doi: 10.3969/j.issn.1007-7375.240100

基于门控循环单元强化学习的晶圆光刻区实时调度方法研究

吴立辉[1]，石津铭[1]，金克山[1]，张　洁[2]

(1. 上海应用技术大学 机械工程学院，上海 201400；2. 东华大学 人工智能研究院，上海 201620)

摘要: 为求解具有动态性、实时性、多约束、多目标特点的晶圆光刻区调度问题，提出一种基于门控循环单元强化学习的晶圆光刻区实时调度方法。设计引入门控循环单元学习光刻区历史调度决策与状态的时序信息，为双深度强化学习模型提供辅助决策信息；设计双深度强化学习模型的输入状态空间、输出动作集，并面向晶圆最小化最大完工时间和晶圆准时交货率指标设计多目标奖励函数，为智能体优化调度输出；设计设备专用性约束与掩模版约束的解约束规则与调度方法相结合，提高调度方案实施的实用性。通过某晶圆制造企业实际算例，将该方法与传统双深度强化学习和光刻区启发式规则方法比较，该方法均为最优，证明了其解决此问题的有效性。

关键词: 晶圆制造系统；光刻区调度；深度强化学习；门控循环单元 (GRU)；多目标

中图分类号: F426.4; TP391　　**文献标志码:** A　　**文章编号:** 1007-7375(2024)03-0012-10

Real-time Scheduling of Wafer Photolithography Area Based on Reinforcement Learning with Gated Recurrent Unit

WU Lihui[1], SHI Jinming[1], JIN Keshan[1], ZHANG Jie[2]

(1. School of Mechanical Engineering, Shanghai Institute of Technology, Shanghai 201400, China;
2. Institute of Artificial Intelligence, Donghua University, Shanghai 201620, China)

Abstract: To address the scheduling problem of wafer photolithography area, characterized by dynamic nature, real-time requirements, multiple constraints, and multiple objectives, a real-time scheduling method based on gated recurrent unit (GRU) reinforcement learning is proposed. This method incorporates GRU to learn the temporal information of historical scheduling decisions and states in the photolithography area, providing auxiliary decision-making information for the double deep reinforcement learning (DDRL) model. The input state space and output action set of the DDRL model are designed, and a multi-objective reward function is established with the objective of minimizing the maximum completion time of wafers and maximizing the on-time delivery rate, optimizing the scheduling output by intelligent agents. Additionally, constraint relaxation rules and scheduling methods are proposed combining equipment-specific constraints and mask constraints, to enhance the practicality of scheduling strategies. Through empirical evaluation using real-world cases from a wafer manufacturing enterprise, this method is compared with traditional double deep reinforcement learning and heuristic rule methods for photolithography area, demonstrating its superiority and verifying its effectiveness in solving this problem.

Key words: wafer fabrication system; scheduling of photolithography area; deep reinforcement learning; gated recurrent unit (GRU); multi-objective

晶圆制造光刻区因光刻设备昂贵、在制品规模大、重入加工次数多等特点，成为晶圆制造系统的主要瓶颈[1]。对光刻区进行调度，能有效改善光刻区整体运行性能[1]，对提高晶圆制造企业效益等具有十分重要意义。

光刻区调度是指将从氧化、薄膜等工艺区搬运来的晶圆 lot 合理配套掩模版并指派到合适光刻机上加工的过程。因光刻区生产过程受晶圆 lot 动态

收稿日期: 2024-03-19

基金项目: 国家重点研发资助项目 (2022YFB3305003)；上海应用技术大学引进人才科研启动项目 (YJ2022-33)

作者简介: 吴立辉 (1981—)，男，湖南省人，副教授，博士，主要研究方向为智能制造、复杂制造系统调度。

到达、优先级变化等动态因素影响，为提高光刻设备利用率、减少空闲等待时间，光刻区调度需要具有动态性与实时性的特点；同时因光刻区加工工艺要求，调度过程存在设备专属性约束、掩模版约束及序列相关准备时间约束等多约束特点；晶圆光刻区调度通常需综合考虑晶圆完工时间、准时交货率等晶圆制造系统的核心性能指标，因此光刻区调度又具有多目标的特点。

近年来，国内外学者针对光刻区调度问题展开大量研究，主要集中于启发式调度规则[2]、数学规划方法[3-4]与群智能算法[1,5-6]，且以周期性调度为主。文献[3-4]采用的数学规划与文献[1,5-6]采用的群智能算法均能获得小规模问题的最优解，但在大规模问题上求解时间增加，调度实时性较差。文献[2]采用的启发式规则求解速度快，实时性较强，但其在动态变化环境中适应性较差。

为实现动态、实时性调度需求，文献[7-8]采用的 R-Learning 与 Q-Learning 强化学习方法将实时调度转化为连续强化学习决策问题，能够在小规模环境变化下实时选择最优调度策略，满足动态变化下实时性调度需求。然而，类似的强化学习方法在光刻区大规模状态下容易引起状态空间规模爆炸，导致模型的学习效率降低，学习效果变差。双深度强化学习 (double deep Q-network, DDQN) [9]算法将深度神经网络与强化学习相结合，能够处理大规模、高维复杂状态输入，更适用于大规模光刻区调度问题，动态适应性优于启发式规则。相较于普通 DQN (deep Q-network, DQN) ，该方法有效解决了过估计问题[10]，学习效率高，求解稳定性好。Liu 等[11]与 Wang 等[12]采用 DDQN 方法解决动态到达的柔性作业车间实时调度问题。Lee 等[13]与 Paeng 等[14]将 DDQN 方法应用于非等效并行机调度问题，均能在不同场景下取得较优的效果。

然而，光刻区调度决策过程及其状态变化过程具有显著的时间序列特征。在这个过程中，当前调度决策受到当前状态信息、上一时刻的调度动作以及上一时刻的状态信息的影响，以上研究往往忽略这些因素。此外，光刻区设备专属性和掩模版等约束限制的考虑不够充分，导致方法缺乏实用性。为了解决这些问题，本文提出了基于门控循环单元强化学习 (double deep Q-network with gated recurrent unit, GDDQN) 的光刻区实时调度方法。鉴于门控循环单元 (gated recurrent unit, GRU) 是一种适合处理时间序列数据的循环神经网络变体[15-16]，相较于其他变体如长短期记忆网络 (long short-term memory, LSTM) ，结构简单，参数更少，因此训练和调优效率更高，更加契合光刻区实时调度环境。该方法具有以下特点：1) 将光刻区前一时刻调度决策动作与状态信息引入 GRU 进行编码、记忆与处理，提取前一时刻调度与状态的关键信息，为当前时刻 DDQN 调度网络提供辅助信息进行优化决策，从而提高 DDQN 调度优化效果；2) 设计解约束规则将光刻区的设备专属性约束、掩模版约束与 GDDQN 相结合，以提高调度方案的实用性；3) 围绕最小化最大完工时间和晶圆准时交货率目标设计奖励函数，以实现 GDDQN 方法的多目标性能优化。

