em que há *setups* longos, costuma haver equipamentos dedicados a cada família de produtos o que efetivamente, em teoria, elimina a necessidade de *setups*, tornando negligenciável esta diferença entre RPT e RTT. Como sabemos, as restantes diferenças entre estes dois tempos são as operações administrativas de *setup* que não são efetuadas em todos os processos, tendem a ser consideravelmente menores do que o tempo de processo e nos piores casos não ultrapassam os 20 minutos (comparado com a média de 1h de todos os processos), pelo que estes dois termos são geralmente intercambiáveis.

Porém, convém ter em conta que nem todos os equipamentos com *setup* existem em quantidade suficiente para poder estar um dedicado a cada família de produtos, o que obriga a fazer *batching* em série (sequenciamento) e os equipamentos que estão dedicados são, normalmente, em número reduzido pelo que qualquer avaria (embora raras) pode obrigar a conversão de *setup* de outro equipamento, o que força que se faça um sequenciamento improvisado.

Existem várias formas de dividir os produtos por família. A primeira distinção pode ser feita entre produtos *fanin* e *fanout*, diferentes tipos de produto com bases técnicas distintas e vários processos independentes. Porém também se pode distinguir famílias de produtos pelas camadas de metal que lhe são aplicadas, ou por um tipo de processo que apenas alguns produtos partilham, ou ainda por outros fatores. Por esta razão, torna-se excessivamente complexo pormenorizar em demasia a família de produtos, sendo suficiente diferenciar os produtos por *fanin* e *fanout*.

3.2 Descrição do processo

Para melhor compreensão da situação atual da empresa, é útil conhecer algumas particularidades técnicas dos produtos e do processo. Há duas famílias distintas de produtos que esta empresa produz – *fanin* e *fanout*. Os primeiros são sempre fornecidos em *wafers* de 300mm (figura 6), ao passo que os segundos são fornecidos em *wafers* de silício de 200 ou 300mm, podendo cada *wafer* conter milhares de circuitos integrados. Existem algumas diferenças técnicas que obrigam os produtos *fanout* a ser transferidos para *wafers* normais de 300mm, o que resulta num processo com passos iniciais diferentes. Depois de estarem "normalizadas", ambas as famílias de produtos são processadas de forma relativamente semelhante. As únicas diferenças são o número de *wafers* por lote, geralmente 25 nas *fanin* e 12 nas *fanout*, e os processos de deposição de camadas que são mais extensivos para as primeiras.

Foquemo-nos então numa fase inicial no processo das *wafers fanout*, pois englobam mais processos. Foram elaborados fluxogramas do processo geral desta família de produtos, que podem ser consultados nos anexos A e B, para auxílio desta descrição. Este produto é recebido em *wafers* de silício de 200 ou 300mm e começa o seu caminho na área de preparação das *wafers* (*wafer prep*) onde os *dies* são protegidos por uma camada plástica e cortados individualmente. Apesar de aparentemente simples, estes processos são dos mais demorados devido ao tamanho reduzido dos *dies* e a sua quantidade por *wafer*. De seguida os produtos são transportados para a área de reconstrução (*reconstitution*) onde os *dies* são transpostos noutra *wafer* de um material com melhor condutividade elétrica e tratados termicamente para colar os *dies* na *wafer*.

Figura 6 - *wafer* de silício 300mm não processada

As *wafers* iniciam então o seu processo de deposição de camadas de metal na área de RDL. É neste ponto que as *wafers fanin* começam o seu processo e convergem com as *wafers fanout*, seguindo processos semelhantes com diferenças apenas em pormenores técnicos, números de camadas de metal e tipos de camadas. A figura 7 apresenta um processo básico de deposição e revelação o qual pode ser repetido várias vezes consoante o número de camadas que o produto requerer. Efetivamente, o produto volta à mesma célula de trabalho várias vezes criando um ciclo no seu fluxo de processos. Este processo é geralmente feito por fotolitografia (passos 2 a 5) e tem o objetivo de criar, dentro e sobre a superfície da *wafer*, um padrão com as dimensões estabelecidas na fase de desenho e colocar corretamente o padrão do circuito na *wafer* (Van Zant 1997).

Num processo típico de deposição de camadas, é inicialmente depositada uma camada de material dielétrico[1] sobre a superfície da *wafer* (a tracejado na figura 7). Um material metálico foto-resistente é depositado sobre toda a camada de material dielétrico e o padrão requerido é coberto por uma máscara opaca enquanto o restante material é exposto a radiação de modo a ficar polarizado. Finalmente é retirado, sequencialmente, o material não polarizado (passo 5), o metal dielétrico no padrão do circuito da *wafer* (passo 8) e o material foto-resistente excedente (passo 9), intercalando passos de cura e inspeção quando necessários.

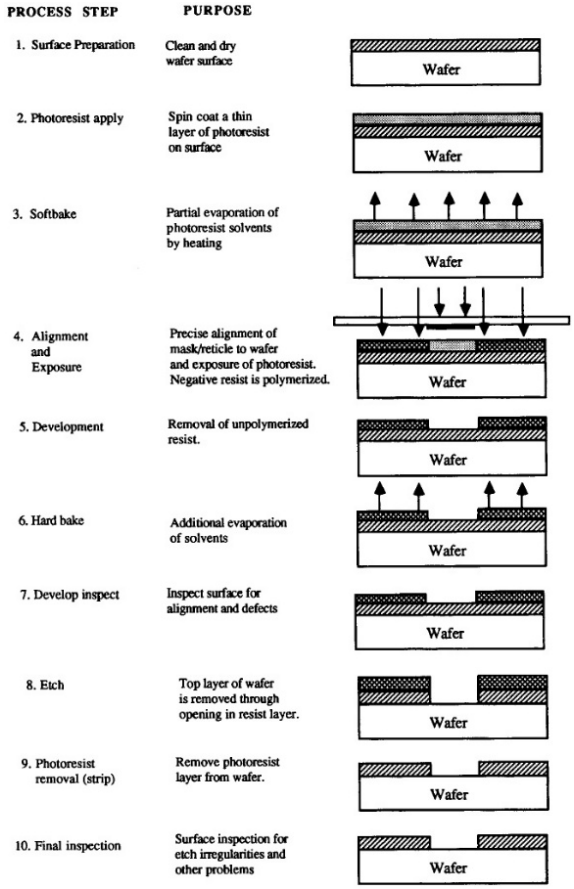

Figura 7 - processo de deposição e revelação de uma camada dielétrica, imagem adaptada de Van Zant (1997)

Para padrões mais complexos, a fotolitografia é geralmente repetida com padrões sobrepostos diferentes. A cobertura da *wafer* com o material foto-resistente pode ser feita depois de haver já um padrão como exemplificado na figura 8.

[1] Material dielétrico é todo o material que não conduz eletricidade sob condições normais, também conhecidos como materiais isolantes.

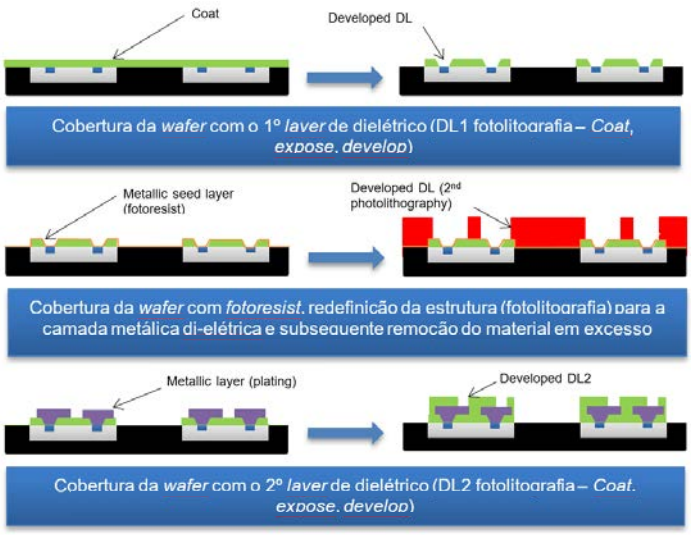

Figura 8 - processo de deposição de várias camadas numa *wafer*, imagem adaptada de Nanium

Este processo é frequentemente repetido com alterações técnicas consoante a complexidade nos padrões dos circuitos e o número de camadas da *wafer* e, como tal, é o processo principal da área de RDL sendo que os outros processos desta área apenas o complementam ou auxiliam (inspeções, correções, limpezas, curas, etc).

No fim de cada deposição de camadas há um processo de inspeção na área de metrologia onde se faz a verificação de cada *die* para confirmar a sua integridade. Estas inspeções são feitas por equipamentos automáticos próprios e servem apenas de "filtro" para acelerar o processo de inspeção. Apesar de demoradas, as inspeções feitas por equipamentos automáticos conseguem ser mais rápidas do que as inspeções manuais, porém estas devolvem demasiados resultados falso-positivos, pelo que são sempre necessárias as inspeções manuais nos *dies* marcados. Nestas inspeções automáticas, os equipamentos marcam e fotografam os *dies* que consideram defeituosos, sendo depois responsabilidade dos operadores da área verificar individualmente esses *dies* e finalmente identificar os *dies* que são, de facto, defeituosos, segregando-os quando possível para retrabalhar a *wafer* e corrigir o defeito.

São ainda feitas inspeções visuais entre cada processo, para verificar que a *wafer* não foi danificada pelo equipamento. Embora sejam muitas, estas inspeções são mais rápidas, não ocupando mais do que alguns segundos por *wafer*.

O produto final deste processo é uma *wafer* com circuitos impressos nos *dies*.

São, no fim, depositadas as bolas de solda para fazer a ligação elétrica entre os *dies* e o exterior na área LBS. O cliente pode fazer a encomenda dos *dies* em *wafers* ou individualizados em tabuleiros (*frames*) ou rolos (*tape reel*) pelo que o passo final será a possível transladação dos *dies* para o método de acondicionamento escolhido pelo cliente.

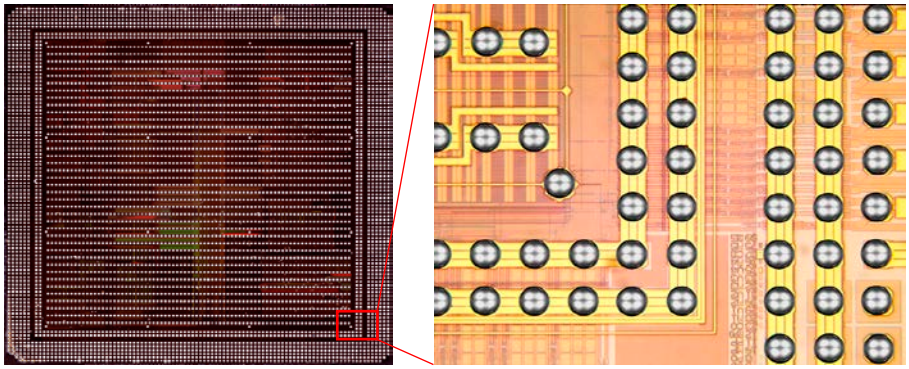

Figura 9 - *Die* processado

3.3 *Value Stream Mapping*

De seguida, foi realizado um VSM para melhor compreensão do sistema global apresentado na figura 10.

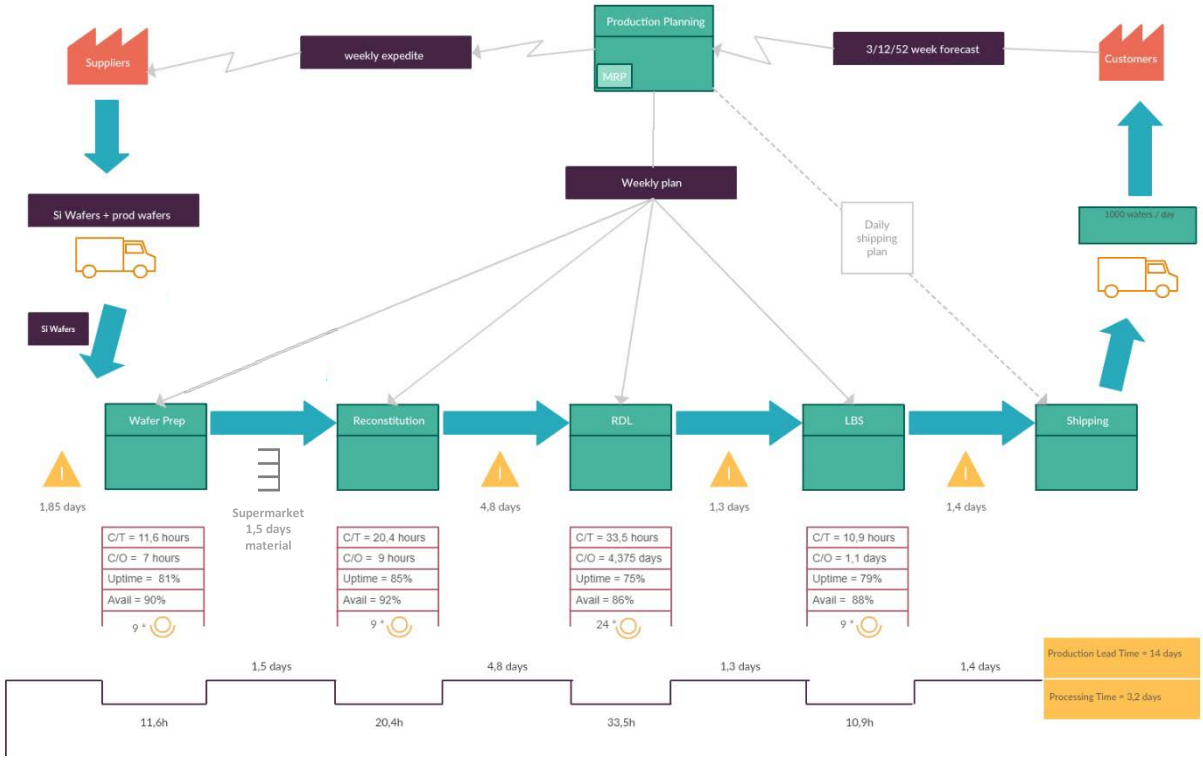

Figura 10 - VSM dos produtos *fanout*

Nesta figura apresentamos o percurso geral de um produto classificado como *fanout*. Este tipo de produto requer um processo mais moroso, com mais passos, ao contrário dos produtos classificados como *fanin* que não requerem os dois primeiros processos apresentados (*wafer prep* e *Reconstitution*), no entanto, as *wafers fanin*, têm um tempo de ciclo em RDL superior, devido a terem mais camadas metálicas. Ainda assim o tempo de ciclo médio dos produtos *fanin* é de apenas 10 dias em contraste com os 14 dias de *fanout*.

Devido ao extenso número de processos existentes foram agrupados na figura acima alguns processos nas áreas apresentadas de modo a simplificar e facilitar a visualização da linha de produção. Assim, os valores apresentados representam a soma dos valores das operações em cada tipologia de processo. Quando se regista que existe um inventário de 4,8 dias em RDL, tal significa que a soma dos tempos entre operações dessa área é de 4,8 dias. A área de *Wafer Prep* funciona de forma algo desligada das restantes, alimenta um "supermercado" ou *buffer* que liga à área seguinte – *reconstitution* (ou simplesmente *recon*). Este buffer é normalmente mantido de modo a haver materiais para um ou dois dias e alimenta diretamente os processos iniciais de *recon*.

Como se pode verificar neste mapa, existe uma grande possibilidade de melhoria nos tempos de ciclo. O tempo de processamento médio dos produtos é de apenas ~4 dias tanto para *fanout* como para *fanin* (sabendo que os produtos *fanin* têm maior variabilidade) e o *lead time* para estes produtos é de 14 e 10 dias, respetivamente. Isto implica um tempo de espera entre processos de 6 a 10 dias. Como se pode concluir pelo VSM, os produtos permanecem cerca de 1/3 do seu *lead time* em RDL. É, portanto, nesta área que se focará o estudo do tempo de ciclo, com especial atenção aos tempos de espera, sendo estes tão ou mais consideráveis do que os tempos de processo.

3.4 Restrições da produção

Numa fase inicial, a realização deste projeto englobou várias sessões de formação dentro do departamento de planeamento e logística da ATEP com os diferentes intervenientes da área, incorporando visitas à linha de produção e infraestruturas da empresa. Estas sessões tiveram como objetivo conhecer os diferentes processos de produção bem como a logística, planeamento e os elementos chave para a compreensão da situação atual da ATEP e foram cruciais no mapeamento dos tempos de espera. Ao ver fisicamente o processo e acompanhar os operadores no seu trabalho, foi possível localizar vários fatores que podem eventualmente contribuir para o aumento do tempo de ciclo e a sua variabilidade.

Como já foi referido, a indústria de semicondutores tem um processo de produção particularmente complexo. As datas de entrega das encomendas são definidas pelos próprios clientes, pelo que o cálculo exato e *à priori* do tempo de ciclo dos produtos não tem sido prioritário, sendo suficiente saber se a entrega consegue ser feita em tempo útil. O excesso de capacidade da fábrica permite controlar este tempo de ciclo de modo a não falhar entregas.

Ainda assim, conseguimos verificar que grande parte do tempo de ciclo dos produtos é derivada dos tempos em espera, mais especificamente a área de RDL que tende a criar *buffers* consideráveis. Este grupo engloba vários processos, entre os quais vários processos distintos de litografia que podem ser realizados pelos mesmo equipamentos e variam consoante os produtos. Por esta razão, o produto terá de regressar à mesma máquina várias vezes, criando efetivamente um *loop* o que incrementa uma complexidade acrescida na previsão de tempos de espera, pois a máquina efetivamente recebe produtos em bruto e já trabalhados ao mesmo tempo, o que cria incertezas sobre qual produto priorizar. A diferente combinação de produtos em linha, juntamente com a camada que está a ser trabalhada (DL1, RD1, DL2, etc), com a capacidade instalada a cada semana e com os outros fatores seguidamente apresentados dificulta a imposição de regras específicas para automatizar esta priorização.

Os principais casos particulares desta indústria que afetam o tempo de ciclo são os seguintes:

- *Batching em paralelo*
- Sequenciamento
- Janelas temporais *(Time Windows)*
- Transporte
- Atividades de engenharia
- Casos especiais
- Planeamento da produção

3.4.1 *Batching* em paralelo

Existem processos, entre os quais os fornos de cura, que são críticos para o cálculo do tempo de ciclo por fazerem *batching* em paralelo. Estes processos aceitam vários lotes de *wafers*, pelo que o primeiro lote terá efetivamente de aguardar que o forno seja completamente preenchido para o processo poder começar. Este facto cria uma discrepância entre os tempos de ciclo do primeiro e último lote caso o processo não se encontre constantemente preenchido.

