

Zhang, Ming, Lu, Yang ORCID

logoORCID: <https://orcid.org/0000-0002-0583-2688>, Liu, Chao and Xu, Yuchun (2022) Dynamic Scheduling Method for Job-shop Manufacturing Systems by Deep Reinforcement Learning with Proximal Policy Optimization. In: Connected Everything Conference 2022, 18/05/2022, Liverpool, UK.

Downloaded from: <https://ray.yorks.ac.uk/id/eprint/6401/>

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. [Institutional Repository Policy Statement](#)

# RaY

Research at the University of York St John

For more information please contact RaY at [ray@yorks.ac.uk](mailto:ray@yorks.ac.uk)





RECLAIM

@reclaim\_FoF

RECLAIM  
Manufacturing

# Dynamic Scheduling Method for Job-Shop Manufacturing Systems by Deep Reinforcement Learning with Proximal Policy Optimization

Ming Zhang, Yang Lu, Chao Liu and Yuchun Xu

<https://doi.org/10.3390/su14095177>



## 1. Summary

## 2. Objectives

For increasingly complex modern manufacturing production systems, operational decision making encounters more challenges in terms of having **sustainable manufacturing to satisfy customers and markets' rapidly changing demands**. Nowadays, the efficiency of decision making could not be guaranteed nor meet the **dynamic scheduling requirement** in the job-shop manufacturing environment based on the traditional knowledge-based method. We propose using **AI-enhanced deep reinforcement learning methods** to tackle the dynamic scheduling problem in the **job-shop manufacturing system with unexpected machine failure**. The proximal policy optimization algorithm was used in the DRL framework to accelerate the learning process and improve performance.