1　问题与模型

1.1　问题描述

晶圆制造光刻区内通常有数台至十几台并行光刻机。由于各光刻机台类型不同，工艺能力不同，加工精度各异，加工效率差异等，导致光刻区设备加工具有典型的非等效特性。从长周期来看，具有多约束、多目标等特点的光刻区调度问题是典型的非等效并行机调度问题，是 NP-hard 难题。在光刻区加工过程中，为充分发挥宝贵的加工资源，减少光刻机空载等待时间，光刻区调度需实时响应加工区的随机动态事件，并进行快速优化求解。因此，晶圆光刻区调度问题亦是动态因素影响环境下的具有多约束、多目标的马尔可夫贯序决策问题。对该问题进行实时优化求解，可在满足机台利用率要求的基础上，减小光刻区的晶圆完工时间并提高晶圆准时交货率等性能指标，对提高晶圆制造企业效益等具有十分重要意义。

1.2　符号变量定义

对光刻区调度问题进行建模，定义 J 为待光刻的晶圆的集合，$J=\{1,2,\cdots,j,\cdots,n_j\}$；$M$ 为所有光刻机集合，$M=\{1,2,\cdots,m,\cdots,n_m\}$，$M_j$ 为晶圆 j 当前可用的光刻机集合；f 为晶圆类型编号，l 为晶圆类型总量，$f=1,2,\cdots,l$；w_{jf} 为当前晶圆 j 所属的权重；t_{aj} 为晶圆 j 的到达时间；t_{bj} 为晶圆 j 的开始处理时间；t_{oj} 为晶圆 j 的调整时间；t_j 为晶圆 j 光刻前所有晶圆的完工时间；当一台光刻机处理上一个晶圆 i 与下一个待处理的晶圆 j 为同一加工工艺 ϕ_{ij} 为 1，否则为 0；p_{jm} 为晶圆 j 在光刻机 m 上的光刻时间；

p_{ejfm}为晶圆j在光刻机m上的预期光刻时间；v_j为晶圆j的专用设备；y_1、y_2分别为两个优化指标。

定义a_x为智能体的可选动作，其中，$x=1,2,3,\cdots,n$；A为一个无限大的正数；d_j为晶圆光刻处理的交货时间，交货时间定义为式 (1)，其中，K为影响因子，用于控制晶圆交货时间的紧急程度；Z_{jm}为晶圆j在光刻机m上的完工时间；当晶圆的完成时间在其交货时间之后，晶圆就会出现拖期时间T_j，J中的晶圆j的拖期时间定义为式 (2)。

$$d_j = t_{aj} + Kp_{ejfm}; \tag{1}$$

$$T_j = \max(Z_{jm} - d_j, 0)。 \tag{2}$$

定义决策变量如下。若晶圆j在晶圆i之前在光刻机上加工，则x_{ijm}为 1，否则，为 0，$i,j \in J,\ m \in M$；若晶圆j在光刻机m上加工，则x_{jm}为 1，否则，为 0，$j \in J,\ m \in M$。

1.3 目标函数与约束

以最小化晶圆最大完工时间以及最大化晶圆准时交货率为优化目标，构建目标函数模型如式 (3)、(4) 所示。

$$\min\ y_1 = \max(t_{aj} + p_{jm} + t_{oj})。 \tag{3}$$

$$\begin{cases} \max\ y_2 = \dfrac{n_j - \sum\limits_{j\in J} U(T_j)}{n_j}。 \\ U(X) = \begin{cases} 1, & X > 0; \\ 0, & X = 0。 \end{cases} \end{cases} \tag{4}$$

其中，y_1表示最大完工时间；y_2表示晶圆准时交货率。相关约束定义如式 (5) ~ (11) 所示。

$$\sum_{j\in J, j\neq i}\sum_{m\in M} x_{ijm} = 1,\ \forall i \in J,\ i \neq 0; \tag{5}$$

$$\sum_{j\in J, j\neq i}\sum_{m\in M} x_{jim} = 1,\ \forall i \in J,\ i \neq 0; \tag{6}$$

$$\sum_{j\in J, j\neq i}\sum_{m\in M, m\notin M_j} x_{jim} = 0,\ \forall i \in J,\ i \neq 0; \tag{7}$$

$$t_{bj} + p_{jm}x_{jm} = Z_{jm},\ \forall j \in J,\ j \neq 0,\ m \in M; \tag{8}$$

$$\sum_{j\in J, j\neq i} x_{ijv_i} = 1,\ \forall i \in J,\ i \neq 0,\ v_i \neq 0; \tag{9}$$

$$t_i \geqslant t_{ai} + \phi_{ij}t_{oi} + p_{im} + \sum_{j\in J, i\neq j}(\max(t_j - t_{ai}, 0))x_{ijm},$$
$$\forall i \in J,\ i \neq 0,\ \forall m \in M; \tag{10}$$

$$t_{bj} + A(1 - x_{jm}) \geqslant t_{aj},\ \forall j, m; \tag{11}$$

$$x_{ijm} \in \{0,1\}; \tag{12}$$

$$x_{jm} \in \{0,1\}。 \tag{13}$$

式 (5)、(6) 表示每台光刻机同一时间只能处理一个晶圆 lot，而且紧前和紧后也只有一个晶圆 lot；式 (7) 表示当前的晶圆 lot 只能被分配到可以加工当前晶圆 lot 的其中一台光刻机上；式 (8) 表示每个晶圆 lot 光刻加工完成时间等于开始处理时间与处理时间之和；式 (9) 代表晶圆的关键层必须在同一台光刻机上加工；式 (10) 代表每个晶圆 lot 只有在到达且所需掩模版可用之后才可以开始加工；式 (11) 代表光刻机中每个晶圆 lot 的开始处理时间大于或等于晶圆 lot 的到达时间；式 (12) 和 (13) 均为决策变量取值。

