

Deep reinforcement learning-based approach for dynamic disassembly scheduling of end-of-life products with stimuli-activated self-disassembly

Muyue Han, Lingxiang Yun, Lin Li *

Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, IL, 60607, USA



ARTICLE INFO

Handling Editor: Cecilia Maria Villas Boas de Almeida

Keywords:
End-of-life management
Dynamic scheduling
Multi-agent deep reinforcement learning
Stimuli-activated self-disassembly

ABSTRACT

Remanufacturing is one of the most critical strategies for end-of-life product management to promote a circular economy; however, it has been seen very limited implementation due to the labor-intensive and time-consuming disassembly processes for component retrieval. The newly emerged 4D printing technology enables the fabrication of stimuli-responsive reconfigurable structures, outlining new ways to achieve non-destructive and simultaneous self-disassembly of components with different geometry. However, large uncertainties and increased process dynamics have also emerged directly pertaining to the real-time scheduling in disassembly lines with self-disassembly workstations, which the existing scheduling methods are not equipped to handle. In this study, a constrained multi-agent deep reinforcement learning approach is proposed to maximize the disassembly profit by dynamically changing the batch mixing ratios of different-sized components in self-disassembly workstations and adapting real-time scheduling to stochastic product quality, changes in operational sequences, and self-disassembly failures. The proposed approach is validated on a disassembly line for hand pulse detectors that contain heat-activated self-disassembly components. Numerical results show that the proposed achieves stable convergence under uncertainties, and the implementation of a dynamic batch mixing scheme in self-disassembly operations yields a substantial improvement in disassembly profit over the scheduling period. In addition, sensitivity analyses are conducted to evaluate the impacts of system uncertainties on the profitability of the disassembly line.

1. Introduction

The last two decades have witnessed rising productivity growth in the manufacturing sector with rapid product replacement. Meanwhile, significant concerns have arisen from the disposal of large quantities of end-of-life (EOL) products in landfills, which often contain valuable components and recoverable materials (Bockholt et al., 2020; Jyothi et al., 2020). In response to these concerns and regulatory pressures towards a circular economy, remanufacturing has emerged as a crucial closed-loop cycle for the retrieval, restoration, and reuse of high-value components (Glöser-Chahoud et al., 2021; Ullah et al., 2021). Recent research has focused on promoting the effective adoption of remanufacturing, incorporating various perspectives, including the determination of value-oriented reprocessing strategies for EOL components (Jiang et al., 2019; Meng et al., 2020), closed-loop supply chain management with remanufacturing (Konstantaras et al., 2021; Reddy and Kumar, 2020), as well as life cycle assessment and economic evaluation of remanufactured products (Xiong et al., 2020; Zhang et al., 2020).

However, remanufacturing has seen limited implementation, primarily due to the time-consuming disassembly processes required to retrieve core components from the returned products (Cong et al., 2019; Ong et al., 2021).

Disassembly processes often involve a high proportion of manual operations. Recovery profit may be compromised due to the stochastic conditions of returned products and insufficient coordination at the operational level in disassembly lines. In the current literature, there has been a significant focus on disassembly sequence planning (Feng et al., 2019; Ren et al., 2020; Xu et al., 2020) and disassembly line balancing problems (Gao et al., 2020; Guo et al., 2022), which primarily aim to determine optimal disassembly sequences for given products (Guo et al., 2018, 2021) and minimize the total number of workstations required in a disassembly system, as well as to find the best combinations of disassembly tasks for each workstation to ensure a balanced allocation of disassembly steps (Liang et al., 2022; Wang et al., 2020). In practice, disassembly lines allow valuable component retrieval at different disassembly levels (Zhou et al., 2022). There is a critical need to

* Corresponding author.

E-mail address: linli@uic.edu (L. Li).

schedule and coordinate disassembly tasks and resource allocation that maximize the profit and process efficiency of the disassembly line over the given time horizon.

In the current literature, various mathematical programming and (meta)heuristic approaches have been proposed for solving disassembly scheduling problems, incorporating capacity constraints (Yuan et al., 2022; Zhou et al., 2022), uncertain quality levels of used product (Darghouth and Abdel-Aal, 2021; Tian and Zhang, 2019), stochastic processing time (Sathish, 2019; Slama et al., 2021), environmental and safety hazards (Lee et al., 2022; Malekkhouyan et al., 2021). For example, Slama et al. (2020) develop a mixed-integer programming method to maximize the average profit in disassembly lines with defective parts. Fu et al. (2021) propose a discrete fruit fly optimization-based algorithm to minimize the total tardiness in an integrated disassembly-reprocessing system with stochastic task time. However, most existing methods are static scheduling, which often requires prior knowledge of disassembly task characteristics and uses probability distributions or fuzzy values for uncertainty representation. These methods may lack flexibility in dynamically adapting the optimal scheduling solution to unforeseen events during disassembly operations (Gao et al., 2020; Wurster et al., 2022). Additionally, the benefits from these approaches are restricted to the nature of “one-to-one” disassembly, where one disassembly action can only separate one connection at a time.

In recent years, additive manufacturing of stimuli-responsive materials, known as 4D printing, has emerged as a promising technology for manufacturing smart structures that can change shape or properties over time in response to external stimuli, such as heat, electricity, and light (Han et al., 2022b; Zhao et al., 2021). 4D printing finds potential applications in various fields, such as flexible electronics (H. Zhang et al., 2022), self-locking devices (Tahouni et al., 2021), and self-assembly furniture (J. Zhang et al., 2022). Inspired by the stimuli-response mechanisms, 4D printing offers a new approach for achieving non-destructive self-disassembly for component retrieval from EOL products. The self-disassembly mechanisms can be embedded in product components and activated by predetermined external stimuli. For example, Mazurek-Budzyńska et al. (2022) demonstrate the self-disassembly of a 4D-printed fastening device upon exposure to temperature. Unlike labor-intensive “one-to-one” disassembly, the presence of stimuli-responsive reconfigurable elements allows for the simultaneous disassembly of multiple sub-assemblies of returned products. The stimuli-activated self-disassembly allows for minimal or no physical contact between the operator or machine tool and the assemblies, eliminating the need for additional positioning and tool changes. This approach theoretically enables a more efficient “one-to-many” disassembly process.

Despite the advantages of stimuli-activated self-disassembly, it also raises new complexities and uncertainties in disassembly scheduling. The self-disassembly feature adds additional dynamics and uncertainty in the flow of sub-assemblies among workstations and task assignments, making the existing methods less effective. Specifically, since self-disassembly components at different disassembly levels may have similar or identical stimuli-triggering mechanisms, parts of various sizes can be grouped as a mixed batch and undergo stimuli-triggering simultaneously, allowing for parallel processing. This introduces new challenges in disassembly scheduling, particularly with regard to managing the batch mixing conditions (e.g., types and quantities) of heterogeneous sub-assemblies and adapting to potential self-disassembly failures in real-time through the re-assignment of the failed parts to manual disassembly workstations.

To fill the abovementioned research gaps, this study proposes a deep reinforcement learning-based dynamic scheduling approach for EOL product disassembly with stimuli-activated self-disassembly. Specifically, a constrained multi-agent deep reinforcement learning (C-QMIX) approach is developed to maximize the cumulative profit over a scheduling horizon, incorporating the dynamic batch mixing schemes in

self-disassembly processes, uncertainties in returned product quality and residual values, failures and sequence changes in disassembly operations, and line-side buffer capacity constraints. A dynamic scheduling problem is designed based on a disassembly line of the hand pulse detector, and numerical experiments are conducted to evaluate the performance of the proposed approach in generating real-time schedules in response to unanticipated events in product disassembly. The proposed method will facilitate the disassembly scheduling toward efficient and non-destructive mass disassembly of EOL products while responding adaptively to process uncertainties. The outcome of this work will help promote the adoption of emerging manufacturing technology to aid a smooth transition from the traditional “take-make-dispose” linear regime to a circular economy.

The rest of this paper is organized as follows. The problem formulation is presented in Section 2, followed by the demonstration of the proposed C-QMIX algorithm in Section 3. Numerical experiments and results are discussed in Section 4. Finally, conclusions and discussions of future work directions are presented in Section 5.

2. Problem formulation

2.1. Problem description

In this study, a flexible disassembly line with stochastic characteristics is designed to investigate the dynamic scheduling problem involving stimuli-activated component self-disassembly. Specifically, the disassembly line consists of traditional manual (MD) workstations and self-disassembly (SD) workstations. The manual workstations can be assigned one of their feasible disassembly operations as needed. In the self-disassembly workstation, the stimuli-activation is assumed to follow a heat-activated shape-morphing mechanism, which is one of the most representative stimuli-response mechanisms demonstrated in the literature (Demoly et al., 2021; Han et al., 2020, 2022a). The self-disassembly workstation can process parts of varying sizes in mixed batches (i.e., a combination of products/sub-assemblies with different geometry in a single batch), as long as the total volume of the assigned parts does not exceed the workstation’s capacity. In addition, the scheduling problem takes into account uncertainties in the residual values of disassembled components, the possibility of missing components, and the potential failure of the self-disassembly processes.

