**ИССЛЕДОВАНИЕ МНОГООБЪЕКТНОЙ ОПТИМИЗАЦИИ РАСПИСАНИЯ ПАРАЛЛЕЛЬНЫХ МАШИН С ПОМОЩЬЮ ПЕРЕМЕННОГО ПОИСКА СОСЕДЕЙ**

Robert Kohn

Oliver Rose

Christoph Laroque

Universität der Bundeswehr München

Institut für Technische Informatik

Fakultät für Informatik

Neubiberg, 85577, GERMANY

University of Paderborn

Heinz Nixdorf Institut

Business Computing, esp. CIM

Paderborn, 33102, GERMANY

**АННОТАЦИЯ**

Управление несколькими целями является критически важной проблемой, связанной с планированием решений по изготовлению пластин. В этой статье представлены результаты вычислений для решения задач параллельной пакетной машины (PBMSP) с помощью поиска переменной окрестности (VNS), обогащенные опытом отрасли. На основе экспериментов мы представляем факторы корреляции между наиболее распространенными ключевыми показателями эффективности (KPI), которые рассматриваются как цели, оценивая силу и направление их взаимосвязей. Мы обсуждаем эксперименты для парето-целевых функций и взвешенных целевых функций, состоящих из важных KPI. Мы придаем большое значение специфической роли критических ограничений в системе планирования, усиленной оптимизацией, например, временные рамки и минимальные размеры партии. Чистое существование критических ограничений обязательно требует многоцелевой оптимизации функций. Экспериментально эта статья исследует иерархические целевые функции, управляющие максимальными временными границами и минимальными размерами партий, обсуждая стратегии решения и ловушки.

**1 ВВЕДЕНИЕ**

Производство пластин как сложная мастерская с многочисленными отраслевыми характеристиками и различными специфическими для области ограничениями широко считается одной из самых трудных в управлении систем производства. С самого начала преобладающей является система оперативного управления цехами в этом промышленном секторе, которая постоянно совершенствуется в своей способности контролировать незавершенное производство (WIP). Несмотря на то, что современные системы диспетчеризации достигли впечатляющего уровня сложности, предлагая хорошие рабочие решения для большинства задач управления, система управления цехом следующего поколения будет значительно отличаться мощными стратегиями планирования. Несколько авторов из отрасли представляют системы планирования, успешно внедренные на операционном уровне, сообщая о существенных преимуществах, даже по сравнению с высокоразвитыми системами диспетчеризации (см. Bixby et al. 2006; Yurtsever et al. 2009).

В процессе замены диспетчеризации системами планирования обычно первым шагом является применение оперативного планирования к ряду параллельных машин, которые принадлежат одной и той же области. В результате декомпозиции весь цех делится на дизъюнктивные наборы параллельных машин, которые обычно связаны друг с другом с помощью поступающих материальных потоков. В общем, снижение сложности с помощью методов декомпозиции позволяет нам частично решить проблему, которая раньше рассматривалась как слишком большая, или слишком сложная, или и та, и другая. В частности, отдельное планирование наборов параллельных машин одного и того же типа имеет смысл как минимум в двух направлениях, что снижает сложность. С одной стороны, структурная сложность снижается за счет исключительно сосредоточения внимания на количестве машин одной технологической области, что облегчает жизнь специалистов в условиях существования многочисленных специфических для данной области характеристик. Следовательно, решение для определенной задачи планирования параллельной машины (PMSP) намеренно разрабатывается для удовлетворения исключительно специфических для конкретной области требований, возникающих в целевой области. С другой стороны, мы уменьшаем сложность вычислений, которая увеличивается в геометрической прогрессии и приводит к миллиардам альтернативных решений, которые в конечном итоге формируют график. Решение PMSP считается одним из наиболее перспективных вариантов использования в области планирования на оперативном уровне. PMSPs кажутся технически осуществимыми и обещают принести существенную выгоду.

Особенно PMSP с пакетной обработкой находятся в центре внимания исследователей и экспертов из отрасли, в ходе дальнейшей работы, упомянутой как Задачи планирования параллельной пакетной машины (PBMSP). PBMSP тщательно изучаются, потому что машины для производства тонких кристаллических пластин существенно влияют на длительность цикла партии, что является ключевым показателем эффективности (KPI) для производственных мощностей. PBMSP, обычно расположенные в области печи, охватывающей процессы диффузии и окисления, имеют различные ограничения, которые имеют большое значение и требуют специальной обработки при разработке системы планирования. Некоторые (критические) ограничения могут привести к невозможным графикам в определенных ситуациях, например, минимальные размеры партии и максимальные сроки. Особенно те критические ограничения, которые могут привести к неверным решениям во время процесса оптимизации, требуют внимания в каждой концепции системы планирования. Несмотря на эти проблемы, сообщество по планированию ожидает значительных улучшений для KPI при применении планирования для PBMSP; особенно в сочетании с системой моделирования, которая обеспечивает прогнозы задач.

Поскольку доказано, что большинство PMSP являются NP-сложными, метаэвристика, проверенная на наличие проблем планирования, обеспечивает большую часть исследовательской деятельности. Помимо метаэвристики важную роль играют смешанное целочисленное программирование (MIP) и динамическое программирование (DP), которые представляют точные методы, даже если они служат только для доказательства глобальной оптимальности для небольших проблемных случаев. Но, учитывая, что технологическое развитие не должно замедляться, точные методы в сочетании с методами разложения имеют хорошие шансы утвердить влияние на проблемные случаи с практическими размерами. В области популяционной метаэвристики была проделана большая работа для изучения эволюционных алгоритмов (EA) и подходов, основанных на Оптимизации муравьиной колонии (ACO). Что касается метаэвристики траектории, мы видим, что значительное количество литературы сосредоточено вокруг поиска переменной окрестности (VNS) и жадного случайного адаптивного поиска (GRASP). До настоящего времени на сложность вычислений совместно отвечают точные методы, методы декомпозиции и мощная метаэвристика.