2 基于 GDDQN 的晶圆光刻区实时调度方法

针对具有动态性、实时性、多约束、多目标特点的晶圆光刻区调度问题，设计基于 GDDQN 的晶圆光刻区实时调度方法如图 1 所示。该方法包括 GRU 模块、DDQN 模块及光刻区约束处理模块。图 1 中的红色版块表示 GRU 模块在经过时序关键信息提取后，向 DDQN 模块输出的状态时序变化信息。GRU 模块考虑光刻区调度决策过程受过去状态、动作变化影响的特性，设计引入 GRU 网络提取前一时刻关键信息并输出至 DDQN 网络，为当前时刻 DDQN 调度网络提供辅助信息优化决策。DDQN 模块从光刻区生产环境获取晶圆、光刻机及光刻区状态特征变量，并从 GRU 模块获取历史决策与状态的时序信息，构建 DDQN 网络模型进行实时调度决策规则输出，实现光刻区加工过程的最小化晶圆完工时间与最大化准时交货目标性能优化。DDQN 网络由两个参数相同的 Q 网络构成，采用优先经验回放机制训练，利用损失函数计算预测值与真实值之间的差异并更新网络参数[9]。针对稀疏奖励问题，设计面向多目标的奖励函数指导 DDQN 智能体决策。光刻区约束处理模块针对设备专属性约束、掩模版约束构建解约束处理流程，以提高调度方案的实用性。GRU 模块、DDQN 状态空间、动作空间、奖励函数及光刻区约束处理模块具体设计如下。

2.1 GRU 模块

在光刻区实时调度过程中，存在当前时刻的决策与环境状态会受到前一时刻影响的特性。因此本文设计引入 GRU 模块提取前一时刻光刻区调度的

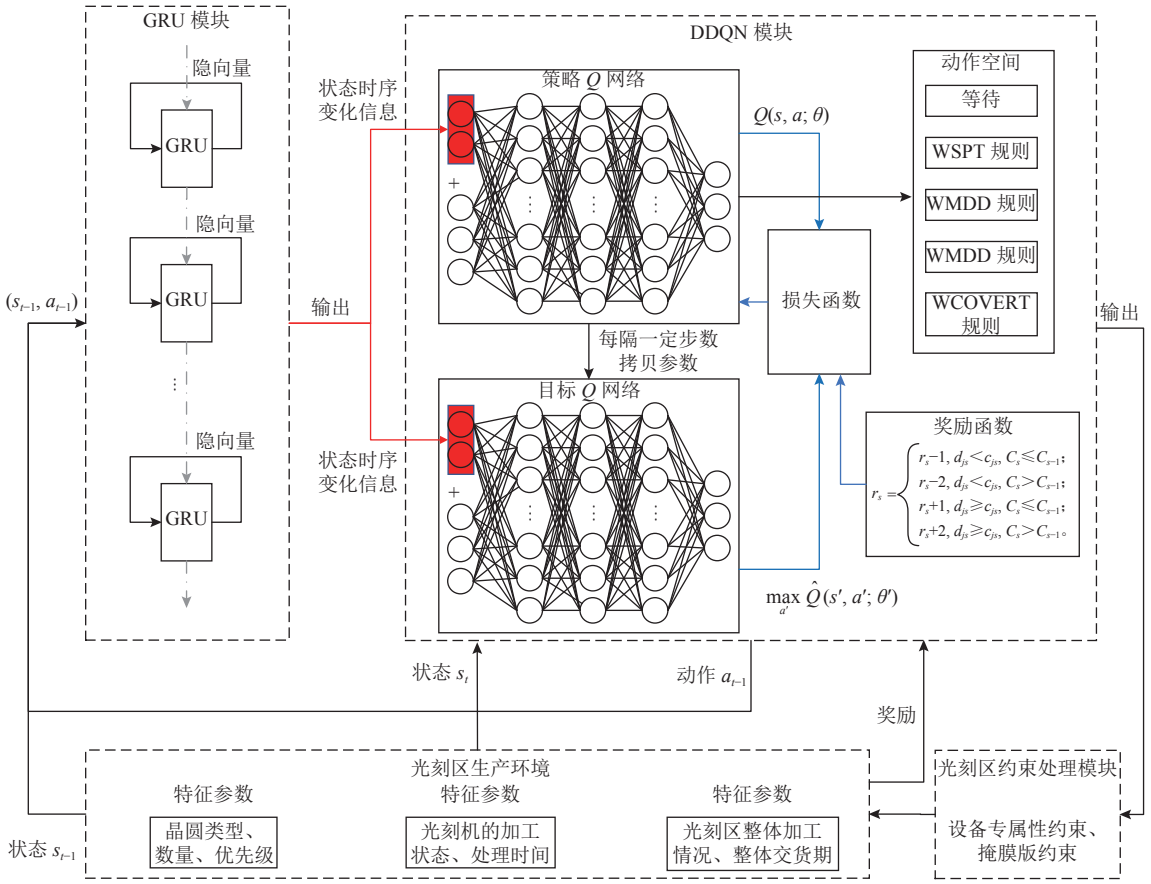

图 1　基于 GDDQN 的晶圆光刻区实时调度方法框架

Figure 1　Framework of GDDQN-based real-time scheduling method for wafer lithography area

关键信息输出给 DDQN 模块，为当前时刻调度提供辅助信息，提高调度优化效果。通过门控机制控制输出并允许内部隐藏层之间的传递，使得 GRU 在处理时间序列数据时效果较优[16]。GRU 内部的计算模型如式 (14) ~ (17) 所示。

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]); \tag{14}$$

$$u_t = \sigma(W_u \cdot [h_{t-1}, x_t]); \tag{15}$$

$$\tilde{h}_t = \tanh(W_h \cdot [u_t h_{t-1}, x_t]); \tag{16}$$

$$h_t = (1 - z_t) h_{t-1} + \tilde{h}_t z_t。 \tag{17}$$

其中，h_{t-1}是隐藏层输入信息，表示记忆信息；x_t是当前输入信息；z_t为更新门；u_t为重置门；$\tilde{h}_t$为输入信息x_t与隐藏层信息h_{t-1}融合所得；W_z、W_u与W_h分别表示其对应的权重。GRU 利用内部两个门控单元可以将当前输入样本与过去输入样本进行选择性记忆与遗忘，从而获取输入样本之间的时序相关性。

GRU 模块由 GRU 单元组成循环网络，将 GRU 网络输出层与 DDQN 网络的输入层相连，使得 GRU 为 DDQN 网络的前馈网络。GRU 网络的输入为上一调度时刻对应的状态及调度决策，输出为学习获得的时序信息G_k，可表示为

$$G_k = \mathrm{GRU}([s_{k-1}, a_{k-1}], h_{k-2}; \lambda)。 \tag{18}$$

其中，$[s_{k-1}, a_{k-1}]$为进行$k-1$次调度的车间状态s_{k-1}与智能体动作a_{k-1}组成的向量；h_{k-2}为 GRU 网络内部的隐状态；λ为 GRU 网络参数。考虑到 GRU 为 DDQN 的前馈网络，在 DDQN 网络更新阶段将相应状态及决策信息回传至 GRU 网络并更新网络参数，实现整体网络优化。