Para maximizar a utilização do equipamento, este processo apenas se inicia quando o *batch* estiver completo. Para completar um *batch*, não é sempre necessário que se usem apenas lotes do mesmo produto, apenas têm de fazer o mesmo processo e fazer parte da mesma família de produtos. Alguns casos permitem o processamento de um *batch* incompleto, tais como quando um lote tem prioridade alta, quando não é expectável que haja um lote da mesma família por algum tempo (tempo esse que não é definido) ou quando se está em risco de falhar um prazo de entrega para um certo lote.

No caso particular de uma máquina que aceita quatro lotes e tem uma cadência de quatro lotes a cada seis horas, o primeiro produto tem um tempo de processamento de 6 horas mais o tempo de espera, enquanto o último tem apenas o tempo de processamento de 6 horas. Por outro lado, se o processo se encontrar frequentemente com uma taxa de utilização alta, este *batching* não teria um efeito tão agravante pois os lotes teriam sempre de aguardar pelos produtos que estão a ser processados em vez de esperarem pelo *batch* completo. Neste caso, o tempo de espera devido ao WIP desse equipamento sobrepor-se-ia ao tempo de espera devido ao *batching*, no entanto, os produtos ainda experienciariam tempos de espera diferentes, e desequilibraria o fluxo de produtos a jusante causando inevitavelmente tempos de espera nos processos seguintes.

3.4.2 Sequenciamento

A linha também tem processos bem definidos onde se faz sequenciamento, o que incrementa outra variável para o cálculo dos tempos de espera. Quando um lote chega a uma máquina que está a ser utilizada para outra família de produtos, para minimizar a mudança excessiva de *setups* da máquina, esse lote fica em espera até terminar o processamento de todos os lotes que estavam em espera daquela família de produtos. Há dois processos onde o efeito do sequenciamento no tempo de ciclo é mais visível, nos restantes este efeito não é tão explicito.

Alguns equipamentos consideram famílias de produtos diferentes com o mesmo *setup*, isto é, um equipamento processa os produtos A e B e apenas faz a conversão de *setup* para produtos C (ou vice-versa), outros equipamentos processam produtos A e C e apenas fazem a conversão de *setup* para produtos B. Esta dessincronização de *setups* é particularmente problemática pois cria um grande desequilíbrio na linha de produção, uma vez que, para evitar perder tempos com conversões de *setup*, o primeiro equipamento naquele exemplo envia para o segundo produtos A e B, mas este só pode processar os produtos A, mais uma vez para minimizar conversões de *setup*, obrigando o produto B a ficar à espera.

Devido a esta complexidade, não há uma regra geral para a escolha dos produtos a processar. Em alguns casos, o equipamento tem de fazer conversão de *setup* para todos os produtos, no qual se procura fazer um sequenciamento mínimo de 4 lotes, ou seja, apenas se faz a conversão de *setup* quando acabou de ser processado um tipo de produtos e há 4 lotes da mesma família em fila de espera. Caso estes não existam, verifica-se os produtos nos processos anteriores e o gestor de produção toma a decisão *in loco* sobre se é preferível manter o *setup* ou fazer alguma conversão. Por vezes, para evitar falhar uma entrega a tempo, são criadas exceções e faz-se a conversão para processar apenas um produto.

3.4.3 Janela de tempo

De notar ainda a particularidade de os lotes terem de aguardar um certo tempo posteriormente a certos processos, característica conhecida por janelas de tempo mínimas (*time windows*), para efeitos de arrefecimento, estabilização, entre outros. Este pormenor pode ser considerado como um processo em si, de tempo fixo por produto, e capacidade infinita e acontece principalmente na área de litografia e de reconstrução da *wafer* com tempos que podem varia entre meia hora e 15 horas.

Existem também janelas de tempo máximas, ou seja, o produto não pode aguardar mais do que um determinado tempo entre alguns processos. O incumprimento implica que o produto seja processado novamente desde um ponto pré-definido, no entanto esta ocorrência não é comum e é raramente problemática, pois as janelas de tempo máximas são, em regra, bastante elevadas e quando um lote está em risco de a sua janela ser quebrada é dada prioridade a esse lote.

3.4.4 Transporte

Existe ainda o contributo do transporte do produto para o aumento dos tempos de ciclo. A maior parte dos transportes dentro de cada área é feito manualmente pelos operadores. Quando um equipamento termina o processamento do lote, este é transportado manualmente para a prateleira do processo seguinte onde fica à espera para ser processado.

No entanto, para fazer a ligação entre as diferentes células de produção utiliza-se o *mizusumashi*. A distância entre estas células é sempre curta, não excedendo algumas dezenas de metros. Neste caso em particular, o *mizusumashi* tem um tempo de ciclo de uma hora, significando que começa o trajeto sempre de hora em hora. Este tipo de transporte otimiza o trabalho dos operadores, impedindo-os de perder tempo com o transporte e cria um abastecimento contínuo de produtos nas diferentes células. O fator humano que afetaria a variabilidade da cadência de transporte é, então, reduzido. Um caso particular desta variabilidade que merece referência é o "*batching*" de transporte que os operadores tendem a fazer para evitar viagens excessivas, isto é, do ponto de vista de um operador, se houver vários lotes no fim de um processo é preferível esperar e levar todos esses lotes do que fazer o mesmo trajeto várias vezes. São estes fatores humanos que são principalmente combatidos pelo *mizusumashi* ao segmentar o trabalho do transporte.

Por outro lado, o *mizusumashi* cria uma flutuação dos tempos de espera pelo transporte dos lotes que pode variar entre poucos minutos (tempo supérfluo comparado com alguns tempos de processo) e uma hora, sendo que se pode considerar um tempo de espera médio de 30 minutos. Esta variabilidade pode ser particularmente prejudicial nos processos que requerem *batching*, pois pode criar o atraso de um ou poucos lotes que depois atrasarão um *batching* ou causarão um *batching* incompleto. Alguns produtos podem ser transportados pelo *mizusumashi* mais do que 25 vezes ao longo do processo produtivo, ou seja, o tempo de espera por este transporte pode facilmente ultrapassar as 10 horas durante o processo inteiro de produção.

Ainda assim, o efeito da espera por este tipo de transporte pode eventualmente não ser representativo. Se os processos para onde o *mizusumashi* for transportar os produtos já se encontrarem com fila de espera, o tempo de espera por este transporte torna-se irrelevante pois o lote permaneceria em fila de espera mesmo que o tempo de espera pele transporte fosse nulo.

3.4.5 Atividades de engenharia

Para além de encomendas recorrentes, a ATEP também processa lotes onde se pretendem testar inovações ou processos diferentes. Estes processos são intitulados atividades de engenharia e podem ser pedidos pelos clientes ou ser parte de um projeto interno de melhoria contínua. As atividades são sempre de prioridade inferior aos produtos comerciais, mas é sempre reservado um tempo fixo por equipamento por semana para se fazerem estas atividades e cabe aos gestores de produção decidir quando alocar estes tempos a cada equipamento.

Há frequentemente mais capacidade nos equipamentos do que a necessária para a produção, o que significa que há o potencial de haver equipamentos temporariamente parados sem produtos. Este tempo é calculado no planeamento da produção e disponibilizado para as atividades de engenharia (designados corredores de engenharia) e é neste tempo que os gestores de produção tentam alocar as atividades de engenharia, no entanto nem sempre é possível pois a variabilidade da cadência de produtos faz com que estes corredores não sejam nem contínuos nem fixos e por vezes são demasiado pequenos para completar uma atividade completa. Em alguns casos as atividades são iniciadas e, ou por atraso ou devido à variabilidade da cadência de produtos, colocam outros produtos comerciais à espera.

É ainda relativamente frequente a conversão de *setup* de equipamentos para estas atividades, o que faz com que estes equipamentos tenham de ser de novo convertidos para o

material de produção. Apesar de este fator ser considerado pelo gestor de produção, que tenta alocar as atividades às máquinas já dedicadas àquele tipo de produto, nem sempre é possível fazê-lo.

3.4.6 Casos especiais

Como em qualquer linha de produção, existem ainda casos particulares pouco comuns que podem afetar o funcionamento da produção. Todas as indústrias lidam com problemas de qualidade de forma diferente, geralmente quando se trata de produtos de baixo valor e alto volume os produtos tendem a ser descartados quando não seguem os padrões de qualidade especificados. No entanto, no caso da indústria de semicondutores, isto apenas acontece quando não é possível corrigir os produtos com defeito. De um modo geral, quando uma *time window* é violada ou quando um processo causa defeitos numa *wafer* ou *dies*, o lote inteiro ou parte dele fica em estado *hold*, o que significa que fica à espera de ser retrabalhado ou corrigido. Lotes neste estado não costumam ser corrigidos em pouco tempo pois têm de ser feitos processos específicos que por vezes requerem conversões prolongadas de *setup* de equipamentos. Como tal, nestes casos, o lote apenas é corrigido quando os equipamentos estiverem disponíveis, portanto em *stand-by* e sem atividades de engenharia. Por esta razão, lotes em *hold* tendem a ficar neste estado vários dias, o que cria um volume considerável de lotes neste estado (entre 20 e 30% dos lotes totais em WIP), apesar da frequência deste evento ser consideravelmente inferior.

Todos os produtos *fanout* são recebidos em lotes de 25 *wafers*, no entanto após poucos processos iniciais são divididos (fazem *split*) em 2 lotes de 12 e 13 *wafers*. Este *split* acontece devido às características dos equipamentos. A maior parte dos equipamentos dedicados aos produtos *fanout* está convertido para lotes de 12 ou 13 *wafers* (a posição das *wafers* no lote afetam o funcionamento do equipamento, entre outros fatores), pelo que este *split* é necessário nos equipamentos que assim o requererem. Assim, quando há apenas um equipamento disponível e os lotes acabaram de ser divididos, um dos lotes vai obrigatoriamente esperar pelo próximo equipamento a ficar disponível. Por vezes há apenas um equipamento que pode processar aquele produto, o que significa que um dos lotes vai sempre ter de esperar pelo menos o tempo de processo do seu lote gémeo. Neste caso, os processos em questão são bastante longos, entre 3 a 10 horas, pelo que o tempo em espera causado por este fator parece ser considerável, por outro lado é algo que apenas acontece uma vez durante a produção pelo que a sua influência no tempo de espera total é algo reduzida.

Convém também referir que alguns processos não são feitos para todos os produtos ou para todas as *wafers* de um lote. Por vezes, principalmente nos processos de metrologia e inspeção de qualidade, apenas uma pequena parte do volume de produção necessita de medição ou inspeção e esta pequena parte está quantificada *à priori* para cada lote. Nestes processos diz-se que os produtos fazem amostragem. As restantes *wafers* dos lotes que fazem amostragem mantém-se à espera dos mesmos. Como apenas alguns lotes fazem estes processos, esses vão de facto ter um acréscimo ao seu tempo de ciclo.

3.5 Planeamento da produção

É elaborado semanalmente um plano de produção onde se define a produção planeada e a utilização das máquinas para a próxima semana. Este plano advém das encomendas dos clientes, neste sentido a empresa funciona num sistema *make-to-order*. Estas encomendas contêm a quantidade de produtos pretendidos num espaço de 3 semanas, assim como a sua data de entrega. Por outro lado, na fábrica, cada área recebe o seu plano e apenas produz a quantidade definida nesse plano, independentemente das outras áreas. Neste sentido, a fábrica funciona num sistema *push*.

Os clientes também fornecem os planos para os próximos 3 meses e para o próximo ano, porém estes estão sujeitos a alterações e servem apenas de orientação para a produção. Por vezes a data de entrega pedida excede folgadamente o *lead time* e é possível completar a produção com alguma margem de tempo, no entanto os produtos apenas podem ser expedidos na data especificada como data de entrega. Estas datas de entrega são, no entanto, referentes à quantidade do produto e não aos lotes específicos, o que concede à empresa alguma flexibilidade para escolher quais lotes enviar consoante o seu estado na linha de produção.

Existem duas formas de controlar o *lead time* de modo a coincidir com a data de entrega quando esta oferece alguma folga para o tempo de processamento: atrasar a sua produção ou armazenar os produtos no fim da sua produção. Como os clientes têm sempre acesso ao estado do produto em qualquer ponto da sua produção, estes têm a capacidade de avaliar o desempenho da linha de produção e por vezes definem objetivos de tempo de ciclo, outras vezes não querem que o produto fique armazenado. Por estas razões, a utilização das máquinas também tem de ser controlada para gerir o tempo de ciclo. Neste caso, o tempo de ciclo é tratado como um objetivo a atingir mais do que uma oportunidade de melhoria. Todas as semanas é calculado o número de equipamentos necessários para a próxima semana e são desligados os restantes de modo a garantir a data de entrega com o mínimo de equipamentos possíveis por uma questão de poupança de energia e materiais consumíveis. A fábrica esta dimensionada para momentos de produção mais alta (a indústria segue uma ligeira sazonalidade trimestral), o que significa que de um modo geral a capacidade da fábrica é superior às suas ordens de produção, pelo que este planeamento é fundamental durante grande parte do ano.

Esta informação é apresentada aos gestores de produção que se reúnem diariamente para decidir o tempo que alocam para produção ou atividades de engenharia. A manutenção é feita, geralmente, de um modo preventivo, como tal é feita com base no tempo de utilização, pelo que os gestores têm de ter em atenção este fator.

A decisão para mudar ou manter o *setup* da máquina é também discutida nessa reunião onde é considerada a quantidade de produtos dos diferentes tipos que têm à espera, os produtos que esperam receber nesse dia (WIP precedente) e a data de entrega dos produtos, priorizando os produtos com *critical ratio* (CR) mais reduzido quando o plano ainda não foi cumprido, mas sem recurso a qualquer tipo de cálculo auxiliar que minimize o tempo de ciclo dos lotes em geral.

Devido à produção semanal se processar com este formato de objetivos, em que se deve processar apenas o plano semanal apresentado, os operadores por vezes evitam processar produtos mesmo tendo equipamentos disponíveis, não só porque já atingiram o plano semanal, mas também para garantirem que têm WIP para fazerem o plano da semana seguinte. Este é outro fator que afeta o tempo de espera de alguns produtos, embora geralmente apenas aconteça no fim de cada semana produtiva e apenas em alguns produtos.

A priorização dos produtos que estão em fila de espera para serem processados é também da responsabilidade dos gestores de produção. Esta priorização é bastante complexa, baseando-se, como já foi explicado, no *setup* atual da máquina, produtos em fila de espera a jusante da produção, objetivos diários e semanais, entre outros. Em casos onde todos estes fatores são semelhantes ou irrelevantes, o CR é o fator padrão que diferencia a prioridade dos lotes, sendo que um CR menor impõe maior prioridade ao lote. O CR é usado quando toma valores maiores que um, ou seja, quando é possível o lote terminar a sua produção a tempo, no entanto quando CR é igual ou menor que um significa que o lote vai invariavelmente chegar atrasado, e caso haja algum lote em linha com baixa margem de tempo (CR tende para um), o lote atrasado perde a prioridade para que o seu processamento não cause a falha da entrega de outros lotes. O valor exato desta “baixa margem de tempo” não esta definido, esta é uma heurística decidida pelos gestores de operação de cada área e varia ligeiramente consoante a pessoa responsável. A

subjetividade nesta decisão em particular cria uma instabilidade nos tempos de espera difícil de calcular e torna esta decisão propícia a erros.

3.6 *Bottleneck*

Devido a este tipo de planeamento a capacidade instalada da fábrica varia semanalmente, o *bottleneck* também vai variando e pode até haver mais do que um[2], tornando complicado salvaguardar o WIP nessa área. A solução utilizada passa por aumentar a utilização dos equipamentos anteriores (incluindo fazer mais do que o plano de produção se necessário) de modo a garantir que o fluxo dos produtos é mais célere nessas áreas do que no *bottleneck*, salvaguardando assim WIP elevado de produtos na área do *bottleneck*. Por exemplo, se for calculada a necessidade de haver apenas 2,5 equipamentos para cumprir o plano da semana, obviamente vão ser ligados 3, mas isto vai causar uma margem de tempo de processamento equivalente a meio equipamento. Isto significa que, na prática, podem-se usar os 3 equipamentos na sua máxima capacidade, produzindo mais do que o plano previa, de modo a salvaguardar WIP na área seguinte.

A identificação deste *bottleneck* é também mais complexa do que o normal, com dezenas de produtos com RPTs, *batching* e fluxos de produção diferentes a utilizarem os mesmos equipamentos (como já vimos, por vezes mais do que uma vez), a definição clássica de *bottleneck* apresentada no capítulo 2 que se refere ao *bottleneck* como o processo com cadência mais baixa não representa necessariamente o processo com menos margem de tempo para produção. Como tal, na indústria de semicondutores usa-se a definição de *bottleneck* para os equipamentos com menos margem de tempo disponível num determinado horizonte temporal (usualmente por semana), como explicado por Zhou e Rose (2009).

No caso específico desta linha, o *bottleneck* costuma estar em RDL ou *Wafer Prep*, dependendo dos fatores já referidos, com taxas de utilização dos equipamentos de certos processos perto dos 100% para os equipamentos disponíveis.

Apesar dos RPTs maiores em *Wafer Prep*, esta área não costuma ser a mais difícil de controlar pois há mais disponibilidade de máquinas, processa menos produtos (*fanin* não são processados por *Wafer Prep* nem *Recon*) e o *buffer* entre *Wafer Prep* e *Recon* fornece alguma margem de planeamento. Há também um aparente *bottleneck* em LBS, no entanto este apenas existe por uma questão de gestão de recursos, num equipamento que faz sequenciamento. Este equipamento, para além do sequenciamento, está divido para dois processos, o que efetivamente divide a sua capacidade pelos dois processos. Quando os produtos correm o risco de falhar os prazos, um segundo equipamento é disponibilizado aumentando a capacidade drasticamente e efetivamente elimina este *bottleneck*. Concluindo, a área com o *bottleneck* mais crítico é a de RDL.

[2] Quando há dois tipos de produtos com *bottlenecks* distintos, mas que partilham os mesmos equipamentos, dependendo da mistura de produtos em linha, o *bottleneck* do produto A pode atrasar outros produtos. Mais à frente na linha, os *bottlenecks* dos outros produtos podem atrasar o produto A. Temos então um caso em que a produção efetivamente "entope" em dois pontos distintos, obrigando os equipamentos a serem utilizados na sua máxima capacidade e mesmo assim não terem folga de tempo de processamento.

3.7 Análise da linha de produção

Na sequência da observação feita na secção 3.4, onde se encontram fatores que podem influenciar o tempo de ciclo, foi feita uma pré-análise de modo a reduzir a quantidade de fatores relevantes para o cálculo do tempo de ciclo.

