반도체디스플레이기술학회지 제23권 제2호(2024년 6월)
Journal of the Semiconductor & Display Technology, Vol. 23, No. 2. June 2024.

강화학습과 시뮬레이션을 활용한 Wafer Burn-in Test 공정 스케줄링

권순우[*,**]·오원준[*,**]·안성혁[*,**]·이현서[*,**]·이호열[***]·박인범[*†]
[*†]명지대학교 산업경영공학과, [**]명지대학교 반도체장비공학 연계전공, [***]㈜뉴로코어

Scheduling of Wafer Burn-In Test Process Using Simulation and Reinforcement Learning

Soon-Woo Kwon[*,], Won-Jun Oh[*,**], Seong-Hyeok Ahn[*,**], Hyun-Seo Lee[*,**], Hoyeoul Lee[***] and In-Beom Park[*†]**
[*†]Department of Industrial and Management Engineering, Myongji University,
[**]Semiconductor Equipment Engineering Program, Myongji University,
[***]Neurocore Co., Ltd

ABSTRACT

Scheduling of semiconductor test facilities has been crucial since effective scheduling contributes to the profits of semiconductor enterprises and enhances the quality of semiconductor products. This study aims to solve the scheduling problems for the wafer burn-in test facilities of the semiconductor back-end process by utilizing simulation and deep reinforcement learning-based methods. To solve the scheduling problem considered in this study. we propose novel state, action, and reward designs based on the Markov decision process. Furthermore, a neural network is trained by employing the recent RL-based method, named proximal policy optimization. Experimental results showed that the proposed method outperformed traditional heuristic-based scheduling techniques, achieving a higher due date compliance rate of jobs in terms of total job completion time.

Key Words : Scheduling, Semiconductor, Wafer Burn-in Test, Reinforcement Learning, Proximal Policy Optimization

1. Introduction

지난 2022년을 기준으로 전세계 반도체 시장 규모는 총 6,000억 달러로 추산되며, 특히 2015년의 3,500억 달러였던 시장 규모 대비 약 70% 상승했다. 시장 규모와 더불어 의료, 교통산업을 비롯해 반도체를 사용하는 분야의 증가로 인해 생산해야 하는 반도체 제품의 종류 또한 매우 다양해지고 있다. 이러한 상황에서는 반도체 공정에서 생산 일정 계획을 수립하여 다양한 종류의 반도체 제품과 많은 생산량을 주어진 환경에서 효율적으로 생산할 수 있다. 따라서 반도체 공정 라인의 스케줄링은 시간 및 자원적 측면에서 매우 중요하다. 하지만 반도체 공정 스케줄링 문제는 한정된 수의 장비를 여러 제품을 위해 공유해야 하며, 공정간 순서가 순차적이지 않고, 제품의 레시피에 따라 달라지는 등의 제약이 많기에 최적화 관점에서 NP-hard이다. 그럼에도 불구하고 현재까지 반도체 생산 공정을 위한 스케줄링 문제에 관한 연구가 다양하게 진행되었다. 다양한 공정들 중 후공정에 해당되는 Die attach, Wire bonding, Molding, Wafer sawing 들을 비롯하여 다양한 테스트 공정들의 중요도가 최근들어 떠오르는 추세인데, 전

[†]E-mail: inbeom@mju.ac.kr

공정을 통해 성능이 우수하고 크기가 작은 칩을 만들어 내더라도, 후공정 과정에서 패키징 기법과 재료를 비롯하여 테스트의 적합성 및 정밀도에 의해 최종 반도체 제품으로 완성되기 때문이다.

특히 반도체 후공정 중에서도 Wafer burn-in test 공정은 Wafer level로 Stress test를 진행해 초기 불량률을 감소시킬 수 있는 주요한 공정으로 손꼽힌다. 하지만 기존 반도체 후공정 스케줄링 문제를 다룬 연구들 중에서 Wafer burn-in test 공정의 스케줄링 문제를 다룬 연구는 상대적으로 그 수가 적다. 따라서 본 논문에서는 Wafer burn-in test 공정을 위한 스케줄링 연구를 진행하였다. 본 연구에서는 Wafer burn-in test 공정 스케줄링 문제 해결을 위한 상태, 행동, 보상함수를 고안하고, 심층강화학습 분야 최신 알고리즘 중 하나인 proximal policy optimization(PPO) 기법을 문제에 적용하였다. 실제 테스트 공정의 일부를 모사한 스케줄링 문제에 대한 실험 결과를 통해, 납기 만족률 측면에서 기존 휴리스틱 기반 스케줄링 기법과 비교해 성능이 우수함을 확인하였다.