2.2　状态描述

DDQN 模块的输入为当前环境状态信息以及 GRU 模块输出的光刻区时序信息，因此将 GRU 的输出G_k与光刻区环境状态共同作为智能体的状态参数。考虑到光刻区当前晶圆、光刻机及光刻区状态等对光刻区调度的影响，面向 DDQN 智能体，设计 10 种特征状态空间，如表 1 所示。其中，$\boldsymbol{F}_{1,f}$、$\boldsymbol{F}_{5,f}$、$\boldsymbol{F}_{6,f}$、$\boldsymbol{F}_{7,f}$表示晶圆属性特征向量，$\boldsymbol{F}_{2,f}$、$\boldsymbol{F}_{3,f}$、$\boldsymbol{F}_{4,f}$表示设备属性特征向量、$\boldsymbol{F}_{8,f}$、$\boldsymbol{F}_{9,f}$表示优化目标特征向量；$\boldsymbol{F}_{10,f}$为 GRU 网络输出的状态时序向

表 1 状态特征列表

Table 1 State characteristics

序号	状态特征公式	参数含义	状态特征描述
1	$\boldsymbol{F}_{1,f}=\begin{cases}0, & N_f=0;\\ 2^{-1/N_f}, & N_f>0。\end{cases}$	N_f为当前晶圆类型为f的等待调度的晶圆数量	表示光刻区缓冲区中各晶圆类型的晶圆数量情况
2	$\boldsymbol{F}_{2,f}=\begin{cases}0, & x_m=1;\\ f/A, & x_m=0。\end{cases}$	A为一个常数；x_m为一个决策变量，如果当前光刻机m空闲，即为 1，否则，为 0	表示当前各光刻机上正在加工的晶圆类型
3	$\boldsymbol{F}_{3,f}=\begin{cases}\dfrac{T_{fm}-t}{2\sum\limits_{m\in M}p_{jm}}, & x_m=1;\\ 0, & x_m=0。\end{cases}$	T_{fm}为当前机器m上处理晶圆类型为f的计划完成时间；t为当前系统时间	表示各光刻机剩余加工时间情况
4	$\boldsymbol{F}_{4,f}=\begin{cases}\dfrac{d_{fm}-t}{2\sum\limits_{m\in M}p_{jm}}, & x_m=1;\\ 0, & x_m=0。\end{cases}$	d_{fm}为当前机器m上处理晶圆类型为f的交货时间	表示各光刻机加工晶圆的剩余交货期情况
5	$\boldsymbol{F}_{5,f}=\dfrac{\min D_{fm}-t}{p_{\mathrm{e}jfm}}$	D_{fm}表示当前机器m上要处理晶圆类型为f的交货时间队列；$\min D_{fm}$表示取队列中的最小值	表示f类晶圆类型具有最短交货期晶圆所对应紧急程度系数
6	$\boldsymbol{F}_{6,f}=\dfrac{\max D_{fm}-t}{p_{\mathrm{e}jfm}}$	$\max D_{fm}$表示取队列中的最大值	表示f类晶圆类型具有最大交货期晶圆所对应紧急程度系数
7	$\boldsymbol{F}_{7,f}=\dfrac{\frac{1}{N_f}\sum\limits_{i=1}^{N_f}D_{fm}-t}{p_{\mathrm{e}jfm}}$	$\dfrac{1}{N_f}\sum\limits_{i=1}^{N_f}D_{fm}$表示取队列的均值	表示f类晶圆类型的晶圆所对应平均交货期紧急程度系数
8	$g=\begin{cases}1,\left(D_{fm}-t\right)\in\left(\max\limits_{1\leqslant i\leqslant m}p_{jmi},+\infty\right);\\ 2,\left(D_{fm}-t\right)\in\left(\min\limits_{1\leqslant i\leqslant m}p_{jmi},\max\limits_{1\leqslant i\leqslant m}p_{jmi}\right);\\ 3,\left(D_{fm}-t\right)\in\left(0,\min\limits_{1\leqslant i\leqslant m}p_{jmi}\right);\\ 4,\left(D_{fm}-t\right)\in(-\infty,0)。\end{cases}$ $\boldsymbol{F}_{8,f}=\begin{cases}0, & \mathrm{num}(g)=0;\\ \mathrm{num}(g), & \mathrm{num}(g)>0。\end{cases}$	$\max\limits_{1\leqslant i\leqslant m}p_{jmi}$表示队列中晶圆在当前光刻机上正在光刻的时间的最大值；同理，$\min\limits_{1\leqslant i\leqslant m}p_{jmi}$表示最小值;$\mathrm{num}(g)$表示缓冲区排队晶圆剩余交货期落在间隔$g$的个数	表示光刻区缓冲区中f类晶圆整体交货期的分布情况
9	$\boldsymbol{F}_{9,f}=\dfrac{t_m}{\sum\limits_{m=1}^{n_m}t_m}$	t_m为当前机器m正在处理晶圆的预计完工时间	表示当前车间完工时间分布情况
10	$\boldsymbol{F}_{10,f}=G_k$	G_k见式 (18)	表示 GRU 网络输出

量。这 10 种特征向量共同组成状态空间输入该方法智能体，通过对状态参数的观测分析输出决策动作。

2.3 动作集

基于规则的调度方法虽然在固定场景下实时性与优化求解能力较好，但任何一种规则都无法适应动态变化的调度场景[17]。因此，本研究设计采用 4 种在晶圆最大完工时间和晶圆准时交货率目标上表现较优的调度规则，以及不选择任何规则的等待动作组成 DDQN 智能体的动作集合，实现 DDQN 面对光刻区状态变化作出调整，选择合理的规则输出。动作集描述如表 2 所示。

为智能体设计动作时考虑了以下 3 种触发场景：1) 当前状态下有多台空闲的光刻机，此时晶圆 lot 随机到达，智能体对该晶圆 lot 选择合适的光刻机与掩模版；2) 当前状态下仅有一台空闲的光刻机，排队队列中等待的晶圆 lot 数量大于 1，智能体会对该光刻机选择合适的晶圆 lot 以及齐套掩模版；3) 其余情况智能体不做任何调度，执行等待。

2.4 奖励函数

强化学习的奖励设计包含即时奖励和累计奖励。本研究需同时优化晶圆最大完工时间和晶圆准时交货率两个目标，但考虑到 DDQN 智能体完成所有调度任务才能获取全局目标奖励，无法对智能体单步决策作出指导，因此存在稀疏奖励问题。故本研究将最大完工时间这一指标细分为每一单步状态下的平均最大完工时间指标，并将其作为准时交货率指标的约束，与单步状态下晶圆的交货情况共同作为智能体的即时奖励。即时奖励设计为