A experiência dos operadores e a movimentação e documentação do lote podem ser excluídos por terem um impacto reduzido e pontual no tempo de ciclo quando comparados com outros fatores. Estes tempos raramente ultrapassam os 20 minutos, sendo que a maioria deles não chega aos 10 minutos. Embora sejam bastantes, o somatório de todos estes tempos não ultrapassa poucas horas e são processos manuais necessários.

As atividades de engenharia e manutenção são fatores previstos no planeamento semanal da produção e são alocados ao tempo disponível em que os equipamentos não estão produtivos (tempo em *stand-by*), pelo que estes fatores à partida têm um efeito reduzido no tempo de ciclo se as atividades estiverem a ser devidamente controladas. As atividades de engenharia, em particular, têm sempre uma alocação própria mínima por equipamento, geralmente nos 6%, mas que pode ser superior, e, como já foi referenciado, por vezes ocupa equipamentos mais tempo do que o planeado ou obriga a fazer a conversão de *setup* do equipamento, no entanto, de um modo geral, não costuma acontecer com frequência.

As inspeções automáticas na área de metrologia recorrentemente apresentam excesso de alarmes (falso-positivos). O tempo perdido com o despiste destes erros, embora significativo, apenas afeta uma área pelo que a sua resolução não terá grande impacto a nível de tempo de ciclo, porém é decididamente um ponto a corrigir futuramente.

Seria interessante procurar reduzir as inspeções manuais, não só em tempo de processo como em quantidade. O somatório destes tempos é diminuto comparado com o tempo de ciclo, variam entre 5 e 20 minutos por inspeção e por lote e são feitas cerca de 30 inspeções ao longo da produção, no entanto é decididamente superior ao tempo perdido com o excesso de alarmes em metrologia.

O sequenciamento está diretamente ligado ao *setup* das máquinas e está previsto no planeamento de produção diário, juntamente com a gestão de recursos, pelo que estes fatores podem ser analisados em conjunto. Este será provavelmente um dos fatores mais importantes a examinar para um melhor controlo do tempo de ciclo. Embora seja absolutamente necessário, é inegável que a dependência neste método cria tempos de espera extremamente elevados que, por vezes, podem chegar a dias. A falta de sincronização dos produtos nos diferentes equipamentos e de preocupação com o tempo de ciclo apenas agrava o efeito do sequenciamento nos tempos de espera.

Juntamente com este fator, o *batching* também deverá influenciar negativamente o tempo de ciclo, ao agrupar os lotes em *batchs* de um certo número vai criar obrigatoriamente tempos de espera nos processos seguintes que não requerem *batching*.

O *mizusumahsi* é usado frequentemente em vários processos ao longo da linha pelo que se a sua chegada a uma área não coincidir com a saída do produto dessa área pode causar alguma espera. Se um produto tiver sido processado precisamente no momento em que o *mizusumashi* sai da área, tem de ficar à espera durante uma hora. Este tempo pode ou não ser considerado *muda* do *mizusumashi*, caso não haja disponibilidade de processamento no equipamento seguinte, o tempo de espera pelo *mizusumashi* é irrelevante pois o produto teria de ficar à espera posteriormente pela disponibilidade do equipamento. Por outro lado, se o equipamento estiver disponível, o produto fica efetivamente à espera quando não é necessário. Este efeito é proporcional ao número de vezes que cada produto usa o *mizusumashi*, que pode ultrapassar as 30 vezes, e ao tempo de ciclo do mesmo, que é de uma hora, pelo que pode causar um atraso considerável.

O WIP é naturalmente um fator decisivo que afeta o tempo de ciclo. Como já foi referido, o tempo de ciclo médio de um produto varia proporcionalmente ao WIP em linha de acordo com a lei de Little. Assim, uma linha muito cheia irá sempre ter tempos de ciclo elevados comparados com o seu RPT. O WIP encontrado em linha é sempre gerido consoante os objetivos semanais de cada célula de trabalho, sendo que cada célula "alimenta" a célula posterior com produtos. O WIP tende a estar sempre a níveis baixos nas primeiras áreas, em *wafer prep* e *reconstitution*, mas à medida que há atrasos na produção, este pode aumentar nas áreas posteriores. Se, por exemplo, por alguma razão, houver um atraso na célula de fotolitografia, o WIP pode acumular nessa zona e as zonas posteriores podem falhar o plano semanal e até entrar em *starving* devido a esse atraso, no entanto estes casos são raros. Em condições normais, o WIP de cada área corresponde aproximadamente ao plano para essa semana, de modo a evitar falhas de entregas, pelo que o WIP varia de semana para semana.

Finalmente, as avarias e manutenções corretivas influenciam também o tempo de ciclo, no entanto, visto que no geral a fábrica tem uma alta capacidade de produção, este fator é relativamente irrelevante, pois geralmente basta ativar outro equipamento nesse processo ou nos seguintes para compensarem o tempo perdido devido à avaria.

3.8 Análise de Pareto

Para evitar obter resultados irrealistas com produtos de baixo volume nas análises seguintes, foi feita uma análise de Pareto de modo a encontrar os produtos que representam o maior volume de saídas. Descobriu-se que cerca de 40% dos produtos processados durante os 3 primeiros meses de 2018 correspondem a 95% das saídas, pelo que é nesses produtos onde se vai focar a análise. Para esta análise retiraram-se os materiais em *hold*, processos logísticos, de *rework* e processos onde os dados são mais escassos para evitar obter resultados que não refletem a realidade normal.

Como demonstra a figura 12, de 21 produtos, 5 representam 80% da produção, dos quais apenas um é *fanin* e 9 representam 95% da produção dos quais 2 são *fanin* e os restantes *fanout*. As seguintes análises foram então restringidas aos produtos A e B.

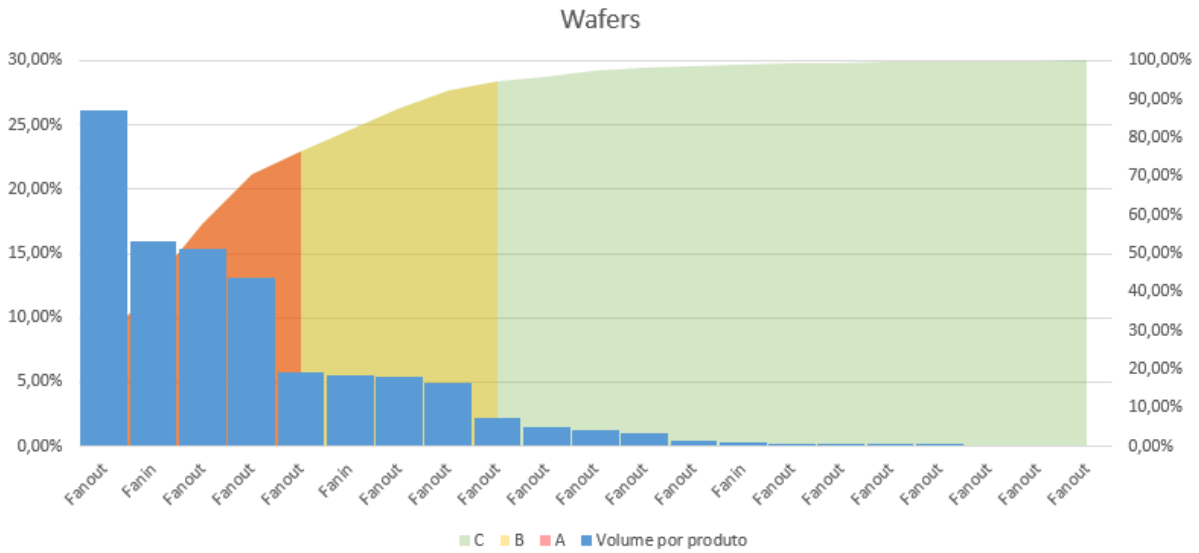

Figura 11 - Análise de Pareto

3.9 Tempo de ciclo

O tempo de ciclo total varia de produto para produto, porém está definido que os objetivos gerais são de 14 dias para produtos *fanout* e 10 dias para *fanin*. Como já vimos, devido à forma como o planeamento da produção é feita, o tempo de ciclo pode ser manipulado de modo a atingir os objetivos definidos no planeamento, porém este tende a ser estável e dentro dos objetivos gerais. Foram retirados os dados de 3 meses de 2018 dos tempos de ciclo de produtos *fanin* e *fanout* e são apresentados na seguinte figura, onde podemos observar no eixo das abcissas o tempo de ciclo em dias e o número de *wafers* no eixo das ordenadas:

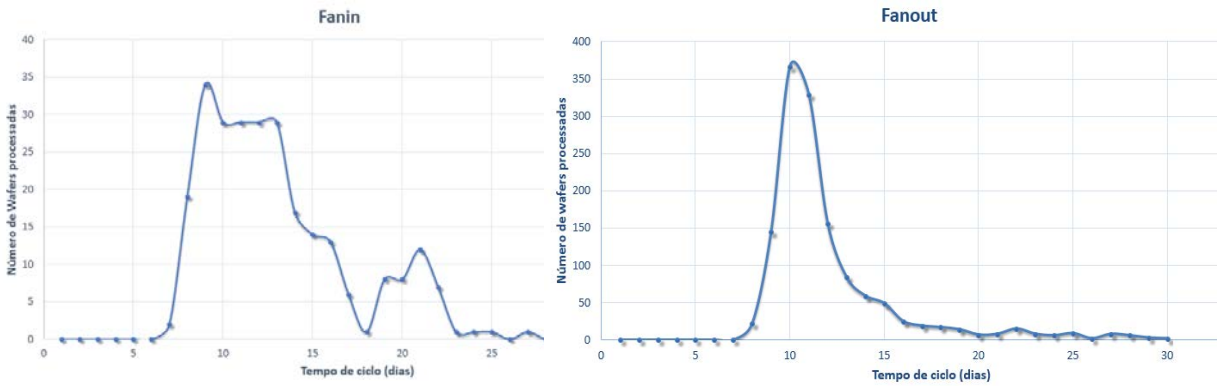

Figura 12 - Distribuição do tempo de ciclo de 3 meses produtivos

O tempo de ciclo dos produtos *fanout* parece seguir uma distribuição de baixa variabilidade, com a média a localizar-se nos 11 dias, bem abaixo dos 14 dias considerados como objetivos gerais, por outro lado os produtos *fanin* apresentam uma maior dispersão de tempo de ciclo com uma média situada nos 12,5 dias, claramente superior aos 10 dias definidos como o objetivo.

Esta dispersão deve-se a vários fatores, entre os quais o simples facto de a empresa ser conhecida pela sua produção de produtos *fanout* e, como tal, ter uma linha de produção mais adaptada e experiente com este tipo de produtos. Para além disso, a família de produtos *fanin* engloba apenas 2 produtos distintos com tempos de ciclo também distintos (9 e 13 dias), como se pode inferir do gráfico. O volume de produtos *fanin* é também consideravelmente inferior aos produtos *fanout*, compreendendo apenas 10% do volume de vendas, o que aumenta o efeito de dispersão por ser uma amostra mais reduzida, e são também produtos com maior complexidade, precisando de mais passagens pela área de RDL do que os produtos *fanout*.

Apesar destes tempos, os RPT destes produtos são consideravelmente inferiores, variando entre 2,5 dias e 3,8 dias para a maior parte dos produtos *fanout* e 3,8 a 4,4 dias para produtos *fanin*. Assim sendo podemos verificar os *flow factors*:

$$3 \leq Flow\ factor_{fanout} \leq 4{,}8$$

$$2 \leq Flow\ factor_{fanin} \leq 3{,}4$$

Isto significa que, em média, os produtos *fanout* demoram 3 a 5 vezes mais tempo a ser processados do que o tempo de processamento em máquina. Este *flow factor*, por definição corresponde aos tempos entre processos, portanto tempos perdidos em espera, transporte ou UD. Como já se desvalorizou os tempos UD e os tempos de espera são consideravelmente superiores aos tempos de transporte, este fator evidencia a importância dos tempos de espera no tempo de ciclo dos produtos.

É também interessante verificar que apesar do tempo de ciclo dos produtos *fanin* ser mais variável (como se verifica pelo gráfico da figura 11), devido ao seu elevado RPT têm um *flow factor* inferior, significando que têm uma produção mais eficiente.

4 Estudo do tempo de ciclo

4.1 Abordagem ao problema

O problema proposto pode então ser sumariado como a integração do cálculo de tempos de processo, incluindo o mapeamento e estudo dos tempos de espera, numa ferramenta automática para ser aplicada no projeto SEMI40.

Neste capítulo são apresentadas as análises feitas aos tempos entre processo, com estas análises pretende-se descobrir as razões do valor elevado de *flow factor*. Estas análises usam tempos históricos reais obtidos automaticamente pelo sistema *online* já implementado que recolhe os dados de *move in* e *move out* dos equipamentos. Efetivamente, o tempo de processo pode ser calculado pela diferença entre o *move out* e o *move in* do lote no equipamento, enquanto que os tempos entre processos podem ser calculados através da diferença entre o *move out* daquele lote do equipamento anterior e o *move in* do próprio equipamento.

Num primeiro passo é feita a análise dos tempos entre processos de um produto de alto volume e a sua comparação com os tempos de processo, de modo a observar de forma objetiva o seu valor relativo e variabilidade. Com esta análise pretende-se não só mostrar a influência no tempo de ciclo dos tempos de processo e dos tempos entre processos, bem como alguma tendência que possa existir.

Com estes dados podemos então escalar esta análise para os produtos com maior volume e encontrar os processos que mais afetam o tempo de ciclo. Tendo quantificado o tempo médio entre processos para cada processo, efetivamente subdivide-se o *flow factor* por processo e produto e faz-se um mapeamento dos processos em função do tempo em produção, isto é, com esta análise consegue-se localizar o produto nos processos onde deveria estar através do tempo de produção que já teve, o que permite avaliar o seu atraso ou avanço na produção de forma mais precisa. Os resultados desta análise podem já ser utilizados na ferramenta de planeamento do projeto SEMI40.

Posteriormente pretende-se encontrar as razões para estes tempos entre processos, quantificando o seu efeito para cada processo e para cada família de produtos. Este mapeamento vai essencialmente dividir os tempos entre processos em parcelas referentes às suas razões, o que permite fazer uma análise mais detalhada sobre as razões para estes tempos existirem e quais destas razões mais afeta o tempo de ciclo tanto nos processos estudados como no total da linha de produção.

4.2 Tempo de ciclo por camada

De acordo com Leachman (2015), o tempo de processamento médio por camada na indústria de semicondutores tende a ser 0,6 a 0,9 dias enquanto que o tempo de ciclo médio é 2 a 4 vezes superior, ou seja, o *flow factor* por camada típico da indústria tende a variar entre 2 e 4. Não é claro a que tecnologia de produtos Leachman se refere, mas sendo *fanin* e *fanout* as tecnologias mais comuns comercializadas, é seguro assumir que se referia a uma destas ou mesmo às duas. Estes valores não podem ser comparados diretamente com os encontrados para a ATEP pois estes englobam processos posteriores e anteriores à deposição de camadas, como tal é necessário fazer uma análise por camada (figura 13 e tabela 1).

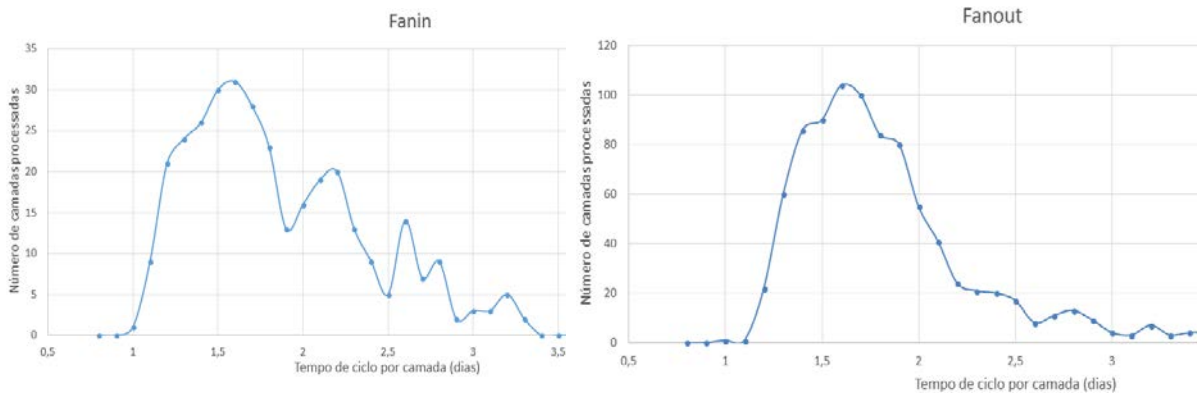

Figura 13 - Distribuição do tempo de ciclo por camada dielétrica de 3 meses produtivos

	RPT	CT	Flow factor
Fanin	0,69-0,82	1,57-2,17	1,92-2,97
Fanout	0,48-0,68	1,72-2,00	2,82-3,84
Indústria	0,6-0,9	1,2-3,6	2-4

Tabela 1 – Comparação das amplitudes das médias dos tempos de processamento, ciclo e *flow factor* da ATEP e do resto da indústria

Nos gráficos da figura 13 podemos ver nas abcissas o tempo de ciclo médio por camada dos produtos *fanin* e *fanout*, em dias, e nas ordenadas o número de camadas processadas. Nos produtos *fanin* consegue-se reparar em 3 picos distintos que pertencem a 3 camadas diferentes. Cada camada tem um tempo de ciclo próprio que depende principalmente de questões técnicas e da complexidade da camada, por estas razões não tem interesse estudar cada camada em separado. Os dados disponíveis publicamente sobre o tempo de ciclo por camada na indústria de semicondutores, pelas mesmas razões, apenas se referem às médias do tempo de ciclo por camada. Não obstante, conseguimos reparar num tempo de ciclo médio por camada de 1,85 dias para produtos *fanin* e 1,88 dias para os *fanout*.

A tabela 1 mostra a amplitude das médias do tempo de processamento, ciclo e *flow factor* de todos os produtos e dos valores encontrados da indústria de semicondutores. Leia-se o produto *fanout* com menor tempo de ciclo médio demora em média 1,72 dias por camada a ser processado e aquele com maior tempo de ciclo médio demora em média 2 dias por camada a ser processado.

Como se pode verificar, todos os valores encontram-se bem correlacionados com o típico da indústria, com uma ligeira vantagem no RPT dos produtos *fanout*, o que é compreensível visto ser a tecnologia pela qual a empresa é reconhecida mundialmente. Apesar dos melhores tempos de processamento dos produtos *fanout*, o *flow factor* é, na realidade, superior ao dos produtos *fanin* o que indica um desempenho de planeamento inferior quando comparado com os produtos *fanin*. No geral ambos os tipos de produto apresentam um tempo de ciclo e RPT por camada bastante positivo comparado com o normal da indústria, ainda que, novamente, com um *flow factor* relativamente alto.