More precisely, it is supposed that the disassembly line consists of N^{MD} manual disassembly workstations (the i^{th} manual workstation is denoted as W_i^{MD}), and N^{SD} self-disassembly workstations (the j^{th} self-disassembly workstation is denoted as W_j^{SD}). A returned product needs to undergo N^{op} operations to be completely disassembled (the k^{th} operation is denoted as O_k). The sub-assemblies S_k of the returned products that need to undergo the k^{th} operation are temporarily stored in the line-side buffer, denoted as B_k . The disassembly operation is assumed to follow the first-in-first-out (FIFO) rule, i.e., the first sub-assembly that is placed into the buffer is the first to be disassembled. If a sub-assembly has missing components and thus can skip the k^{th} operation, this sub-assembly is said to be ready for succeeding operations. In addition, if a self-disassembly failure occurs at W_j^{SD} , the sub-assembly is then reassigned to a MD workstation for manual disassembly (see Table 1). Based on the above descriptions, the objective of the scheduling problem is to maximize the profit within a scheduling horizon T under buffer capacity constraint, which is determined by subtracting the disassembly costs from the total residual value of all disassembled components. To mathematically characterize the dynamic scheduling problem, a list of notations is presented as follows.

The framework of the proposed C-QMIX scheduling algorithm is shown in Fig. 1, which contains training and testing phases. In the training phase, the scheduling problem is repetitively solved in the stochastic simulation environment to optimize the parameters for all the agents, which enables the agents to learn the best action under each

Table 1
Table of notations.

$a_{i,t}^{MD}$, $a_{j,t}^{SD}$	The action taken by the MD agent i /SD agent j at time t	$r_{i,k}, r_{j,k}$	Expected profits for MD agent i /SD agent j if it chooses O_k
$b_{k,t}$	Number of S_k stored in B_k	S_k	The root item/sub-assembly needs to be disassembled through operation O_k
B_k	The line-side buffer for operation O_k	s_t	The system state at time step t
b_k^{cap}	The maximum capacity of B_k	$v_{k,r}$	Residual value obtained after disassembling S_k
d_i, d_j	Operation costs if MD agents i /SD agents j choose non-idle action	W_i^{MD}, W_j^{SD}	The manual workstation i /self-disassembly workstation j
D_k	Sub-assemblies that require further operations after O_k	$\delta_{i,k}^{MD}, \delta_{i,k}^{SD}$	Number of S_k stored in the B_k that the W_i^{MD}/W_j^{SD} can disassemble within one time step
D_k^{alt}	The index of the alternative MD workstation if the self-disassembly operation O_k fails	μ_k	Mean value of residual value of the retrieved components after O_k
M_j	The index set of feasible operations for W_j^{SD}	σ_k	Standard deviation of residual value of the retrieved components after O_k
N^{MD}, N^{SD}	Number of MD/SD workstations	φ_k	Probability that a sub-assembly can skip O_k due to missing parts
N^{op}	Number of operations required to fully disassemble a product	ω_k	The probability of self-disassembly failure for S_k in W_j^{SD}
$o_{i,t}, o_{j,t}$	The observation of MD agent i /SD agent j at time t	e	The probability that an agent randomly choosing an action
O_k	The k^{th} operation	γ	Discount factor in C-QMIX
$p_{r,i,k}$	Disassembly rate of W_i^{MD} when it performs operation O_k	τ	Update rate in target network parameter soft update

state. In the testing phase, the performance of these well-trained networks is evaluated in terms of the effectiveness in solving the dynamic scheduling problem in a flexible disassembly line with stochastic behaviors. Specifically, the simulation environment is formulated based on the constrained Markov decision process (CMDP), which imitates the dynamic scheduling environment and the interactions among multiple agents (note that each agent represents one workstation) in the disassembly line with flexible task assignment and capacity constraints of line-side buffers. The details of the CMDP are discussed in the following section.

2.2. Constrained markov decision process

The CMPD is defined as a 6-tuple $\langle \mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{C} \rangle$. Specifically, \mathcal{S} represents the state of the entire system, \mathcal{O} stands for the set of all possible observations, and \mathcal{O}_i represents the observation of agent i . The term \mathcal{A} defines the set of actions for all agents, and the set of actions for agent i is denoted as \mathcal{A}_i . Each agent selects its action based on its own observation \mathcal{O}_i and policy $\pi_i: \mathcal{O}_i \mapsto \mathcal{A}_i$, which results in the following system state \mathcal{S}' based on the Markov transition function $\mathcal{T}: \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}'$. At the end of each time step, all agents receive a reward \mathcal{R} indicating the quality of their collaboration, which is a function of the system state and action: $\mathcal{S} \times \mathcal{A} \mapsto \mathcal{R}$. In addition to the Markov decision process, the CMPD consists of a constraint set \mathcal{C} , and the constraint set for agent i , denoted as \mathcal{C}_i , is defined as:

$$\mathcal{C}_i = \{c_{i,v} : \mathcal{O}_i \times \mathcal{A}_i \mapsto \mathbb{R} | v \in [1, V_i]\} \quad (1)$$

where $c_{i,v}$ represents the v^{th} constraint for agent i , and V_i is the number of constraints for agent i . To avoid constraint violation, the safe action set for agent i , denoted as \mathcal{A}'_i , is defined as:

$$\mathcal{A}'_i = \{a_i \in \mathcal{A}_i | c_{i,v}(o_i, a_i) \leq \beta_v, \forall v \in [1, V_i]\} \quad (2)$$

where o_i and a_i respectively represent a specific observation and action for agent i , and β_v is the boundary of the v^{th} constraint. Specifically, the CMDP in the disassembly scheduling problem is formulated as follows.

2.2.1. Action, state, and observation

There are two types of agents in the proposed C-QMIX algorithm, representing the manual workstations and the self-disassembly workstations, respectively. The action set for the first type of agent contains all its feasible operations and idle. An example is presented in Fig. 2 to better clarify the descriptions of action, state, and observation involved in C-QMIX. As shown in Fig. 2, the returned product (i.e., the root item, denoted as S_1) can be disassembled into sub-assemblies (i.e., S_2 to S_4) and further dismantled to five main components (i.e., the leaf items, signified as circles containing numerical numbers). To visualize the disassembly procedure, the required operations for disassembling S_1 to S_4 are represented by rounded rectangles, and the precedence relations among disassembly operations are highlighted by blue elbow arrow connectors. In addition, the rounded rectangles with yellow infill indicate the self-disassembly operations that can take place under the same stimuli-triggering condition.

Note that the operation for disassembling S_k is denoted as O_k , and the buffer for storing S_k is denoted as B_k . In the above example, O_1 and O_2 are manual disassembly operations, whereas O_3 and O_4 represent self-

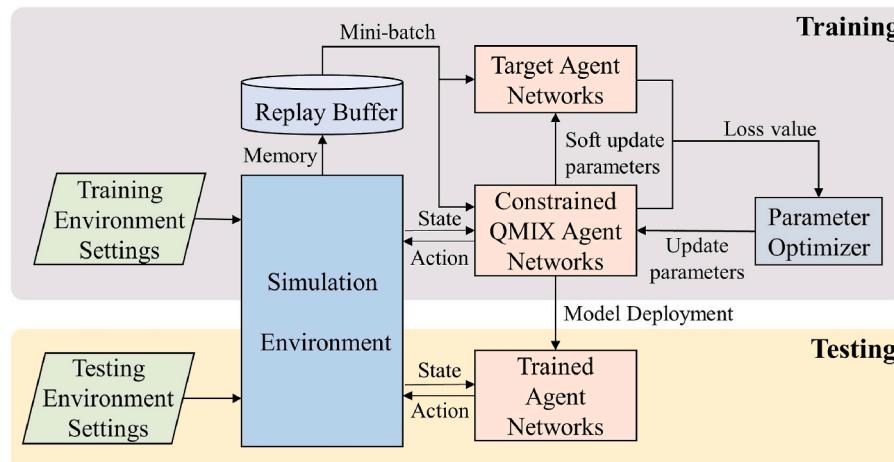


Fig. 1. The framework of the proposed DRL-based scheduling method.