表 2　DDQN 动作集

Table 2　DDQN action set

序号	动作	数学模型	描述
1	等待	a_0	智能体等待
2	WSPT (weighted shortest processing time)	p_{jm}/w_{jf}	选择带权总加工时间最短的工件或机台
3	WMDD (weighted modified due date)	$\dfrac{\max\{p_{jm}, d_j - t\}}{w_{jf}}$	根据交货时间与处理时间的比例选择工件或机台
4	WCOVERT (weighted cost over time)	$\dfrac{w_{jf}}{p_{jm}}\left[1-\dfrac{\left(d_j-p_{jm}-t\right)^+}{\eta p_{jm}}\right]^+$	根据拖期时间与处理时间最大比率选择工件或机台，η为 WCOVERT 规则中的调整系数
5	ATC (apparent tardiness cost)	$\dfrac{w_{jf}}{p_{jm}}\exp\left[-\dfrac{\left(d_j-p_{jm}-t\right)^+}{\mu p_{jm}}\right]$	根据拖期成本选择工件或机台，μ为 ATC 规则中的调整系数

$$r_s=\begin{cases} r_s-1, & d_{js}<c_{js},\ C_s\leqslant C_{s-1};\\ r_s-2, & d_{js}<c_{js},\ C_s>C_{s-1};\\ r_s+1, & d_{js}\geqslant c_{js},\ C_s\leqslant C_{s-1};\\ r_s+2, & d_{js}\geqslant c_{js},\ C_s>C_{s-1}。\end{cases} \tag{19}$$

其中，r_s表示在阶段s下智能体完成相应动作之后获得的即时奖励；d_{js}表示交货时间；c_{js}表示光刻完成时间，晶圆j的交货时间只要小于光刻完成时间即产生拖期时间，晶圆无法准时交货；C_s与C_{s-1}分别表示在阶段s与$s-1$下车间的平均最大完工时间。C_s的计算式为

$$C_s=\max(t_{ajs}+p_{jms}+t_{ojs})/n_s,\ \forall j\in J,\forall m\in M。 \tag{20}$$

其中，t_{ajs}、p_{jms}、t_{ojs}分别表示阶段s下晶圆j的到达时间、处理时间与包含序列切换时间和掩模版调度时间的调整时间；n_s则表示已完成的晶圆数。

智能体在即时奖励函数下，学习实现单步最大化晶圆准时交货率的同时考虑最小化完工时间，从而累计获得以上目标的全局最大奖励。因此通过即时奖励的累加处理，构建累计奖励函数为

$$R_s=\sum_{j=1}^{n_j} r_s。 \tag{21}$$

2.5　光刻区约束处理模块

在设备专属性约束与掩模版约束影响下，为提高该方法在光刻区调度上的实用性，设计解约束规则与 GDDQN 算法结合，实现多约束的处理。解约束流程图如图 2 所示。

针对设备专属性约束，在初始化光刻区调度环境时，每一种晶圆类型会初始化可用机台编号，以 3 种晶圆类型为例，表 3 展示了对应的专用机台。1) 晶圆 lot 随机到达时，智能体根据表 3 判定选出该晶圆候选空闲状态专用机台集合，并在该集合上进行后续调度任务，若集合为空，智能体执行等待；2) 光刻机空闲时，智能体根据表 3 判定选出该机台专用晶圆 lot 集合，并在该集合上进行后续调度任务，若集合为空，智能体执行等待。

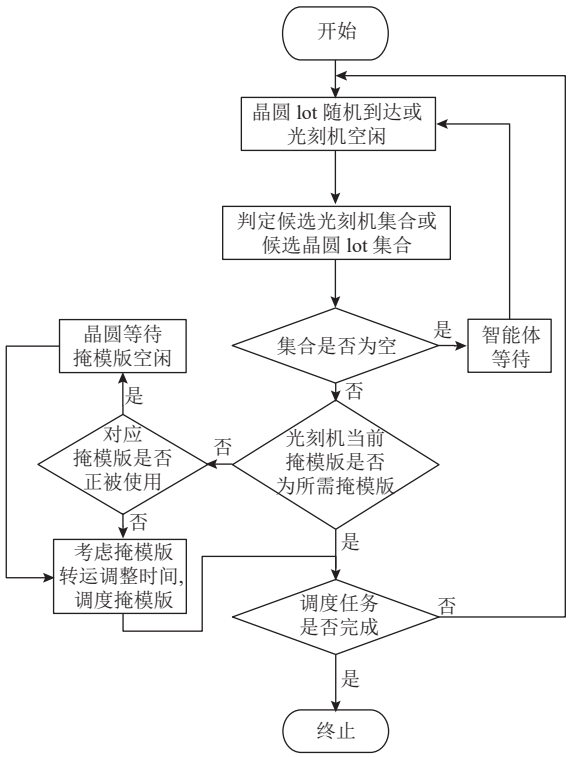

图 2　考虑设备专属性、掩模版约束的解约束流程图

Figure 2　Flowchart of constraint relaxation considering equipment specificity and mask constraints

针对掩模版约束，假设每台光刻机上有且仅有一个专用掩模版，不同晶圆类型对应两种不同的专用掩模版。初始化光刻区环境时，不同晶圆类型初始化专用掩模版编号，如表 3 所示，并在后续调度针对掩模版约束进行判定。智能体基于调度规则选出合适晶圆 lot 或光刻机时，若当前掩模版与所需掩模版相同，则无需考虑掩模版转运与调整时间。

表 3　3 种晶圆类型专用光刻机与掩模版示例表

Table 3　Examples of dedicated photolithography machines and masks for 3 types of wafers

晶圆类型编号	专用光刻机编号	专用掩模版编号
2	1,2,3,4,5	1,5
5	1,2,3,5,6,7	1,6
1	3,4,6,7,8	7,8

如果当前掩模版与所需掩模版不同，则需要判断该掩模版是否正在被使用。若正在被使用，则需要等待该掩模版空闲状态后才可进行掩模版调度；若此时掩模版处于空闲状态，则直接将掩模版转运到该机台上并完成调整工作。