4.3 Variabilidade dos tempos de entre processo

Numa fase inicial, foram estudados os tempos entre processo do produto A durante o ano de 2017, devido ao seu alto volume de produção e consequente facilidade de acesso a dados. Este produto faz parte da família dos produtos *fanin*, tem um tempo de processo total de cerca de 4,3 dias e tempo de ciclo médio de 9 dias (Fevereiro 2018), tornando-o um produto ideal para o estudo na zona crítica de RDL. Como já foi explicado, devido à natureza da recolha destes dados, o tempo de espera obtido refere-se ao tempo entre processos, incluindo todas as operações desde que é retirado do equipamento até ser inserido no equipamento do processo seguinte.

Filtrando os resultados obtidos para cada processo descobriu-se não só um valor médio elevado de tempo entre processos como também uma elevada variabilidade comparado com os tempos de processo, que por sua vez apresentam uma estabilidade relativamente boa. Na figura 14 apresenta-se esta comparação entre os tempos entre processo (a azul) e tempo de processo (a laranja) de um processo de litográfica ao longo de 2 anos.

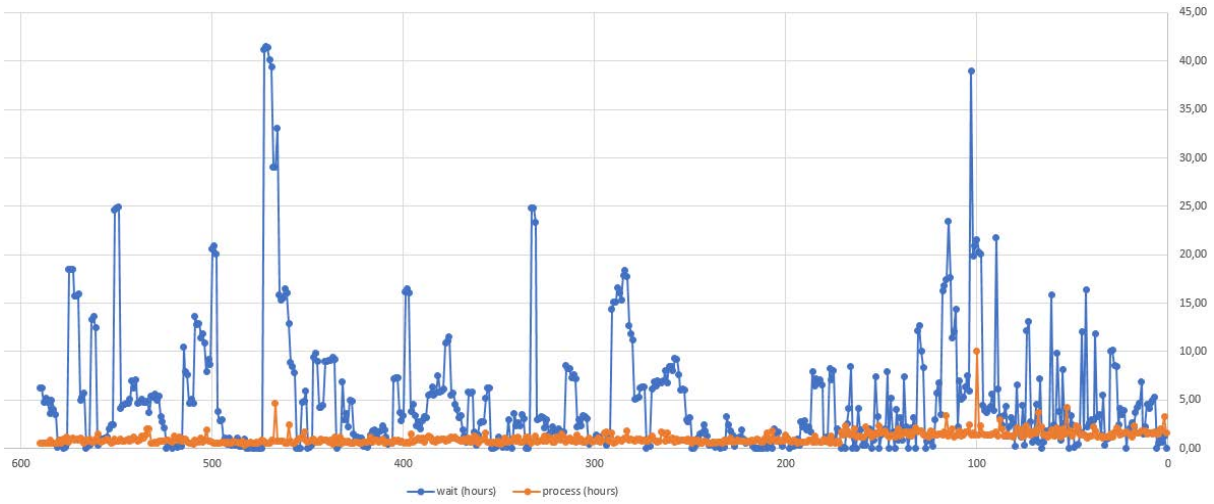

Figura 14- Tempos de espera e processo de uma operação de litografia

Este exercício foi repetido para vários processos que apresentam gráficos muito semelhantes a este, em particular em processos das áreas de RDL e LBS, havendo mais de 15 processos com tempos médios entre processos compreendidos entre 2h e 8h e desvios-padrão de 3h a 14h, sendo que apenas 2 processos são de LBS. A causa principal destes elevados tempos de espera e variação dos mesmos é, potencialmente, devida à gestão dos equipamentos a utilizar, como os equipamentos não necessários são desligados de acordo com o planeamento semanal, a capacidade da fábrica efetivamente varia com o tempo. No entanto, é também de notar que os processos com valores mais elevado fazem sequenciamento, o que pode ser uma causa deste tempo entre processos alto e todos os processos de *batching* estão incluídos nestes 15 casos, o que é outro indicador de potenciais causas desta variabilidade nos tempos de espera.

Para os processos de LBS, estes tempos elevados de espera são conhecidos por várias razões. Os equipamentos utilizados no primeiro processo desta área, apesar de serem apenas dois, têm uma capacidade bastante elevada, no entanto requerem uma conversão de *setup* para todos os produtos diferentes, o que leva a fazer sequenciamento. Na realidade, têm tanta capacidade que normalmente um dos equipamentos é desligado para poupar consumíveis, efetivamente eliminando um equipamento do processo. Para além disso, o equipamento é também utilizado no processo seguinte, pelo que a sua utilização tem de alternar entre os dois processos, o que limita a sua capacidade. Estes factos explicam a grande variação dos tempos de espera - quando o produto A chega a este processo e este acontece estar a processar esta família de produtos, o seu tempo de espera vai ser muito reduzido pois o equipamento tem uma

cadência alta, no entanto se o equipamento estiver a processar outros produtos, o produto A terá de esperar que o equipamento termine todos os outros produtos e seja feita a conversão de *setup* para os produtos A para poder começar a ser processado. Com uma mistura de produtos alta, como é o caso, não é raro alguns lotes ficarem em fila de espera mais do que dois dias.

O planeamento nestes dois processos tende a sincronizar bem os produtos, isto é, quando o primeiro processo faz lotes A, o seguinte também os vai fazer. No entanto, o equipamento utilizado é o mesmo e, por questões técnicas, apenas faz um processo de cada vez, ou seja, os lotes fazem o primeiro processo, são retirados do equipamento e ficam sempre a aguardar que o equipamento seja disponibilizado para o segundo processo. Normalmente, o equipamento faz vários lotes com o primeiro processo e só depois muda para o segundo processo, onde volta a processar vários lotes, o que leva os lotes a ficarem sempre em fila de espera para o segundo processo. Não existe uma regra específica para fazer esta mudança de processo, deixando ao critério do gestor de produção da área que tenta gerir o tempo de utilização do equipamento com o plano de produção da semana.

Para além disso, problemas de qualidade e outros atrasos podem causar produtos a perderem esta sincronização, o que, ocasionalmente, faz com que um lote fique dessincronizado e com isto fique em fila de espera mais tempo do que o normal. Estes tempos, como se pode ver no gráfico da figura 14, conseguem facilmente ultrapassar as dezenas de horas e até dias.

Os 3 processos que fazem *batching* são os de cura e apresentam uma dispersão ligeiramente diferente, com maior variação nos tempos entre processo e tempos de processo superiores, mas mais estáveis. Sendo um processo que implica o processamento de vários lotes simultaneamente, este aumento de variação na espera é expectável (Hopp e Spearman 2001), no entanto é de notar que estes processos não requerem sequenciamento, mas obtém um resultado de variação semelhante, apenas mais acentuado, o que é explicado pelo facto de ambos os tipos de produção implicarem, obrigatoriamente, que os produtos aguardem que certos requisitos sejam cumpridos para poderem ser processados.

O último caso particular nesta análise é um processo de metrologia, que faz análise dos semicondutores individualmente e procura problemas e imperfeições, pelo que é relativamente normal observar-se uma variação alta tanto nos tempos de processo como, consequentemente, nos tempos de espera. A somar à variação, existe o facto de este ser um processo relativamente novo, pelo que os dados disponíveis são escassos. Em raras ocasiões houve problemas com a máquina deste processo, o que causou um atraso considerável a vários lotes e, aliado à escassez de dados, causou uma inflação irrealista dos tempos de espera. Eliminando esses dados, o tempo médio de espera reduz de 5h para 1h e o desvio-padrão reduz abruptamente de 15,5h para 1,2h.

Os restantes processos com tempos de espera superiores a 2h têm um janelas temporais mínimas, em alguns casos de 2h, pelo que é obrigatório que estes tempos entre processos sejam superiores ao valor da janela de tempo. Isto contribui para o aumento do tempo médio de espera, porém não deverá contribuir para a variabilidade do mesmo, pois os valores das janelas de tempo são constantes.

Devido aos valores elevados de tempos de espera encontrados, é inconcebível que os tempos de operações administrativas, que não chegam a ultrapassar os 20 minutos nos processos mais morosos, sejam os fatores mais relevantes destes resultados. Esta análise sugere, portanto, uma forte relação entre sequenciamento, *batching*, planeamento e gestão de recursos como principais fatores que afetam a variabilidade e dimensão dos tempos de espera. No entanto, após a visita à área de produção ficou claro que o uso elevado do *mizusumashi* para transporte de lotes, dada a sua frequência de passagem de uma hora, poderia afetar o tempo de ciclo total dos produtos. Seria interessante fazer uma análise ao número de vezes que cada produto necessita de o usar. O efeito do *mizusumashi* nos tempos de espera não é claro na análise dos tempos de espera por processo já feita, pois os valores deverão ser relativamente reduzidos por

processo (média de 30 minutos, como explicado no capítulo 3), porém devido à sua utilização elevada podem ter um efeito considerável no tempo de ciclo total. Relembrando, os produtos são transportados pelo *mizusumashi* sempre que mudam de célula de produção, estando estas afastadas menos do que algumas dezenas de metros. Como os produtos retornam à mesma célula de produção várias vezes durante o seu processamento para fazerem processos diferentes, é expectável que sejam transportados pelo *mizusumashi* várias vezes.

Foi escolhido o mesmo produto e mapeado no seu fluxo de processo a utilização do *mizusumashi*. Verificou-se que o produto é transportado pelo *mizusumashi* nada menos do que 28 vezes, o que implica, teoricamente, uma média de 14h à espera do transporte. Este tempo é, de facto, considerável e não pode ser ignorado, no entanto, convém referir que em alguns casos pode ser redundante devido a estar englobado no tempo de espera do processo seguinte, isto é, por vezes, depois de aguardar pelo *mizusumashi*, o lote ainda tem de aguardar pela disponibilidade do equipamento seguinte, o que efetivamente invalida o tempo de espera pelo *mizusumashi*. Este efeito tem de ser estudado para definir o impacto efetivo deste método no tempo de espera.

4.4 Tempos de processo e tempo entre processos

Repetiu-se a análise para um produto *fanout* com alto volume de produção de modo a observar também as áreas iniciais (*wafer prep* e *reconsitution*) e pode comparar graficamente a diferença entre tempos de processo e entre processo. De modo a filtrar os *outlyers*, removeram-se os dados dos tempos entre processo quando o lote ficou em *hold* por qualquer motivo. Obteve-se o gráfico da figura 15 que demonstra a diferença entre os tempos médios de entre processo (a azul) e os tempos de processo (a laranja) em horas:

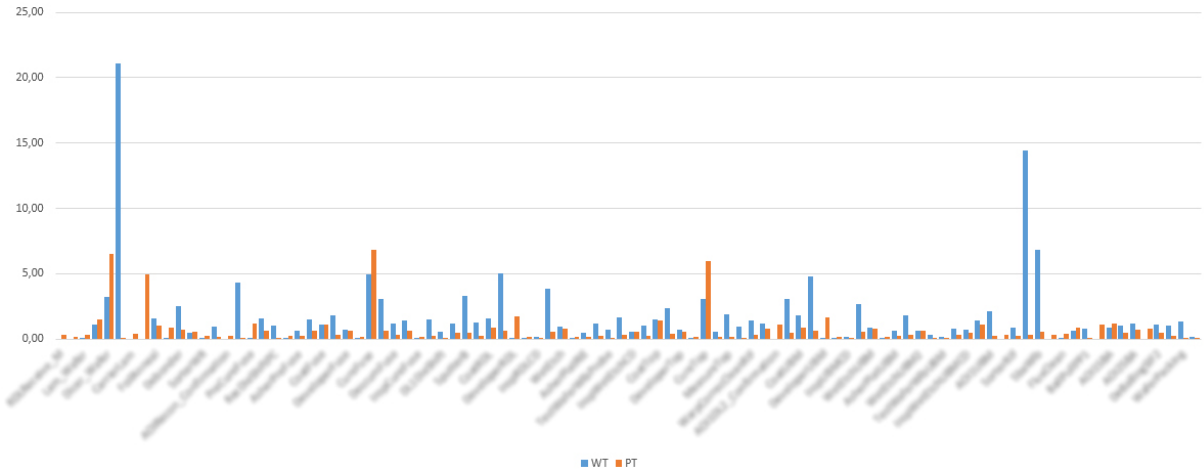

Figura 15 - Tempos médios de processo e entre processos

Como se pode verificar, existem 3 pontos onde os tempos entre processo excedem largamente o próprio processo. O primeiro, com o valor de 21,07 horas, corresponde ao supermercado que existe entre as áreas de *wafer prep* e *reconstitution*, é um *buffer* necessário e propositado, o que explica o seu valor elevado. Os restantes dois, com tempos de 14,4 horas e 6,82 horas, pertencem à área de LBS e, como explicado anteriormente, são os processos onde se faz sequenciamento. Pode-se ainda reparar que no geral, salvo poucas exceções, os tempos entre processo excedem largamente os tempos de processo.

4.5 Tempo de ciclo por processo

Seguidamente expandiu-se a análise a todos os produtos de modo a poder quantificar todos os tempos para cada produto e processo. Visto já haver em base de dados todos os valores de RPT e conhecendo a cadeia de processos de um produto, ao adicionar o RPT dos processos com o tempo total entre processos, é expectável obter-se o tempo de ciclo médio por produto. Alocando estes tempos a cada processo de cada produto consegue-se ainda obter o tempo de processo esperado de cada produto para qualquer ponto na sua cadeia de processos. Esta informação é vital para perceber o atraso ou adiantamento de um produto na linha. Na figura 16 apresenta-se um excerto do fim e ínicio da cadeia de processos de dois produtos *fanin*.

PRODUCT	OPERATION GROUP	OPERATION_SEQ	OPERATION	REAL_TIME(h)/lot	IDEAL_LotSize	WT (hours)	PT+WT (days)
A	SBA	177		0,70	25	8,73	8,92
A	SBA	179		1,08	25	6,08	9,21
A	SBA	180		0,70	25	0,00	9,24
A	SBA	181		0,77	25	0,02	9,28
A	SBA	182		0,57	25	0,32	9,31
A	BALLinsp_clean	184		0,25	25	0,95	9,36
A	BALLinsp_clean	190		1,44	25	5,47	9,66
A	BALLinsp_clean	191		1,73	25	0,00	9,73
A	BALLinsp_clean	192		0,25	25	2,36	9,84
A	SGL	195		0,25	25	5,03	10,06
A	SGL	199		0,20	25	0,13	10,08
B	DL1_Prep	5		0,08	25	0,39	0,04
B	DL1_Prep	6		0,46	25	0,73	0,09
B	DL1_Prep	7		0,35	25	1,04	0,15
B	DL1_Prep	8		0,25	25	0,00	0,16
B	Litho_DL1	10		2,28	25	1,42	0,31
B	Litho_DL1	11,01		0,75	25	0,00	0,41
B	Litho_DL1	13		1,32	25	2,75	0,58
B	Litho_DL1	14		0,50	25	0,00	0,58
B	Litho_DL1	15		0,25	25	2,17	0,68
B	Cure_DL1	20		6,83	25	5,79	1,21

Figura 16 - Excerto da cadeia de processos de dois produtos

Como se pode verificar pela figura, o produto A chega, por exemplo, ao grupo de processos de SBA passados, em média, 7,68 dias, no primeiro processo dessa área tem um tempo de espera médio de 8,73 horas para um processo que demora 42 minutos (0,7 horas) e um tempo de ciclo total médio de 9,23 dias. Estes dados foram então incluídos na ferramenta de planeamento do SEMI40 com sucesso e está agora em utilização na empresa.

De notar que, sendo dados históricos, é necessária uma grande quantidade dos mesmos para se obter resultados fiáveis. Esta problemática já se contornou, em parte, ao limitarmos esta análise aos produtos com maior volume (e, por definição, com mais dados históricos), no entanto para os produtos novos ou de menor volume esta análise pode não ser muito precisa. Aumentar o espaço temporal dos dados históricos pode resolver este problema, no entanto, torna os dados menos viáveis pois a situação da fábrica a nível de produção podia ser diferente, pelo que não convém usar dados demasiado antigos.

Para produtos novos é recomendado que se use os dados de tempos entre processos de produtos com trajetos semelhantes. Os resultados nestas situações não serão os mais precisos, mas serão uma aproximação razoável.

A fórmula utilizada anteriormente pela empresa para calcular o tempo entre processo para cada processo envolvia multiplicar o *flow factor* do produto pelo respetivo RPT, no entanto, estes valores variam muito da realidade. A soma total, por definição, resulta no tempo de ciclo, porém os valores individuais não refletiriam a realidade como demonstra a figura 17.

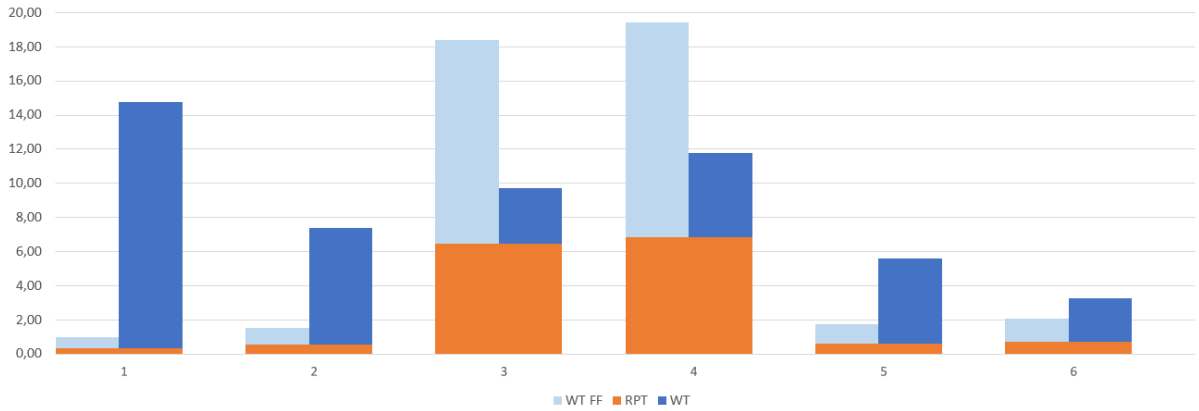

Figura 17 - Diferença entre tempos entre processos reais e calculados através do *flow factor*

Em que:
WT FF (azul claro) = tempo entre processos calculado através do *flow factor*;
RPT (laranja) = tempo de processo
WT (azul) = tempo médio entre processos em horas.

Para este produto em particular, o *flow factor* tem o valor de 2,84, o que significa que RPT+WT FF=2,84*RPT.