Исследователи были в первую очередь сосредоточены на вопросах выполнимости и производительности при разработке моделей оптимизации, независимо от того, выбрали ли они точный или эвристический подходы. В частности, для PBMSP, возможность оптимизации моделей, которые будут окончательно применены в цехе, зависит от различных ограничений (например, несовместимых семейств работ, преданности процессу, временных ограничений, минимальных размеров партии и т. Д.), Которые должны быть должным образом рассмотрены соответствующими реализациями. Что касается эффективности методов оптимизации, исследователи работали над процедурами планирования, характеризующимися приемлемым соотношением между качеством решения и временем вычислений. Таким образом, сообщество в основном борется с проблемами, связанными с производительностью и ограничениями, большинство в литературе решает проблемы планирования с единственными целями компонента. Тем не менее, до сих пор существует множество открытых вопросов для планирования задач с отдельными целями, представление слегка перемещается к многоцелевой оптимизации (MOO). Для некоторых PBMSP, включающих критические ограничения, которые могут привести к неосуществимым решениям, MOO возникает как необходимость, а не как дополнительная функция, иметь которую хорошо.

**2 STATE-OF-THE-ART**

В качестве руководства к обширному разнообразию литературы, касающейся планирования в области производства полупроводников, мы хотели бы сослаться на недавний обзор, представленный в (Mönch et al. 2011). Далее мы дадим краткий обзор наиболее важных эвристических методов пакетной обработки. Для более подробного обзора пакетного планирования (и диспетчеризации) можно найти подробный обзор в (Mathirajan et al. 2006).

Glassey et al. (1991) представляют динамическую пакетную эвристику (DBH), которая впервые включает в себя информацию о будущих поступлениях на работу, которая в следующем тексте называется прогнозом. Фаулер и др. (1992) представляют более сложную эвристику управления следующим поступлением (NACH), которая исключительно учитывает следующую задачу (см. Фаулер и др. 2000; Соломон и др. 2002). Ван дер Зи предлагает эвристику динамического назначения задачи (DJAH), которая также учитывает прогнозную информацию (см. Van der Zee 2007). Habenicht и Mönch (2003) обсуждают эвристику пакетирования в сочетании с генетическими алгоритмами (GA), чтобы минимизировать опоздания. Balasubramanian et al. (2004) представляют эвристику пакетной очевидной опоздания (BATC) для минимизации взвешенного опоздания, усиленную схемой поиска, основанной на GA. Gupta и Sivakumar (2006) предлагают стратегию управления, ориентированную на своевременность, с использованием прогнозной информации, чтобы минимизировать максимальное опоздание и количество запоздалых заданий. Ша также обсуждает эвристику прогнозирования (см. Sha et al. 2007). Более того, некоторые эвристические методы пакетной обработки, направленные на минимизацию общего взвешенного опоздания (TWT), обсуждаются в (Kim et al. 2010).

В области планирования с использованием методов оптимизации существует множество подходов к решению задач PBMS с учетом различных подмножеств ограничений. Reichelt и Mönch (2006) изучают мультипопуляционный генетический алгоритм (MPGA) и многоцелевой генетический алгоритм (MOGA) для PBMSP с целью минимизации TWT и промежуточного периода времени. В (Klemmt et al. 2011, Wang et al. 2010, Klemmt et al. 2008) можно найти составы MIP для PBMSP, которые направлены на минимизацию запаздывания или общего взвешенного опоздания. Wang and Chou (2010) представляют сравнение только между MIP и подходами, основанными на имитационной нормализации (SA) и GA, и оба усовершенствованы с помощью многоступенчатого динамического программирования (MSDP). Также были опубликованы различные эволюционные подходы, соответственно вдохновленные природой схемы поиска, основанные на концепции GA, например, Chiang et al. (2010), Kashan et al. (2008), Malve and Uzsoy (2007), Mönch et al. (2005), Balasubramanian et al. (2004) и Habenicht and Mönch (2003). Damodaran и Velez-Gallego (2011) представляют GRASP и реализацию поиска SA в (Damodaran и Velez-Gallego 2012), чтобы минимизировать время выполнения. Существуют подходы, использующие концепцию ACO для минимизации TWT для PBMSP (см. Mönch et al. 2009, Li et al. 2008, Raghavan and Venkataramana 2006). В Cakici et al. (2013) обсуждаются эвристические алгоритмы, использующие схемы VNS, и сравниваются с математической моделью, разработанной для PBMSP с динамическими поступлениями и несовместимыми семействами заданий. Almeder и Mönch (2011) изучают популярную метаэвристику, применяемую к PBMS с несовместимыми семействами задач, чтобы минимизировать TWT; они изучают варианты ACO, GA и VNS и сравнивают их показатели. Klemmt et al. (2009) изучали PBMS с несовместимыми семействами задач и динамическим прибытием задач, сравнивая MIP и VNS с точки зрения минимизации TWT. Jula и Leachman (2010) представляют алгоритм, основанный на линейном программировании (LP), подход, основанный на целочисленном программировании (IP), и алгоритм, основанный на эвристике, для решения неоднородных PBMSP с неодинаковыми размерами заданий и несовместимыми семействами заданий, Югма и соавт. (2008) представляют подход решения, основанный на имитации отжига (SA), применяемый для PBMSP, улучшающий пропускную способность, эффективность партии и коэффициент времени потока.