2.6　基于 GDDQN 的光刻区实时调度方法流程

基于 GDDQN 的晶圆光刻区实时调度方法流程如图 3 所示。具体步骤如下。1) 初始化光刻区调度环境状态。2) 当晶圆 lot 随机到达或光刻机空闲，根据当前光刻区状态信息等触发 GRU 与 DDQN 智能体 (DDQN 算法伪代码如下所示)，DDQN 智能体决策输出优化调度规则。3) 基于调度规则选择晶圆 lot 或光刻机，在此过程中，对晶圆 lot 的设备专属性约束、掩模版的齐套性约束进行判定。4) 执行当前调度任务，安排晶圆在相应光刻机上加工。5) 对调度任务完成情况进行判定，如未完成，返回步骤 2；否则调度方法结束。

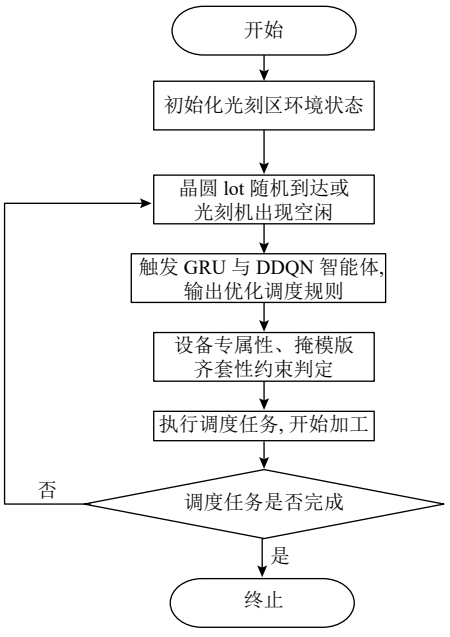

图 3　基于 GDDQN 光刻区实时调度流程图

Figure 3　Flowchart of real-time scheduling for lithography area based on GDDQN

DDQN 算法伪代码

初始化：最小训练批量 mini_batch，经验回放池D，初始化 Q 网络的网络参数θ、目标网络的参数θ'并将θ复制给θ'，初始化 GRU 网络参数λ，目标网络的更新周期N^-，决策步数 step；
　for 回合数 $e \in \{1,2,\cdots,E_m\}$ do
初始化调度状态s_0,a_0,GRU 输入(s_0,a_0)；
for t=1，最大时间步 do
　根据贪婪策略从可选动作集中选择动作a_t；
　执行动作a_t，环境与智能体交互获得奖励r_t与下一个状态s_{t+1}；
　观察所得奖励r_t与下一个状态s_{t+1}，判断所有调度是否已完成；
　将当前的(s_t,a_t,s_{t+1},r_t)存储至经验回放池D；
　if 经验回放池的容量到达既定值 then
　　根据优先经验回放抽取概率最大的小批量 mini_batch 样本进行训练，计算每个样本的 TD 目标差，并执行梯度下降法更新θ,λ；
　　step+=1
　end if
当前状态$s_t = s_{t+1}$, $a_t = a_{t+1}$；
if$\text{step}\%N^- = 0$ 并且$\text{step} \neq 0$then
　将 Q 网络的参数θ重新复制到目标网络参数上$\theta' \leftarrow \theta$；
　end if
　break
end for
　end for

假设该方法神经网络深度为H，输入数据维度δ_{in}，GRU 输入数据序列长度为L，输出数据维度δ_{out}，S与A'分别为状态空间与动作空间大小。由 DDQN 算法伪代码可知，该方法总体时间复杂度为$O(\delta_{\text{in}}H+\delta_{\text{out}}H+SA'+LH^2)$，因此在光刻区调度场景中该方法的求解时间不会随着问题规模的增大显著增加。

3　实验验证

为验证基于 GDDQN 的晶圆光刻区实时调度方法的有效性，基于上海某晶圆制造企业数据进行实验验证。运行本实验的工作站软硬件配置为 Windows 10 64 位操作系统、Intel (R) Xeon (R) W-2235 CPU @ 3.80GHz 处理器、32G 运行内存、 NVIDIA RTX A4000 显卡处理器。程序在搭载了 pytorch2.0 版本框架的 pycharm 上运行，编程语言为 python。实验总体分为实验设计、参数优化及实验分析 3 个部分。

3.1　实验设计

该晶圆厂的光刻工艺 A 区有 8 台光刻机，从该区取连续 130 个时间周期的生产数据，保证单个时间周期内有 60 个晶圆随机到达。该晶圆厂晶圆类

型共有 30 种，晶圆到达时间t_{aj}服从泊松分布，每种晶圆的光刻时间p_{jm}服从区间$[1, p_{\max})$上的均匀分布。其中，$p_{\max}$为光刻时间的最大值，同一类型的晶圆在不同光刻机上的光刻时间各不相同。每一种晶圆类型的权重系数w_{jf}都服从区间为$[1, w_{\max})$上的均匀分布，$w_{\max}$为晶圆权重最大值。为了模拟光刻区晶圆交货期分布，设定影响因子 K 服从区间[1,2) 与 [4,5) 的均匀分布，使得车间内同时包含交货期紧张和盈余的晶圆 lot。基于 130 个时间周期数据，随机选择 100 个周期用作训练集，剩下 30 份数据集作为测试集。

3.2　参数优化

超参数的设计对 GDDQN 的性能影响较大，基于训练集数据分别对 GDDQN 的超参数：最小训练批量、学习率、折扣率以及目标网络的更新周期进行灵敏度分析。其中，最小训练批量分为 3 个实验：128、256、512；学习率分为 3 个实验：0.001、0.000 1、1×10^{-5}；折扣率分为 3 个实验：0.99、0.95、0.90；目标网络更新周期分为 3 个实验：100、200、400。对某一超参数进行灵敏度分析实验时采用其他超参数不变的原则进行，所有超参数均在训练集上进行训练，并在迭代过程中分别记录两个目标优化过程。图 4 展示了超参数灵敏度分析实验结果。学习率对算法性能的影响体现在训练时参数迭代的步长大小，从图 4 (a) 中可以看出，学习率取值为 1×10^{-5} 时算法性能最优;图 4 (b) 表明选取的折扣率γ为 0.90 时算法最优；图 4 (c) 和 (d) 表明当最小训练批量为 128 以及目标网络更新周期为 200 时算法性能最优。其余的超参数设置如表 4 所示。