Como se pode verificar, para o processo 1 o tempo de espera calculado é muito inferior ao real enquanto que nos processos 3 e 4 o oposto acontece. O somatório de todos os tempos dos processos de um produto, por definição deve ser igual em ambos os casos, porém, para perceber o estado de um produto a meio da sua produção, a vantagem de se usar dados reais é incontornável. Com o método das médias simples, obter estes valores é simples, pode ser automatizado e acima de tudo é bastante mais preciso. Consegue-se obter valores bastante próximos da realidade quando comparado com o método utilizado anteriormente, no entanto, devido à alta variabilidade dos processos, este método pode ser melhorado.

Estes dados foram adotados pela empresa e estão agora a ser utilizados para definir o plano de produção por área de forma mais realista.

Com estes dados podemos também comparar o tempo médio de ciclo calculado com o tempo de ciclo médio real (tabela 2) de forma a confirmar que a soma total dos tempos de processamento com os tempos entre processos é igual ao tempo de ciclo real dos produtos.

		CTq	RPTT	CT calculado	CT real	Δ	flow factor
Fanout	A	10,64	3,19	12,90	13,03	-0,13	4,09
Fanout	B	8,96	4,32	11,86	12,42	-0,56	2,87
Fanin	C	5,13	5,27	10,08	10,65	-0,57	2,02
Fanin	D	7,17	4,44	11,26	10,87	0,39	2,45
Fanout	E	9,58	3,95	13,17	11,78	1,39	2,99
Fanout	F	6,59	3,62	9,76	10,30	-0,54	2,84
Fanout	G	6,60	3,62	9,77	10,15	-0,38	2,80
Fanout	H	6,90	3,54	10,05	10,26	-0,21	2,89
Fanout	I	7,51	4,72	11,88	13,40	-1,52	2,84

Tabela 2 - Tempos de ciclo calculados, comparação com os reais e respetivo *flow factor*

Onde:
CT_q = soma do tempo entre processos para cada produto;
RPTT = tempo total de processamento por produto;
$CT_{calculado}$ = soma dos dois dados anteriores;
CT_{real} = valores reais de tempo de ciclo de cada produto;
Δ = diferença entre o CT_{real} e o $CT_{calculado}$.

A diferença do tempo de ciclo calculado e o tempo de ciclo real varia entre -1,52 dias e 1,39 dias, ou seja, há um erro máximo de ~+/-12% para os produtos escolhidos. Este erro é provavelmente derivado de três fontes:

- Na obtenção dos tempos entre processos foram retirados os dados de quando um produto fica em *hold* em qualquer processo, enquanto que os tempos de ciclo reais foram obtidos a partir de outro sistema que não tem essa informação disponível, pelo que apenas foram removidos da amostra os lotes que ficaram em *hold* sob determinados códigos que implicavam tempos de ciclo excessivamente altos e, portanto, se consideraram *outlyers*;
- Alguns equipamentos têm vários *loading ports* (locais de carregamentos dos lotes), mesmo que o equipamento apenas processe os lotes individualmente. Isto implica que haja ocasiões onde o operador faz *move in* ao lote e este fica em fila de espera no *loading port* até o equipamento acabar de processar o lote que já estava a processar;
- Os valores tabelados de RPTT não representam exatamente o mesmo horizonte temporal que os restantes dados, como tal podem haver pequenas variações quando comparados com os dados reais.

Apesar de tudo, um erro de +/-12% no tempo de ciclo total é aceitável considerando as vantagens que este método proporciona na visualização dos tempos de espera por processo.

4.6 *Queueing theory*

A média dos tempos históricos entre processos tem uma boa precisão ao longo de um período extenso de tempo, no entanto, devido à diferente mistura de produtos na linha e ao planeamento da produção, este valor pode variar ao longo do tempo. Hopp e Spearman (2001) apresentam a teoria das filas de espera como uma forma analítica de calcular tempos de espera baseados na variabilidade dos tempos de chegada e partida a um processo e na utilização dos equipamentos nesse processo. No entanto, de acordo com Schelasin (2011), esta abordagem não tem obtido resultados aceitáveis nesta indústria devido à dificuldade em obter os fatores de variabilidade requeridos, pelo que Schelasin propõe obter-se o valor de variabilidade através de dados históricos em que a mistura de produtos e a taxa de utilização dos equipamentos é semelhante. Os resultados desta análise deverão ser ainda mais precisos do que a média pois incluem a variabilidade dos tempos, fator que não é contabilizado no cálculo da média.

A equação utilizada é a equação de Kingman, assumindo que os tempos de chegada e de processo seguem uma distribuição normal, que os equipamentos funcionam todos em paralelo (uma simplificação explicada à frente em mais detalhe) e que o número de produtos em fila de espera não tem limite. Para calcular a variabilidade do sistema utiliza-se o método de regressão de Schelasin (2011).

Embora teoreticamente mais preciso, este método implica um esforço maior de recolha de dados. Há várias máquinas do mesmo modelo dedicadas a produtos diferentes, no entanto esta dedicação varia consoante a mistura de produtos na linha e esta informação nem sempre é disponibilizada. Por exemplo, há a informação de existirem 4 equipamentos iguais ligados capazes de processar um produto, no entanto apenas 2 têm a correta conversão de *setup* para processar esse produto, mas não há informação de quais são os 2 equipamentos com esse *setup* para uma dada semana. Há, de facto, a informação de quais equipamentos foram usados, no entanto os equipamentos utilizados numa semana podem não ser os mesmos que os utilizados na semana para a qual pretendemos fazer o cálculo dos tempos de espera. Dos 2 equipamentos com *setup* para um certo produto, pode ser utilizado um equipamento para 90% dos lotes e o outro para os restantes. Por esta razão, não é possível ter os dados de taxa de utilização mais

corretos para todos os processos, a melhor aproximação que se pode fazer é usando a taxa de utilização média dos 2 equipamentos, o que pode incrementar uma certa inexatidão no cálculo.

Para se obter estes valores é necessário cruzar manualmente os dados de utilização dos equipamentos com os dados dos processos e com os dados históricos dos equipamentos utilizados, pois o código do modelo do equipamento que está ligado aos produtos não é o mesmo que está ligado à taxa de utilização. Por esta razão testou-se esta metodologia em apenas um produto de alto volume e alguns processos específicos de cada área e comparou-se com os dados reais (tabela 3):

PRODU	AREA	OPERATION_SEQ	OPERATION	REAL_TIME(h)/lot	Equipment_Model	EQP dedicados	WT (hours)	nr eqp ligados	u teorico	u historico	Vb	CTq	CTq reais
B	Sawing	4	[illegible]	0,33	[illegible]	1	0,06	3	13%	18,0%	0,82	0,00	0,02
B	Sawing	6	[illegible]	1,50	[illegible]	2	1,09	3	59%	38,7%	3,54	1,62	1,17
B	RDL	57	[illegible]	6,83	[illegible]	3	4,92	2	43%	50,4%	1,93	3,37	6,28
B	RDL	82	[illegible]	0,61	[illegible]	4	5,00	1	83%	62,1%	5,03	15,35	4,63
B	BallApply	183	[illegible]	0,34	[illegible]	2	14,40	2	41%	50,4%	114,75	9,00	23,21
B	BallApply	186	[illegible]	0,54	[illegible]	2	6,82	1	50%	25,6%	36,69	19,56	5,53

Tabela 3 - CTq calculado vs CTq real

Apesar dos resultados positivos de Schelasin (2011), o mesmo não se pode dizer nesta situação. Não só o CTq calculado é diferente do real em todos os processos como não parece existir qualquer correlação entre os mesmos. Neste caso, os tempos em fila de espera reais parecem ser mais próximos dos tempos médios de espera do que dos tempos calculados de espera.

Os resultados parecem ser altamente influenciáveis pela taxa de utilização, nos processos onde a taxa de utilização real histórica é superior à taxa de utilização teórica esperada, os tempos de espera calculados tendem a ser inferiores à média e vice-versa. Esta relação é espectável visto que se a taxa de utilização real é de facto superior à esperada, o equipamento deverá ter uma fila de espera inferior e à partida os produtos ficam menos tempo à espera. Como já foi explicado anteriormente, esta taxa de utilização não está corretamente calculada pelo que a diferença pode vir ou desses dados ou da taxa de utilização de esperada. Para confirmar se o erro advém apenas desta particularidade, foram fixados à posteriori os tempos de processamento reais e as taxas de utilização previstas da semana para a qual se pretendia fazer o estudo. Com estes valores, resolveu-se a equação anterior em função da utilização histórica de modo a encontrar os valores que deviam existir para a equação obter os resultados reais e compará-los com a utilização real obtida à posteriori. Os resultados obtidos são apresentados na tabela 4:

OPERATION	u histórico resolvido	u real	CTq reais	Δ
[illegible]	2,5%	14,90%	0,02	12%
[illegible]	45,0%	58,47%	1,17	13%
[illegible]	38,2%	49,88%	6,28	12%
[illegible]	84,5%	71,21%	4,63	-13%
[illegible]	32,4%	63,2%	23,21	31%
[illegible]	54,9%	43,21%	5,53	-12%

Tabela 4 - Comparação utilização calculada *vs* utilização real

Como se pode verificar, os 3 primeiros processos parecem obter uma diferença consistente de valores, porém os 3 processos seguintes não obtêm a mesma diferença. Com estes dados, resta apenas concluir que não parece haver qualquer correlação entre a taxa de utilização calculada e a taxa de utilização real, o que sugere que o erro não é apenas deste fenómeno. De modo a verificar se o erro reside então no cálculo da taxa de utilização teórica esperada, fez-se a mesma validação para a taxa de utilização teórica com os valores reais de

taxa de utilização histórica e tempos de espera, desta forma, eliminam-se as outras potenciais fontes de erro. Os resultados são apresentados na tabela 5:

OPERATION	u real	CTq reais	u teórico resolvido	u teórico	Δ
[illegible]	14,90%	0,02	30%	13%	-17%
[illegible]	58,47%	1,17	70%	59%	-12%
[illegible]	49,88%	6,28	55%	43%	-12%
[illegible]	71,21%	4,63	70%	83%	14%
[illegible]	63,2%	23,21	72%	41%	-31%
[illegible]	43,21%	5,53	38%	50%	12%

Tabela 5 - Comparação utilização teórica esperada vs utilização teórica calculada

Mais uma vez, não parece existir qualquer correlação entre os valores de utilização teórica esperada e os de utilização teórica calculada. A variação dos valores foi semelhante aos do teste anterior, apenas com os valores inversos, devido ao facto de o efeito destes valores na equação de Kingman serem semelhantes e inversos. No cálculo da variabilidade usam-se os valores de taxa utilização histórica no denominador da equação, enquanto que no cálculo dos tempos de espera usam-se os valores da taxa de utilização teórica esperada na multiplicação.

Dado estes resultados, é manifestamente preferível utilizar os dados médios de tempos de espera aos calculados através desta metodologia. Estes valores podem ter sido desviados devido aos problemas em encontrar as taxas de utilizações corretas, mas também devido ao facto de a fábrica estar sobredimensionada e como tal não estar a operar com taxas altas de utilização, o que incrementa variabilidade e viola as suposições da metodologia.

4.7 Análise dos tempos de espera

Continuando então a usar os dados médios de tempo entre processo, podemos fazer a análise dos tempos entre processos mais elevados, de onde se obtiveram os resultados apresentados na tabela 6.

Área	Processo	Tempo entre processos
LBS	1	12,52
LBS	2	8,21
LBS	3	7,18
WP	4	6,70
RDL	5	5,10
RDL	6	5,02
RECON	7	4,61
RDL	8	4,59
RECON	9	4,50
RDL	10	4,49
RDL	11	4,28
RDL	12	3,82
RDL	13	3,58
RDL	14	3,50
LBS	15	3,48
WP	16	3,45
RDL	17	3,42
RDL	18	3,16
RDL	19	3,02

Tabela 6 - Tempo entre processos médios mais elevados em horas

Estes 19 processos são responsáveis por 38% dos tempos entre processos acumulados. Mais uma vez, os tempos entre processos mais elevados fazem parte da área de LBS, seguidos por um processo de *Wafer Prep* e os restantes são principalmente de RDL. Para estudar estes tempos e a razão do seu elevado valor, decidiu-se observar os produtos na linha e questionar os operadores sobre quaisquer movimentações do produto assim como marcar o tempo que estas demoram a ser efetuadas. Como estamos a lidar com tempos muito elevados, foram ignoradas pequenas movimentações como o transporte do lote para o equipamento ou preenchimento dos *kanbans*, entre outras tarefas. Para simplificação dos dados e depois de discutir a abordagem do

estudo com engenheiros mais experientes da fábrica, também se decidiu dividir o estudo nas áreas antes e depois de RDL, visto que são áreas consideravelmente diferentes e os produtos *fanin* apenas começam na área de RDL e agrupou-se também os produtos por 4 famílias distintas, como demonstra a seguinte tabela:

	Fanin	Fanout			
WP Recon	-	A com LG	A sem LG	B	C
RDL LBS	Fanin	A		B	C

Tabela 7 - Agrupamento de produtos e áreas para o estudo dos tempos entre processo

Os processos estudados foram também filtrados e agrupados em 8 grupos, novamente para simplificação da análise. Os processos com menos ocorrências (13, 15 e 19) ou que representavam processos de correção de deformação das *wafers* (7 e 8) foram eliminados da análise. O seu estudo é pouco interessante devido a serem processos específicos de poucos produtos ou processos de correção que, por definição, são fora do espectro normal da produção ou ainda porque são processos com amostragem, ou seja, apenas parte do lote ou uma percentagem de lotes o faz. Os processos 5, 8, 10, 11 e 12 são todos processos semelhantes de litografia efetuados nos mesmos equipamentos e produtivamente idênticos, apenas variando a nível técnico e com alguma dedicação de equipamentos, pelo que estes processos foram agrupados. O mesmo acontece com os processos 6, 14, 17 e 18, desta vez referindo-se a curas, pelo que estes também foram agrupados. No final obtemos a tabela 8 que representa os processos que vão ser analisados em detalhe e mapeadas as razões que levam aos elevados tempos entre processo.

Área	Processo	Tempo entre processos
LBS	1	12,52
LBS	2	8,21
LBS	3	7,18
WP	4	6,70
RDL	5	4,46
RDL	6	3,78
WP	7	3,45

Tabela 8 - Processos a serem analisados e respetivos tempos de espera

Onde:
Processo 5 = grupo de processos de litografia;
Processo 6 = grupo de processos de cura;
Processo 7 = processo anteriormente com o número 16
Todos os restantes processos mantêm a mesma numeração.

4.8 Quantificação dos tempos de espera

Seguindo então para a quantificação dos tempos de espera, de modo a obter dados estatisticamente significativos tentou-se obter o maior número de observações possíveis, no entanto, devido a diferentes fatores tais como tempos elevados de processo, baixo volume de algumas famílias de produtos, irregularidades de produção (problemas com equipamentos ou produtos), entre outros, apenas foi possível obter uma amostra relativamente baixa, pelo que é sugerido elaborar no futuro um estudo mais prolongado e completo.

Antes de se apresentar a análise convém referir que a maioria destes processos contém alguma particularidade específica que pode explicar pelo menos parte destes tempos de espera elevados. O processo 1 e 2 partilham o mesmo equipamento embora sejam processos distintos, o que efetivamente limita a sua produção para metade do seu potencial e, tendo apenas um equipamento disponível, necessita de fazer conversão de *setup* para praticamente todos os produtos, o que leva a ser sempre feito um sequenciamento de pelo menos 4 lotes. Este número

provém de um estudo interno feito no passado que indicou que 4 é o número mínimo de lotes que se devia processar naquele equipamento, de modo a melhorar o tempo de ciclo dos produtos, antes de se converter o *setup* para outro produto. Dado que este é o processo que tem o tempo de espera mais elevado, a fiabilidade deste estudo é questionável, porém é uma realidade da fábrica que é agravada pelo facto de existirem fisicamente 2 equipamentos que não são utilizados simultaneamente de modo a poupar consumíveis. Apesar de ser um dos processos mais céleres da fábrica (apenas 40 minutos por lote), esta vantagem é totalmente desperdiçada com estes tempos de espera.

No grupo de processos 5, pelo menos 2h das suas 4,46h de tempo de espera são devidas à janela de tempo precedente ao processo. Antes de ser processado, o produto tem de cumprir 2h de estabilização ao ambiente obrigatoriamente e, como já referimos, estas janelas de tempo podem ser consideradas processos sem equipamento, pelo que o tempo real de espera nestes processos é de 2,46h.

Finalmente o grupo de processos 6 implica *batching*, dependendo do equipamento esse *batching* pode ter o tamanho de 100 *wafers* ou 48, o que significa que até existirem em fila de espera esse número de *wafers*, o processo não se inicia. A agravar esta situação, nem sempre se podem agrupar lotes no mesmo *batch* em processos diferentes, ou seja, por vezes há mais do que 100 *wafers* em fila de espera apesar de haver um equipamento disponível porque os seus processos ou produtos diferem.