**3 КОНСТРУКЦИЯ ЭКСПЕРИМЕНТОВ**

**3.1 Экспериментальная система**

Экспериментальная среда в основном состоит из основанной на моделировании среды оптимизации, разработанной для решения PBMSP, существующих в интерфейсных средствах изготовления пластин. Базовая система охватывает различные PBMSP, которые различаются по своему набору ограничений и целей. Для поиска улучшенных расписаний мы реализовали обобщенную концепцию VNS, предлагая многочисленные варианты VNS, включая большинство из упомянутых в соответствующей литературе. Разница между этими вариантами в основном заключается в балансе между эксплуатацией и исследованием пространства поиска, помимо множества различных комбинаций параметров на выбор.

Мы рассматриваем разработанную систему планирования как готовую к запуску на операционном уровне, что означает, что мы можем загрузить (и проверить) существующий в настоящий момент экземпляр проблемы с фактическими данными из потрясающих баз данных. Затем следует процедура планирования, которая создает улучшенное расписание, которое немедленно записывается обратно в систему управления производством (MES) для выполнения. В дополнение к реализованным функциям реального времени, генератор моделей предлагает создавать пользовательские экземпляры PMSP с определенными характеристиками. Механизм генерирования модели генерирует определенные экземпляры проблемы с конкретными характеристиками, в основном описываемыми генератором моделей, устанавливая важные переменные модели, используя случайные числа и следуя стандартным статистическим распределениям, чтобы генерировать набор независимых экземпляров модели, которые в основном показывают одинаковые характеристики, но в то же время немного отличаются друг от друга. Мы используем базу данных для организации управления данными, необходимого для эффективной (и удобной) работы экспериментальной системы. Все данные эксперимента, представляющие входные параметры (включая экземпляры модели), а также результаты и выходные данные доступны через соединения с базой данных. Вся система ограничена экземплярами PMSP, размер которых не превышает 5 ГБ в сжатом размере, что соответствует максимальному размеру для крупногабаритного символа (CLOB). Поскольку крупные эксперименты по моделированию (оптимизации) часто страдают от недостатка вычислительной мощности и времени, мы делегируем обширные исследования кластеру высокопроизводительных вычислений (HPC) с 64 ядрами, подключенными к базе данных. Вся система написана на C#, и пока что больше внимания уделяется понятности кода, а не скорости вычислений.

**3.2 Задача планирования параллельной пакетной машины (PBMSP)**

Мы опишем задачи планирования, рассмотренные в этом исследовании, с использованием схемы классификации α | β | γ, предложенной Грэмом (1979). Рассматриваемая базовая PBMSP включает в себя неравное время обработки, назначение машины для конкретной работы, параллельное пакетирование с несовместимыми семействами работ и произвольно максимальные размеры партий для семейства работ на машине. В соответствии с α | β | γ-нотацией и с учетом того факта, что цель в наших исследованиях меняется и подвергается анализу, поэтому она просто обозначается как цель, мы описываем проблему с Rm | Mj, p-пакет, несовместимым, bmaxj | цель. Набор реализованных показателей эффективности, которые в основном определяют целевые функции, относится к общим KPI, соответствующим пропускной способности, времени потока и запаздыванию. Целевая функция составляет, по крайней мере, одну из этих мер эффективности, но в основном их множество, с весами или иерархической структурой. В следующем списке представлены α | β | γ-нотации, используемые в остальной части статьи; где

* Rm: несвязанные параллельные машины (с неравным временем обработки),
* Mj: посвящение машин (работа посвящена ограниченному набору машин),
* p-пакет: параллельное пакетирование (на компьютере одновременно обрабатывается несколько заданий),
* incompatible: несовместимые семьи рабочих мест (рабочие места разных семей не могут обрабатываться вместе),
* bmaxj: произвольно максимальный размер партии для семейства работ на машине,
* bminj: произвольно минимальный размер партии для семейства работ на машине,
* tbj: произвольное ограничение по времени для задания.

Разработанная система планирования позволяет нам исследовать PBMSP с произвольно выбранным подмножеством перечисленных ограничений и с многокомпонентными целевыми функциями, объединяющими общие KPI вместе. Ранее в этой статье мы упоминали критические ограничения, соответственно минимальные размеры пакетов (bminj) и максимальные временные границы (tbj). При расширении базовой модели мы также рассмотрим PBMSPs Rm | Mj, p-пакет, несовместимые, bmaxj, bminj, | target и Rm | Mj, p-пакет, несовместимые, bmaxj, tbkj, | target. Специальные намерения этих критических ограничений, которые могут вызвать недействительные решения, и их неявное влияние на целевую функцию, которая обязательно заканчивается многокомпонентной целевой функцией, обсуждается в разделе 5.3.

Несмотря на то, что мы можем вычислить PBMSP с динамическим поступлением, соответственно Rm | Mj, rj, p-пакет, несовместимы, bmaxj, | target, где rj обозначает ненулевую дату выпуска задания и даже с учетом динамического прибытия (заблаговременного) заметно способствует потенциалу оптимизации, мы опускаем их в этом исследовании. Причина в том, что мы частично выбрали общую минимизацию рабочего времени для цели, что не имело бы особого смысла при работе с динамическими поступлениями. Чтобы избежать путаницы и сохранить эксперименты в ясной и сопоставимой форме, мы не исследовали динамические поступления в любом эксперименте, упомянутом в этой статье. Следовательно, также является излишним учитывать различные уровни использования или ограничения минимального размера партии; это утверждение верно для PBMSP без динамических поступлений.