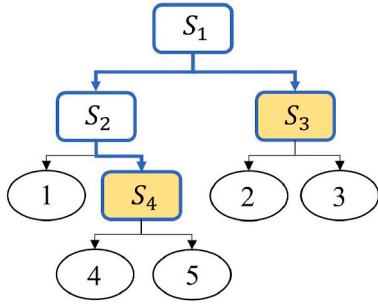


Fig. 2. Demonstration of disassembly sequence.

disassembly operations. In the case of an unsuccessful stimuli-triggering of self-disassembly, the undissembled S_3 and/or S_4 are forwarded to designated MD workstations for further manual disassembly, denoted by O_5 and O_6 , respectively. Suppose that W_i^{MD} can perform O_1 , O_2 , and O_5 , the action set of the agent controlling W_i^{MD} is $\mathcal{A}_i^{MD} = \{0, 1, 2, 5\}$, where action 0 represents the idle option, and other actions represent executing O_1 , O_2 , and O_5 , respectively. For the agent representing W_j^{SD} , its action set contains all possible batch mixing ratios of root item/sub-assemblies containing self-disassembly parts. For example, assuming that the percentages of W_j^{SD} 's available space filled with a single S_3 or S_4 are 20% and 30%, respectively, and let $M_j = \{3, 4\}$ denote the index set of feasible operations for W_j^{SD} , thus the action set of the agent controlling W_j^{SD} is $\mathcal{A}_j^{SD} = \{0, (0, 3), (2, 2), (3, 1), (5, 0)\}$. In this case, action 0 represents the idle option, and the other four actions represent the mixing quantities of S_3 and S_4 with respect to different mixed batch schemes to which W_j^{SD} is filled to its maximum capacity. The actions taken by all agents at time t can be represented by a vector a_t , which is defined as:

$$a_t = (a_{1,t}^{MD}, \dots, a_{i,t}^{MD}, \dots, a_{N^{MD},t}^{MD}, a_{1,t}^{SD}, \dots, a_{j,t}^{SD}, \dots, a_{N^{SD},t}^{SD}) \quad (3)$$

where $a_{i,t}^{MD}$ represents the action taken by the MD agent controlling W_i^{MD} at time t , $a_{i,t}^{MD} \in \mathcal{A}_i^{MD}$, and $a_{j,t}^{SD}$ represents the action taken by the SD agent controlling W_j^{SD} at time t , $a_{j,t}^{SD} \in \mathcal{A}_j^{SD}$.

In the simulation environment of the disassembly line, the stochastic characteristics can be reflected as the uncertain quantity and quality of S_k stored in the k^{th} line-side buffer. The system state contains the information on the quantity of S_k in the disassembly line, which is defined as:

$$s_t = (b_{1,t}, \dots, b_{k,t}, \dots, b_{N^{op},t}) \quad (4)$$

where s_t represents the system state at time step t , and $b_{k,t}$ represents the number of S_k stored in B_k . In addition, each agent can also observe the expected profit corresponding to each non-idle action. The observations of agent i is defined as:

$$o_{i,t} = (s_t, r_{i,1}, \dots, r_{i,|\mathcal{A}_i^{MD}|-1}) \quad (5)$$

where $o_{i,t}$ represents the observation of MD agent i at time t ; $r_{i,1}$ to $r_{i,|\mathcal{A}_i^{MD}|-1}$ represent the expected profits for agent i 's all non-idle actions. Similarly, the observation of SD agent j is defined as:

$$o_{j,t} = (s_t, r_{j,1}, \dots, r_{j,|\mathcal{A}_j^{SD}|-1}) \quad (6)$$

2.2.2. Markov state transition

In the Markov decision process, the system state at time $t+1$ only depends on the state and actions at time t . Specifically, the number of S_k stored in the line-side buffer B_k that the W_i^{MD} can disassemble within one time step, denoted as $\delta_{i,k}^{MD}$, can be calculated as:

$$\delta_{i,k}^{MD} = \begin{cases} \min(b_{k,t}, pr_{i,k}), a_{i,t}^{MD} \neq 0, k = a_{i,t}^{MD} \\ 0, \text{otherwise} \end{cases} \quad (7)$$

where $pr_{i,k}$ stands for the disassembly rate (unit per time step) of W_i^{MD} when it performs operation O_k . Similarly, the number of S_k in line-side buffer B_k that W_j^{SD} can process in one time step, denoted as $\delta_{j,k}^{SD}$, can be calculated as follow.

$$\delta_{j,k}^{SD} = \begin{cases} \min(b_{k,t}, a_{j,t}^{SD}(m)), a_{j,t}^{SD} \neq 0, \forall m \in \{1, \dots, |M_j|\}, k = M_j(m) \\ 0, \text{otherwise} \end{cases} \quad (8)$$

After all the agents select their actions, items in B_k with respective quantities of $\delta_{i,k}^{MD}$ and $\delta_{j,k}^{SD}$ are transferred to workstations W_i^{MD} and W_j^{SD} for disassembly. Hence, the system state at the beginning of time t is updated by:

$$b_{k,t} = b_{k,t} - \sum_{i=1}^{N^{MD}} \delta_{i,k}^{MD} - \sum_{j=1}^{N^{SD}} \delta_{j,k}^{SD} \quad (9)$$

At the end of time t , the disassembled components are removed from the disassembly line, and the sub-assemblies that require further operations are moved into downstream buffers. Let D_k denote the index set of downstream buffers after the operation O_k . Referring to the example shown in Fig. 2, the downstream buffers for O_1 are B_2 and B_3 , i.e., $D_1 = \{2, 3\}$; and there is no downstream buffer for operation O_4 , i.e., $D_4 = \emptyset$.

Considering the EOL products with missing components, it is supposed that the probability that a sub-assembly can skip the operation O_k due to missing parts is φ_k , and thus it can be placed directly into the downstream buffer D_k . Hence, at the end of time t , the quality status of each item in W_i^{MD} can affect the inventory level of downstream buffers through the following rules.

With probability $1 - \varphi_k$:

$$b_{k,t+1} = b_{k,t} + 1, D_k \neq \emptyset, k' \in D_k \quad (10)$$

With probability φ_k :

$$b_{k',t+1} = b_{k',t} + 1, D_k, D_{k'} \neq \emptyset, k' \in D_k, k'' \in D_{k'} \quad (11)$$

As an example, suppose that a product is disassembled through O_1 and with the probability $\varphi_2 = 0.1$ and $\varphi_3 = 0$, then at the end of time t , the number of sub-assemblies in B_2 is increased by one with a probability of 0.9, and the number of sub-assemblies in B_4 is increased by one with a probability of 0.1. In either case, the number of sub-assemblies in B_3 is increased by one.

In terms of the self-disassembly operations, let ω_k denote the probability of self-disassembly failure for S_k in W_j^{SD} , and let D_k^{alt} denote the index of the alternative manual disassembly workstation if the self-disassembly operation O_k fails. Referring to the example shown in Fig. 2, $D_3^{alt} = 5$. Hence, at the end of time t , the disassembly status of each item in W_j^{SD} can affect the inventory level of downstream buffers through the following rules.

With probability $1 - \omega_k$:

$$b_{k,t+1} = b_{k,t} + 1, D_k \neq \emptyset, k' \in D_k \quad (12)$$

With probability ω_k :

$$b_{k',t+1} = b_{k',t} + 1, k' = D_k^{alt} \quad (13)$$

Finally, the system state at time $t+1$ can be updated as follow.

$$s_{t+1} = (b_{1,t+1}, \dots, b_{k,t+1}, \dots, b_{N^{op},t+1}) \quad (14)$$

In summary, the Markov state transition $\mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}'$ is presented in Algorithm 1. Lines 1 to 7 show state changes at the beginning of time t

when items are transferred from buffers to workstations. Lines 8 to 18 show state changes at the end of time t when items are moved from workstations to downstream buffers.

Algorithm 1 Markov state transition

```

Input:  $s_t, a_t$ 
Output:  $s_{t+1}$ 
1: for all agents of  $W_i^{MD}$  do
2:   Calculate  $\delta_{i,k}^{MD}$  through (6)
3: end for
4: for all agents of  $W_j^{SD}$  do
5:   Calculate  $\delta_{j,k}^{SD}$  through (7)
6: end for
7: Update buffer occupancy through (8)
8: for all agents of  $W_i^{MD}$  do
9:   for each assembly processed by  $W_i^{MD}$  do
10:    Update buffer occupancy through (9) and (10)
11:   end for
12: end for
13: for all agents of  $W_j^{SD}$  do
14:   for each assembly processed by  $W_j^{SD}$  do
15:     Update buffer occupancy through (11) and (12)
16:   end for
17: end for
18: calculate  $s_{t+1}$  through (13)

```

2.2.3. Reward

The objective of the scheduling problem is to maximize the profit over the entire scheduling horizon. Thus, the unified reward of all agents at time t is set as the profit obtained within that time step. The residual value of the retrieved components after O_k is assumed to follow a normal distribution with mean μ_k and standard deviation σ_k . In addition, let d_i and d_j denote the operation costs if MD and SD agents choose non-idle action. Hence, the expected profits of MD and SD agents can be calculated as follows.

$$r_{i,k} = \sum_{r=1}^{\delta_{i,k}^{MD}} (v_{k,r} - d_i) \quad (15)$$

$$r_{j,k} = \sum_{r=1}^{\delta_{j,k}^{SD}} (v_{k,r} - d_j) \quad (16)$$

where $v_{k,r}$ is the residual value obtained after disassembling S_k , and $v_k \sim N(\mu_k, \sigma_k^2)$. The unified reward of all agents at time t , denoted by r_t , is defined as follows.

$$r_t = \sum_{k=1}^{N^{op}} \left(\sum_i^{N^{MD}} r_{i,k} + \sum_j^{N^{SD}} r_{j,k} \right) \quad (17)$$