Depois de feito o estudo, obtiveram-se os seguintes resultados globais (em minutos):

Processos	Sequenciamento	WIP (equipamentos dedicados/desligados)	Batching	TW	Engenharia	Tempos administrativos	Split	setup	UD	Inspeção	WIP (todos equipamentos a funcionar)	Count
1												
A	1043,7	28,9	0,0	0,0	3,1	0,0	0,0	7,5	0,0	0,0	0,0	16,0
B LG	432,6	85,2	0,0	0,0	162,0	0,0	0,0	0,0	0,0	3,5	0,0	10,0
B UBM	717,6	105,2	0,0	0,0	77,1	0,0	0,0	0,0	0,0	0,0	0,0	7,0
Fanin	748,9	28,5	0,0	0,0	0,0	0,0	0,0	5,0	0,0	0,0	0,0	6,0
1 Total	**783,1**	**57,0**	**0,0**	**0,0**	**56,7**	**0,0**	**0,0**	**3,8**	**0,0**	**0,9**	**0,0**	**39,0**
2												
A	264,5	246,3	0,0	0,0	0,0	1,2	0,0	11,5	0,0	0,0	0,0	13,0
B LG	106,4	227,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	11,0
B UBM	823,6	150,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	12,0
Fanin	534,9	169,3	0,0	0,0	0,0	0,0	0,0	8,6	6,9	0,0	0,0	7,0
2 Total	**424,1**	**202,2**	**0,0**	**0,0**	**0,0**	**0,3**	**0,0**	**4,9**	**1,1**	**0,0**	**0,0**	**43,0**
3												
A	0,0	405,5	0,0	0,0	0,0	1,1	0,0	0,0	0,0	0,0	0,0	7,0
3 Total	**0,0**	**405,5**	**0,0**	**0,0**	**0,0**	**1,1**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**7,0**
4												
A com LG	0,0	272,0	0,0	0,0	0,0	14,0	0,0	0,0	0,0	0,0	0,0	5,0
B LG	0,0	300,0	0,0	0,0	11,3	11,4	0,0	0,0	0,0	0,0	0,0	16,0
4 Total	**0,0**	**293,3**	**0,0**	**0,0**	**8,6**	**12,0**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**21,0**
5												
A	0,0	20,8	0,0	120,0	0,0	2,8	0,0	0,0	0,0	0,0	0,0	4,0
B LG	0,0	67,6	0,0	120,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	3,0
B UBM	0,0	92,2	0,0	120,0	1,0	0,0	0,0	0,0	0,0	0,0	0,0	10,0
Fanin	0,0	50,3	0,0	120,0	0,0	3,8	0,0	0,0	0,0	0,0	0,0	4,0
5 Total	**0,0**	**67,1**	**0,0**	**120,0**	**0,5**	**1,2**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**21,0**
6												
A	0,0	0,0	196,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	6,0
B LG	0,0	92,0	146,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	3,0
B UBM	0,0	54,4	189,8	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	17,0
Fanin	0,0	39,3	196,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	6,0
6 Total	**0,0**	**44,9**	**188,1**	**0,0**	**0,0**	**0,1**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**32,0**
7												
A com LG	0,0	176,7	0,0	0,0	0,0	11,7	0,0	0,0	0,0	0,0	0,0	3,0
A sem LG	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	120,0	0,0	1,0
B LG	0,0	235,5	0,0	0,0	0,0	12,5	0,0	0,0	0,0	0,0	0,0	2,0
B UBM	0,0	277,5	0,0	0,0	0,0	24,6	48,2	0,0	0,0	3,3	0,0	12,0
7 Total	**0,0**	**240,6**	**0,0**	**0,0**	**0,0**	**19,7**	**32,1**	**0,0**	**0,0**	**8,9**	**0,0**	**18,0**
Grand Total	**269,48**	**149,68**	**33,26**	**13,92**	**13,26**	**3,65**	**3,19**	**1,99**	**0,27**	**1,077**	**0,00**	**181**
% total	**55%**	**31%**	**7%**	**3%**	**3%**	**1%**	**1%**	**0%**	**0%**	**0%**	**0%**	**100%**

Tabela 9 - Resultados da análise do mapeamento dos tempos de espera

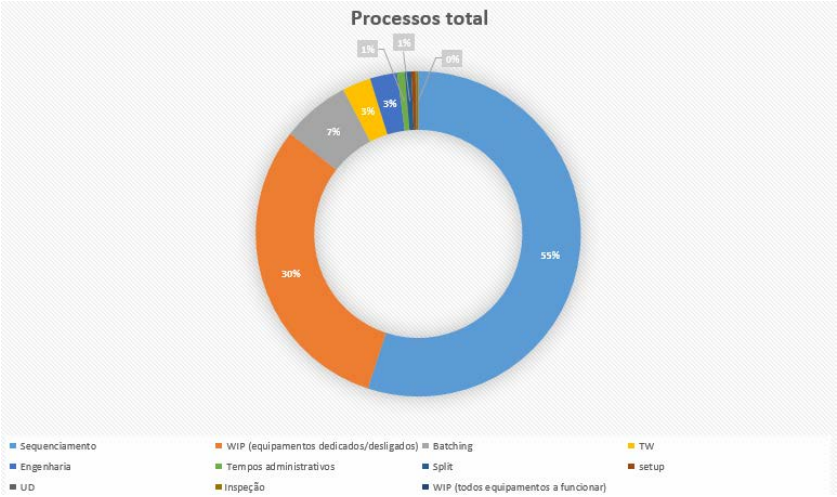

Figura 18 - gráfico representativo da quantificação das razões dos tempos de espera no total dos processos estudados

Os gráficos específicos a cada processo estão presentes nos anexos C. A tabela 9 mostra a média em minutos dos tempos anteriores a cada processo em que um lote fica em fila de espera para ser processado com as suas respetivas razões, assim como a quantidade de lotes estudados (última coluna). Para melhor compreensão de como estes tempos foram quantificados, segue uma legenda explicativa das mesmas:

- Sequenciamento: O tempo total em que um processo que requer mudança de *setup* de produto para produto tem todos os equipamentos ocupados com outros produtos (por exemplo está a processar produtos A e *fanin* com produtos B UBM em espera)
- WIP (equipamentos dedicados/desligados): Tempo total em que estão a ser processados produtos em todos os equipamentos (estando estes limitados por estarem dedicados a um certo produto/família e/ou existirem equipamentos desligados) e a única razão pela qual o produto está em espera é devido a ter menor prioridade. Esta razão é, portanto, apenas contabilizada quando todas as outras não o puderem ser.
- *Batching*: Tempo em que o produto está à espera de outros produtos para completar o *batch*, estando todos os equipamentos indisponíveis ou não. Este tempo deixa de contar quando o *batch* do produto está completo e em fila de espera, passando a razão de estarem em fila de espera a ser qualquer outra.
- TW: Tempo durante o qual o produto está a cumprir a janela temporal. Durante este tempo o produto não pode ser processado, como tal qualquer outra razão que o possa fazer estar em espera é irrelevante.
- Engenharia: Tempo durante o qual as máquinas disponíveis estão a ser utilizadas para atividades de engenharia e não há outra razão aparente para o produto ficar em espera.
- Tempos administrativos: Tempo durante o qual o operador está a efetuar operações administrativas (*move in* e *move out* no sistema, atualizar os *kanbans*, movimentação do lote, etc).
- *Split*: Tempo que o lote leva a ser inspecionado e divido em dois lotes mais pequenos.
- *Setup*: Tempo durante o qual o produto aguarda apenas que o equipamento disponível complete o seu *setup*.
- UD: Tempo durante o qual o produto não pode ser processado devido a uma avaria (*unschedule down*).
- Inspeção: Tempo durante o qual o produto não está a ser processado devido a uma inspeção. Se o produto, depois de inspecionado, ficar à espera por qualquer outra razão, esta não é contabilizada visto que o produto efetivamente estaria em espera com ou sem a inspeção.
- WIP (todos os equipamentos a funcionar): Igual ao "WIP (equipamentos dedicados/desligados)", com a diferença de neste caso estarem todos os equipamentos existentes disponíveis para produção. Dos casos estudados, como se pode verificar, tal nunca sucedeu.

Dos 7 processos observados e 181 lotes seguidos, verificamos que 55% dos tempos de espera se refere a sequenciamento, o WIP com equipamentos dedicados ou desligados refere-se a outros 30% do tempo de espera total e 7% deve-se ao *batching*. Estes resultados por si só não são muito reveladores visto que apenas 2 processos fazem sequenciamento e um grupo de processos faz *batching*. De notar que, apesar de terem sido observados 181 lotes, este número é relativamente baixo para se obter uma análise estatisticamente significativa devido ao alto número de combinações de processos e famílias de produtos. É, no entanto, uma amostra suficientemente grande para perceber a origem destes tempos de espera elevados, ainda que a sua quantificação não seja bem conseguida.

Seria de esperar encontrar também alguma influência do *mizusumashi* nos tempos de espera visto que este é o método usado para o transporte de produtos para os processos 1 e 6, no entanto, como se pode verificar, não teve qualquer influência nos tempos de espera. No primeiro processo é simples de explicar o sucedido, como os tempos de espera são tão elevados devido a outros fatores, o produto estaria em fila de espera durante o mesmo tempo mesmo que o tempo de transporte fosse nulo. Como tal, mesmo havendo algum tempo de espera pelo *mizusumashi*, a razão principal do seu tempo de espera não é essa.

No grupo de processos 6, o *mizusumashi* já poderia ter algum efeito, no entanto, os operadores estão formados para ir buscar os produtos finalizados à área anterior quando necessário, o que efetivamente elimina tempos de espera pelo *mizusumashi*. Contudo, ao fazerem este trabalho estão a incluir tempos de transporte normais, se bem que muito reduzidos. Estes tempos, por serem de apenas alguns segundos num processo com tempos de espera médios de 4-5h, não foram contabilizados, apenas foram contabilizados os tempos administrativos que pertencem a estes produtos.

Analisando os resultados por processo verificamos que no 1º processo, o sequenciamento é responsável por 86% dos tempos de espera enquanto que no 2º é responsável por 67%. Estes resultados indicam uma grande dependência nesta heurística nestes dois processos, algo que seria de esperar dado as limitações a nível de equipamentos e *setup* já mencionadas. O primeiro processo tem um tempo de processamento consideravelmente inferior, o que explica o valor inferior de tempos de espera devidos ao WIP (média de 57 minutos, ou 7% do tempo médio total). Os restantes 7% existiram devido a uma atividade de engenharia com a duração de 3 horas que atrasou um sequenciamento inteiro. Neste caso, como há apenas um equipamento em utilização para os dois primeiros processos, apenas se considerava que a razão do tempo em espera era WIP quando produtos iguais estavam a ser processados, ou seja, apenas quando se iniciava um sequenciamento de um certo produto (o *setup* do equipamento era convertido para esse produto) é que se considerava que o produto estava em espera devido ao WIP.

Neste processo podemos ver uma diferença considerável de valores entre as famílias de produtos. Os produtos B UBM são os de maior volume, pelo que normalmente são os que têm mais tempo de espera devido a WIP. O mesmo não acontece com os produtos A e *Fanin*, que são os que têm menor volume e como tal não atingem uma fila de espera tão longa como a dos B UBM. Consegue-se notar uma certa correlação entre os tempos de espera devido ao WIP e o volume dos produtos, assim como com os tempos de espera devido ao sequenciamento. Aparentemente, quanto maior o volume de um produto, maior o seu tempo de espera devido ao WIP e menor o seu tempo de espera devido ao sequenciamento.

No 2º processo os valores são muito semelhantes, com o sequenciamento a ser responsável por 67% do tempo total de espera, 32% devido ao WIP e um valor residual devido a outras razões. Como o processo imediatamente anterior é o 1º processo e este tem uma cadência de produção muito elevada, é de esperar que o WIP detenha uma responsabilidade maior nos tempos de espera, pois o 1º processo é capaz de alimentar a fila de espera do 2º mais rápido do que a sua fila é reposta e como usam o mesmo equipamento, efetivamente acumula os produtos atrás do 2º processo.

O 3º processo, também da área de LBS, é um processo posterior aos dois primeiros com um tempo de processamento bastante alargado, sempre com um ou dois equipamentos disponíveis (dependendo do planeamento semanal) e é um processo que é feito apenas pelos produtos A. Depois da observação dos lotes neste processo e segundo a opinião dos gestores de produção da área, a única razão para o tempo de espera elevado neste processo é devida ao WIP. Depois dos dois primeiros processos, os produtos chegavam aos processos sempre em *batch* em série, ou seja, normalmente chegavam sempre vários produtos seguidos do mesmo tipo. Por esta razão, e por este ser um processo mais vagaroso que os anteriores, estes produtos tendem a acumular neste processo.

O 4º processo é também limitado a 2 famílias de produtos, A com LG e B LG. Embora tenha tempos de processo muito elevados, o facto de não ter um alto volume de produtos faz com que não sejam necessários muitos equipamentos disponíveis pelo que normalmente apenas estão disponíveis para produção um ou dois equipamentos (dependendo da semana). Este facto explica o alto tempo de espera neste processo, praticamente a única razão encontrada é relativa ao WIP em ambas as famílias de produtos.

O grupo de processos 5 são processos semelhantes de litografia que tem uma janela temporal mínima antecedente de 2h, pelo que, naturalmente, em todos os casos o tempo de espera é sempre superior a 2h. Incrementado a este tempo, temos apenas uma média de uma hora a mais de tempo de espera, a vasta maioria dela devido a WIP. Mais uma vez consegue-se verificar a correlação entre volume de produtos e tempo de espera devido ao WIP, os produtos B UBM têm novamente os maiores tempos de espera que vai gradualmente diminuindo até aos produtos A que têm menor volume. Neste caso, como os produtos *fanin* têm mais passagens por estes processos, o seu tempo de espera aproxima-se mais dos B LG que no primeiro processo.

O grupo de processos 6 refere-se a processos de cura onde os equipamentos têm capacidade para vários lotes, pelo que, para otimizar o equipamento apenas se faz o processo quando há produtos suficientes para completar um *batch*. Esta regra cria tempos de espera consideráveis, no entanto não tanto quanto o sequenciamento pois neste caso há poucas conversões de *setup*, ou seja, o *batch* pode ser composto por produtos diferentes. Neste processo, a espera para completar o *batch* é a razão mais significativa dos tempos de espera elevados, como se pode verificar corresponde a 81% desses tempos, e o WIP é de novo a segunda razão mais significativa ocupando os restantes 19% dos tempos de espera. Devido ao número de medições não ser suficientemente grande, os produtos A aparentam nunca perder tempo devido ao WIP, no entanto repare-se que houve apenas 6 medições neste processo desta família de produtos, o que é insuficiente para um *batch* completo. Neste caso, para não falhar ao plano da semana, os operadores tomaram a decisão deliberada de processar estes produtos num *batch* incompleto, pelo que efetivamente os produtos nunca ficaram à espera por outra razão.

Também convém denotar que o efeito das atividades de engenharia neste grupo de processos é algo complicado de quantificar, mas é certamente inferior aos outros processos pela seguinte razão. Como é possível processar vários produtos simultaneamente, por vezes completa-se um *batch* com atividades de engenharia, o que reduz o seu efeito no atraso dos produtos.

Finalmente, o 7º processo refere-se a um processo longo da área de *Wafer Prep*. Todos os produtos *fanout* são processados, e, apesar do existirem 16 equipamentos disponíveis, o tempo de processo é simplesmente demasiado longo, como tal é provável que qualquer produto fique sempre bastante tempo em espera. Os produtos B UBM por vezes fazem *split* antes deste processo, o que também causa algum tempo de espera à metade de lote que não é processado imediatamente depois do *split*.

Sumariando, os resultados obtidos permitiram perceber que cada processo tem normalmente uma ou duas razões principais para os produtos ficarem à espera de ser

processados. Como seria de esperar, o sequenciamento e o *batching* englobam grande parte dos tempos de espera, no entanto, enquanto heurísticas, não são necessariamente os responsáveis pelos tempos altos de espera. A fraca sincronização de produtos na linha cria problemas a estas heurísticas difíceis de contornar. Algumas formas de combater este problema são apresentadas no capítulo 5.

O WIP é também, previsivelmente, um dos fatores que mais influencia os tempos de espera. Como já foi evidenciado pela lei de Little, a partir do ponto em que a capacidade de produção de uma linha é ultrapassada pelo número de materiais (*jobs*) em processamento, o tempo de ciclo dos materiais é afetado. Pela diferente mistura de produtos em linha com processos e RPTs diferentes que podem ou não partilhar o mesmo equipamento, pelo diferente número de equipamentos disponibilizado por semana assim como outros fatores particulares desta produção, a lei de Little torna-se difícil de implementar matematicamente, no entanto a sua premissa mantém-se e os resultados apontam claramente para uma gestão de WIP pouco focada nos tempos de ciclo.

4.9 Melhorias propostas

As heurísticas encontradas responsáveis pelos altos tempos de espera não são, por si só, prejudiciais ao tempo de ciclo de um produto, como tal é necessário perceber onde reside a responsabilidade real destes tempos de espera. Aplicando a metodologia dos 5 "porquês" ajuda a chegar à fonte real do problema. Neste caso já teríamos respondido a 3 "porquês":

1. Tempos de ciclo muito elevados – porquê?
2. Tempos entre processo muito elevados – porquê?
3. Sequenciamento, WIP e *batching* estão a causar tempos de espera elevados – porquê?

A resposta deve-se agora dividir para as heurísticas e gestão de WIP visto que, apesar de estarem correlacionadas, a sua origem pode divergir. Tanto o sequenciamento como *batching* requerem que um grupo elevado de produtos seja processado simultaneamente ou sequencialmente, pelo que quando tal não acontece os produtos tendem a ficar bastante tempo à espera. Esta é a principal razão que leva a tempos de espera elevados nos processos onde se usam estas heurísticas, como tal é aconselhável que os produtos sejam mais agregados no início e durante a sua produção. Um WIP mais agregado por tipos de produtos ao longo da linha deverá ser benéfico para a variabilidade encontrada nos tempos de espera pois diminui a variação da cadência de saída dos lotes de cada processo, assim como diminui o tempo de espera para completar um *batch* ou para fazer sequenciamento. Devido à extensão da linha de produção é compreensível que esta gestão seja mais complicada do que aparenta, com os problemas de qualidade, reintrodução dos materiais em linha, etc, garantir que um grupo de produtos atinge as áreas finais sequencialmente torna-se praticamente impossível. Assim respondemos aos 5 "porquês" para o sequenciamento e *batching*:

4. Falta de sincronização dos lotes na linha – porquê?
5. Processamento longo e com várias fontes de variabilidade no sistema – porquê?

Podemos continuar a metodologia, porém o processamento é longo devido à natureza da indústria – um circuito integrado requer obrigatoriamente vários processos por ser um produto complexo – e as várias fontes de variabilidade no sistema existem devido novamente à complexidade do produto, algo que tende a aumentar.

Assim sendo, a melhor opção pode passar por diminuir o recurso a estas heurísticas. Como já foi referido, o sequenciamento acontece principalmente em dois processos em que existe normalmente apenas um equipamento disponível para poupar consumíveis, não permitindo fazer os dois processos em simultâneo. Num processo onde o tempo de *setup* é considerável (meia hora a uma hora), é crítico haver mais do que um equipamento para minimizar o número de conversões de *setup*. De facto, existem no total dois equipamentos, que vão intercalando a sua utilização com este mesmo propósito, porém é claramente insuficiente para diminuir os tempos de espera nestes processos que representam quase 10% do tempo de ciclo de alguns produtos. Existe então um certo equilíbrio entre tempo de ciclo e custo de produção. Como a maior parte dos clientes da ATEP impõe uma data de entrega dos produtos específica, o tempo de ciclo não precisa de ser muito menor do que o necessário.

Relativamente aos tempos de espera devido ao *batching*, para além da sincronização já referida, Killian (2010) defende que este tipo de equipamentos causa invariavelmente tempos de ciclo superiores devido precisamente ao tempo de espera para completar um *batch*. Nos estudos que realizou comparou os tempos de ciclo entre vários tipos de *batch* e sem *batch* e os resultados que obteve parecem indicar que quanto menor for o tamanho do *batch*, menor é o seu tempo de espera. Assim conclui que é preferível, a nível de tempo de ciclo, optar por vários equipamentos menores que não fazem *batching* ou que requerem um *batch* reduzido.