В рамках экспериментов мы последовательно исследовали три размера проблемных экземпляров, обозначенных как S (5 | 150), M (10 | 300) и L (15 | 450), где первый параметр представляет количество машин, а второй - число задач. Настройки модели по умолчанию приведены в таблице 1.

Таблица 1 – Настройки модели по умолчанию

|  |  |  |  |
| --- | --- | --- | --- |
| Модель (машины|задачи) | S (5|150) | M(10|300) | L(15|450) |
| # Машины | 5 | 10 | 15 |
| # Задачи | 150 | 300 | 450 |
| # Семейства задач | 5 | 10 | 15 |
|  | | | |
| Плотность посвящения | 0.7 | | |
| Время процесса [мин] | U~(240,360) | | |
| Начальное опоздание [час] | N~(0,24) | | |
| Веса  задач | U~(1,2,3,4,5) | | |
| Максимальный размер пакета [лот] | Const~(8) | | |

**3.3 Поиск переменной окрестности (VNS)**

Младенович и Хансен (1997) предложили эвристику VNS, основанную на соседских структурах, используемых для решения крупномасштабных комбинаторных задач. Среда оптимизации на основе моделирования, которую мы используем для решения PBMSP, использует VNS для создания оптимизированных графиков с учетом поставленных целей. Мы реализовали VNS как абстракцию предложенных схем, что позволяет нам свободно настраивать два вложенных уровня поиска. Оба уровня могут быть параметризованы независимо друг от друга, где каждый уровень поиска определяет набор структур соседства, локальную процедуру поиска (первое улучшение или лучшее улучшение) и политику колебаний, управляющую диапазоном колебания (постоянным или возрастающим). Эта обобщенная реализация VNS охватывает широкий диапазон вариантов VNS, описанных в литературе, а именно: уменьшенный VNS (RVNS), переменный спуск соседства (VND), обобщенный VNS (GVNS), поиск разложения переменной окрестности (VNDS) и перекос VNS (SVNS). Для подробного описания см. Хансен и Младенович (2009). Комбинируя стратегии и параметры, мы получаем сотни схем поиска VNS, детерминированные варианты, которые используют только локальный поиск, а также стохастические варианты, которым удается уйти от локальных оптимумов. Эти варианты в основном отличаются по балансу между поиском и использованием пространства поиска. Кроме того, система поддерживает MOO, в то время как несколько целей объединяются иерархически или взвешенно, или одинаково объединяются для улучшения границ Парето.

Шесть структур окрестностей создают подпространства всего пространства поиска, инкапсулируя определенный набор операций, используемых для изменения расписания. Реализованные окрестности определяются следующим образом:

* Объединить две партии: найдите две партии, чтобы объединить одну из них (и новую позицию).
* Разделить партию: найдите партию для разделения и вставьте новую партию в новую позицию.
* Поменяйте местами две партии: найдите партии и поменяйте их местами.
* Переместить партию: переместить партию в другую позицию.
* Поменяйте местами две задачи: найдите две задачи из разных партий и поменяйте их местами.
* Переместить задачу: найти задачу и переместить ее в другой пакет.

Поскольку эвристические процедуры поиска работают с данными решениями, мы должны предоставить начальное расписание в качестве начального решения для каждого экземпляра проблемы. Мы используем правила диспетчеризации, выполняемые в системе моделирования, для создания начальных расписаний, которые также предоставляют эталонные объективные показатели для анализа улучшений, полученных в результате оптимизации. В качестве простых правил отправки мы используем «первым пришел - первым вышел» (FIFO) и «Самый ранний срок оплаты» (EDD). Интервал отправки устанавливается равным трем минутам, что означает, что каждые три минуты выполняется процедура отправки, что, вероятно, приводит к запуску нового задания / пакета. Что касается более сложной, оптимизирующей эвристики диспетчеризации (BATC, NACH), описанной в литературе, мы предпочитаем использовать простые правила диспетчеризации, по крайней мере, для целей этой статьи. Поскольку основная цель этой статьи не состоит в том, чтобы ни оценить производительность метода поиска, ни оценить точные потенциалы оптимизации для определенных особенностей модели; речь идет о MOO для PBMSP.

Для метода оптимизации мы выбрали VND в качестве подходящей стратегии поиска, чтобы продемонстрировать влияние MOO на две разные цели вместе взятые. В связи с этим единственный исключительный эксперимент обсуждается в разделе 4, где мы изучаем факторы корреляции между набором KPI, где каждый рассматривается как однокомпонентная цель. VND можно рассматривать как детерминированную (наилучшее улучшение) локальную стратегию поиска, работающую в ограниченном наборе окрестностей, специально разработанную для решения проблемы параллельной пакетной машины. Есть две причины, оправдывающие наше решение применить детерминированный вариант VNS вместо того, чтобы выбрать стохастический вариант, который, как было доказано, превосходит детерминированные подходы. С одной стороны, детерминированное поведение VND уменьшает общее количество экспериментальных прогонов, поскольку нет необходимости запускать множественные репликации, которые гарантируют определенный уровень статистической надежности, в отличие от стохастических производных VNS. С другой стороны, детерминированный локальный поиск обеспечивает лучшее понимание сложности планирования с учетом размера проблемных экземпляров. Изучение измеренного времени вычислений и / или количества ходов поиска в сочетании с анализом улучшений путем оптимизации, связанных с показателями производительности, дает понять, можно ли найти локальные оптимумы для определенных задач или нет в соответствии с заданными вычислительными сроками. Кроме того, чтобы минимизировать бремя анализа средних значений и дисперсий в экспериментах, одновременно увеличивая понятность и достоверность результатов экспериментов, мы отказались от использования стохастического поиска. В свою очередь, детерминированный поиск избегает анализа отклонений, возникающих в результате стохастических эффектов, вызванных множественными репликациями. Мы также хотели бы отметить тот факт, что потенциалы в оптимизации значительно зависят от характеристик модели. На основе обширных исследований мы обнаружили, что существуют существенные отклонения для результатов моделирования (оптимизации) среди множества независимых экземпляров, принадлежащих к одному и тому же типу модели.