AI-enhanced Data-driven Method



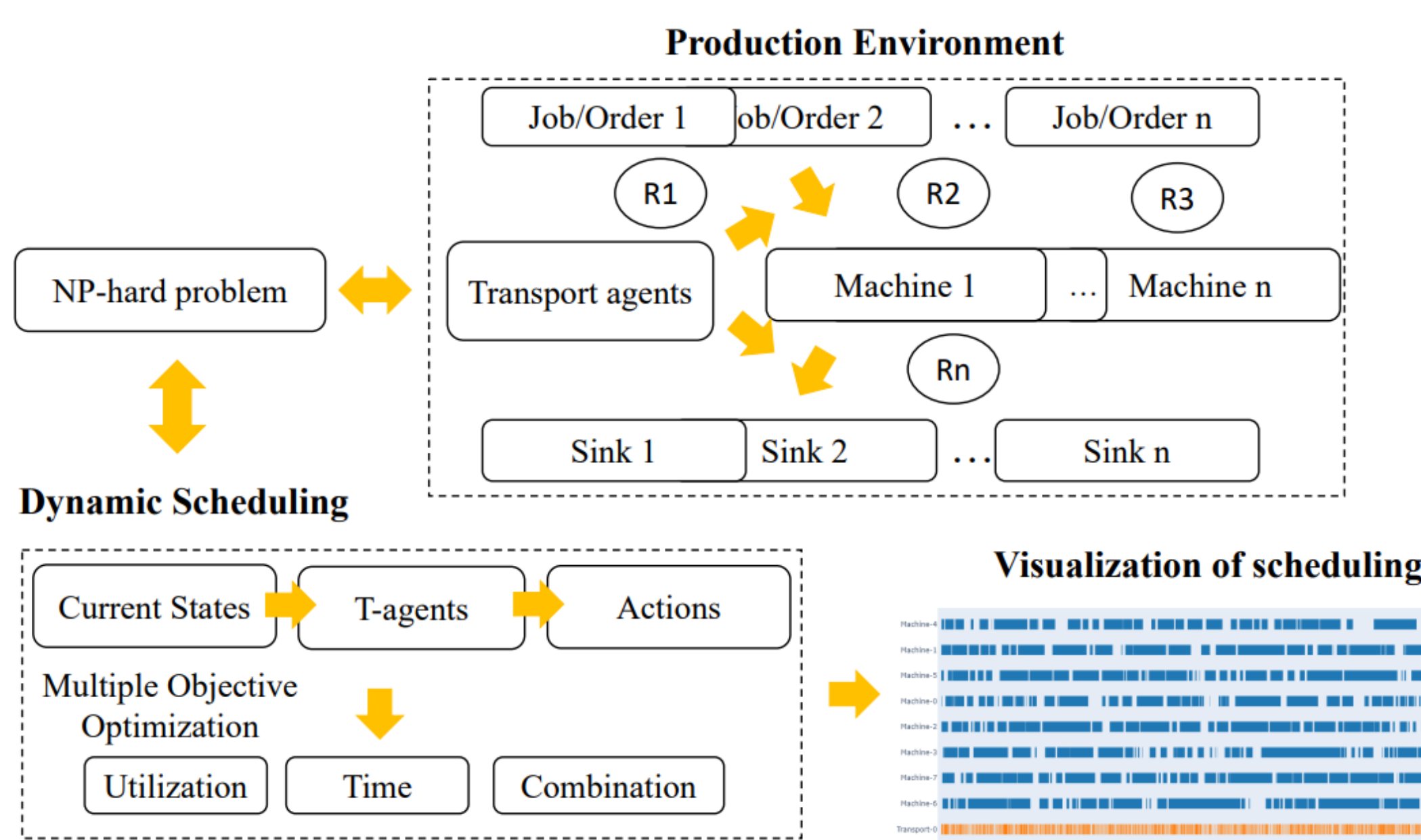
Sustainable Manufacturing



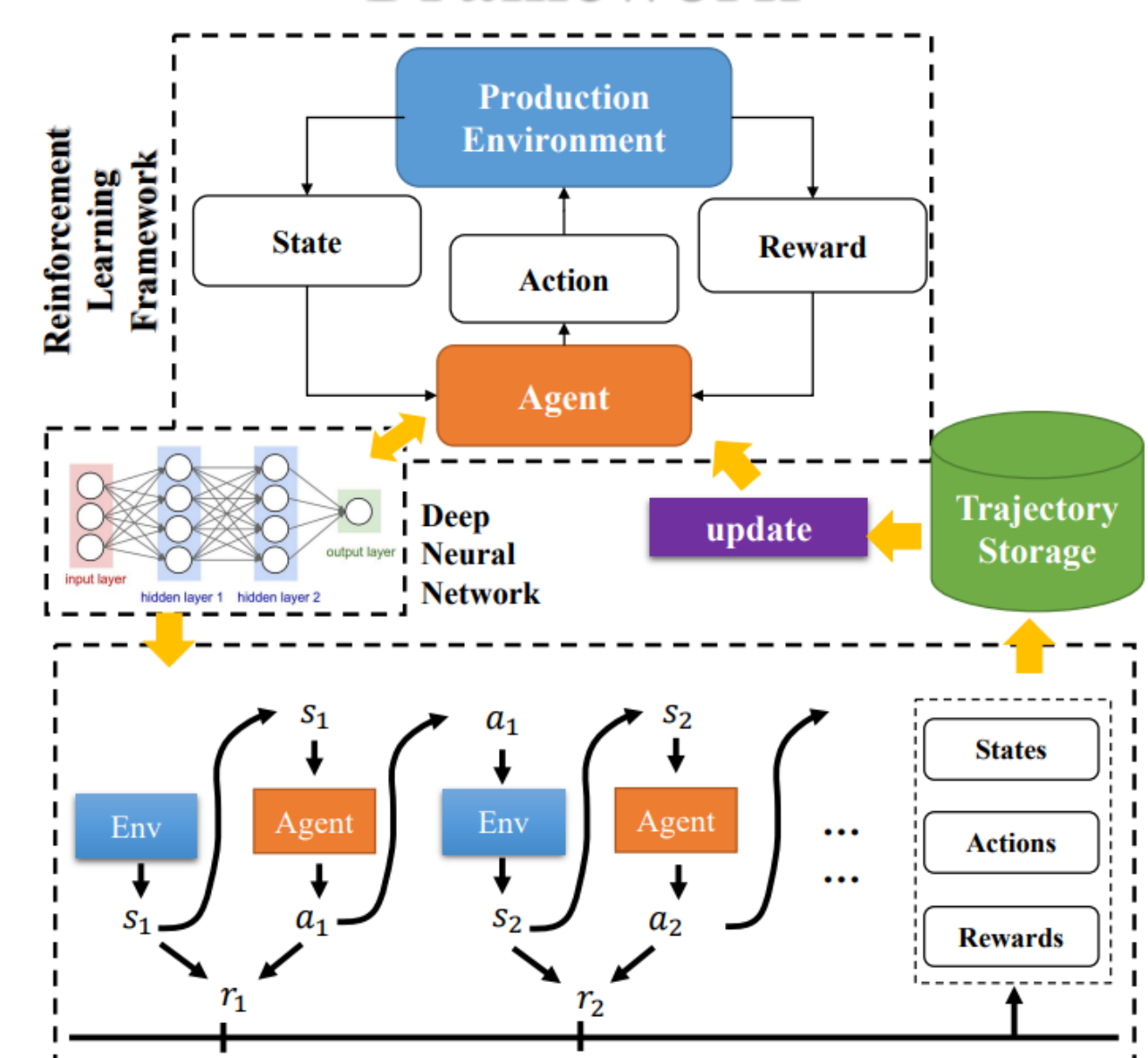
Production High Efficiency

## 3. Method

### Dynamic Job-Shop Scheduling Problem

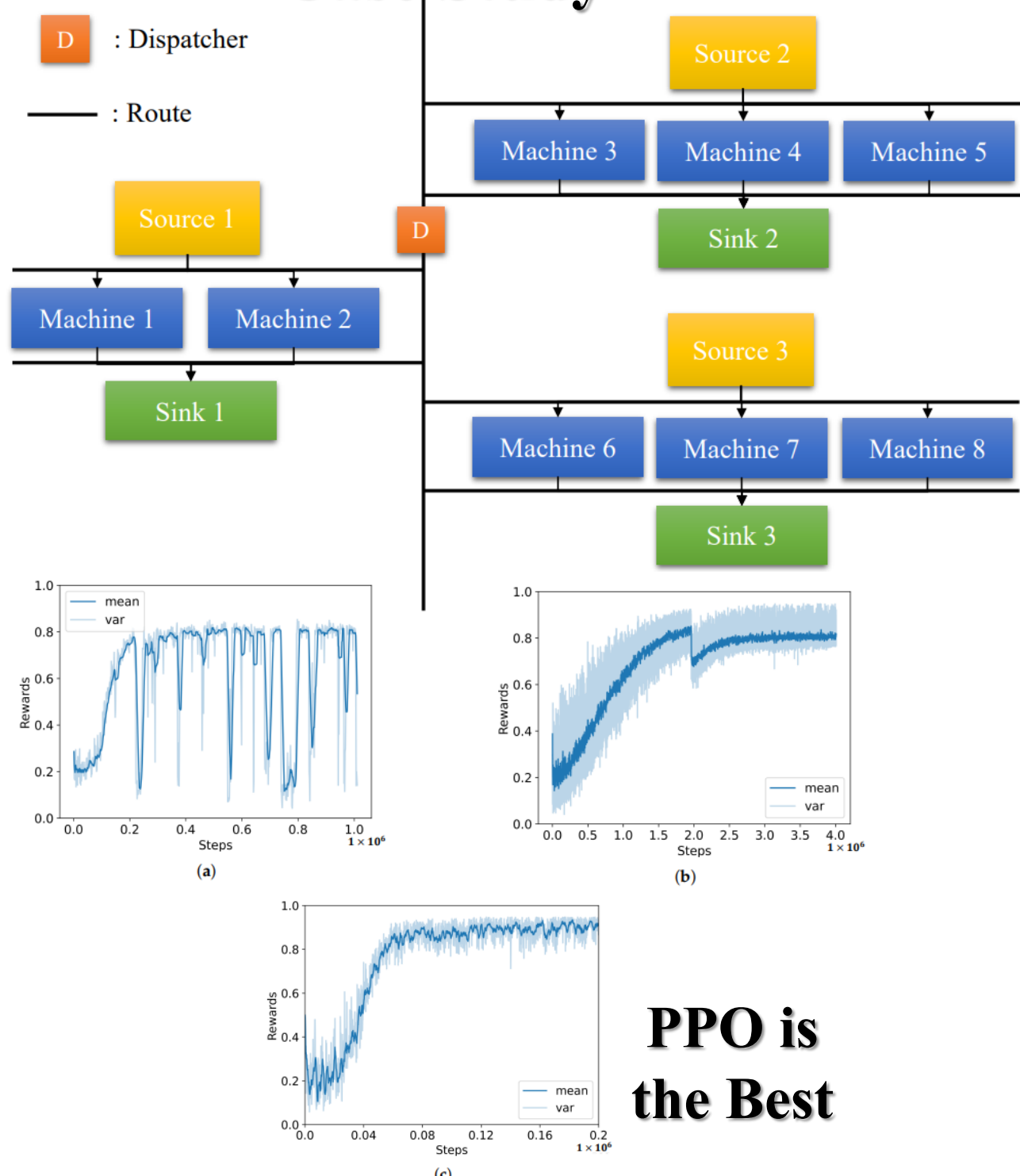


### Deep Reinforcement Learning Framework



## 4. Results

### Case Study



PPO is the Best

Figure 4. Learning process of different algorithms: (a) policy gradient (PG), (b) trust region policy optimization (TRPO), and (c) proximal policy optimization (PPO).

