

CIRO7.2: A Material Network with Circularity of -7.2 and Reinforcement-Learning-Controlled Robotic Disassembler

Federico Zocco^{1*}, Monica Malvezzi¹

Received: 15 April 2026 / Accepted: 21 August 2026 / Published: 27 August 2026

© The Author(s) 2026

Abstract

The competition over natural reserves of minerals is expected to increase in part because of the linear-economy paradigm based on take-make-dispose. Simultaneously, the linear economy considers end-of-use products as waste rather than as a resource, which results in large volumes of waste whose management remains an unsolved problem. Since a transition to a circular economy can mitigate these open issues, in this paper we begin by enhancing the notion of circularity based on compartmental dynamical thermodynamics, namely, λ , by defining a time-window formulation to derive a physics-inspired value from the time-varying definition. Then, we model a thermodynamical material network (TMN) processing a batch of 2 solid materials of criticality coefficients of 0.1 and 0.95, with a robotic disassembler compartment controlled via reinforcement learning (RL), and processing 2–7 kg of materials contained in a PC desktop. The tested RL algorithms are the soft-actor critic (SAC), the truncated quantile critics (TQC), and the twin delayed deep deterministic policy gradient (TD3), and each algorithm is enhanced through hindsight experience replay (HER). Subsequently, we focus on the design of the robotic disassembler compartment using state-of-the-art RL algorithms and assessing the algorithm performance with respect to λ (Figure 1). The simulations show that the highest circularity is -2.1 achieved in the case of disassembling 2 parts of 1 kg each, whereas it reduces to -7.2 in the case of disassembling 4 parts of 1 kg each contained inside a chassis of 3 kg, which is the most complex task of those considered. Finally, a sensitivity analysis highlights that RL performance has a positive correlation with circularity, with the impact increasing with material criticality and mass. In contrast with the data-intensive material flow analysis (MFA), we approach circularity as a design of material flows to be optimized with respect to λ and create a compartmental network tunable through 12 parameters depicting its physical and temporal properties. Furthermore, our approach captures the impact that very fast dynamics occurring in less than 1 second has on the whole system circularity, whereas MFA studies usually depict only slow dynamics, e.g., variations from year n to year $n + 1$. This work lies in the emerging research topics of circular intelligence and robotics (CIRO). Source code is publicly available.²

Keywords Circular systems · Sustainable consumption · Sustainability

1. Introduction

The traditional (linear) economy paradigm extracts non-renewable minerals from natural reserves, uses them to manufacture products, sells them to the end user, and eventually disposes of the products at the user discretion. Since this practice is the most common worldwide, it is not surprising that there are two main consequences of increasing concern: the supply uncertainties of critical raw materials used to make green (e.g., photovoltaic panels, wind turbines, batteries for electric vehicles) and digital technologies (CRM Alliance, 2024; European Commission, 2023; Federal Register, 2023), and the management of waste (Eurostat, 2024).

* Corresponding author: federico.zocco.fz@gmail.com

¹ Department of Information Engineering and Mathematics, University of Siena, Via Roma 56, Siena, 53100, Italy

² <https://github.com/ciroresearch/ciro7RL>

The linear economy also fosters competition over natural reserves (Zhou and Månberger, 2024) and highlights the stark reality of resources constraints (Nygaard, 2023; Hagelüken and Goldmann, 2022; Lieder and Rashid, 2016; CE-HubTeam, 2022; Henckens et al., 2014). As observed by (Henckens, 2021), key factors affecting the access to non-renewable materials are the world population size and the welfare of each individual, with citizens of developed countries currently enjoying the highest consumption rates to satisfy their demands. The scarce raw materials considered by (Henckens, 2021) are antimony, bismuth, gold, indium, copper along with other eight materials. Enabling future generations to access these materials is at the core of sustainability as defined by the United Nations Brundtland Commission back in 1987 (United Nations, n.d.).

To mitigate these fundamental issues existing both at national and international levels, the circular economy (CE) paradigm promotes practices such as reuse, reduce, and repair, with the goal of keeping products and materials in use for as long as possible (Ellen MacArthur Foundation, 2020). Since high operational costs are among the major barriers to the implementation of circularity, the cost reductions achievable with robotic and autonomous systems make automation a key enabler of the transition to CE (Barteková and Börkey, 2022). Moreover, robots can avoid the contamination risks resulting from handling waste (Zocco, Sleath, et al., 2025).