2.2.4. Constraints and safe action set

In practice, the line-side buffers have limited buffer capacities. Hence, the workstations should not be assigned to disassemble only the most valuable components, allowing the other sub-assemblies to pile up in the buffer with no limit. Specifically, if an action taken for a workstation can cause an exceedance of maximum capacity in its downstream buffer, the selection of this action should be avoided. Let b_k^{cap} denote the maximum capacity of B_k , the constraint and safe action sets for the MD agent are defined as follows.

$$\mathcal{C}_i^{MD} = \left\{ c_{i,k}^{MD} : b_k^{cap} - b_{k,t} - pr_{i,k} \mid \forall k' \in D_k, k = a_{i,t}^{MD} \right\} \quad (18)$$

$$\mathcal{A}_i^{MD'} = \left\{ a_{i,t}^{MD} \in \mathcal{A}_i^{MD} \mid c_{i,k}^{MD} \leq 0 \right\} \quad (19)$$

Similarly, the constraint and safe action sets for the SD agent are defined as follows.

$$\mathcal{C}_j^{SD} = \left\{ c_{j,k}^{SD} : b_k^{cap} - b_{k,t} - a_{j,t}^{SD}(m) \mid a_{j,t}^{SD} \neq 0, \forall m \in \{1, \dots, |M_j|\}, k = M_j(m) \right\} \quad (20)$$

$$\mathcal{A}_j^{SD'} = \left\{ a_{j,t}^{SD} \in \mathcal{A}_j^{SD} \mid c_{j,k}^{SD} \leq 0 \right\} \quad (21)$$

3. Constrained multi-agent deep reinforcement learning

In this study, the architecture of QMIX (Rashid et al., 2018) is adapted for the proposed C-QMIX, as presented in Fig. 3. In particular, QMIX is designed for decentralized execution and centralized training for agents with discrete action space. The decentralized execution allows each agent to have its own network to respond to the stochastic environment, and the centralized training ensures that all agents can learn to cooperate and maximize the profit of the entire disassembly line.

Specifically, the right panel of Fig. 3 shows the Q-network for each agent, which contains three fully connected (FC) layers (i.e., two hidden layers and one output layer), and the ReLU activation function is applied after the hidden layers. The ϵ -greedy is used as training policy, i.e., in the training phase, the agent would randomly choose one action from the safe action set with probability ϵ , and it would choose the safe action associated with the largest Q value with probability $(1 - \epsilon)$. The Q-network outputs the action of the agent's choice and a constrained Q value associated with that action, denoted as Q^c .

As shown in the middle panel of Fig. 3, the calculated Q^c 's of each agent are combined with the current system state s_t as inputs into a mixing network to calculate a single constrained Q value, denoted as Q_{tot}^c , representing the overall performance of all agents in the current state. The structure of the mixing network is presented in the left panel of Fig. 3. The absolute values of state s_t after passing through an FC layer are used as weight coefficients and are multiplied by the Q values of all agents. A bias, i.e., the output of state s_t after passing through another FC layer, is then added to the results, followed by the ELU activation function. A similar procedure is repeated afterward to calculate the Q_{tot}^c , whereas the only difference is that the final bias is calculated through two FC layers with a ReLU non-linearity. It can be proved that the Q_{tot}^c is monotonic in the per-agent Q^c 's, which allows tractable optimization of the entire network in off-policy learning. Detailed proofs are available in (Rashid et al., 2018).

In the training phase, a mini-batch of memories is extracted from the replay buffer \mathcal{D} . The parameter set θ in all agents' Q-networks and mixing network are optimized by minimizing the loss function \mathcal{L} , which is defined as follows.

$$\mathcal{L}(\theta) = \mathbb{E}_{o,a,r,o' \sim \mathcal{D}} \left[(Q_{tot}^c(o, a | \theta) - y_{tot})^2 \right] \quad (22)$$

$$y_{tot} = r + \gamma \bullet \max_{a \in \mathcal{A}} Q_{tot}^c(o', a' | \theta^-) \quad (23)$$

where (o, a, r, o') represents the memory consisting of current observation o , action a , reward r , and the succeeding observation o' ; γ is a discount factor, a' and \mathcal{A} represent the succeeding action and the safe action sets, and θ^- is the parameter set in the target network. The parameters in target network are updated as follows, where τ is the soft update rate. The pseudocode of the C-QMIX is provided in Algorithm 2.

$$\theta^- \leftarrow \tau \theta + (1 - \tau) \theta^- \quad (24)$$

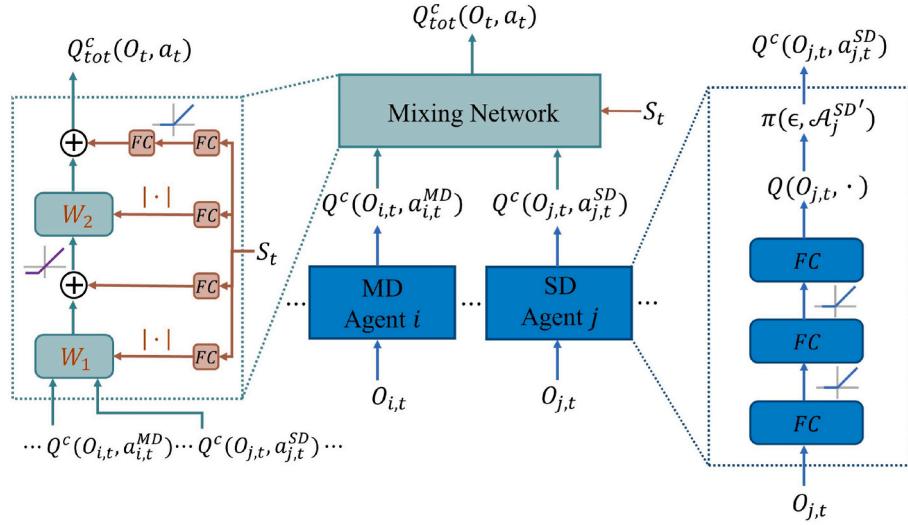


Fig. 3. Architecture of the C-QMIX network.

Algorithm 2 C-QMIX for dynamic disassembly scheduling

Input: Hyper-parameters of C-QMIX
Output: Parameter set θ in C-QMIX

- 1: Randomly initialize constrained Qmix network Q_{tot}^c
- 2: Initialize target constrained Qmix network: $\theta^- \leftarrow \theta$
- 3: Initialize replay buffer \mathcal{D}
- 4: for episode=1,2,... do
- 5: Reset environment
- 6: Initialize state and obtain observations
- 7: for $t=1$ to T do
- 8: for all agents do
- 9: Calculate safe action set $\mathcal{A}_i^{MD'}$ or $\mathcal{A}_j^{SD'}$
- 10: Choose action according to its policy π_i
- 11: end for
- 12: Execute actions, obtain rewards and next state
- 13: Store the memory in replay buffer \mathcal{D}
- 14: for all agents do
- 15: Sample mini-batch from \mathcal{D}
- 16: Calculate safe action set $\mathcal{A}_i^{MD'}$ or $\mathcal{A}_j^{SD'}$
- 17: Set $y_{tot} = r + \gamma \max_{a'_i \in \mathcal{A}'} Q_{tot}^c(o', a' | \theta^-)$
- 18: Update θ by minimizing $\mathcal{L}(\theta)$
- 19: Update target network parameters θ^-
- 20: end for
- 21: end for
- 22: end for

4. Numerical experiments and results

4.1. Construction of numerical experiments

To evaluate the effectiveness of the proposed C-QMIX method for dynamic disassembly scheduling for products containing self-disassembly components, numerical experiments are constructed based on a disassembly line for the EOL hand pulse detector (the device design is adopted from (Sept, 2018)). The exploded view of the main components in the test device and the corresponding disassembly procedures are shown in Fig. 4(a) and (b), respectively.

To investigate the impacts of self-disassembly components on disassembly scheduling, the following assumptions are formulated in the numerical experiment: (1) item 3 is connected with item 1 through a heat-activated snap-fit structure; (2) subassembly S_3 is attached to item 1 and fixed with two shape memory screws; (3) item 7 is a self-disassembly button with a clamping rim and can be released from item 6 upon heat activation. The component descriptions and respective residual values are listed in Table 2. Note that the residual values are estimated based on the geometry design and material types of core components, and methods for residual value estimation for EOL consumer electronics previously described (Tian et al., 2018; Wang et al., 2022).