Continuando a análise dos 5 "porquês" relativa à gestão de WIP, as razões pela qual existem tempos de espera tão elevados são mais difusas e numerosas. É necessário um estudo

mais focado neste fator para se obter respostas concretas, no entanto os resultados obtidos permitem concluir que o facto de haver equipamentos dedicados a certos produtos e outros desligados é uma das principais razões para as quais estes tempos de espera existem. Esta é uma realidade difícil de evitar, a dedicação de equipamentos existe com o propósito de evitar fazer sequenciamento e como a mistura de produtos em cada área é diferente todos os dias não é simples adaptar os equipamentos consoante os diferentes volumes de produtos. Como já foi referido, a necessidade de poupar em consumíveis e energia é uma política importante da empresa pelo que disponibilizar mais equipamentos do que os estritamente necessários não é uma alternativa muito apelativa.

Outras formas de evitar estes tempos de espera elevados passam por aplicar metodologias de SMED (*single minute exchange of die*) para melhorar os tempos de conversão de *setup*. Estas ideias já estão a ser aplicadas noutros equipamentos, possivelmente seria útil focarem-se nestes equipamentos.

5 Conclusões e trabalhos propostos

Os objetivos desta dissertação consistiam em fazer uma estimativa eficaz do tempo de ciclo de qualquer produto em qualquer ponto da produção através da quantificação dos tempos entre processo e tempos de processo, assim como implementar estes resultados no projeto do SEMI40 e conhecer os fatores relevantes que causam o tempo de espera dos produtos e a sua quantificação. Estes objetivos foram cumpridos e os resultados obtidos dos tempos entre processo estão a ser utilizados na ferramenta do SEMI40.

As particularidades desta indústria apresentaram alguns problemas, principalmente com a reintrodução de produtos a meio da linha de produção, a dimensão anormal de número de processos assim como o planeamento e logística de produção. Ainda assim, foi possível demonstrar de forma clara a importância não só do valor, mas também da variabilidade dos tempos entre processo. Ficou demonstrado que a vasta maioria destes tempos se refere unicamente a tempos em fila de espera, pelo menos nos processos onde estes tempos são os mais elevados. O peso da variabilidade e valor destes tempos quando comparados com os tempos de processo é absolutamente inquestionável e é algo que merece mais atenção de futuro.

O cálculo destes tempos era anteriormente feito usando o *flow factor*, algo que no somatório total dos processos apresenta resultados razoavelmente precisos, porém não permite encontrar os processos com maior tempo de ciclo e potencia uma imagem incorreta dos processos que ocupam de facto mais tempo na linha de produção quando se considera as filas de espera. Os resultados do cálculo destes tempos através das médias históricas permitem ter uma visão mais correta do trajeto de um produto na linha. A análise é limitada no sentido em que não consegue individualizar e prever os tempos de ciclo dos lotes, no entanto é bastante precisa quando se observa os resultados de vários produtos e permite ter uma referência que não existia anteriormente sobre o estado de qualquer produto na linha de produção, o que auxilia a priorização dos mesmos. Estes resultados estão a ser utilizados pela ATEP para o auxílio do planeamento de produção.

Aplicar a teoria das filas de espera usando a fórmula desenvolvida por Schelasin para encontrar os valores de variabilidade, ao contrário das conclusões do autor, não permitiu obter resultados precisos o suficiente para serem usados. Parte destes resultados podem ser explicados pelas premissas desta teoria que não foram cumpridas, tais como a fábrica não se encontrar com alto volume e taxas de utilização, existirem ações deliberadas da parte dos operadores para priorizarem certos produtos consoante várias heurísticas mal definidas e a forma diferente de cálculo da taxa de utilização prevista dos equipamentos. Também convém denotar que os resultados positivos de Schelasin referem-se a um conjunto de processos, o autor apenas considera os resultados que obteve para o tempo de ciclo por camada, o que implica vários processos. Olhando para os resultados obtidos, embora insuficientes para se fazer uma análise correta, de facto o somatório dos tempos de espera aparenta ser semelhante ao somatório dos tempos de espera reais (48,84 horas de espera totais calculadas *vs* 40,84 horas de espera reais para os processos estudados), porém estes resultados não são tão úteis para o projeto como obter os tempos de espera por processo.

A utilização dos valores de tempo de espera médios por processo implica que sejam atualizados recorrentemente e deve-se ter em atenção o número de amostras recolhidas. Uma amostra baixa tende a fornecer resultados variáveis, pelo que para produtos com baixo volume será preciso aumentar o espaço temporal da amostra. Para produtos com alto volume, dados de uma amostra de 3 meses é manifestamente suficiente.

Também é necessário evitar dados fora do normal (*outliers*), tais como de lotes que ficaram em *hold* com certos códigos que não permitem o processamento do produto durante mais do que alguns dias. Este valor não é específico pois varia de produto para produto e mesmo da utilização da fábrica.

Produtos novos ou em avaliação necessitam de uma atenção especial. Numa fase inicial, como este estudo requer uma base de dados histórica já desenvolvida, deve-se usar os dados de produtos semelhantes. Caso estes não existam, existe uma forte correlação entre os tempos de espera dos produtos, no sentido em que o *ranking* de tempos de espera por processo e por produto tende a ser semelhante, pelo que se pode usar a média geral dos tempos de espera de cada processo para todos os produtos e incrementar um multiplicador consoante as características do produto – mais *dies* por *wafer* deverá implicar um maior tempo de processo, o que por sua vez deve implicar tempos de espera maiores. Um estudo futuro sobre a correlação entre o número de *dies* por *wafer* e os tempos de processo e espera poderia auxiliar esta análise.

Com a quantificação dos tempos de espera consegue-se obter uma imagem ainda mais precisa dos tempos de espera, repartindo-os nas suas razões. Devido à falta de tempo e recursos apenas foi possível analisar os 7 processos/grupos de processos com os tempos de espera mais elevados, dos quais se seguiu individualmente 181 lotes de 4 famílias de produtos diferentes e se anotou as razões pelas quais ficaram em espera. Embora aparente ser uma amostra de tamanho razoável, é necessário um estudo mais completo para se obter resultados mais precisos. Ainda assim o efeito do sequenciamento, WIP e *batching* nos tempos de espera é óbvio. Com esta análise não se pretende transmitir que as heurísticas estão a ser mal aplicadas ou que a gestão de WIP está a ser mal planeada, mas permite concluir que estes são os fatores a ter em atenção para futuras melhorias de tempo de ciclo.

Seria interessante de futuro fazer um estudo sobre o aumento de custo por produto ao diminuir o seu tempo de ciclo. Apesar de não ser absolutamente necessário diminuir o tempo de ciclo, seria uma vantagem competitiva atraente que poderia compensar o aumento do custo dos serviços prestados e talvez abrir portas a novos clientes. Os estudos de Killian (2010) no tamanho do *batch* são também interessantes e sugerem que seria preferível, a nível de melhoria do tempo de ciclo, nos equipamentos que fazem *batching*, utilizar vários equipamentos de menor porte (que façam *mini-batch* ou mesmo sem *batching*) em vez de um equipamento com um tamanho do *batch* elevado.

6 Referências e bibliografia

Chung, Shu-Hsing e Hung-Wen Huang. 2002. "Cycle time estimation for wafer fab with engineering lots". *IIE Transactions* no. 34 (2):105-118. https://doi.org/10.1023/A:1011935728647.

Coimbra, E. 2013. *Kaizen in Logistics and Supply Chains*. McGraw-Hill Education.

FabTime. 2018. "Cycle time estimation formulas". Acedido a 6/4/2018. http://www.fabtime.com/cycle-time-queuing.php#.

Goldratt, E.M. e J. Cox. 1992. *The Goal: A Process of Ongoing Improvement*. North River Press.

Goldratt, E.M., I. Eshkoli e J. Brownleer. 2009. *Isn't it Obvious?*: North River Press.

Hopp, Wallace J. e Mark L. Spearman. 2001. *Factory physics: foundations of manufacturing management*. 2nd ed. Boston: Irwin McGraw-Hill.

Ishikawa, K. 1976. *Guide to quality control*. Asian Productivity Organization.

Killian, Stubbe. 2010. "Development and Simulation Assessment of Semiconductor Production System Enhancements for Fast Cycle Times". Tese de doutoramento, Technischen Universität Dresden - Fakultät Informatik. http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-27343.

Leachman, Rob. 2015. Cycle Time Management. [Material de apoio da Universidade de Berkeley, Califórnia, acedido em 26/3/2018, disponível em: http://ieor.berkeley.edu/~ieor130/CT%20Management%20Nov%202015.pdf].

Li, Minqi, Feng Yang, Hong Wan e J. W. Fowler. 2014. "Simulation-based experimental design and statistical modeling for lead time quotation". *Journal of Manufacturing Systems* no. 37 (1):362-374. https://doi.org/10.1016/j.jmsy.2014.07.012.

Ohno, T. 1988. *Toyota Production System: Beyond Large-Scale Production*. Taylor & Francis.

Reis, L., L. Varela, J. Machado e J. Trojanowska. 2016. "Application of lean approaches and techniques in an automotive company". *The Romanian Review Precision Mechanics, Optics & Mechatronics* (50):112-118.

Rose, Oliver. 2003. "Comparison of Due-date Oriented Dispatch Rules in Semiconductor Manufacturing". Em *Proceedings of the 2003 Industrial Engineering Research Conference*.

Rother, Mike e John Shook. 1999. *Learning to See - Value Stream Mapping to Create Value and Eliminate Muda*. Lean Enterprise Institute.

Schelasin, R. 2011. "Using static capacity modeling and queuing theory equations to predict factory cycle time performance in semiconductor manufacturing". Comunicação apresentada em Proceedings of the 2011 Winter Simulation Conference (WSC). 11-14 Dec. 2011.

Sha, D. Y. e S. Y. Hsu. 2004. "Due-date assignment in wafer fabrication using artificial neural networks". *The International Journal of Advanced Manufacturing Technology* no. 23 (9-10):768–775.

Shanthikumar, J. George, Shengwei Ding e Mike Tao Zhang. 2007. "Queueing Theory for Semiconductor Manufacturing Systems: A Survey and Open Problems". *IEEE Transactions on Automation Science and Engineering* no. 4 (4):513-522.

Tirkel, I. 2013. "Forecasting flow time in semiconductor manufacturing using knowledge discovery in databases". *International Journal of Production Research* no. 51 (18):5536–5548.

Toyota Motor Corporation. 2006. "Ask 'why' five times about every matter". Acedido a 2/3/2018. http://www.toyota-global.com/company/toyota_traditions/quality/mar_apr_2006.html.

Tracy, Dan e Clark Tseng. 2016. "Fan-Out Wafer Level Packaging Takes Off". Acedido a 16/3/2018. http://www.semi.org/en/fan-out-wafer-level-packaging-takes.

Van Zant, P. 1997. *Microchip fabrication: a practical guide to semiconductor processing*. McGraw-Hill. https://books.google.pt/books?id=gylTAAAAMAAJ.

Wang, Junliang, Jie Zhang e Xiaoxi Wang. 2018. "A Data Driven Cycle Time Prediction With Feature Selection in a Semiconductor Wafer Fabrication System". *IEEE Transactions on Semiconductor Manufacturing* no. 31 (1):173-182.

Zhou, Zhugen e Oliver Rose. 2009. "A Bottleneck Detection and Dynamic Dispatching Strategy for Semiconductor Wafer Fabrication Facilities".1646-1656. https://doi.org/10.1109/WSC.2009.5429181.