By definition, the notion of circularity requires to consider the whole material life cycle, from extraction to disposal, and hence, circularity cannot be measured by merely considering a single process. The holistic perspective is fundamental as pointed out by the Ellen MacArthur Foundation: “A lack of holistic thinking is one of the reasons industries have been stuck in linear business models up to now; it’s only through rethinking at the system level that we’re likely to get unstuck. [...]” (Ellen MacArthur Foundation, 2022b).

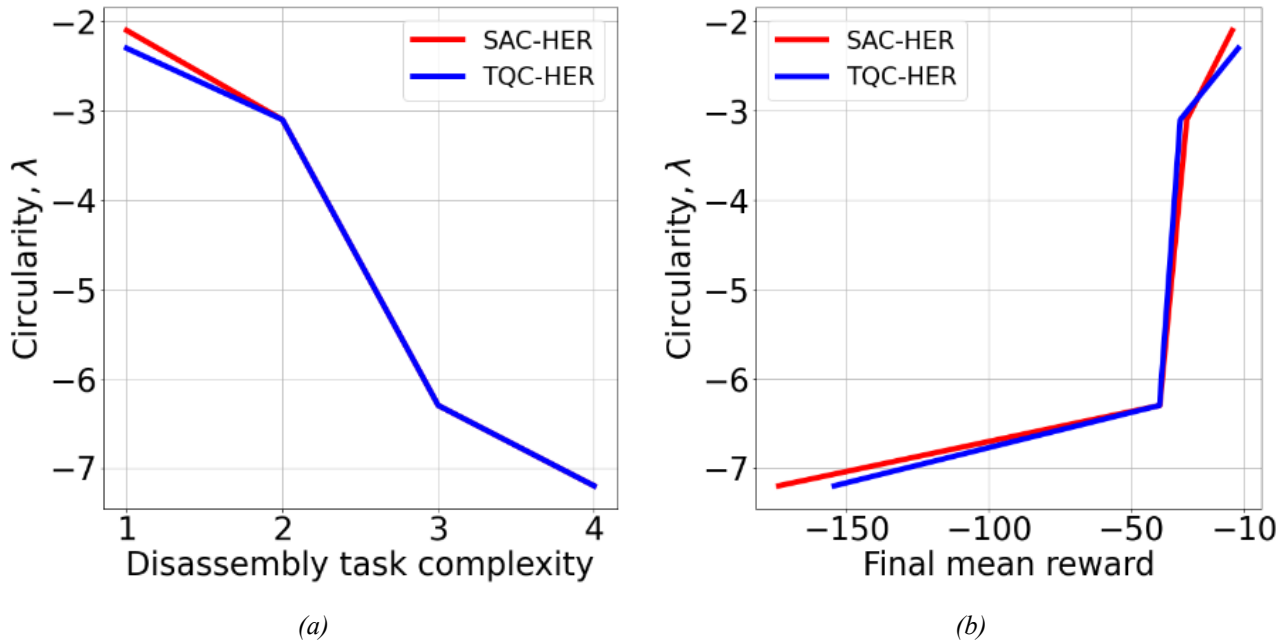


Figure 1. (a) Disassembly task complexity vs. λ . (b) Final mean reward vs. λ . SAC-HER and TQC-HER are reinforcement-learning algorithms. An higher circularity $\lambda \in (-\infty, +\infty)$ indicates a more sustainable use of natural resources.

In this context, this paper makes the following main contributions:

- It enhances the mathematical formulation of circularity in (Zocco and Malvezzi, 2025), which is predicated on compartmental dynamical thermodynamics (Haddad, 2019), and hence, has the advantage of being based on fundamental laws of physics in contrast to other circularity metrics (Calzolari et al., 2022). Specifically, we define the *time-window* circularity via integration of the formulation in (Zocco and Malvezzi, 2025) (see Section 3).
- It quantifies and analyzes the impact of the reinforcement-learning (RL) robot controllers on the system circularity as summarized in Figure 1(b) (details in Sections 4.1 and 4.2).
- Overall, we clarify the contribution that the robotics community could make for improving circularity, and hence, a sustainable use of natural resources. Indeed, this paper enables, for the first time, the quantitative assessment of how robotic disassembly performance affects system-level material circularity.