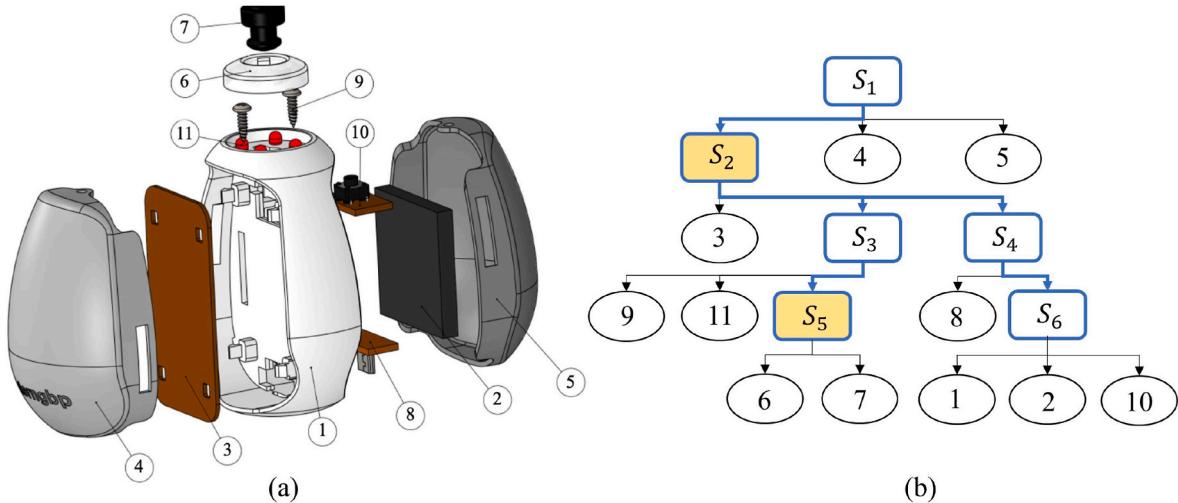


Fig. 4. Numerical study of a hand pulse detector: (a) exploded view of major components and (b) multi-level disassembly structure.

Table 2

General information of the hand pulse detector.

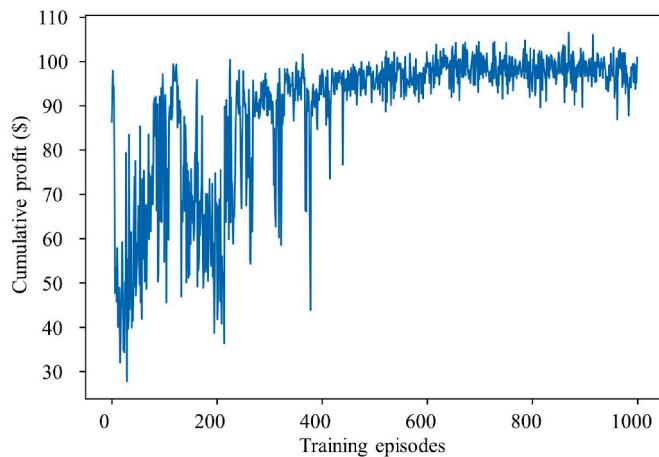
Item No.	Description	Residual value (\$)
1	Main casing	$N(0.3, 0.02)$
2	Battery module	$N(0.6, 0.03)$
3	Printed circuit board	$N(0.4, 0.03)$
4	Top plate	$N(0.2, 0.02)$
5	Bottom plate	$N(0.2, 0.02)$
6	Semi-transparent cover	$N(0.3, 0.02)$
7	Power push button	$N(0.3, 0.02)$
8	Micro USB port	$N(0.3, 0.02)$
9	Pan head screws	$N(0.1, 0.01)$
10	4-pin button switch	$N(0.1, 0.01)$
11	LED indicator	$N(0.1, 0.01)$

It is assumed that the disassembly line consists of four manual workstations and one self-disassembly workstation. Referring to Fig. 4 (b), sub-assemblies S_2 and S_5 contain 4D printed components, enabling simultaneous self-disassembly upon activation by the thermal stimulus. Suppose that S_2 and S_5 fill up 50% and 25% of W_1^{SD} 's available space, respectively, such that the action set of the agent controlling W_1^{SD} is $\mathcal{A}_1^{SD} = \{0, (2, 0), (1, 2), (0, 4)\}$. The action sets of the agents representing manual workstations W_1^{MD} to W_4^{MD} are set as follows: $\mathcal{A}_1^{MD} = \mathcal{A}_2^{MD} = \{0, 1, 6, 8\}$ and $\mathcal{A}_3^{MD} = \mathcal{A}_4^{MD} = \{0, 3, 4, 7\}$, where O_7 and O_8 are the alternative manual operations when failures occur in the self-disassembly of S_2 and S_5 , respectively. The disassembly cost corresponding to executing manual operations and stimuli-triggering self-disassembly are set as \$0.1 and \$0.2, respectively, adopted from (Tian et al., 2018). To incorporate the stochastic features in the dynamic disassembly scheduling, the skipping rate of operation O_1 or O_4 due to component missing is set as 10%, and the failure rate of the self-disassembly operations is set as 30%, adopted from (Ilgin and Gupta, 2011; Riggs et al., 2015), respectively.

4.2. Simulation results

4.2.1. Performance evaluation of the proposed C-QMIX

The convergence performance of the proposed C-QMIX is shown in Fig. 5, where the cumulative profit over the pre-defined scheduling horizon is plotted as a function of training episodes. Specifically, the scheduling horizon in each episode is set as 50 time steps, and the number of EOL products waiting to be disassembled at the beginning of the scheduling horizon is assumed to be 80. In the training phase, the Adam optimizer is used for optimizing the parameters in the QMIX networks with a learning rate of 5×10^{-4} . Each agent's Q-network has two hidden layers (with 64 and 32 neurons per layer, respectively) and

**Fig. 5.** Convergence performance of the proposed C-QMIX.

one output layer. In the mixing network, the hidden layer has 32 neurons to calculate the last bias. The respective sizes for mini-batch and replay buffer are set as 1024 and 5120, and the discount factor γ is set to 0.99. The ϵ is gradually reduced from 0.9 to 0.1 in the training phase, and the soft update rate τ is set as 0.01.

As shown in Fig. 5, the cumulative profit oscillates at the beginning of the training process and gradually converges to around \$100 after 500–600 episodes. The significant fluctuations during the early training phase are primarily attributable to the ϵ -greedy exploration, which allows the agents to randomly choose actions to explore the state-action space, learn policies, and maximize the received rewards during their interactions with the simulation environment. The slight oscillation of the cumulative profit near the end of the training phase is mainly due to the uncertainties in stochastic dynamic environment. Specifically, the part residual value, the occurrence of missing components, and self-disassembly failure are randomly and independently generated for each EOL product in each training episode. Hence, the total cumulative profit fluctuates in each episode even if the DRL agents have learned the best policies.

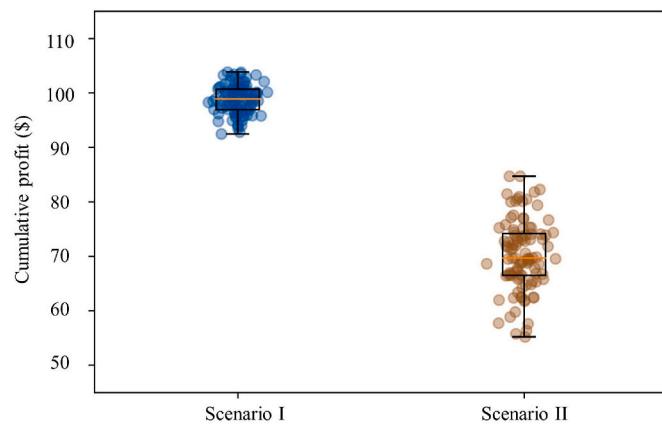
4.2.2. Analysis of dynamic batch mixing in stimuli-triggered self-disassembly

To explore the unique characteristics of simultaneous self-disassembly of multiple subassemblies by stimuli-triggering, comparative numerical studies are performed to examine the effects of batch mixing of different subassemblies. More specifically, the proposed C-QMIX method is used to generate disassembly schedules under two scenarios, i.e., (I) disassembly scheduling incorporating dynamic changes in the batch mixing ratios of subassemblies S_2 and S_5 in W_1^{SD} throughout the scheduling horizon, i.e., $\mathcal{A}_1^{SD} = \{0, (2, 0), (1, 2), (0, 4)\}$; (II) scheduling with a fixed batch mixing ratio of S_2 and S_5 in W_1^{SD} , i.e., $\mathcal{A}_1^{SD} = \{0, (1, 2)\}$. The comparisons of cumulative profits under these two scenarios are shown in Fig. 6. Note that a set of 100 replication experiments are conducted under each scenario, and cumulative profits are obtained over each generated schedule.

As shown in Fig. 6, within the same scheduling horizon, there is a significant increase in the cumulative profit when the dynamic batch mixing scheme is adopted in the disassembly scheduling (the median of the cumulative profit changes from \$69.66 to \$98.86, increased by 41.90%). The results indicate that by adopting the dynamic batch mixing, the respective mixing quantities of sub-assemblies in the self-disassembly workstation are actively adjusted in response to the interaction dynamics and the effects of other agents' actions along the disassembly line, leading to a more efficient disassembly schedule.