본 논문의 나머지 부분은 네 부분으로 나누어 구성된다. 2장에서는 본 연구를 진행한 방법론과 대상공정을 위해 탐구한 기존 연구들을 소개한다. 3장에서는 연구 대상 문제를 정의하고, 알고리즘에 대한 적용 및 사용된 알고리즘을 설명한다. 4장에서는 실험 환경을 비롯해 실험에 사용된 데이터셋을 소개하고 실험 결과를 설명하며, 마지막으로 5장에서는 본 연구의 성과와 한계점 및 추후 연구 방향을 제시한다.

2. Previous Research

2.1 Semiconductor Packaging Line Scheduling

반도체 후공정에 대한 스케줄링 연구는 기존에도 활발히 진행되었다. 이는 반도체 산업에서 후공정 자체의 중요도가 높다는 것을 의미한다. Park은 반도체 후공정에 대한 생산계획 일정을 수립하기 위한 스케줄링에 대한 연구를 진행하였으며[1], 또한 후공정에서 장비 변경으로 인한 기대효과를 도출해내기 위한 연구를 진행하였다[2]. Klemmt는 복잡한 반도체 패키징 공정을 위해 다목적 파레토 동적 스케줄링 개발에 대한 연구를 진행하였다[3]. 그 외에도 다중 칩 제품이 패키징 공정에서 반복적으로 재 투입되는 점으로 인해 공정 장비의 셋업이 빈번하게 일어나는 현상을 최소화하기 위해 심층 Q 네트워크를 활용하여 강건성을 가진 공정 정규화에 대한 연구[6]에서는 규칙 기반 등의 다른 방식에 비해 우수한 결과를 도출하였다[4].

본 논문에서 진행하고자 하는 테스트 공정을 위한 스케줄링에 대한 연구 또한 다양하게 진행되었다. Wafer Burn-In Test 공정에서 다른 장비로 이동 가능한 상황에 사이클 타임을 최소화하는 연구[5]를 비롯해 반도체 테스트 장비에서 단일 프로세서를 가지는 사용방법을 사용자 실행방식에서 daemon으로 변경하는 연구[6]가 존재하며, Kress는 자주 발생하는 고객 요구 변동 및 장비 오류로 인해 변화되는 공정 환경에서의 테스트 공정을 위한 스케줄링 기법 개발을 진행하였다[7]. Song은 군집 최적화 방법을 이용해 테스트 공정에서의 병목 장비를 위한 스케줄링 기법 개발을 진행하였다[8].

2.2 Reinforcement learning-based scheduling

강화학습은 스케줄링을 비롯해 많은 최적화 문제를 해결하기 위한 기법으로 알려져 있다. 실제로 반도체를 비롯하여 많은 산업군에서 강화학습 기법을 사용하여 생산계획을 수립하기 위한 다양한 연구들이 제안되었다.

특히, 다양한 강화학습 알고리즘들 중 PPO 알고리즘은 학습데이터를 재사용하고 step 단위로 데이터를 생성하는 알고리즘으로 스케줄링 문제 해결에 있어서 다른 알고리즘들에 비해 성능이 우수하게 나온다고 알려져 있다[9]. PPO 알고리즘을 사용하여 스케줄링 기법을 연구한 기존 사례들 또한 다양하게 존재한다. Liu가 진행한 PPO 알고리즘 기반 저탄소 데이터 센터 운영을 위한 스케줄링 기법 연구[10]를 비롯하여 Zhao의 PPO 알고리즘을 사용한 flexible job shop 스케줄링에 대한 연구[11]가 존재한다. 이 외에도 PPO 알고리즘을 사용해 공정 처리 시간이 불확실한 dynamic job shop을 위한 스케줄링에 대한 연구[12]와 PPO 알고리즘을 사용해 비행기의 운행 테스트에 대한 dynamic task 스케줄링에 대한 연구[13]가 있으며, Felder가 진행한 PPO 알고리즘을 비롯한 심층 강화 학습을 사용한 energy-flexible job shop 스케줄링에 대한 연구 또한 진행되었다[14]. 기존 연구들 에서도 알 수 있듯, PPO 알고리즘은 스케줄링 기법 개발에 관한 연구에서 활발히 사용되고 있고, 우수한 성능을 보장하기에, 본 연구는 PPO 알고리즘을 활용해 Wafer burn-in test 공정을 위한 스케줄링 기법을 개발하였다.

3. Proposed Method

3.1 Problem Definition

본 논문에서 다루고자 하는 문제는 Parallel machine scheduling이며 납기일을 만족하는 작업의 수량을 최대화하는 것이다. 이 문제는 혼합 정수 계획의 형태로 모델링할 수 있으며, 사용된 기호는 Table 1에 소개되어 있다.