Таблица 2 – Стандартные настройки VNS.

|  |  |
| --- | --- |
| VNS метод | VND |
| Крайний срок [мин] | 10 |
| Интервал отправки [мин] | 3 |
| Правило распределения работ | FIFO / EDD |

**4 СООТНОШЕНИЕ ЦЕЛЕЙ**

Показатели эффективности взаимосвязаны, некоторые тесно связаны друг с другом, а некоторые нет. Мы знаем о существовании этих отношений, но мы очень мало знаем о величинах. Чтобы лучше понять, мы разработали эксперимент для анализа корреляции между выбором важных KPI, соответственно показателями эффективности или целями. Исследуемыми целями являются пропускная способность (THP) и рабочий диапазон (Cmax), общее время цикла (TCT), общее взвешенное время цикла (TWCT), общее время ожидания в очереди (TQT), общее взвешенное время очереди (TWQT), общее запаздывание (TT) и общее взвешенное опоздание (TWT). Идея состоит в том, чтобы оптимизировать для конкретной цели, а затем рассчитать коэффициенты корреляции после метода Пирсона между KPI в качестве цели и оставшимися KPI. Результирующая матрица коэффициентов корреляции обеспечивает точные значения, которые указывают направление и силу взаимозависимостей между любой комбинацией двух KPI. Рисунок 1 показывает результаты.

Из результатов эксперимента, рассчитанной матрицы коэффициентов корреляции по Пирсону, мы получаем несколько наблюдений, которые представляют самые сильные взаимосвязи, приведенные в порядке убывания их величин. THP является эквивалентом Cmax, и оба, как и ожидалось, напрямую и прочно связаны. Коэффициент корреляции равен -1 и подтверждает экспериментом, что увеличение THP как цели непосредственно приводит к снижению Cmax, и наоборот. Мы наблюдали очень сильную зависимость (почти 1) между QT и CT, а также между их невзвешенными версиями WQT и WCT. Это имеет смысл, так как оба отличаются только в части времени обработки. TWT и TT показывают высокие значения около 0,8 для их коэффициентов корреляции друг с другом, в обоих направлениях. Очень похоже, мы наблюдали коэффициенты корреляции около 0,7 между TWCT и TCT, а также между TQT и TWQT. Максимизация THP приводит к снижению TCT и TQT, о чем свидетельствует значение корреляции около -0,5. Мы наблюдали немного меньшее влияние на TWCT и TWQT, тогда как коэффициент корреляции составляет около -0,4. Соответственно, минимизация Cmax приводит к снижению TCT и TQT и менее непосредственно к снижению TWQT и TWCT.



Рисунок 1 - Факторы корреляции между показателями эффективности

**5 МУЛЬТИОБЪЕКТНАЯ ОПТИМИЗАЦИЯ**

**5.1 Парето объекты**

Для этого эксперимента мы анализируем двухкритериальные целевые функции, которые объединяют два критерия, эквивалентных друг другу. Результирующая целевая функция Парето с двумя критериями рассматривает решения как улучшенные, если обе меры показывают улучшения, принимая, что одна из них остается неизменной. Мы рассмотрим три целевые функции, каждая из которых состоит из двух составляющих (Cmax), общего времени цикла (TCT) и общего опоздания (TT). Следовательно, мы рассмотрим целевые функции, обозначенные через Cmax | TCT, Cmax | TT и TCT | TT. В качестве исходного решения и для справки мы используем графики, полученные путем моделирования правил диспетчеризации FIFO или EDD. Результаты, представленные в таблице 4, приведены относительно исходного решения.

Как правило, мы не видим почти никакой разницы между результатами, полученными для маленькой модели S (5 | 150) и средней модели M (10 | 300), тогда как для большой модели L (15 | 450) значительно меньше улучшений становятся очевидными. Причина в том, что детерминистическая схема локального поиска (VND) не достигает локального оптимума в течение заданного срока для более крупной модели. Интересно, что по сравнению с FIFO и EDD результаты показывают улучшение Cmax в среднем на 6%. Для TCT мы наблюдаем улучшение на 3%, тогда как самые низкие улучшения проявляются при одновременной фокусировке на TT. Этот эффект подтверждает предположение, что TCT и TT противоречат друг другу. Улучшения в TT в среднем составляют около 2% по сравнению с EDD и до 14% по сравнению с FIFO. При оптимизации для TT, начиная с исходного решения, полученного с помощью FIFO, мы наблюдаем уменьшение улучшений с увеличением размера модели. Это наблюдение также объяснимо концептуальной связью между вычислительной сложностью, вызванной размером модели, и вычислительным крайним сроком, ограничивающим процедуру поиска. На рисунке 2 показаны результаты, отображаемые на диаграмме рассеяния, где каждая точка представляет отдельный экземпляр модели.