4.2.3. Disassembly scheduling under different returned product quantities

Given that the number of received EOL products may vary in the real

**Fig. 6.** Cumulative profits under different batch mixing schemes.

disassembly systems, simulation experiments are performed under two different disassembly settings to demonstrate the effectiveness of the proposed method for dynamic scheduling in disassembly lines. More specifically, in Case A, the number of returned EOL products is limited, and all received products are completely disassembled within the scheduling horizon. Whereas in Case B, the returned products are available in sufficient quantity such that not all the received products can be fully disassembled by the end of the scheduling horizon. The comparisons of scheduling solutions are provided in Fig. 7, where the upper panel, Fig. 7(a) and (b), shows the Gantt chart and the corresponding inventory level changes in line-side buffers under Case A; the lower panel, Fig. 7(c) and (d), shows the scheduling solutions under Case B.

As indicated by the Gantt charts in Fig. 7(a) and (c), in Case A, fewer shifts of batch mixing ratios in the self-disassembly workstation are observed, and workstations tend to remain idle near the end of the scheduling horizon as compared to Case B. The differences in workstation schedules are mainly due to the fact that the scheduling horizon in Case A is sufficiently long to ensure the completion of disassembly operations for all received products. As shown in Fig. 7(b), the inventory levels of all line-side buffers are driven to zero right before the end of the scheduling horizon. In addition, the varying patterns of the inventory levels of line-side buffers also suggest different strategies in dynamic scheduling under two cases. In Case A, since all EOL products can be eventually disassembled, the DRL agents do not show a strong preference for selecting disassembly operations based on the residual values. In comparison, since only partial EOL products can be disassembled within the scheduling horizon in Case B, workstations are scheduled to prioritize the disassembly operations that yield components with higher residual values. As shown in Fig. 7(d), the inventory levels for high residual value sub-assemblies, such as sub-assemblies in B_5 , B_6 , and B_8 , remain below two throughout the scheduling horizon. As a result, the

cumulative profit is higher in Case B (\$99.45) than in Case A (\$93.46), which is also in accordance with the respective scheduling strategies.

4.3. Sensitivity analyses

To investigate the uncertainty factors that may influence the disassembly scheduling, sensitivity analyses are conducted to understand the related impacts on disassembly profit. In particular, three types of quality-related uncertainties of considered: the uncertainty of self-disassembly failure rates, the uncertainty of operation skipping rates, and the uncertainty of the residual value of the retrieved components. It is assumed that a sufficient quantity of EOL products needs to be disassembled for the analyses of quality-related uncertainty factors. In addition, one quantity-related uncertainty, i.e., the uncertainty in the number of returned EOL products, is also considered. The results of 100 replications under each experimental setting are shown in Fig. 8.

The analysis results of quality-related uncertainty factors are shown in Fig. 8 (a)–(c). As shown in Fig. 8 (a), changing the occurrence probability of self-disassembly failure from 10% to 90% results in a 28.08% reduction in the median cumulative profit (decreased from \$108.05 to \$77.71). A major factor behind the decreasing tendency is that when the stimuli-triggering of self-disassembly is unsuccessful, the subassembly is reprocessed at a designated MD workstation, resulting in additional disassembly costs. In the case of varying the skipping rate of manual operation, as depicted in Fig. 8 (b), the median of the cumulative profit shows fewer variations when the skipping rate of manual operation O_1 or O_4 is increased from 10% to 50% and exhibits a significant increase to over \$105 as the skipping rate rises to 90%. When a high skipping probability of manual operation O_1 or O_4 is employed, subassemblies with missing components can be directly forwarded to downstream workstations for high-value components retrieval (such as items 2 and 3 listed in Table 2). Additionally, as shown in Fig. 8 (c), the cumulative

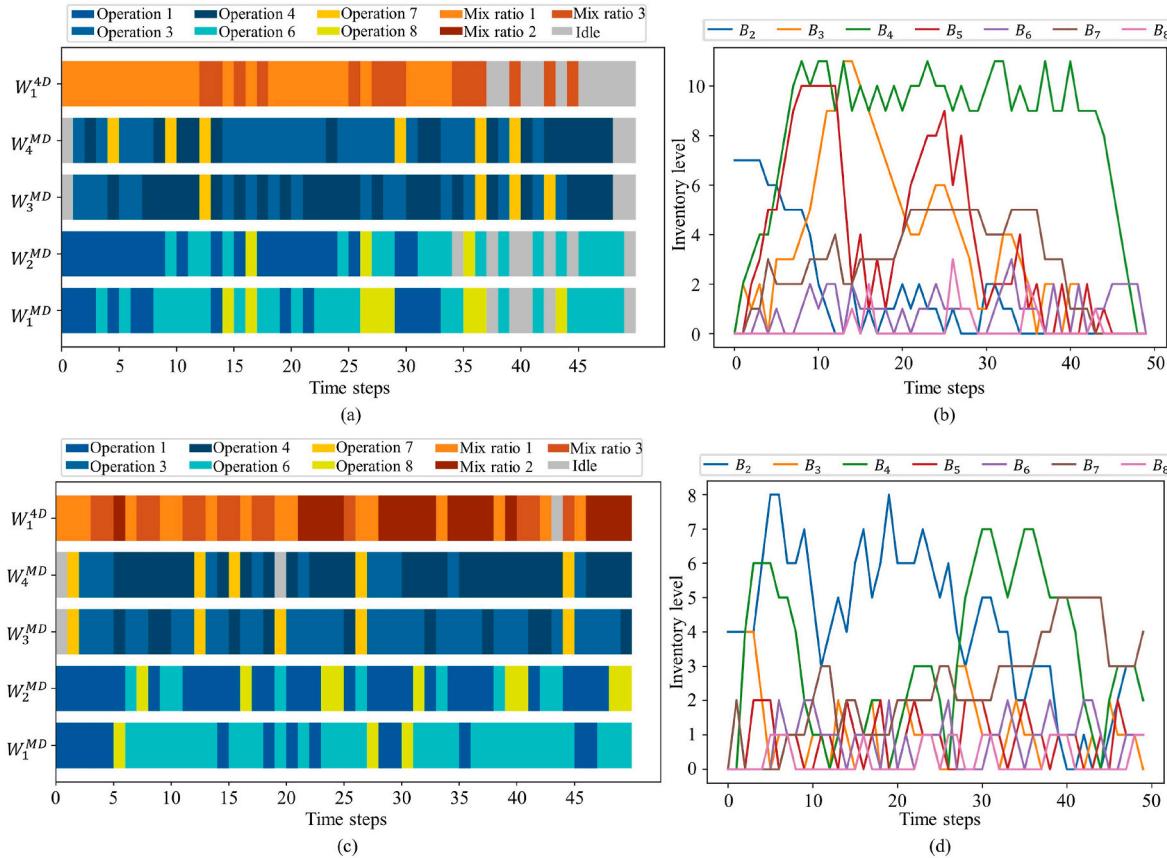


Fig. 7. Gantt charts and the corresponding inventory level changes in line-side buffers under Cases A (a and b) and B (c and d).

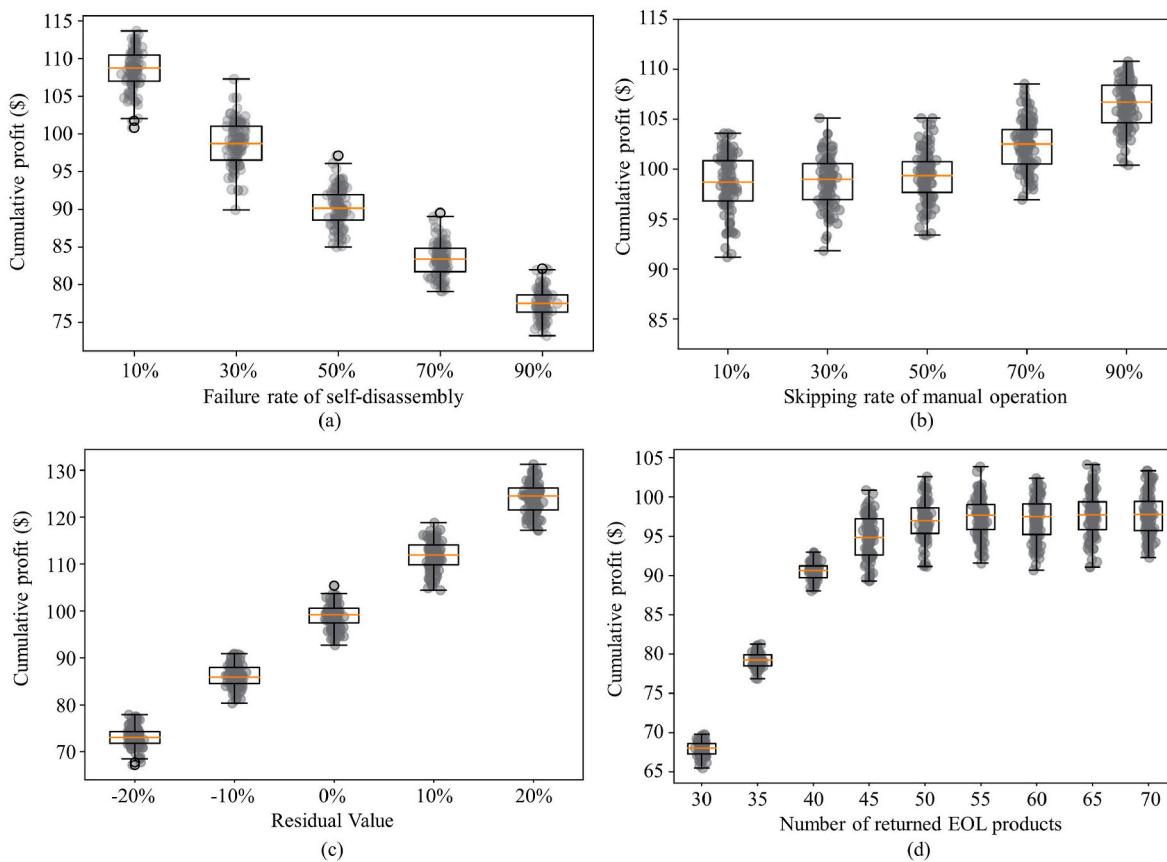


Fig. 8. Impacts of system uncertainties on cumulative profit: (a) failure rate, (b) skipping rates, (c) residual value, and (d) number of returned EOL products.

profit increases linearly with the residual values of core components.