Table 1. Notation for optimization

m	장비의 총 개수
n	작업의 총 개수
t_j	작업 j의 수행 완료 시간
d_j	작업 j의 납기일
r_j	작업 j의 수행 준비 완료 시간
s_{ij}	작업 j의 장비 i 에서의 수행 변경 소요 시간

$$\boldsymbol{Maximize} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{I}(t_j \le d_j) \ \forall i,j \qquad (1)$$

위 목적함수는 납기일 내 수행 완료된 작업의 총합을 최대화하는 것을 의미한다. $\mathbf{I}(t_j \le d_j)$ 는 지시함수 $\mathbf{I}$를 이용해 작업 $\boldsymbol{j}$가 납기일 이전에 완료되었는지 여부를 나타내는 이진 변수이다. 목적함수를 위한 제약 조건들로는 다음과 같은 것들이 있다. 각 작업 $\boldsymbol{j}$에 대해 작업을 수행할 수 있는 시간을 $\boldsymbol{r_j}$로 정의한다. 또한 작업 $\boldsymbol{j}$는 정확히 한 대의 장비 $\boldsymbol{i}$에 할당되어야 한다. 각 작업 $\boldsymbol{j}$는 하나의 장비에서만 수행되어야 하며 각 장비 $\boldsymbol{i}$는 동일한 시간에 하나의 작업만을 처리할 수 있다. 또한 특정 시간에 여러 작업이 동일한 장비에 할당되지 않도록 하며, 이는 각 장비가 주어진 시간 내에 최대 하나의 작업만 수행하도록 한다. 각 작업 $\boldsymbol{j}$ 는 이전 작업이 다른 종류의 작업이었을 경우, 이전 작업이 완료된 후 일정한 작업 변경 소요 시간이 필요한데, 이는 $\boldsymbol{s_{ij}}$로 정의한다. 이를 통해 각 작업이 이전 작업의 완료 시간과 작업 변경 소요 시간을 고려하여 현재 작업의 시작 시간을 결정한다. 각 작업에 대해 지정된 납기일은 $\boldsymbol{d_j}$로 정의된다.

3.2 Interaction between simulator and neural network

본 연구에서 wafer burn-in test공정 스케줄링 문제 해결을 위해 PPO 알고리즘을 시뮬레이션에 적용하여 실험을 진행하는 방식을 채택하였다. 인공지능 신경망의 학습 구조와 시뮬레이터간의 관계를 [Fig. 1]을 통해 요약할 수 있다.

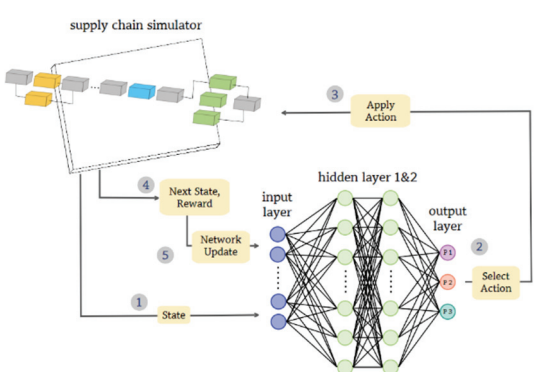

Fig. 1. Learning Process and Neural Network Structure.

시뮬레이터가 인공 신경망에 상태를 제공하면 신경망은 적절한 행동을 선택해 시뮬레이터에 전달하고 시뮬레이터는 해당 행동들을 통해 얻은 보상을 도출한다. 이러한 과정은 사전에 지정한 학습 횟수만큼 누적되어 진행된다.

3.3 State-Action-Reward

본 장에서는 강화학습 기법을 위해 설계한 Markov decision process(MDP) 구조를 소개한다. Wafer burn-in test공정 스케줄링 문제를 해결하기 위한 상태를 구성하는 각 요소는 Table 2와 같으며 각 후보 작업의 계획에 따른 진척 상황 및 전체 작업에 대한 진척 상황이 포함된다. 대표적으로 Order release planning(ORP) 및 Latest possible start time(LPST)을 기반으로 한 재공 비율 및 Setup에 대한 비중 등의 정보가 상태로 정의된다.