The paper contributions are of the following type: *methodological* since it provides a mathematical formulation of material circularity as an integral of mass transfers occurring over a finite-time horizon; *technological* since it integrates RL into a material network to quantify its impact on the network circularity; and *conceptual* since it clarifies how to work in the emerging fields of circular intelligence and robotics (CIRO).

Throughout the paper, matrices and vectors are indicated with bold upper- and lower-case letters, respectively, while sets are indicated with calligraphic letters.

The remainder of the paper is organized as follows. Section 2 discusses the related work, Section 3 covers the preliminary concepts needed to understand what follows, while Section 4 details the design of the system of which the disassembly stage is part of. Finally, Section 5 gives the conclusions and future work.

2. Related Work

2.1. Circular Economy and Thermodynamical Material Networks

A circular economy is based on the so-called “Rs”, e.g., reduce, reuse, repair, remanufacture, and recycle, which are practices to be adopted at local scales to improve the material circularity (Zorpas, 2024). These practices are usually prioritized from the least to the most energy intensive, e.g., *recycling* is one of the last practices in the ranking, whereas *reduce* is one of the first ones (Potting et al., 2017). A fundamental open question in CE is *where, when, and how to implement the “Rs” in order to maximize circularity*, which implies a holistic perspective (Abu-Bakar and Charnley, 2024). Addressing this challenging question without relying on the large amount of data needed for material flow analysis (MFA) has motivated the formalization of thermodynamical material networks (TMNs) (Zocco et al., 2023; Zocco, 2026; Zocco, Lake, et al., 2025), and subsequently, a notion of circularity predicated on TMNs (Zocco and Malvezzi, 2025). MFA has become a standard methodology to analyze the circularity of material flows (Cullen and Cooper, 2022; Brunner and Rechberger, 2016).

Essentially, MFA is a data-analysis approach based on the principle of mass conservation (Brunner and Rechberger, 2016). Although TMNs are more mathematically intensive than MFA, they have three main advantages. First, they enhance MFA by adding energy balances to the mass balances; second, they increase the overall modeling and prediction accuracy by leveraging ordinary differential and difference equations derived from thermodynamic principles; and third, they are less data-intensive than MFA because not relying on the availability of data to be analyzed (Zocco et al., 2023; Zocco, Sleath, et al., 2025). Hence, TMNs are the approach we adopted in this paper as will be detailed in Section 4.

An active stream of research in CE consists of defining circularity metrics, while another one focuses on CE definitions. A recent review on the former was provided by (Munonye, 2025), while a review on the latter can be found in (Kirchherr et al., 2023). Other recent works involving indicators are (Rashid, 2026), which investigates the level of circularity of the linear economy, and (Paletti et al., 2026), which focuses on the circularity of designs of aircraft components. In this paper, in common with (Zocco, 2026; Zocco et al., 2023), we adopt the well-established definition of CE of (Ellen MacArthur Foundation, 2020) and translate it into a physics-grounded problem of mass transfer occurring, in space and time, in a network of thermodynamic compartments. Our approach is a generalization of the one adopted by thermal engineers in designing thermal processes, with the difference that the circularity metric becomes the main performance indicator of the compartmental network. Thus, no social and political factors are considered in our approach in contrast with other CE metrics (Munonye, 2025).

2.2. Robotics for Recovery Chains

As shown in (Zocco, Sleath, et al., 2025) (specifically, in Proposition 1), the standard form of robot dynamics can be derived from the first law of thermodynamics, and hence, a robot can be considered a thermodynamic compartment of a TMN (Zocco et al., 2024). While the integration of robots into TMNs is very recent (Zocco, Sleath, et al., 2025), robots have been proposed in the past to perform recovery-chain tasks such as waste sorting and disassembly. For example, (Chen et al., 2022; Koskinopoulou et al., 2021; Kiyokawa et al., 2022)

focused on waste sorting, whereas (Laili et al., 2022; Beghi et al., 2023; Li et al., 2018; Qu et al., 2023) focused on disassembly. A review on human-robot interaction for disassembly was provided by (Yuan et al., 2025).