The results of the quantity-related sensitivity analysis are shown in Fig. 8 (d). Specifically, when the number of returned EOL products is less than the number that can be processed within the given time horizon (approximately 45–50 products in this case), the cumulative profit exhibits a linear relationship with the number of EOL products. Furthermore, as all products are eventually disassembled, the cumulative profit demonstrates less fluctuation pertaining to the uncertain component quality/residual values, resulting in a relatively narrow profit distribution. Conversely, when the number of EOL products exceeds the processing capacity within the time horizon, the additional products cannot be disassembled, and as such, the average cumulative profit does not exhibit a significant change. In this scenario, the reinforcement learning agents have more freedom to choose which sub-assemblies to be disassembled before the time limit, which leads to a larger profit distribution due to the uncertain quality/residual values of different products/components.

5. Conclusions and future work

This study considers a dynamic scheduling problem in a unique disassembly system involving 4D printing-enabled self-disassembly of components. The disassembly problem is formulated as a constrained Markov decision process considering different forms of uncertainties, including the possibility of missing components in returned EOL products, potential self-disassembly failure, stochastic quality of disassembled parts, and the uncertain quantity of returned products. In addition, a deep reinforcement learning algorithm C-QMIX is proposed for maximizing the disassembly profit over the scheduling horizon. The C-QMIX algorithm is capable of dynamically altering the MD workstation configurations and batch mixing ratios in SD workstations, coordinating disassembly tasks among different types of workstations, and

eventually reaping the advantages of stimuli-triggered self-disassembly as an effective solution for more efficient EOL product disassembly. The effectiveness of the proposed dynamic scheduling approach is validated through numerical experiments on the disassembly line of a hand pulse detector. The results indicate that the C-QMIX agents can respond to the dynamic and stochastic changes in the disassembly environment and achieve higher profits compared to the transitional disassembly line. Additionally, sensitivity analyses are conducted to investigate the impacts of different uncertainty factors on disassembly profits.

As an extension to this work, the proposed method can be extended to address more complex dynamic scheduling problems in disassembly systems involving different types of returned EOL products and/or components possessing different types of stimuli-triggered self-disassembly mechanisms. In addition, it is also worth modifying the proposed method to address the scheduling and real-time control in multi-product partial disassembly or incorporate additional factors related to the scheduling over consecutive stages, such as disassembly, refurbishment, and reassembly operations in the real remanufacturing system.