후보 작업의 계획 대비 진척 현황은 후보 작업 별 계획한 생산 필요량 대비 진척 정도에 대한 정보이며 후보 작업은 장비에서 수행할 수 있는 작업들의 집합을 의미한다. 후보 작업의 전체 작업 대비 진척 현황은 모든 작업을 통틀어 계획한 수행 요구량 대비 후보 작업의 수행 진척 현황을 의미한다. 현재 시점 기준으로 납기일 까지의 기간이 $\boldsymbol{k}$주 이내로 남아 납기가 급한 후보 작업과 비교적 급하지 않은 후보 작업들간 재공의 비율을 후보 작업의 ORP 계획 대비 $\boldsymbol{k}$일 전·후 LPST 재공 비율로 표현한다. 본 연구에서는 $\boldsymbol{k}$를 21로 설정하였다. 후보 작업의 전체 작업 재공 대비 $\boldsymbol{k}$일 전·후 LPST 재공 비율은 모든 작업 대비 후보 작업의 납기가 급한 재공 재고의 개수와 비교적 급하지 않은 후보 작업의 재공 재고간 비율을 의미한다. 전체 장비 대비 후보 작업을 처리할 수 있는 Setup인 장비를 후보 작업 Setup의 전체 Setup 대비 할당 비중으로 표현한다. 후보 작업의 Setup time 비중은 대상 장비의 평균적인 작업 변경 소요 시간 대비 후보 작업을 선택했을 때의 작업 변경 소요 시간을 의미하며 후보 작업의 Setup type 비중은 대상 장비가 수행할 수 있는 작업 대비 후보 작업을 선택했을 때의 효율을 의미한다.

계획 대비 전체 작업 진척 현황은 모든 작업을 통틀어 계획한 수행 요구량 대비 진척 현황을 의미하며 전체 작업의 ORP 계획 대비 $\boldsymbol{k}$일 전·후 LPST 재공 비율은 현재 시점 기준으로 납기가 급한 작업과 비교적 납기가 급하지 않은 작업의 비율을 나타낸다. 가장 높은 Setup time의 비중은 장비의 평균적인 작업 변경 소요시간 대비 가장 긴 작업 변경 소요시간을 의미하며 계획 대비 Setup 건수 비중은 현재 시점까지 Setup을 변경한 횟수를 의미한다.

행동의 구성 요소는 Shortest setup time(SST), Equipment requirement index(ERI), Q-weighted delayed product(QWIP), continuous operation product(COP), Rich stock item(RSI), High equipment profile(HEP)이며, 총 6차원으로 구성되어 있다.

Table 2. Elements of State

State	후보 작업들의 계획 대비 진척 현황
	후보 작업들의 전체 작업 대비 진척 현황
	후보 작업들의 ORP 계획 대비 k일 전·후 LPST 재공 비율
	후보 작업들의 전체 작업 재공 대비 k일 전·후 LPST 재공 비율
	후보 작업 Setup들의 전체 Setup 대비 할당 비중
	후보 작업들의 Setup time 비중
	후보 작업들의 Setup type 비중
	계획 대비 전체 작업 진척 현황
	전체 작업의 ORP 계획 대비 k일 전·후 LPST 재공 비율
	가장 높은 Setup time의 비중
	계획 대비 Setup 건수 비중

SST는 작업이 수행되는 장비 라인에서 필요한 초기 설정 시간이 가장 짧은 작업을 먼저 할당하는 규칙이다. SST는 작업 효율성을 향상시키며 작업 수행 라인의 설정 시간을 최소화하는 데 사용될 수 있다. ERI는 해당 작업 수행에 필요한 장비 요구량이 가장 적은 작업을 먼저 할당하는 규칙이다. ERI 규칙을 반영 시 장비 요구량이 적은 작업부터 처리하게 됨에 따라 작업 중 발생할 수 있는 장비 병목 현상을 줄일 수 있다. 따라서 ERI 규칙은 전체 생산라인에 장비 활용도를 높이고 병목 발생 위험을 낮추는 역할을 한다. QWIP은 특정 작업의 납기 지연이 작업량에 미치는 영향을 가중치로 반영한 작업을 나타내며 납기 지연으로 인한 가중치가 가장 높은 작업으로 납기 준수가 가장 중요한 작업을 먼저 할당하는 규칙이다. COP는 연속적으로 수행 가능한 작업 중 가장 먼저 할당하는 규칙이다. COP는 작업 수행 라인에서 중단 없이 연속적으로 수행될 수 있는 작업을 식별한다. 그에 따라 특정 작업이 완료된 후 바로 다음 작업을 이어서 수행 가능한 경우, 그 작업을 우선적으로 배정한다. RSI는 충분한 여유분이 있는 작업을 나타내 여유분이 가장 많은 작업을 먼저 할당하는 규칙이다. RSI는 재고 관리나 작업 계획 수립 시에 필요한 정보이다. 재고가 부족한 작업은 후에 처리하여 추가 주문이나 긴급 조달 등의 대응 가능한 시간을 확보할 수 있다. HEP는 작업을 수행하기 위해 필요한 장비나 장비의 요구량이 높은 작업을 먼저 할당하는 규칙이다. 장비 가용성이 낮아지기 전에 해당 작업을 먼저 처리하여 작업 지연을 방지할 수 있다.