To overcome the operational errors caused by the unpredictable conditions of end-of-life items processed via disassembly, (Laili et al., 2021) introduced online backup actions to the disassembly sequence, while (Lee et al., 2022) proposed an algorithm for allocating the tasks either to a human operator, a robot, or to a human-robot collaboration taking into account the limits on available resources and the human safety. Note that, as pointed out by (Smith and Behdad, 2025), physical disassembly is only one of five alternative treatment processes, with the other four being pyrolysis, hydrometallurgy, biometallurgy, and supercritical fluid extraction. Reinforcement learning is gaining interest for controlling disassembly operations as reviewed by (Guo et al., 2023); for example, (Kristensen et al., 2019) develop an RL simulation framework based on Gazebo for robot programming in unscrewing tasks, while tool deterioration was considered by (Cai et al., 2022).

Although the literature on robotic waste sorting and disassembly is significant, to the best of our knowledge, this is the first paper to measure the RL-controlled robot performance against the circularity of the whole supply-recovery chain, meaning that we assess the impact of the RL-driven disassembler on the material circularity. This requires to consider the whole material life cycle instead of merely the disassembly stage as done so far in the literature. The details are provided in Section 4, while the main paper result is summarized in Figure 1. Specifically, Figure 1(a) reports the disassembly task complexity vs. material circularity, where task complexity is expressed by a value between 1 and 4 (higher means more complex). In contrast, Figure 1(b) shows the final mean reward of an RL algorithm vs. circularity. Figures 1(a) and 1(b) result from the simulation experiments detailed in Section 4.

2.3. Reinforcement Learning for Robot Control

The capability of RL systems is progressing fueled by the advances in deep learning. While RL control has shown very promising results in simulated scenarios and games, it still lacks the sampling efficiency and safety guarantees required for deployment in real settings (Ju et al., 2022). (Ibarz et al., 2021) encountered other challenges when using RL in real robots such as achieving a reliable and stable learning, designing high-fidelity simulators to train the RL controller, and effectively guiding the exploration phase during learning. Examples of RL uses are (Jia and Ma, 2021) for snake robot control, (Yau et al., 2023) for diabetes management, (Rapetswa and Cheng, 2023) in cognitive radio networks, and (Ghignone et al., 2023) in autonomous racing. A review of robotic applications is provided in (Singh et al., 2022). In this paper, we implement RL in a simulated environment for robotic disassembly tasks and assess its impact on material circularity. Details on the RL algorithms are discussed in Section 3.2.

2.4. Industrial Ecology and Material Flow Analysis

The concept of circularity at the core of a circular economy can be seen as rooted in industrial ecology, which is a field of study aiming to harmonize the production and consumption systems with the limits of the surrounding natural environment (Graedel and Eckelman, 2023). In particular, ongoing geopolitical tensions highlight that physical constraints on the availability of essential natural resources such as the rare-Earth minerals could raise the competition between nations (Baranowski et al., 2025). The traditional economy, i.e., linear, has minimal consideration for the Earth physical limits by assuming an infinite availability of materials and, simultaneously, neglects the alteration of natural ecosystems caused by waste and pollution (Fet and Deshpande, 2023). Increasing the circularity of materials can mitigate the international competition to secure the finite material supplies and the environmental damage caused by mismanagement of waste.

To increase the circularity of materials, a useful industrial ecology tool is material flow analysis (MFA) as it captures the life cycle of materials, and hence, it helps to understand where the circularity of the system is affected (Emami et al., 2026; Mason et al., 2025; Liao et al., 2025). In this paper, we borrow from MFA the focus on material life cycles, while we differ from MFA in the fact that we model the material flow as a dynamical system occurring in a network of thermodynamic compartments as in (Zocco et al., 2023). Indeed, our mathematical formulation of circularity is time-varying and results from an integral over a time window of the mass transfers occurred in the network. This enabled us to quantify the impact that the RL-controlled robotic disassembler has on the circularity of the whole system (summarized in Figure 1(b)). With MFA it is not straightforward to assess the impact that the time needed for disassembly has on circularity since MFA

provides static snapshots of the reality depicted by the available data over a relatively long period of time, e.g., one-year resolution in (Houssini et al., 2025), in (Lipp and Lederer, 2025), and in (Grund et al., 2025). In contrast, with TMNs we can depict dynamics occurring in less than one minute thanks to the use of ordinary differential/difference equations (Zocco, 2026).