Рисунок 2 - Результаты расчетов для двухкритериальной оптимизации Парето

**5.2 Взвешенные цели**

Кажется, что взвешивание целей является правильной стратегией для объединения нескольких целей. Для анализа влияния весов мы подготовили следующий эксперимент. В рамках этого эксперимента мы исследуем взвешенную многоцелевую функцию, состоящую из двух показателей: общего времени цикла (TCT) и общего опоздания (TT). Веса варьируются от нуля до единицы с шагом 0,2 с условием, что сумма всегда равна единице. Сначала мы нормализуем оба улучшения, чтобы сделать их сопоставимыми. Затем нормализованные улучшения умножаются на их веса. Нормализованные и взвешенные улучшения суммируются, и в итоге результат представляет собой значение решения.

В соответствии с другими проведенными экспериментами, мы видим худшие результаты в улучшении для больших моделей L (15 | 450), что приводит нас к заключению, что процедура локального поиска, вероятно, не достигает локального оптимума. В частности, в соответствии с упомянутым выше двухкритериальным экспериментом по оптимизации парето мы видим идентичные результаты для одинаково взвешенного случая TCT (0,5) | TT (0,5). Мы наблюдали максимальное улучшение CT на 5% по сравнению с FIFO и улучшение TT на 2% по сравнению с EDD. Как правило, мы должны сказать, что для целевых функций с весами выше нуля и ниже единицы результаты сильно напоминают друг друга. На рисунке 3 показаны результаты, отображаемые на диаграмме рассеяния, где каждая точка представляет отдельный экземпляр модели.



Рисунок 3 – Результаты расчетов для оптимизации с взвешенными целями

**5.3 Иерархические цели**

Скорее всего, не существует иерархий, включающих KPI, представляющих интерес в качестве целей, которые бы в достаточной мере отражали потребности контроля над цехами. Но существуют особые случаи использования, оправдывающие иерархические структуры для целевых функций. Такой вариант использования, например, имеет дело с критическими ограничениями. Критические ограничения могут привести к повреждению расписаний, которые содержат внеплановые задания, которые в нашей концепции рассматриваются как нежелательные нарушения. Давайте определим термины, используемые в этом контексте, чтобы обеспечить четкое понимание. При решении PBMSP нам, возможно, придется учитывать различные ограничения, и некоторые из них считаются критическими. Критическое означает в этом контексте, что для определенных проблемных случаев некоторые (критические) ограничения могут привести к повреждению расписаний, которые здесь определены как расписания, которые содержат ряд незапланированных заданий. Это означает, что может существовать количество заданий, оставленных незапланированными из-за одного или нескольких неудовлетворенных ограничений, соответственно критических ограничений. Во время наших исследований мы определили два важных критических ограничения: временные рамки для конкретной работы и минимальный размер партии.

При решении проблемных случаев с критическими ограничениями мы сталкиваемся с задачей поврежденных расписаний, которые содержат ряд внеплановых заданий из-за нарушенных ограничений. В случае точного метода оптимизации, используемого для создания расписания, у нас есть доказательство того, что не существует расписания, содержащего весь объем заданий, удовлетворяющих требуемым условиям. В случае процедуры эвристического планирования мы знаем только то, что метод не приводит к расписанию, которое содержит все задания с учетом ограничений. Но оба подхода, точный и эвристический, страдают от одной и той же проблемы: потенциальный риск ситуации, в которой мы не распланировали должным образом все задания.

Обычно задачи планирования определяются таким образом, что весь объем доступных заданий должен планироваться должным образом с учетом данных ограничений без каких-либо исключений. Это требование (не допускать никаких внеплановых заданий) в сочетании с критическими ограничениями, приводит к риску недопустимых решений. Иными словами, мы сталкиваемся с опасностью столкнуться с ситуациями, когда нет действующего решения / расписания. Этот риск неприемлем для систем управления цехами.

Грубо говоря, хоть и очень желательно избегать нарушений с точки зрения незапланированных работ, это не обязательно. Практики предпочли бы плохие графики вместо никаких графиков. Это заставляет нас ослабить требование, которое требует, чтобы все работы были правильно запланированы. В этот момент мы обнаруживаем новую проблему в отношении оптимизации. Представьте себе проблему планирования с критическими ограничениями и целью, которая предназначена для минимизации времени выполнения или времени доставки, связанных с измерением производительности. При оптимизации независимо от того, какой метод применяется, результирующее расписание, скорее всего, будет в значительной степени искажено. В результате поврежденное расписание будет содержать значительное количество внеплановых заданий из-за нарушенных ограничений и очень низкого целевого значения в соответствии с целевой функцией.

Это сводится к тому, что становится неизбежным как-то наказать оптимизатора за создание нарушений. Внедрение системы штрафов в процесс оптимизации является опцией. Аналогичный подход для преодоления этой задачи состоит в том, чтобы рассматривать в качестве цели количество нарушений и, соответственно, количество заданий, которые остаются незапланированными из-за ограничений. Для этой статьи мы ввели показатель эффективности, равный количеству нарушений заданий. В процессе оптимизации мы рассматриваем любое незапланированное задание как нежелательное нарушение. Поскольку предотвращение нарушений является нашей первоочередной задачей, мы изучили комбинированную целевую функцию, которая иерархически связывает две цели, во-первых, количество нарушений и, во-вторых, произвольно выбранную объектную меру.