보상 값은 작업 수행에 대한 성과를 나타내며, Due date compliance percent(DDCP) 과 Time taken to produce all products(TPAP)로 구성되어 있다 즉, 납기일을 만족하고 빠른 시간 내에 작업을 완료하면 높은 보상을 얻는 구조로 설계하였다. DDCP는 수행된 작업의 총 출하건 수 중 납기일 내에 출하된 건 수의 백분율이다. TPAP는 모든 작업 수행이 완료되기까지 걸리는 총 시간을 의미한다TPAP는 강화학습에서 최적해 도출 의사결정 문제에 주로 사용되는 보상 함수 중 하나로, 작업 수행 완료 시간 최소화로 장비 가동의 전체적인 효율성 향상에 기여한다. 보상에 대한 수식은 식 (2)-(4)과 같다.

$$R = \sum_{j=1}^{n}(R_{D_j} - R_{T_j}) \tag{2}$$

$$R_{D_j} = \sum_{j=1}^{n} \mathrm{I}(ct_j \le d_j) \tag{3}$$

$$R_{T_j} = \sum_{j=1}^{n}(ct_j - s_j) \tag{4}$$

DDCP는 클수록 좋고, TPAP는 작을수록 좋기에, 두 항목에 대한 보상 함수인 R_D와 R_T의 차이가 클수록 우수한 성능을 도출하도록 보상 함수를 정의하였다. 작업 j에 대한 보상 값인 R_{D_j}와 R_{T_j}의 차이 값을 작업 종류별로 구하고 그 차이 값을 모두 합한 값을 총 보상 R로 설정하였다.

3.4 PPO algorithm

PPO는 안정적이면서도 빠른 정책 학습을 가능하게 하기 위해 강화학습의 정책을 최적화하는 PG 알고리즘을 발전시킨 방법이다. 일반화된 PPO알고리즘은 아래와 같이 정의된다.

Algorithm 1 PPO Algorithm

1: 입력: 초기 정책 매개변수 θ_0, 추기 함수 매개변수 ϕ_0
2: for $k = 0, 1, 2, \ldots$ do
3: 정책 $\pi_k = \pi(\theta_k)$를 환경에서 실행하여 궤적 집합 $D_k = \{\tau_i\}$를 수집한다.
4: 보상 총합 $\hat{R}_t$를 계산한다.
5: 현재 가치 함수 $V_{\phi k}$를 기반으로 이점 추정치 $\hat{A}_t$ 를 계산한다.

$$A_t = R_t - V_k\phi(s_t) \tag{5}$$

6: PPO-Clip objective 를 최대화하여 정책을 업데이트 한다.

$$\theta_{k+1} = arg\max_{\theta} \frac{1}{|D_k|T}\sum_{\tau\in D_k}\sum_{t=0}^{T}\min\left(\frac{\pi_\theta(\alpha_t|s_t)}{\pi_{\theta_k}(\alpha_t|s_t)}A^{\pi_{\theta_k}}(s_t,\alpha_t),\ g(\epsilon, A^{\pi_{\theta_k}}(s_t,\alpha_t))\right) \tag{6}$$

7: 회귀 분석을 통해 가치 함수를 맞춘다(평균 제곱 오차 기준):

$$\phi_{k+1} = arg\min_{\phi}\frac{1}{|D_k|T}\sum_{\tau\in D_k}\sum_{t=0}^{T}\left(V_\phi(s_t) - \hat{R}_t\right)^2 \tag{7}$$

8: 종료

1번 라인에서 초기 정책 파라미터 $\boldsymbol{\theta}_0$와 초기 가치 함수 파라미터 $\boldsymbol{\phi}_0$를 설정하고 2-4번 라인의 과정을 걸쳐 식 (5)를 통해 $\boldsymbol{A}_t$를 구한다. 다음으로는 PPO알고리즘의 클리핑 기법인 클리핑 함수 $\boldsymbol{g}(\boldsymbol{\varepsilon}, \boldsymbol{A}^T\boldsymbol{\theta}_k(\boldsymbol{s}_t, \boldsymbol{\alpha}_t))$를 이용하여 $\boldsymbol{\theta}$를 안정적으로 업데이트하며, 식은 식 (6)과 같다. $\boldsymbol{\pi}_{kt}(\boldsymbol{\alpha}_t|\boldsymbol{s}_t)$는 $\boldsymbol{s}_t$에서 $\boldsymbol{\alpha}_t$ 행동을 선택할 확률이고, 클리핑 함수 안의 $\boldsymbol{\varepsilon}$는 클리핑 범위를 의미한다. 식 (7)을 사용해 $\boldsymbol{\phi}_{k+1}$를 구하고 반복문을 종료한다.