CRediT authorship contribution statement

Muyue Han: Conceptualization, Investigation, Validation, Writing – original draft. **Lingxiang Yun:** Investigation, Visualization, Software, Formal analysis, Writing – original draft. **Lin Li:** Conceptualization, Methodology, Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Bockholt, M.T., Hemdrup Kristensen, J., Colli, M., Meulengracht Jensen, P., Vejrum Wahrens, B., 2020. Exploring factors affecting the financial performance of end-of-life take-back program in a discrete manufacturing context. *J. Clean. Prod.* 258, 120916 <https://doi.org/10.1016/j.jclepro.2020.120916>.
- Cong, L., Zhao, F., Sutherland, J.W., 2019. A design method to improve end-of-use product value recovery for circular economy. *J. Mech. Des.* 141 <https://doi.org/10.1115/1.4041574>.
- Darghouth, M.N., Abdel-Aal, M., 2021. A capacitated disassembly scheduling problem considering processing technology selection and parts commonality. *J. Remanufacturing* 11, 243–261. <https://doi.org/10.1007/s13243-021-00104-3>.
- Demoly, F., Dunn, M.L., Wood, K.L., Qi, H.J., André, J.-C., 2021. The status, barriers, challenges, and future in design for 4D printing. *Mater. Des.* 212, 110193 <https://doi.org/10.1016/j.matdes.2021.110193>.
- Feng, Y., Zhou, M., Tian, G., Li, Z., Zhang, Z., Zhang, Q., Tan, J., 2019. Target disassembly sequencing and scheme evaluation for CNC machine tools using improved multiobjective ant colony algorithm and fuzzy integral. *IEEE Trans. Syst. Man, Cybern. Syst.* 49, 2438–2451. <https://doi.org/10.1109/TSMC.2018.2847448>.
- Fu, Y., Zhou, M., Guo, X., Qi, L., 2021. Stochastic multi-objective integrated disassembly-reprocessing-assemble scheduling via fruit fly optimization algorithm. *J. Clean. Prod.* 278, 123364 <https://doi.org/10.1016/j.jclepro.2020.123364>.
- Gao, K.Z., He, Z.M., Huang, Y., Duan, P.Y., Suganthan, P.N., 2020. A survey on meta-heuristics for solving disassembly line balancing, planning and scheduling problems in remanufacturing. *Swarm Evol. Comput.* 57, 100719 <https://doi.org/10.1016/j.swevo.2020.100719>.
- Glöser-Chahoud, S., Huster, S., Rosenberg, S., Baazouzi, S., Kiemel, S., Singh, S., Schneider, C., Weeber, M., Miehe, R., Schultmann, F., 2021. Industrial disassembling as a key enabler of circular economy solutions for obsolete electric vehicle battery systems. *Resour. Conserv. Recycl.* 174, 105735 <https://doi.org/10.1016/j.resconrec.2021.105735>.
- Guo, X., Liu, S., Zhou, M., Tian, G., 2018. Dual-objective program and scatter search for the optimization of disassembly sequences subject to multiresource constraints. *IEEE Trans. Autom. Sci. Eng.* 15, 1091–1103. <https://doi.org/10.1109/TASE.2017.2731981>.
- Guo, X., Zhang, Z., Qi, L., Liu, S., Tang, Y., Zhao, Z., 2022. Stochastic hybrid discrete grey wolf optimizer for multi-objective disassembly sequencing and line balancing planning in disassembling multiple products. *IEEE Trans. Autom. Sci. Eng.* 19, 1744–1756. <https://doi.org/10.1109/TASE.2021.3133601>.
- Guo, X., Zhou, M., Abusorrah, A., Alsokhiry, F., Sedraoui, K., 2021. Disassembly sequence planning: a survey. *IEEE/CAA J. Autom. Sin.* 8, 1308–1324. <https://doi.org/10.1109/JAS.2020.1003515>.
- Han, M., Li, L., Zhao, J., 2022a. Volatile organic compound emissions from 4D printing: effects of material composition and external stimulus. *Addit. Manuf.* 56, 102894 <https://doi.org/10.1016/j.addma.2022.102894>.
- Han, M., Yang, Y., Li, L., 2022b. Techno-economic modeling of 4D printing with thermo-responsive materials towards desired shape memory performance. *IIE Trans* 54, 1047–1059. <https://doi.org/10.1080/24725854.2021.1989093>.
- Han, M., Yang, Y., Li, L., 2020. Energy consumption modeling of 4D printing thermal-responsive polymers with integrated compositional design for. *Material. Addit. Manuf.* 34, 101223 <https://doi.org/10.1016/j.addma.2020.101223>.
- Ilgin, M.A., Gupta, S.M., 2011. Performance improvement potential of sensor embedded products in environmental supply chains. *Resour. Conserv. Recycl.* 55, 580–592. <https://doi.org/10.1016/j.resconrec.2010.05.001>.
- Jiang, Z., Wang, H., Zhang, H., Mendis, G., Sutherland, J.W., 2019. Value recovery options portfolio optimization for remanufacturing end of life product. *J. Clean. Prod.* 210, 419–431. <https://doi.org/10.1016/j.jclepro.2018.10.316>.
- Jyothi, R.K., Thenepalli, T., Ahn, J.W., Parhi, P.K., Chung, K.W., Lee, J.-Y., 2020. Review of rare earth elements recovery from secondary resources for clean energy technologies: grand opportunities to create wealth from waste. *J. Clean. Prod.* 267, 122048 <https://doi.org/10.1016/j.jclepro.2020.122048>.
- Konstantaras, I., Skouri, K., Benkherouf, L., 2021. Optimizing inventory decisions for a closed-loop supply chain model under a carbon tax regulatory mechanism. *Int. J. Prod. Econ.* 239, 108185 <https://doi.org/10.1016/j.ijpe.2021.108185>.
- Lee, M.-L., Behdad, S., Liang, X., Zheng, M., 2022. Task allocation and planning for product disassembly with human–robot collaboration. *Robot. Comput. Integrated Manuf.* 76, 102306 <https://doi.org/10.1016/j.rcim.2021.102306>.
- Liang, J., Guo, S., Du, B., Liu, W., Zhang, Y., 2022. Restart genetic flatworm algorithm for two-sided disassembly line balancing problem considering negative impact of destructive disassembly. *J. Clean. Prod.* 355, 131708 <https://doi.org/10.1016/j.jclepro.2022.131708>.
- Malekhouyan, S., Aghsami, A., Rabbani, M., 2021. An integrated multi-stage vehicle routing and mixed-model job-shop-type robotic disassembly sequence scheduling problem for e-waste management system. *Int. J. Comput. Integrated Manuf.* 34, 1237–1262. <https://doi.org/10.1080/0951192X.2021.1963484>.
- Mazurek-Budzyńska, M., Behl, M., Neumann, R., Lendlein, A., 2022. 4D-actuators by 3D-printing combined with water-based curing. *Mater. Today Commun.* 30, 102966 <https://doi.org/10.1016/j.mtcomm.2021.102966>.
- Meng, K., Cao, Y., Peng, X., Prybutok, V., Youcef-Toumi, K., 2020. Smart recovery decision-making for end-of-life products in the context of ubiquitous information and computational intelligence. *J. Clean. Prod.* 272, 122804 <https://doi.org/10.1016/j.jclepro.2020.122804>.
- Ong, S.K., Chang, M.M.L., Nee, A.Y.C., 2021. Product disassembly sequence planning: state-of-the-art, challenges, opportunities and future directions. *Int. J. Prod. Res.* 1–16. <https://doi.org/10.1080/00207543.2020.1868598>.
- Rashid, T., Samvelyan, M., de Witt, C.S., Farquhar, G., Foerster, J., Whiteson, S., 2018. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. <https://doi.org/10.48550/arXiv.1803.11485Focustolearnmore>.
- Reddy, K.N., Kumar, A., 2020. Capacity investment and inventory planning for a hybrid manufacturing – remanufacturing system in the circular economy. *Int. J. Prod. Res.* 1–29 <https://doi.org/10.1080/00207543.2020.1734681>.
- Ren, Y., Meng, L., Zhang, C., Zhao, F., Saif, U., Huang, A., Mendis, G.P., Sutherland, J.W., 2020. An efficient metaheuristics for a sequence-dependent disassembly planning. *J. Clean. Prod.* <https://doi.org/10.1016/j.jclepro.2019.118644>.
- Riggs, R.J., Battaià, O., Hu, S.J., 2015. Disassembly line balancing under high variety of end of life states using a joint precedence graph approach. *J. Manuf. Syst.* 37, 638–648. <https://doi.org/10.1016/j.jmsy.2014.11.002>.
- Sathish, T., 2019. Profit maximization in reverse logistics based on disassembly scheduling using hybrid bee colony and bat optimization. *Trans. Can. Soc. Mech. Eng.* 43, 551–559. <https://doi.org/10.1139/tcsme-2019-0017>.
- Sept, R., 2018. Industrial design of a hand pulse detector. Egg [WWW Document]. URL <https://grabcad.com/library/egg/hand-pulse-detector-1> . accessed 7.22.22.
- Slama, I., Ben-Ammar, O., Dolgui, A., Masmoudi, F., 2021. Genetic algorithm and Monte Carlo simulation for a stochastic capacitated disassembly lot-sizing problem under random lead times. *Comput. Ind. Eng.* 159, 107468 <https://doi.org/10.1016/j.cie.2021.107468>.
- Slama, I., Ben-Ammar, O., Dolgui, A., Masmoudi, F., 2020. New mixed integer approach to solve a multi-level capacitated disassembly lot-sizing problem with defective items and backlogging. *J. Manuf. Syst.* 56, 50–57. <https://doi.org/10.1016/j.jmsy.2020.05.002>.
- Tahouni, Y., Krüger, F., Poppinga, S., Wood, D., Pfaff, M., Rühe, J., Speck, T., Menges, A., 2021. Programming sequential motion steps in 4D-printed hygromorphs by architected mesostructure and differential hydro-responsiveness. *Bioinspiration Biomimetics* 16, 055002. <https://doi.org/10.1088/1748-3190/ac0c8e>.
- Tian, G., Zhou, M., Li, P., 2018. Disassembly sequence planning considering fuzzy component quality and varying operational cost. *IEEE Trans. Autom. Sci. Eng.* 15, 748–760. <https://doi.org/10.1109/TASE.2017.2690802>.
- Tian, X., Zhang, Z.-H., 2019. Capacitated disassembly scheduling and pricing of returned products with price-dependent yield. *Omega* 84, 160–174. <https://doi.org/10.1016/j.omega.2018.04.010>.
- Ullah, M., Asghar, I., Zahid, M., Omair, M., AlArjani, A., Sarkar, B., 2021. Ramification of remanufacturing in a sustainable three-echelon closed-loop supply chain management for returnable products. *J. Clean. Prod.* 290, 125609 <https://doi.org/10.1016/j.jclepro.2020.125609>.
- Wang, K., Li, X., Gao, L., Li, P., 2020. Energy consumption and profit-oriented disassembly line balancing for waste electrical and electronic equipment. *J. Clean. Prod.* 265, 121829 <https://doi.org/10.1016/j.jclepro.2020.121829>.
- Wang, K., Li, X., Gao, L., Li, P., Sutherland, J.W., 2022. A discrete artificial bee colony algorithm for multiobjective disassembly line balancing of end-of-life products. *IEEE Trans. Cybern.* 52, 7415–7426. <https://doi.org/10.1109/TCYB.2020.3042896>.
- Wurster, M., Michel, M., May, M.C., Kuhnlle, A., Stricker, N., Lanza, G., 2022. Modelling and condition-based control of a flexible and hybrid disassembly system with manual and autonomous workstations using reinforcement learning. *J. Intell. Manuf.* 33, 575–591. <https://doi.org/10.1007/s10845-021-01863-3>.
- Xiong, S., Ji, J., Ma, X., 2020. Environmental and economic evaluation of remanufacturing lithium-ion batteries from electric vehicles. *Waste Manag.* 102, 579–586. <https://doi.org/10.1016/j.wasman.2019.11.013>.
- Xu, W., Tang, Q., Liu, J., Liu, Z., Zhou, Z., Pham, D.T., 2020. Disassembly sequence planning using discrete Bees algorithm for human–robot collaboration in remanufacturing. *Robot. Comput. Integrated Manuf.* 62, 101860 <https://doi.org/10.1016/j.rcim.2019.101860>.
- Yuan, G., Yang, Y., Tian, G., Fathollahi-Fard, A.M., 2022. Capacitated multi-objective disassembly scheduling with fuzzy processing time via a fruit fly optimization algorithm. *Environ. Sci. Pollut. Res.* <https://doi.org/10.1007/s11356-022-18883-y>.
- Zhang, H., Huang, S., Sheng, J., Fan, L., Zhou, J., Shan, M., Wei, J., Wang, C., Yang, H., Lu, J., 2022. 4D printing of Ag nanowire-embedded shape memory composites with stable and controllable electrical responsivity: implications for flexible actuators. *ACS Appl. Nano Mater.* 5, 6221–6231. <https://doi.org/10.1021/acsnano.2c00264>.
- Zhang, J., Yin, Z., Ren, Luquan, Liu, Q., Ren, Lei, Yang, X., Zhou, X., 2022. Advances in 4D printed shape memory polymers: from 3D printing, smart excitation, and response to applications. *Adv. Mater. Technol.*, 2101568 <https://doi.org/10.1002/admt.202101568>.
- Zhang, X., Zhang, M., Zhang, H., Jiang, Z., Liu, C., Cai, W., 2020. A review on energy, environment and economic assessment in remanufacturing based on life cycle assessment method. *J. Clean. Prod.* <https://doi.org/10.1016/j.jclepro.2020.120160>.
- Zhao, J., Han, M., Li, L., 2021. Modeling and characterization of shape memory properties and decays for 4D printed parts using stereolithography. *Mater. Des.* 203, 109617 <https://doi.org/10.1016/j.matdes.2021.109617>.
- Zhou, F., He, Y., Ma, P., Lim, M.K., Pratap, S., 2022. Capacitated disassembly scheduling with random demand and operation time. *J. Oper. Res. Soc.* 73, 1362–1378. <https://doi.org/10.1080/01605682.2021.1911603>.