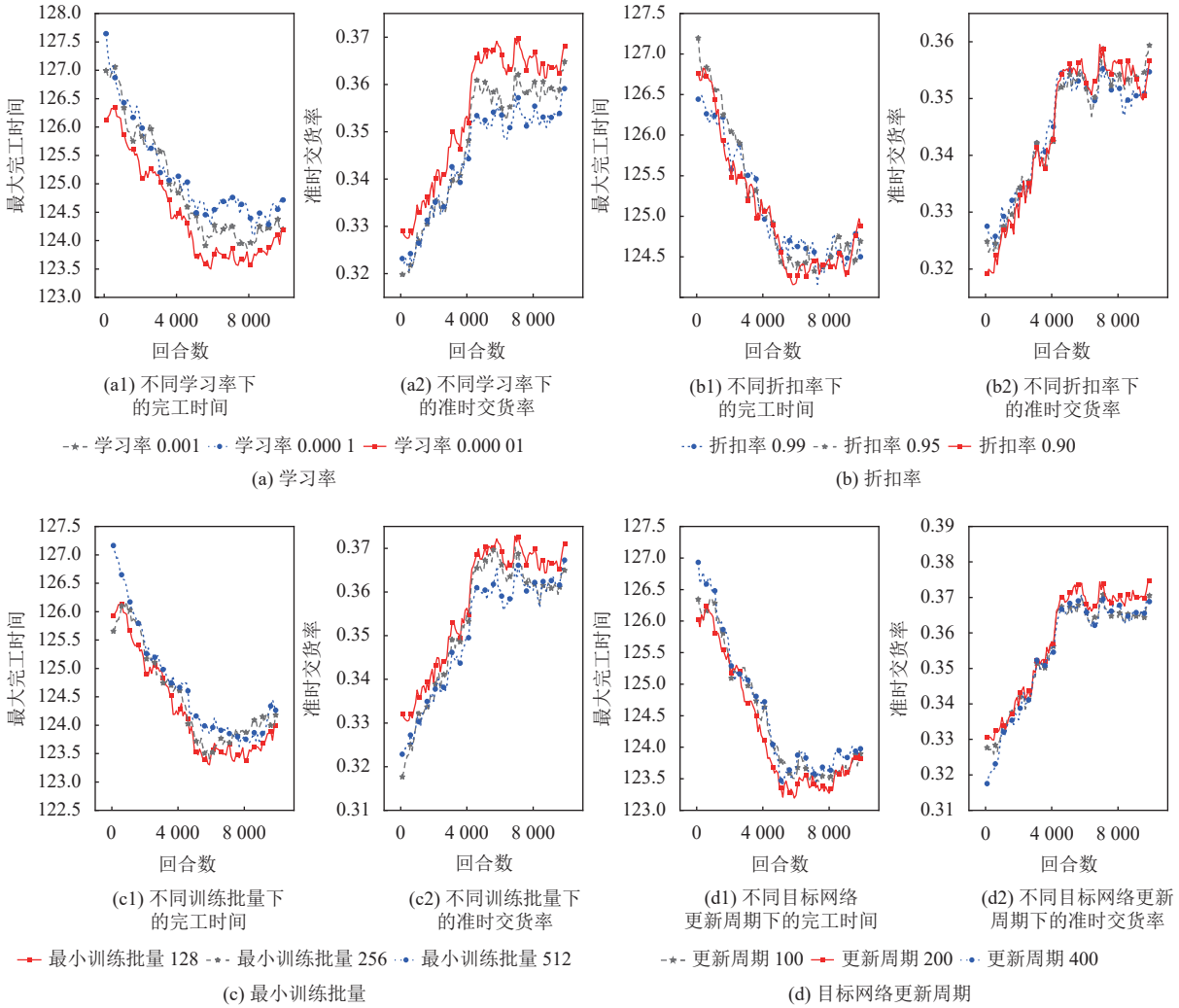

图 4　超参数验证结果

Figure 4　Verification results of hyperparameters

表 4 超参数优化值

Table 4 Optimized hyperparameters

超参数名称	超参数优化值
折扣率γ	0.90
最小训练批量	128
迭代最大值	10 000
经验回放池容量	100 000
探索率ε	0.9→0.1
Adam 学习率	1×10^{-5}
目标网络更新频率	200
优先经验回放α	0.6
优先经验回放β	0.4
Q 网络与目标 Q 网络隐藏层	3
GRU 网络隐藏层神经元数量	128

3.3 实验分析

基于表 4 的超参数设置，并采用训练集利用 GDDQN 算法进行训练。图 5 展示的是 GDDQN 与传统 DDQN[9]在 100 个相同规模的训练集上迭代训练的结果对比。图 5 (a) 、(b) 分别展示了训练过程最大完工时间与准时交货率优化指标对比，GDDQN 与 DDQN 均表现出良好的收敛趋势。随着迭代回合数的增加，曲线趋向平稳，表明在解约束规则方法的结合下，智能体未出现长期无效等待情况，非可行解比例较少，收敛速度较快。GDDQN 在两个优化指标上的性能均比 DDQN 好，这说明本研究设计的方法求解能力与动态优化能力更强，验证了改进的有效性。

在此基础上，以最小化完工时间以及最大化晶圆准时交货率为评价指标，将其与光刻区传统启发式规则 WMDD、WSPT、FIFO 等[17]，以及 DDQN 算法进行对比实验。通过测试算例验证该方法的有效性。测试集与训练集同取于上海某企业制造数据。

图 6、图 7 展示了本方法与 DDQN 和启发式规则在两个不同优化指标的对比情况。实验结果表明，GDDQN 在所有场景下最大完工时间指标上优于传统 DDQN 与启发式规则；在准时交货率指标上 GDDQN 优化效果更明显，且 GDDQN 方法 30 次测试结果的波动范围较小，优化能力稳定。表 5 展示了 30 次测试结果在最优值、均值以及标准差上的对比结果。30 次测试结果中，在最大完工时间以及晶圆准时交货率两个优化指标上，GDDQN 取得的最优值以及均值均优于传统 DDQN 以及启发式规则方法。