4. Experiment

4.1 Dataset

실험에 사용한 데이터셋은 대상 작업의 종류, 장비의 수로 구성되어 있고, 이에 대한 정보는 Table 3과 같다. 표의 각 행은 대상 작업의 종류와 구성, 장비의 수와 구성 그리고 각 작업의 목표 수행 횟수를 보여준다.

Table 3. Dataset description

	Number	Description	
Job type	3	P1, P2, P3	
Machine	5	MC-1, MC-2, MC-3, MC-4, MC-5	
Total number of jobs	24	P1	8
		P2	13
		P3	3

또한 작업과 장비 사이에는 수행 가능 제약이 존재하는데, Fig. 2는 해당 제약을 시각화 하여 보여준다. Product 1은 Product 2, Product 3과 다르게 Machine 1혹은 Machine 2로 작업을 수행해야만 하는 단일 작업 제약 조건이 존재한다. 나머지 작업인 Product 2와 Product 3는 모두 다섯 대의 장비에 범용적으로 투입 가능한 상태임을 보여준다.

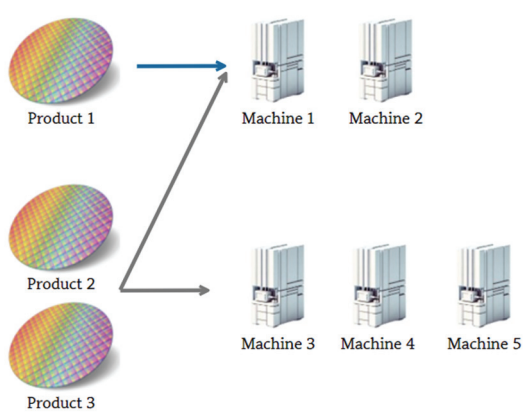

Fig 2. Process relationship of product & machine.

4.2 Experiment setting

앞서 설계한 MDP를 이용하여 PPO 알고리즘 강화학습 모델을 구현하였고, 앞 장에서 정의된 데이터셋을 사용하였으며, 학습은 총 300회 진행하였고, PPO 알고리즘의 하이퍼 파라미터는 Table 4와 같다.

Table 4. Hyper Parameters of PPO algorithm & Neural Network Architecture

Hyper Parameters of PPO algorithm	
Learning rate	0.001
Gamma	0.99
Lambda	0.99
Clip range	0.1
Epoch	12
Batch size	6000
Mini batch size	3000
Neural Network Architectur	
Input layer	142
The number of nodes in hidden layers	256, 256
Activation function	ReLu
Output layer	2

실험은 13th Gen Intel® Core™ i5-13500, NVIDIA GeForce RTX 4060, 64.0GB RAM환경에서 시행하였으며, 강화학습 알고리즘을 사용하여 원활한 시뮬레이션을 위해 Neurocore㈜의 NEMO APS365를 사용하였다. 실험을 통해 각 학습 횟수별로 누적된 보상 값과 손실 값을 각각 Fig. 3과 Fig. 4를 통해 알 수 있다.

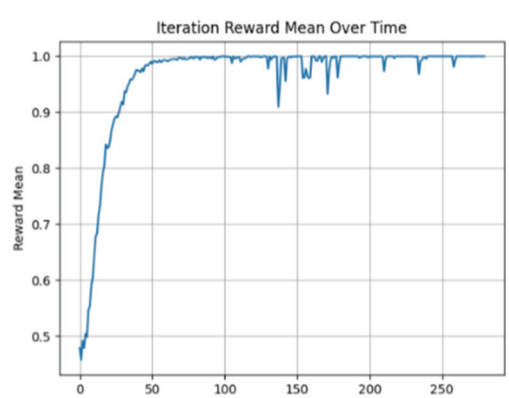

Fig. 3. Reward according to learning progress.