3. Methods

3.1. Theory of Circular Flow Designs

We will begin by giving the definition of a digraph, and then, use it in the definition of a TMN.

- **Definition 1** (Bondy and Murty (1976)). A directed graph D or digraph is a graph identified by a set of n_v nodes $\{v_1, v_2, \dots, v_{n_v}\}$ and a set of n_a arcs $\{a_1, a_2, \dots, a_{n_a}\}$ that connect the nodes.
- **Definition 2** (Zocco et al. (2023)). A thermodynamical material network (TMN) is a set $\mathcal{N}$ of connected thermodynamic compartments, that is,

$$\mathcal{N} = \left\{ c_{1,1}^1, \dots, c_{k_v,k_v}^{k_v}, \dots, c_{n_v,n_v}^{n_v}, \right. \\ \left. c_{i_{n_v+1},j_{n_v+1}}^{n_v+1}, \dots, c_{i_{n_v+k_a},j_{n_v+k_a}}^{n_v+k_a}, \dots, c_{i_{n_c},j_{n_c}}^{n_c} \right\} \quad (1)$$

which transport, store, use, and transform a target material, namely, β . Each compartment is indicated by a control surface (Moran et al., 2010) and is modeled using dynamical systems derived from a mass balance and/or at least one of the laws of thermodynamics (Haddad, 2019).

Specifically, $\mathcal{N} = \mathcal{R} \cup \mathcal{T}$, where $\mathcal{R} \subseteq \mathcal{N}$ is the subset of compartments $c_{i,j}^k$ that store, transform, or use the target material, while $\mathcal{T} \subset \mathcal{N}$ is the subset of compartments $c_{i,j}^k$ that move the target material between the compartments belonging to $\mathcal{R} \subseteq \mathcal{N}$. A net $\mathcal{N}$ is associated with its compartmental digraph $M(\mathcal{N})$, which is a digraph whose nodes are the compartments $c_{i,j}^k \in \mathcal{R}$ and whose arcs are the compartments $c_{i,j}^k \in \mathcal{T}$. For node-compartments $c_{i,j}^k \in \mathcal{R}$ it holds that $i = j = k$, whereas for arc-compartments $c_{i,j}^k \in \mathcal{T}$ it holds that $i \neq j$ because an arc moves the material from the node-compartment $c_{i,i}^i$ to the node-compartment $c_{j,j}^j$. The orientation of an arc is given by the direction of the material flow. The superscript k is the identifier of each compartment. The superscripts k_v and k_a in (1) are the k -th node and the k -th arc, respectively, while n_c and n_v are the total number of compartments and nodes, respectively. Since n_a is the total number of arcs, it holds that $n_c = n_v + n_a$ (Zocco et al., 2023; Zocco, 2026; Zocco, Lake, et al., 2025).

Now, since our goal is to design circular networks $\mathcal{N}$ (1), we will introduce the notion of circularity after the following definition.

- **Definition 3** (Zocco and Malvezzi (2025)). A mass or flow is finite-time sustainable if either it exits a non-renewable reservoir or it enters a landfill, an incinerator, or the natural environment as a pollutant.

The locations mentioned in Definition 3, i.e., reservoirs, landfills, incinerators, and the environment, are thermodynamic compartments $c_{i,j}^k \in \mathcal{N}$ as they obey the laws of thermodynamics (Haddad, 2017). We can now define the time-window circularity of $\mathcal{N}$ (1).