**5.3.1 Ограничения по времени**

Ограничения по времени - это обычная практика при изготовлении пластин, обусловленная проблемами с качеством. Мы чаще всего наблюдаем временные границы между операциями в области влажной химии и работой печи. Ограничения по времени в основном предназначены для ограничения нежелательных процессов окисления субстрата, что происходит, когда задание / партия ожидает следующего процесса в атмосферных условиях. Как только ограничение по времени нарушается, задание требует отдельного рассмотрения экспертами, и эта дополнительная процедура нежелательна. При создании и улучшении графиков для области печи первоочередной задачей является избежать нарушений, связанных с указанным временем: во-первых, поскольку нарушения с указанным сроком увеличивают риск проблем с качеством, во-вторых, поскольку нарушенные временные рамки требуют отдельной обработки, это в частности, связывает человеческие ресурсы. Нам нравится ссылаться на подходы диспетчеризации и составления графиков, которые также связаны с управлением по срокам (Ham and Fowler 2008; Klemmt et al. 2012; Mason et al. 2007).

Для этого эксперимента мы определили целевую функцию, которая содержит две цели в иерархии, в первую очередь, количество или нарушения и, во-вторых, общее время цикла. Мы рассмотрели десять различных временных схем, которые представляют собой равномерно распределенные временные границы между нулем и кратным ожидаемого временного диапазона (EM = 1800 мин). EM рассчитывается с использованием теоретической пропускной способности системы, в основном определяемой количеством машин и их распределением времени обработки, по отношению к количеству заданий. Диспетчеризация FIFO служит исходным решением и ориентиром для улучшений. В таблице 6 показаны результаты экспериментов, где V обозначает общее количество внеплановых заданий из-за нарушения временных границ, ΔV описывает улучшение с точки зрения предотвращения нарушений с ограничением по времени по сравнению с диспетчеризацией FIFO, CT обозначает среднее время цикла, а ΔCT обозначает улучшение среднего времени цикла по сравнению с FIFO. На рисунке 4 показан блок-график, отображающий общее количество нарушений с привязкой по времени в зависимости от размера модели и схем с привязкой по времени. На рисунке 5 показаны предотвращенные нарушения временной привязки, отображаемые на диаграмме рассеяния, где каждая точка представляет отдельный экземпляр модели.



Рисунок 4 - Результаты расчетов для иерархического MOO с временными границами (1)



Рисунок 5 - Результаты расчетов для иерархического MOO с временными границами (2)

На этом этапе мы хотели бы отразить необходимость рассматривать нарушения как часть целевой функции, а затем представить идею для улучшения. Во-первых, мы хотели бы прояснить, почему нарушения должны быть частью целевой функции для планирования задач, характеризуемых временными рамками. Представьте себе систему с единственной целевой функцией, предназначенной для минимизации продолжительности времени обработки, времени ожидания, времени цикла или опоздания и характеризующейся ограничениями по времени. Система оптимизации, независимо от того, является ли она точной или эвристической, будет искать решения с увеличенным числом нарушений, что, как следствие, приводит к снижению объективных значений, обусловленных минимальным количеством запланированных заданий, способствующих вышеуказанным показателям производительности. В этом исследовании мы управляли нарушениями, сводя к минимуму общее количество нарушений.

С нашей точки зрения, это не полностью удовлетворит наши потребности. Мы предлагаем вместо того, чтобы считать общее количество нарушений объективным, выбрать более сложную объективную меру; общее время нарушения. Мы определяем общее время нарушения рассчитывается как сумма всех тайм-аутов нарушения. Идея заключается в том, чтобы различать мягкие и жесткие нарушения, те работы, которые немного отстали от срока, и те, которые явно задержались. Мы хотели бы проиллюстрировать желаемый эффект на простом примере. Предполагая, что нам нужно выбирать между двумя решениями, первое показывает два нарушения с мягким ограничением по времени, а второе страдает от одного задания с нарушением с ограничением по времени. С точки зрения практиков, мы однозначно выбрали бы решение с двумя мягкими нарушениями, потому что чем больше превышен срок, тем больше проблем с качеством. В непосредственном смысле такое разграничение становится возможным при общем времени нарушения в качестве цели, в отличие от общего числа нарушений. Последовательно следуя этой идее и учитывая тот факт, что мы часто имеем дело с приоритетами с точки зрения взвешенных заданий, мы пришли к выводу, что общее взвешенное время нарушения должно быть важной частью целевой функции.

**5.3.2 Ограничение минимального размера пакета**

Ограничения минимального размера пакета представляют нижний порог для количества пластин или заданий в пакете. Они обычно устанавливаются по трем причинам: а) как требование, гарантирующее стабильность процесса, б) как более низкий порог, который предотвращает слишком мелкие партии, обусловленные экономическими проблемами, в) как плановое значение, нацеленное на производительность завода в соответствии с теорией работы управление кривой (OCM).

Чтобы гарантировать стабильность процесса, часто требуется поддерживать постоянным количество субстратов (пластин) внутри реактора на всех участках. Обычно реактор полностью заполнен при максимальной мощности пластинами. В тех случаях, когда предлагаемая партия насчитывает меньше пластин, чем максимум реактора, оставленные свободными пазы реактора заполняются непроизводительной пластиной. В этом случае эти непроизводительные пластины имеют только одно намерение: гарантировать требуемую теплоемкость внутри реактора во время процесса. В машине всегда хранится определенное количество непроизводительных пластин для быстрого доступа без значительной задержки. Если этого недостаточно, необходимо дополнительно запросить непроизводительные пластины, которые будут доставлены в виде непроизводительных партий откуда-то на фабрике.