### Sustainability Comparison

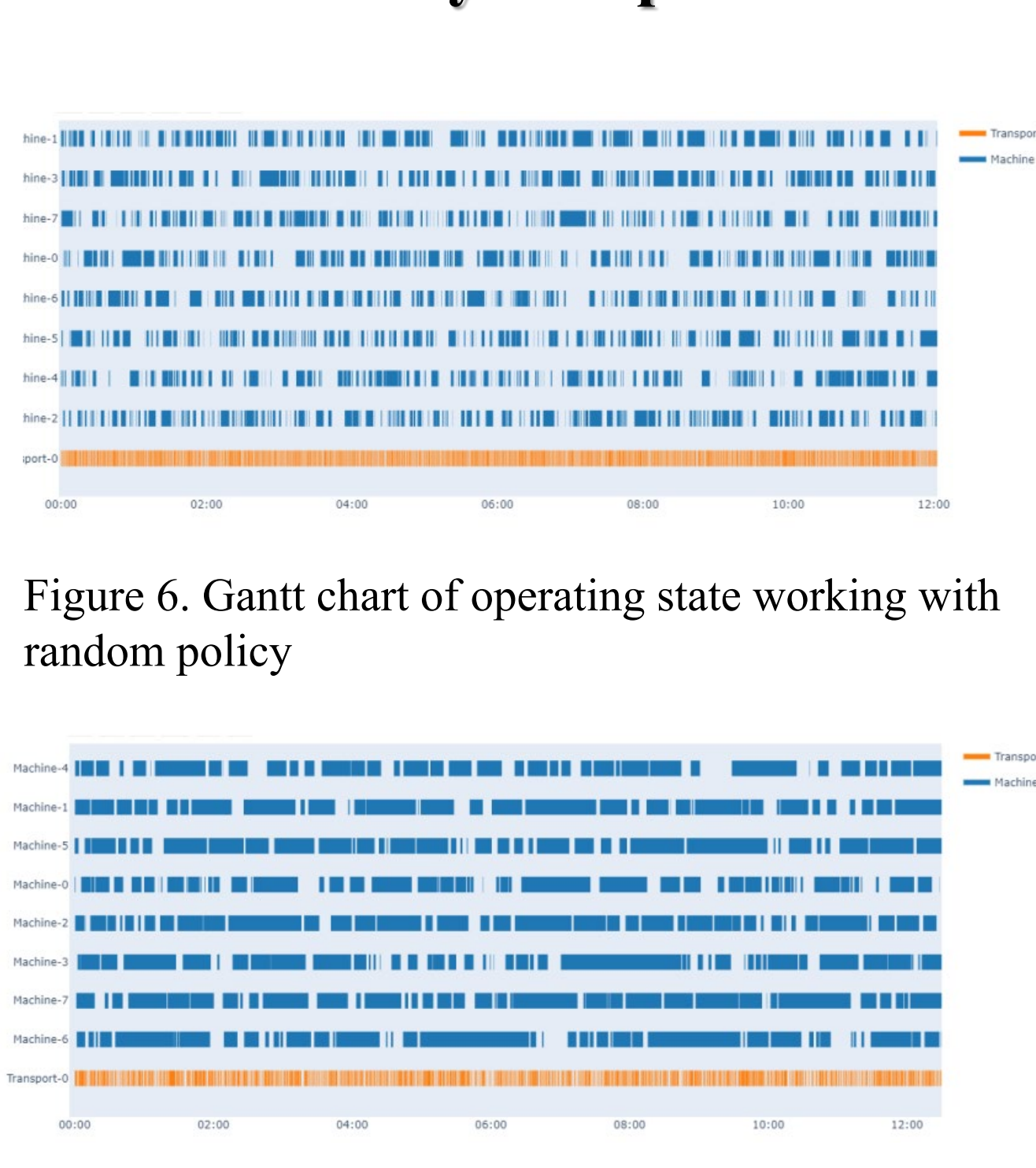


Figure 6. Gantt chart of operating state working with random policy

Figure 7. Gantt chart of operating state working with optimal policy trained by the PPO algorithm