- **Definition 4.** The time-window circularity $\lambda(\mathcal{N}; t) \in (-\infty, +\infty)$ is defined as

$$\lambda(\mathcal{N}; t) = -\frac{1}{t_f} \int_0^{t_f} \mu_{f,b} \bar{m}_{f,b}(t) + \Delta \bar{m}_{f,c}(t) dt. \quad (2)$$

In (2), $m_{f,b}(t)$ is the net finite-time sustainable mass transported in batches (e.g., solids transported on trucks) and $\dot{m}_{f,c}(t)$ is the net finite-time sustainable flow transported continuously (e.g., fluids transported through pipes). The quantity

$$\bar{m}_{fb}(t) = c_{fb}m_{fb}(t) \quad (3)$$

is the *weighted* net finite-time sustainable mass transported in batches, where $c_{fb} \in (0,1]$ is the criticality coefficient indicating to what extent a material is critical, e.g., $c_{fb} > 0.8$ in the case of critical raw materials (European Commission 2023; Federal Register 2023), while μ_{fb} is the functionality coefficient such that

$$\mu_{fb} = 1 + l, \quad (4)$$

where l is the number of *functional* discarded batches, i.e., batches of products or parts sent to incineration even if they are not faulty. The parameter μ_{fb} in (2) takes into account the functionality of the discarded products or parts since one of the key principles of circularity is to retain products and materials *at their highest value* (Ellen MacArthur Foundation, 2022a) (e.g., disposing of a batch of reusable bottles yields $l = 1$, whereas a batch of broken bottles yields $l = 0$; thus, discarding working items reduces circularity more than the disposal of faulty items). Similarly, $\dot{\bar{m}}_{fc}(t) = c_{fc}m_{fc}(t)$ is the *weighted* net finite-time sustainable flow of material transported continuously, where $c_{fc} \in (0,1]$ is the criticality coefficient of the flowing material. Moreover, $\Delta > 0$ is a constant interval of time introduced as a multiplying factor in order to convert the flow $\dot{\bar{m}}_{fc}(t)$ into a mass, and hence, to make the sum with $\bar{m}_{fb}(t)$ physically consistent. The choice of Δ is arbitrary, but its value must be kept the same for any calculation of $\lambda(\mathcal{N}; t)$ in order to make comparisons. Finally, t_f is the final time for the calculation of circularity.

Note that:

- The time-window circularity (2) corresponds to the area under the curve of the weighted net finite-time sustainable masses and flows in the network in the interval $t \in [0, t_f]$.
- A coherent integration of compartmental dynamical thermodynamics and circular economy is realized considering jointly Definition 2 (formulating a network of thermodynamic compartments $\mathcal{N}$ to be designed), Definition 3 (specifying the types of materials impacting circularity), and Definition 4 (formulating a notion of circularity that takes into account the architecture of $\mathcal{N}$ and the mass transfers occurring within it). This coherent integration was firstly proposed by (Zocco and Malvezzi, 2025) with an *instantaneous* notion of circularity, i.e., as a continuous function of t . In this work, since we needed to extrapolate a meaningful quantity from the *instantaneous* formulation, we added the integral to the mathematical expression yielding the *time-window* circularity in Definition 4.

3.2. Deep Reinforcement Learning

Deep RL is one of the most promising learning approaches as it is able to generate learning-driven actions that bring the target system to converge to the desired behavior. The working principle of an RL system is analogous to the learning experience of mammals: by trial and error, the agent interacts with the surrounding environment and takes actions, and then it receives a reward or a penalty depending on how good is the action in achieving a predefined goal, e.g., picking an object or walking in the environment. The goal in RL is expressed as a mathematical function, namely, the reward function, whose formulation is task-dependent and is a key step in designing RL systems (Eschmann, 2021). One or more deep neural networks are used in modern RL algorithms as function approximators (Sutton and Barto, 1998). The quality of learning highly affects the quality of the RL agent actions, and hence, the impressive improvement of deep neural network learning performance of the last decade (Nature, 2024) has led to a step forward in the capability of RL frameworks.

One of the most comprehensive and reliable libraries of RL algorithm implementations is Stable-Baselines3 (SB3) proposed by (Raffin et al., 2021) in 2021 and currently maintained by the German Aerospace Center (DLR). In this paper, we will use the SB3 implementations of three algorithms to control the robotic disassembler, namely, the soft-actor critic (SAC) (Haarnoja et al., 2018), the truncated quantile critics (TQC) (Kuznetsov et al., 2020), and the twin delayed deep deterministic policy gradient (TD3) (Fujimoto et al., 2018). Each learning algorithm was enhanced with hindsight experience replay (HER) (Andrychowicz et al., 2017), thus the notation “SAC-HER” refers to the enhancement of SAC through HER.