ANEXO A: Descrição geral da cadeia de processos

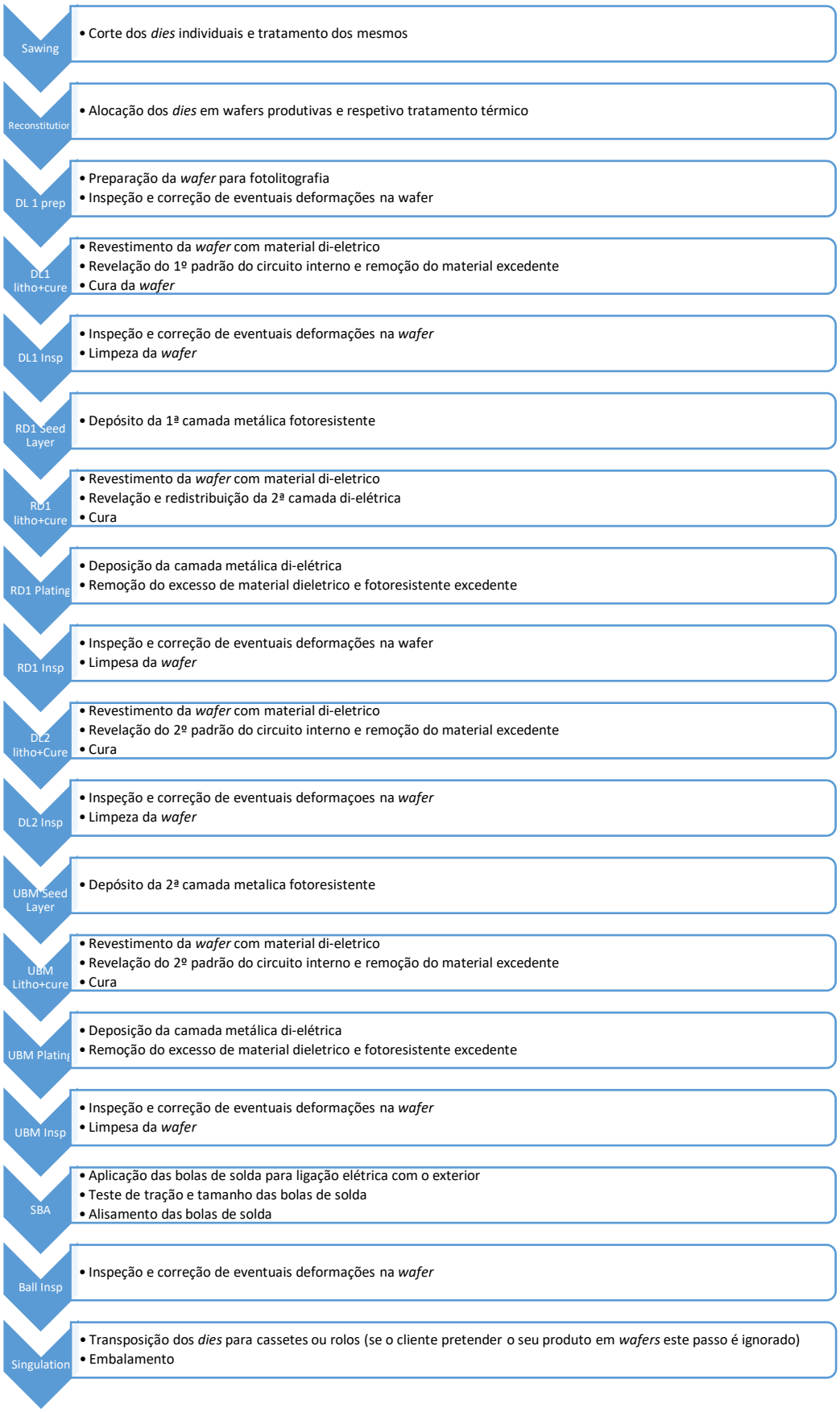

Figura A 1 – Descrição geral da cadeia de processos

ANEXO B: Trajeto de produção de um produto

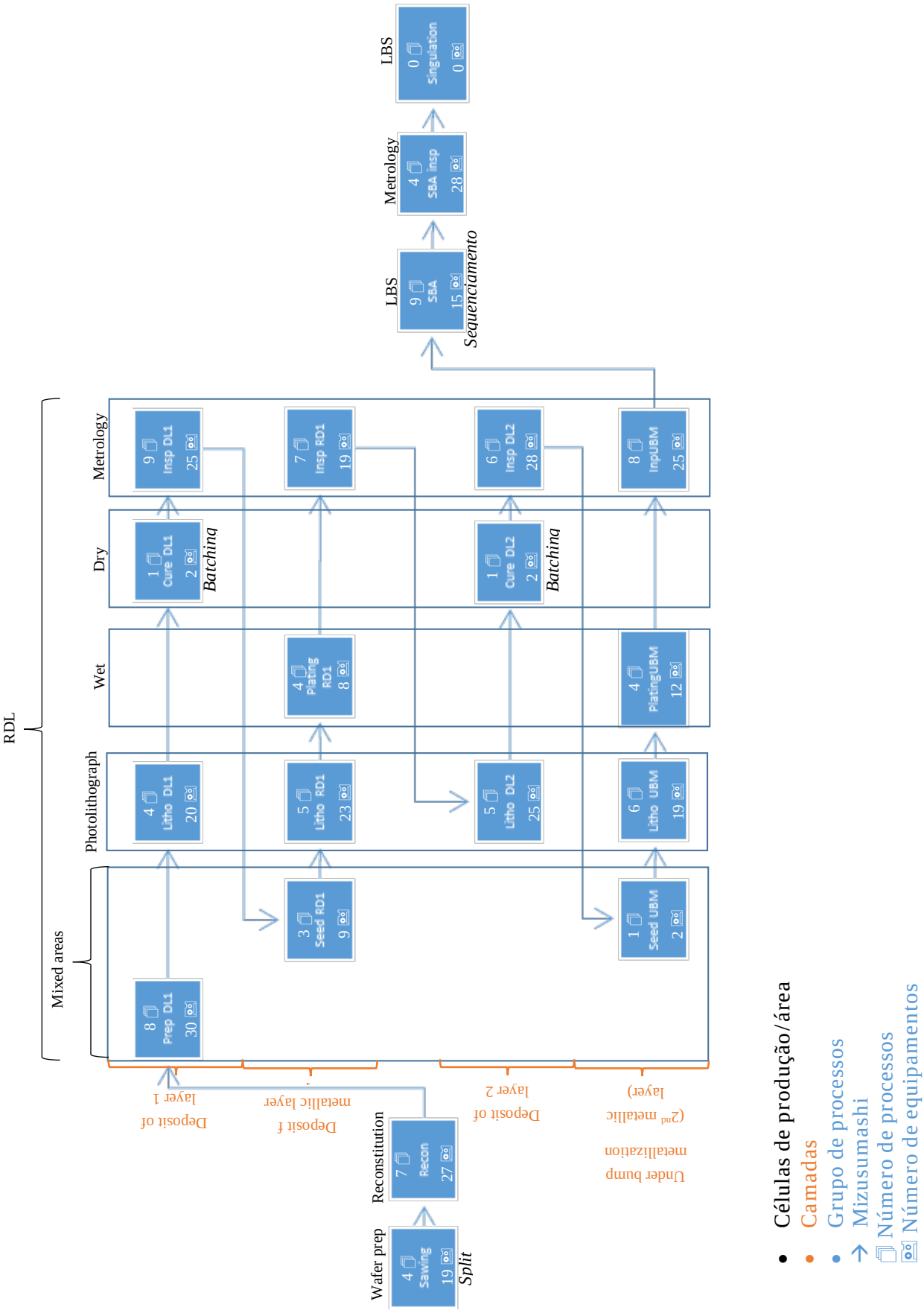

Figura B 1 - Trajeto de produção de um produto

ANEXO C: Resultados quantificação dos tempos de espera

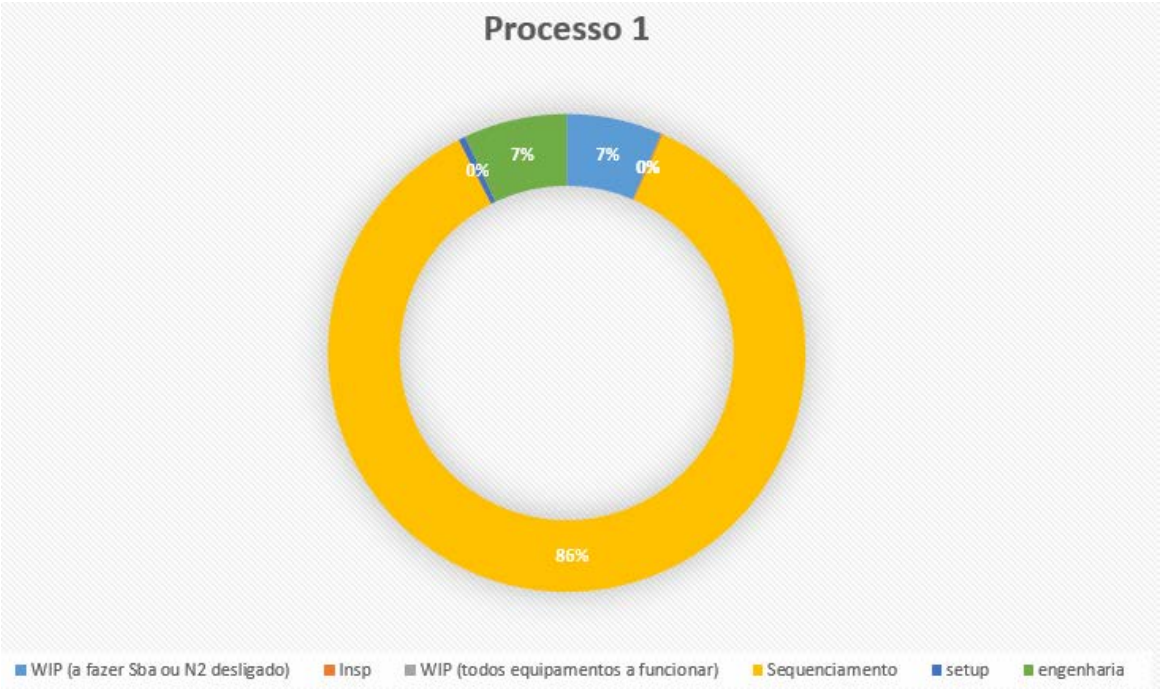

Figura C 1 – Peso relativo das razões dos tempos de espera no processo 1

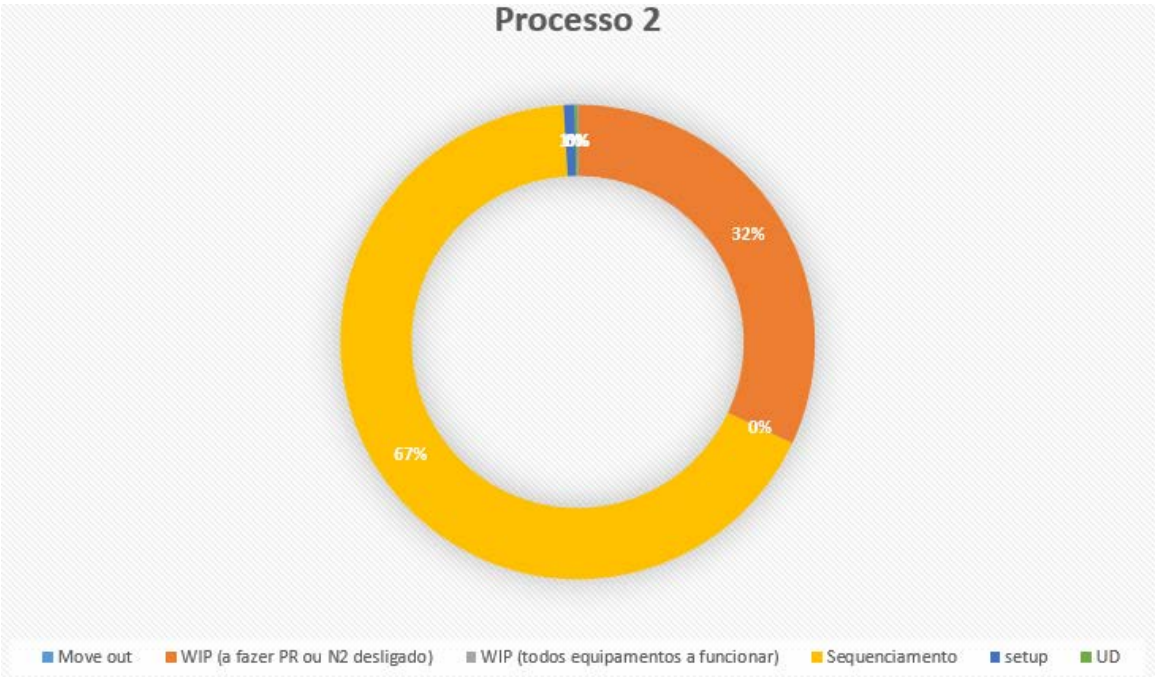

Figura C 2 – Peso relativo das razões dos tempos de espera no processo 2

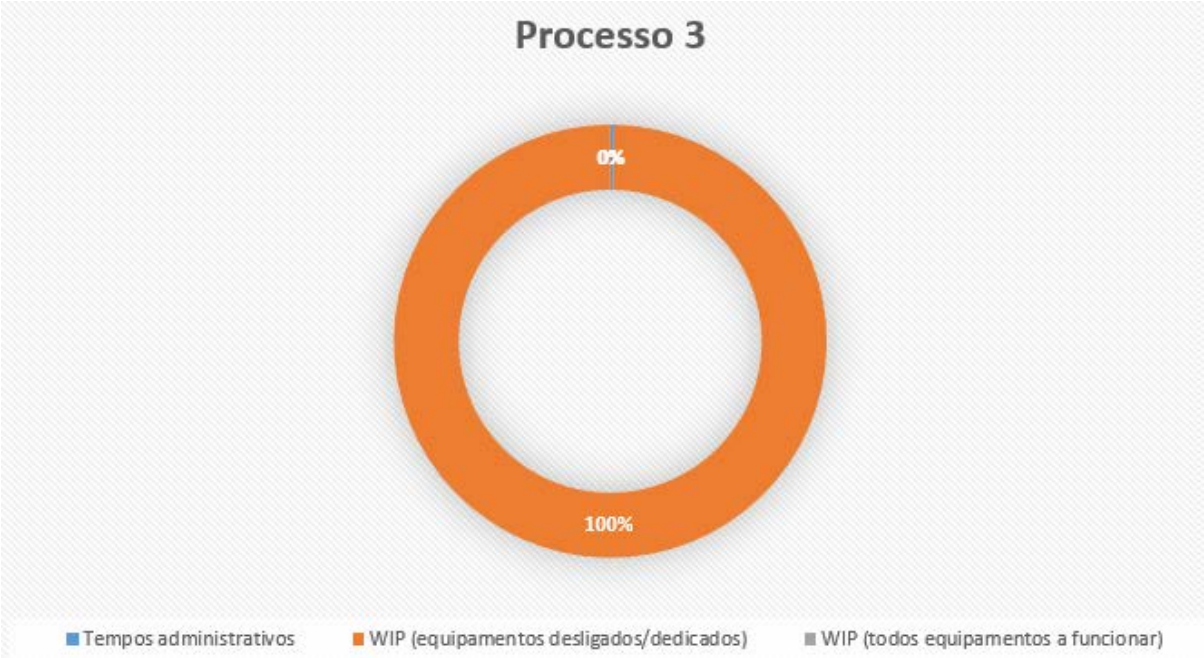

Figura C 3 – Peso relativo das razões dos tempos de espera no processo 3

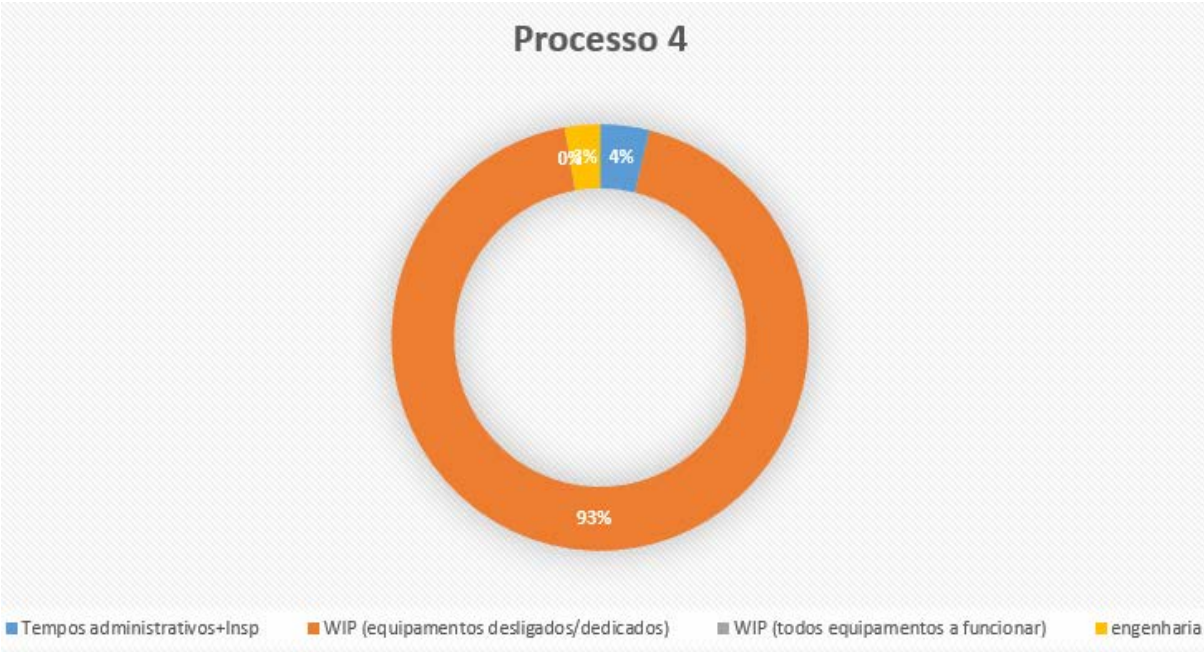

Figura C 4 – Peso relativo das razões dos tempos de espera no processo 4

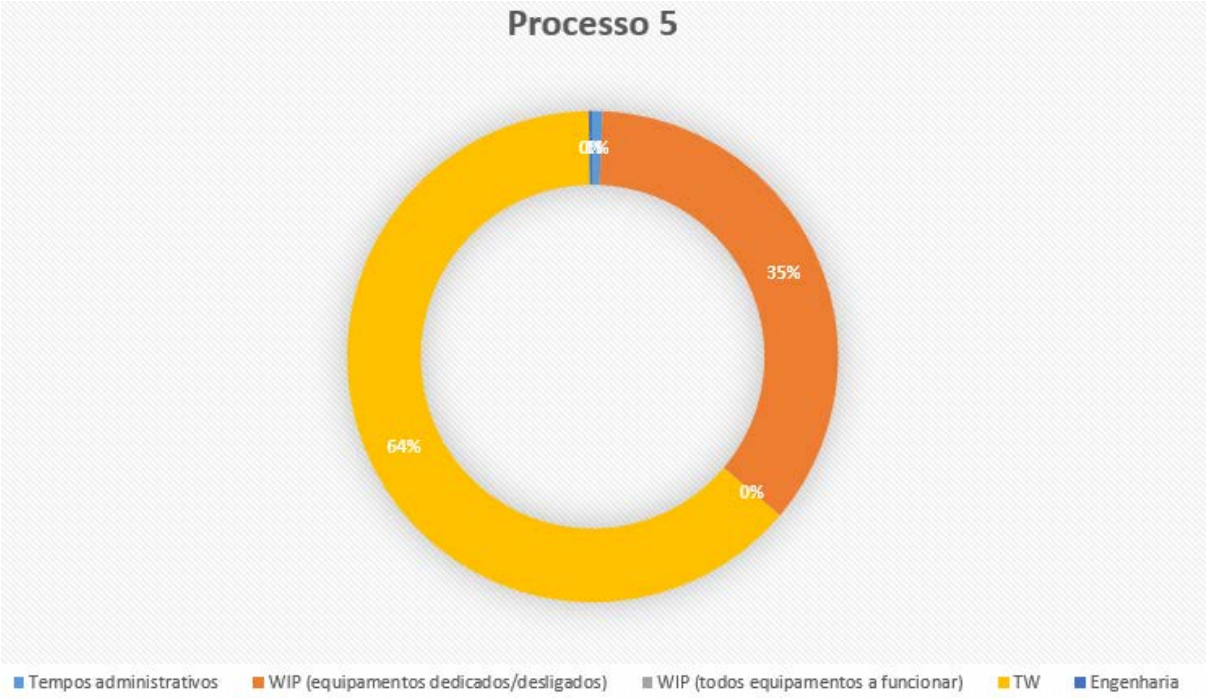

Figura C 5 – Peso relativo das razões dos tempos de espera no processo 5

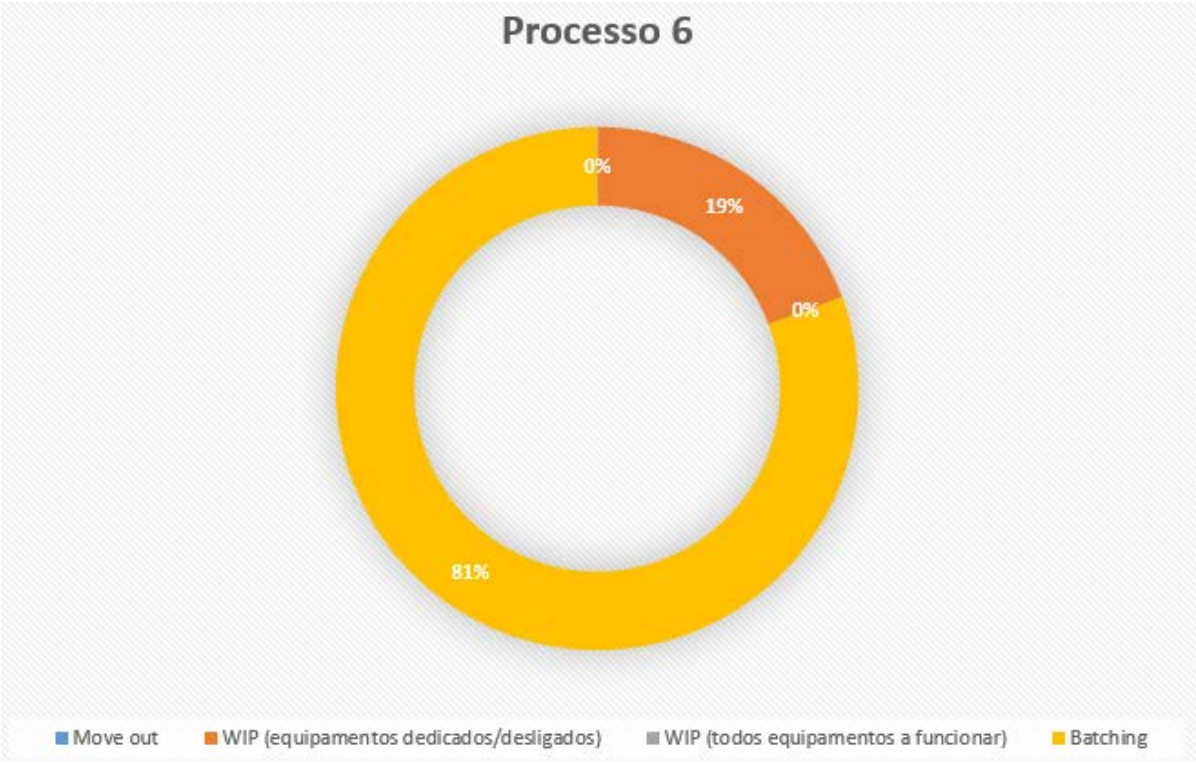

Figura C 6 – Peso relativo das razões dos tempos de espera no processo 6

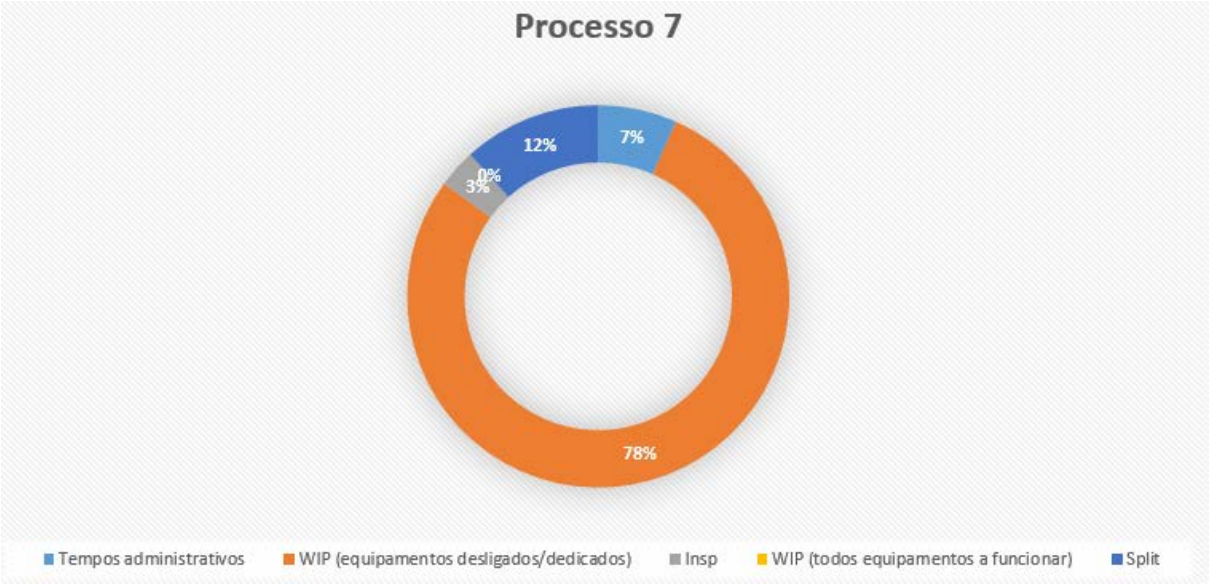

Figura C 7 – Peso relativo das razões dos tempos de espera no processo 7