В этом случае минимальный размер партии как требование, которое намеревается предотвратить процесс, запускается с нежелательным соотношением между производительными и непроизводительными пластинами из-за экономических проблем. Другая причина может заключаться в недостаточной системе поставок непроизводительных пластин. Подводя итог, по некоторым причинам может потребоваться создание пакетов, в которых учитывается минимальное количество пластин в сумме, предоставляемой их партнерами по пакетам.

Однако в ходе наших исследований мы пришли к выводу, что ограничение минимального размера партии редко является необходимым ограничением из-за упомянутых выше фактов, скорее оно используется в качестве значения планирования и контроля. Мы настоятельно рекомендуем проводить различие между пороговыми значениями минимального размера партии, которые считаются необходимыми ограничениями, и запланированными размерами партии, исходя из соображений производительности. Мы поняли, что запланированные размеры партий часто объявляются минимальными размерами партий, что не выгодно для задач планирования. Это правда, что диспетчерским системам и даже системам планирования без какой-либо прогнозной информации о входящих заданиях из восходящего потока действительно необходим запланированный размер пакета, рассматриваемый как минимальный порог. В зависимости от уровня использования существует оптимальный размер пакета, который не должен превышаться диспетчерскими действиями. В отличие от этого, если система диспетчеризации или планирования располагает информацией о будущих поступлениях на работу, больше нет необходимости в планируемых размерах партий.

На этом этапе мы хотели бы обсудить проблемную ситуацию, которая может быть описана как дилемма минимального размера пакета для оптимизации расписания партии. Мы воспринимаем это как факт, что существует ограничение минимального размера пакета, которое предназначено для развития или предотвращения определенной ситуации. Принимая как должное, что весь объем заданий должен планироваться без исключений, и принимая во внимание ограничения минимального размера пакета, существуют проблемы с расписанием, для которых не существует действительного решения. Такие проблемные случаи страдают от ряда незапланированных заданий, которые не вписываются ни в одну из существующих партий и не соответствуют ограничению минимального размера пакета. Чтобы гарантировать правильное решение в любое время и для любой проблемы, необходимо ослабить предпосылку, которая требует нахождения расписания без каких-либо внеплановых заданий. Следовательно, мы вынуждены разрешить работу, которая остается незапланированной. И как следствие этого, необходимо рассматривать количество внеплановых как дополнительную цель. Потому что, если нет, процедура оптимизации будет искать решения с увеличенным количеством внеплановых заданий, чтобы уменьшить общие цели, такие как время цикла или опоздания. И это точка, где возникает дилемма. Принимая количество незапланированных заданий как часть целевой функции, система оптимизации в первую очередь стремится к небольшому количеству незапланированных заданий и, во-вторых, улучшает цели, на которых мы на самом деле ориентируемся.

**6 ВЫВОДЫ**

Наши эксперименты показывают, что граница сложности для детерминированной процедуры локального поиска, которая может продолжаться 10 минут, выходит за рамки проблемных случаев с 10 машинами и 300 заданиями. Мы хотим сказать, что наша реализация VND чаще всего достигала оптимума в течение 10 минут для экземпляра проблемы M (10 | 300). Как показано в предыдущих исследованиях, представленные эксперименты подтверждают, что, к сожалению, PBMS без динамического поступления на работу показывает лишь немного лучшие результаты по сравнению с диспетчеризацией. Это связано с тем, что преимущества пакетного планирования в значительной степени зависят от прогнозов прибытия на работу. Как общее утверждение, мы можем сказать, что мы получаем улучшение на 5% для общих показателей эффективности. Мы наблюдали 6% -ное улучшение в рабочем времени, 4% -ное улучшение среднего времени цикла и 2% -ное улучшение среднего опоздания. Существование критических ограничений, временных ограничений и / или минимального размера пакета требует многоцелевых функций, соответствующим образом учитывающих ситуации с незапланированными заданиями, вызванными нарушением ограничений. В этом случае иерархические структуры поддаются для многоцелевых функций.

Что касается минимального размера пакеты, мы заявляем, что для планирования партии обычно требуются ограничения минимального размера пакета, если нет информации о предварительном просмотре. Чтобы гарантировать действительное расписание в любое время, ограничения минимального размера пакета позволяют разрешить ряд незапланированных заданий. Для разрешения незапланированных заданий требуется целевая функция, которая в значительной степени учитывает незапланированные задания. Наконец, внеплановые задания как часть целевой функции потенциально снижают те улучшения показателей производительности, которые нас действительно интересуют. Это то, что мы называем дилеммой минимального размера пакета для планирования.

Мы также подчеркиваем, что помимо проблем производительности соответствующая целевая функция, тщательно учитывающая важные потребности, будет определять успех системы оперативного планирования в долгосрочной перспективе. Для PBMSP это означает, что мы объединили различные меры в целевую функцию. Помимо общих локальных целей охвата, обычно ориентированных на взвешенное время цикла и взвешенное опоздание, важно учитывать влияние локально оптимизированных решений на последующие операции. Это имеет большое значение для того, чтобы сохранить улучшения в течение следующих операций. Поскольку PBMSP в оптимальных условиях полагаются на прогнозы прибытия на работу, которые всегда сопровождаются ошибками, представляется необходимым включить ошибки прогнозирования в процесс принятия решений как часть многоцелевой функции.

**БЛАГОДАРНОСТЬ**

Мы хотели бы поблагодарить наших колледжей от Infineon Technologies за их поддержку в любое время. Особая благодарность выражается Йенсу Долешалу и Манфреду Бенешу из Дрезденского университета наук, которые поддерживают сервер и системы, необходимые для наших экспериментов, со всей их компетентностью и преданностью. Эта работа была поддержана грантом 13N11588 Федерального министерства образования и науки Германии.