### Reward Function Comparison

Utilization:

$$R_{w-uti}(S_t, A_t) = \begin{cases} \omega_1 R_{uti}(S_t, A_t) & A_t \in A_{S \rightarrow M} \\ \omega_2 R_{uti}(S_t, A_t) & A_t \in A_{M \rightarrow S} \\ 0 & \text{else} \end{cases}$$

Waiting time:

$$R_{w-wt}(S_t, A_t) = \begin{cases} \omega_1 R_{wt}(S_t, A_t) & A_t \in A_{S \rightarrow M} \\ \omega_2 R_{wt}(S_t, A_t) & A_t \in A_{M \rightarrow S} \\ 0 & \text{else} \end{cases}$$

Table 5. Results for PPO dispatching approaches under different reward function in both production scenarios

| PPO          | Scenario 1       |                    |                 |
|--------------|------------------|--------------------|-----------------|
|              | U(%)             | WT(s)              | $\alpha$        |
| $R_{const}$  | 43.20 $\pm$ 3.72 | 119.30 $\pm$ 11.04 | 2.30 $\pm$ 0.63 |
| $R_{w-uti}$  | 44.21 $\pm$ 3.60 | 130.65 $\pm$ 11.51 | 2.37 $\pm$ 0.59 |
| $R_{w-wt}$   | 43.68 $\pm$ 4.11 | 126.61 $\pm$ 12.02 | 2.38 $\pm$ 0.71 |
| $R_{hybird}$ | 43.35 $\pm$ 3.67 | 124.53 $\pm$ 19.15 | 2.32 $\pm$ 0.62 |
| PPO          | Scenario 2       |                    |                 |
|              | U(%)             | WT(s)              | $\alpha$        |
| $R_{const}$  | 62.29 $\pm$ 5.02 | 80.79 $\pm$ 14.87  | 0.56 $\pm$ 0.15 |
| $R_{w-uti}$  | 66.31 $\pm$ 7.09 | 99.87 $\pm$ 20.55  | 0.54 $\pm$ 0.18 |
| $R_{w-wt}$   | 62.03 $\pm$ 5.98 | 80.10 $\pm$ 15.63  | 0.57 $\pm$ 0.18 |
| $R_{hybird}$ | 62.75 $\pm$ 6.99 | 80.56 $\pm$ 17.12  | 0.54 $\pm$ 0.19 |

Multiple-objective:  $R_{hybird}(S_t, A_t) = w_1 R_{uti} + w_2 R_{wt}$

Table 6. Results for different combination of parameters  $\omega_1$  and  $\omega_2$  under reward function  $R_{hybird}$  in production scenario 2.

|                                    | U(%)             | WT(s)              | $\alpha$        |
|------------------------------------|------------------|--------------------|-----------------|
| $\omega_1 = 0.1, \omega_2 = 0.9$   | 61.89 $\pm$ 5.81 | 80.99 $\pm$ 16.14  | 0.57 $\pm$ 0.16 |
| $\omega_1 = 0.25, \omega_2 = 0.75$ | 62.30 $\pm$ 6.08 | 80.35 $\pm$ 14.69  | 0.56 $\pm$ 0.17 |
| $\omega_1 = 0.5, \omega_2 = 0.5$   | 62.75 $\pm$ 6.99 | 80.56 $\pm$ 17.12  | 0.54 $\pm$ 0.19 |
| $\omega_1 = 0.75, \omega_2 = 0.25$ | 68.46 $\pm$ 7.02 | 106.22 $\pm$ 19.30 | 0.48 $\pm$ 0.16 |
| $\omega_1 = 0.9, \omega_2 = 0.1$   | 69.79 $\pm$ 7.16 | 104.88 $\pm$ 20.29 | 0.44 $\pm$ 0.16 |

## 5. Conclusion

The deep reinforcement learning framework with the PPO algorithm has been approved as a suitable solution to the dynamic scheduling problems in the manufacturing environment. This research is still in the initial phase. However, it shows **the powerful potential of data-driven AI-based methods to significantly enhance the manufacturing process**.

Dr Ming Zhang  
m.zhang21@aston.ac.uk

Prof. Yuchun Xu  
y.xu16@aston.ac.uk

Dr Chao Liu  
liuc16@aston.ac.uk