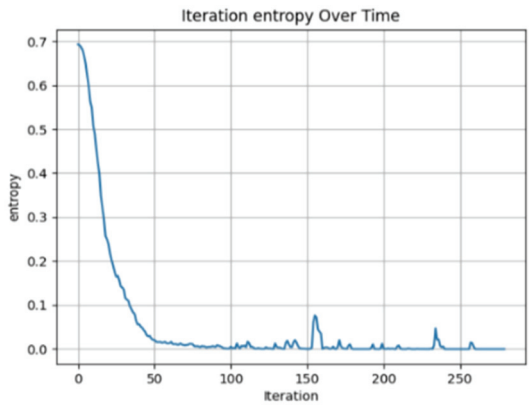

Fig. 4. Loss according to learning progress.

사전에 정의함에 따라 계산되어지는 보상 값은 학습이 진행됨에 따라 최대화되는 방향으로 학습이 진행되어진다. 또한 손실 값은 학습이 진행됨에 따라 최소화되는 방향으로 학습이 진행되며, 학습이 진행됨에 따라 보상이 증가하여 납기 만족률이 100%로 수렴하였다.

제안 기법과 기존 휴리스틱 모델의 성능 평가를 위한 성능 평가 지표는 작업별 납기 만족률과 세 종류의 작업을 수행하기 위한 최후 작업 수행이 끝나는 시간으로 설정하고, 모델에 대한 평가를 진행하였다. PPO 모델과 휴리스틱 모델을 통해 학습한 스케줄링의 결과를 간트 차트로 나타낸 자료를 Fig. 5에서 확인할 수 있다. 또한 Table 5에서 알 수 있듯이 기존 휴리스틱 모델이 한 작업에 대해 납기일을 만족하지 못하는 반면, PPO를 사용해 만든 모델에서는 모든 작업의 납기일을 만족한다. 또한 PPO 모델을 사용해 만든 스케줄링 모델의 작업 소요 시간이 휴리스틱 모델에 비해 약 36.5%의 시간을 절약할 수 있음을 확인할 수 있었고, 이를 통해 제안 강화학습 기법이 본 연구에서 고려한 스케줄링 문제를 해결하기 적합하다고 판단된다.

Table 5. Due date compliance percent by product & Time taken to produce all jobs

Due date compliance percent by product		
MODEL / JOB	Heuristic	PPO
Product 1	**2/8 (20%)**	**8/8 (100%)**
Product 2	**13/13 (100%)**	**13/13 (100%)**
Product 3	**3/3 (100%)**	**3/3 (100%)**
Time taken to produce all jobs		
Heuristic		PPO
2460 minutes		1560 minutes

5. Conclusion

본 논문에서는 반도체 후공정 중 Wafer Burn-In Test에서 사용할 수 있는 공급망 수립 기법을 개발하였다. 동일한 환경에서 휴리스틱 기법으로 생성한 스케줄링 모델 대비

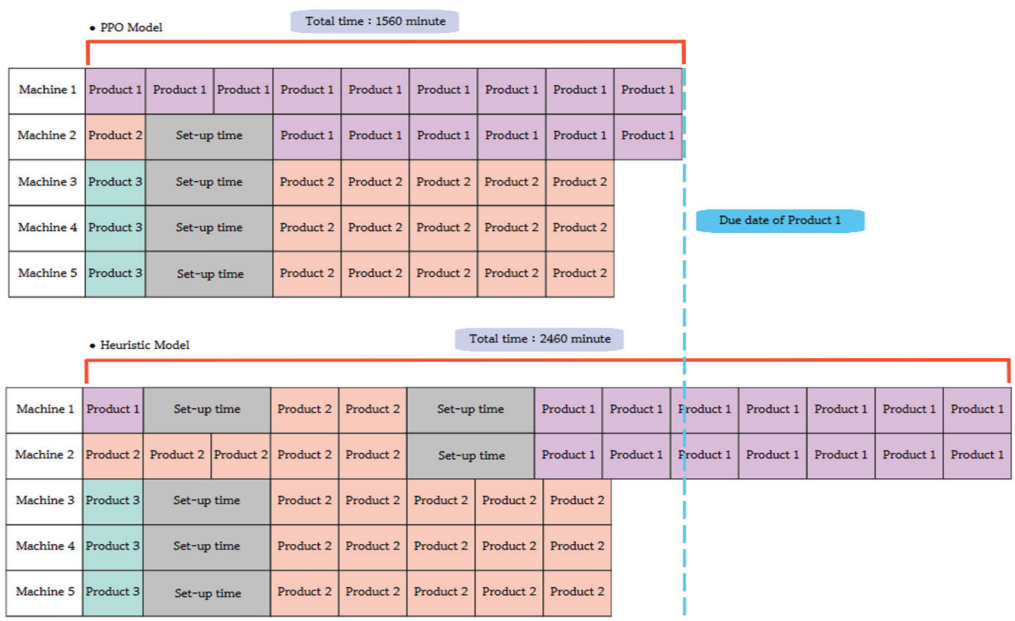

Fig. 5. Scheduling results trained with PPO model and Heuristic model.