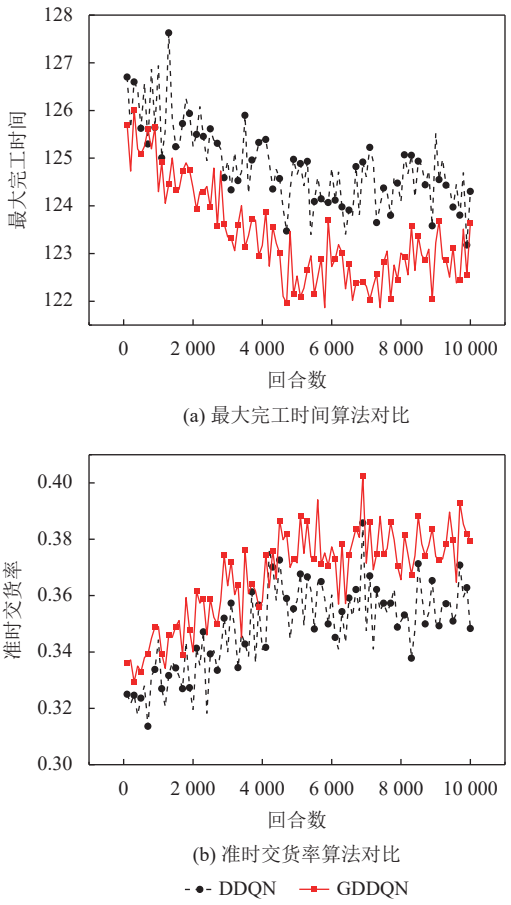

图 5 训练迭代过程

Figure 5 Iteration process of training

表 5 算法性能对比结果

Table 5 Comparison results of algorithm performance

调度方法	晶圆准时交货率			最大完工时间		
	最优值	均值	标准差	最优值	均值	标准差
GDDQN	**0.60**	**0.45**	0.05	**105.52**	**118.70**	6.68
DDQN	0.57	0.42	0.06	113.85	123.81	7.16
WSPT	0.58	0.38	0.10	111.25	121.78	6.75
WMDD	0.32	0.16	0.04	120.07	132.87	7.73
ATC	0.62	0.37	0.11	107.70	120.69	6.66
WCOVERT	0.55	0.34	0.09	111.24	124.48	6.18
WCR	0.22	0.13	**0.02**	130.21	144.46	7.49
WEDD	0.33	0.17	0.05	151.35	161.72	**5.33**
FIFO	0.33	0.22	0.06	152.48	162.19	5.73

以上对比结果表明，本文提出的方法优于传统 DDQN 与启发式规则方法，满足多目标性能优化的同时具有较优的动态调度优化能力。在所有场景中，本文方法单次获得优化调度方案的时间均在

1 s 以内，具有良好的实时性。该方法可以满足单个周期内 8 台光刻机、150 个晶圆 lot 到达的调度场景，满足晶圆厂实际调度需求。因此本文所提方法用于晶圆光刻区实时调度是有效的。

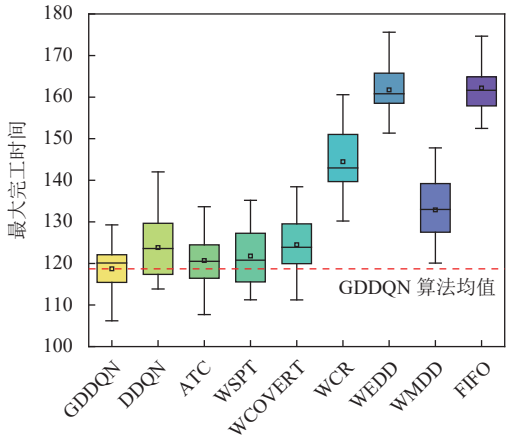

图 6　最大完工时间指标算法性能对比

Figure 6　Comparison of algorithm performance on the maximum completion time

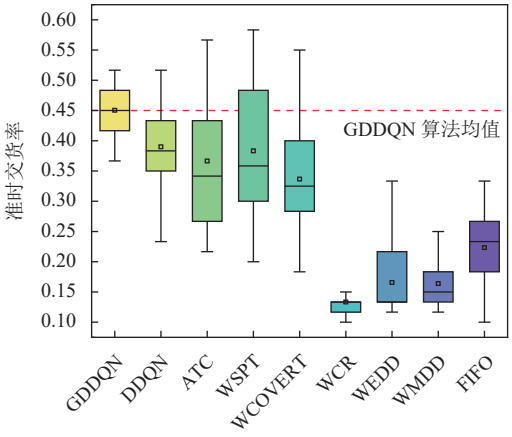

图 7　准时交货率指标算法性能对比

Figure 7　Comparison of algorithm performance on the on-time delivery rate

4　结束语

本研究针对具有动态实时性、多约束与多目标特点的光刻区调度问题，提出一种基于门控循环单元强化学习的方法。该方法将 GRU 与深度强化学习 DDQN 模型结合，记录光刻区车间实时车间与调度状态的同时提取前时序状态，提高方法调度优化能力；通过设计 DDQN 输入状态特征、输出动作集、面向最小化最大完工时间和晶圆准时交货率的奖励函数，设计设备专属性约束、掩模版约束处理流程，实现多目标优化的同时提高了调度方法的实用性；通过某晶圆制造企业实际算例来验证方法的有效性。实验结果表明，动态场景下，基于 GDDQN 的晶圆光刻区调度方法性能均优于传统 DDQN 算法与启发式调度规则。因此本文方法可有效提高光刻区实时调度性能。由于受晶圆制造系统物料搬运的影响，考虑晶圆物料搬运时间的光刻区实时调度是下一步的研究工作。

参考文献：

[1] CHEN H, GUO P, JIMENEZ J, et al. Unrelated parallel machine photolithography scheduling problem with dual resource constraints[J]. IEEE Transactions on Semiconductor Manufacturing, 2022, 36(1): 100-112.

[2] LI X L, HUANG Y L, TAN Q, et al. Scheduling unrelated parallel batch processing machines with non-identical job sizes[J]. Computers & Operations Research, 2013, 40(12): 2983-2990.

[3] MAECKER S, SHEN L, MÖNCH L. Unrelated parallel machine scheduling with eligibility constraints and delivery times to minimize total weighted tardiness[J]. Computers & Operations Research, 2023, 149: 105999.

[4] HAM A. Scheduling of dual resource constrained lithography production: using CP and MIP/CP[J]. IEEE Transactions on Semiconductor Manufacturing, 2017, 31(1): 52-61.

[5] 张朋, 张洁, 王卓君, 等. 基于分解多目标进化算法的光刻区调度方法[J]. 华中科技大学学报 (自然科学版), 2022, 50(4): 26-32.
ZHANG Peng, ZHANG Jie, WANG Zhuojun, et al. Decomposition multi-objective evolutionary algorithm based photolithography area scheduling method[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2022, 50(4): 26-32.

[6] ZHANG P, ZHAO X, SHENG X, et al. An imperialist competitive algorithm incorporating remaining cycle time prediction for photolithography machines scheduling[J]. IEEE Access, 2018, 6: 66787-66797.

[7] ZHANG Z, ZHENG L, LI N, et al. Minimizing mean weighted tardiness in unrelated parallel machine scheduling with reinforcement learning[J]. Computers & Operations Research, 2012, 39(7): 1315-1324.

[8] ZHANG T, XIE S, ROSE O. Real-time batching in job shops based on simulation and reinforcement learning[C]//Proceedings of the 2018 Winter Simulation Conference (WSC) . Gothenburg: IEEE Press, 2018: 3331-3339.

[9] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double q-learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Phoenix: AAAI Press, 2016.

(下转第 30 页)