우리가 제안한 PPO 알고리즘을 사용한 방법의 작업 납기 만족률이 우수하고, 작업 완료 시간 또한 약 36.5% 우수한 점을 확인할 수 있다.

본 연구결과를 통해 Wafer Burn-In Test 공정에서 심층강화학습 기반 스케줄링 기법의 효과를 확인할 수 있었으나, 연구 대상 스케줄링 문제가 작업의 종류가 세 가지이고 작업을 진행하는 장비가 다섯 대인 소규모 작업 환경에서 학습되었기에 한계점이 있다고 판단된다. 따라서, 추후 연구에서는 대규모 환경에서도 강건하게 적용되는 강화학습 기반 스케줄링 연구할 계획이다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단(2022R1G1A101175)과 2024년도 부처협업형 반도체전공트랙 사업을 통해 한국산업기술진흥원(G02P188000 05502)의 지원을 받아 수행된 연구입니다.

참고문헌

1. In-Beom Park, Jonghun Park. “Scalable scheduling of semiconductor packaging facilities using deep reinforcement learning”, IEEE Transactions on Cybernetics, Vol.53, No.6, pp. 3518-3531, 2021.
2. Beom-suk Chung, Junseok Lim, In-Beom Park, Jonghun Park. “Setup change scheduling for semiconductor packaging facilities using a genetic algorithm with an operator recommender”, IEEE Transactions on Semiconductor Manufacturing, Vol.27, No.3, pp. 377-387, 2014.
3. Weigert, Gerald, A. Klemmt, and S. Horn. “Design and validation of heuristic algorithms for simulation-based scheduling of a semiconductor backend facility”, International Journal of Production Research, Vol.47, No.8, pp. 2165-2184, 2009.
4. Kim, J. K. “Enhancing robustness of deep reinforcement learning based semiconductor packaging lines scheduling with regularized training”, Master’s degree thesis, 2019.
5. Akcalt, Elif, Kazunori Nemoto, and Reha Uzsoy. “Cycle-time improvements for photolithography process in semiconductor manufacturing”, IEEE Transactions on Semiconductor Manufacturing Vol.14, No.1, pp. 48-56, 2001.
6. Lim, S., Choi, H., Han, Y., and Lee, C. “Queue Modeling of Semiconductor Test Equipment Using Effective Background Process.” International Conference on Modeling and Simulation. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 15-19, 2012.
7. Kress and Müller. “Semiconductor final-test scheduling under setup operator constraints”. Computers & Operations Research, Vol.138, 2022.
8. Song, Y., Zhang, M. T., Yi, J., Zhang, L., and Zheng, L. “Bottleneck station scheduling in semiconductor assembly and test manufacturing using ant colony optimization.” IEEE Transactions on Automation Science and Engineering, Vol.4, No.4, pp. 569-578, 2007.
9. Zheng, T., Wan, J., Zhang, J., and Jiang, C. “Deep reinforcement learning-based workload scheduling for edge computing.” Journal of Cloud Computing, Vol.11, No.1, pp. 3, 2022.
10. Liu, W., Yan, Y., Sun, Y., Mao, H., Cheng, M., Wang, P., and Ding, Z. "Online job scheduling scheme for low-carbon data center operation: An information and energy nexus perspective." Applied Energy, Vol.338, 2023.
11. Zhao, L., Fan, J., Zhang, C., Shen, W., and Zhuang, J. “A drl-based reactive scheduling policy for flexible job shops with random job arrivals.” IEEE Transactions on Automation Science and Engineering, 2023.
12. Wu, X., Yan, X., Guan, D., and Wei, M. “A deep reinforcement learning model for dynamic job-shop scheduling problem with uncertain processing time.” Engineering Applications of Artificial Intelligence, Vol.131, 2024.
13. Tian, Bei, Gang Xiao, and Yu Shen. “A deep reinforcement learning approach for dynamic task scheduling of flight tests.” The Journal of Supercomputing, pp. 1-36, 2024.
14. Felder, M., Steiner, D., Busch, P., Trat, M., Sun, C., Bender, J., and Ovtcharova, J. “Energy-flexible job-shop scheduling using deep reinforcement learning.” ESSN: 2701-6277, pp. 352-362, 2023.

접수일: 2024년 6월 6일, 심사일: 2024년 6월 21일,
게재확정일: 2024년 6월 21일