A description of the above mentioned algorithms requires to preliminary introduce the deep deterministic policy gradient (DDPG) algorithm, which was initially proposed in 2015 by (Lillicrap et al., 2015). The *DDPG*

algorithm simultaneously learns the value of an action, i.e., a Q-function $Q(s, a)$, and a policy (OpenAI, 2018a; Lapan, 2018), where a Q-function is the total reward obtainable by taking a specific action a while being in a specific state s . To learn the Q-function, DDPG uses the Bellman equation, and then uses the Q-function to learn the policy; the policy generates the action a_k to be taken at the k -th time (Lapan, 2018). DDPG works on continuous action spaces, is model-free (i.e., it does not learn a representation of the environment (Sutton and Barto, 1998)), and is an “actor-critic” method.

TD3 is an improvement of DDPG and was proposed in 2018 (Fujimoto et al., 2018; OpenAI, 2018b). It addresses the issue of DDPG of overestimating values of Q-functions, which leads to errors in policy estimation, i.e., in the actor. TD3 improved DDPG by taking the minimum value between two critics and by delaying the policy updates to reduce update errors.

SAC was also proposed in 2018 as TD3 and, in common with the latter, it incorporates two Q-learning approximators (i.e., the “twins”). A key feature of SAC is that the actor aims to maximize the expected reward while also maximizing entropy, meaning that it aims to successfully perform the task while acting as randomly as possible.

TQC is the most recent algorithm of those used in this paper and was proposed in 2020 (Kuznetsov et al., 2020). It uses quantile regression to predict a distribution for the Q-function instead of a mean value and it performs a truncation of the tails of the learned Q-function distribution to address the above mentioned issue of overestimation; the truncation achieves overestimation control without the need of the multiple critics implemented in TD3 (Stable-Baselines3, 2026).

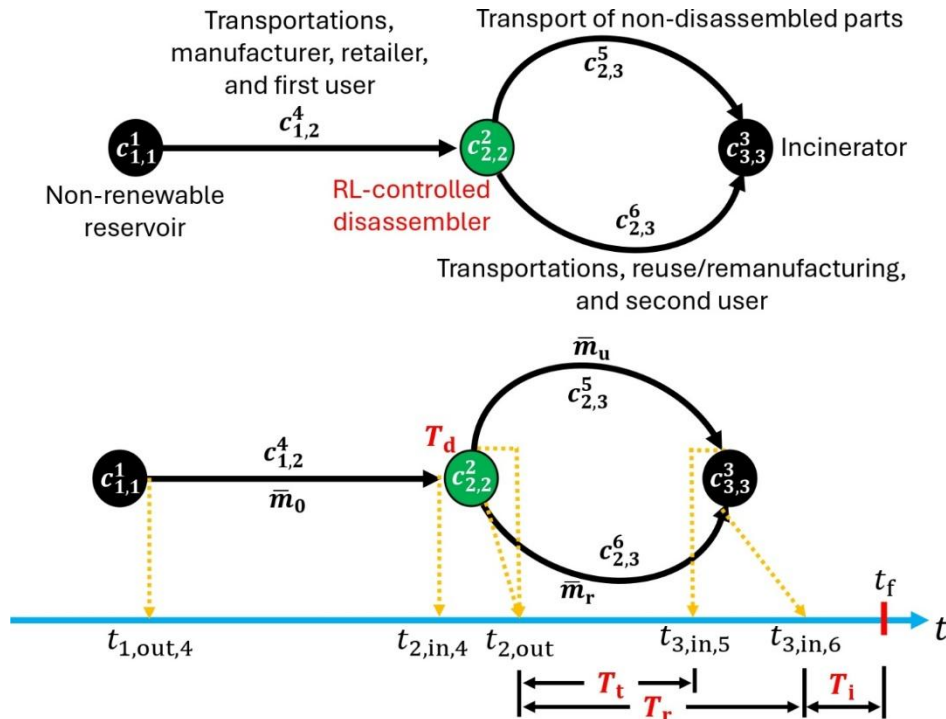


Figure 2. Compartmental diagram of $\mathcal{N}_s(5)$. At the top, the description of each compartment; at the bottom, the masses and times are indicated with respect to the temporal axis. In green the disassembler compartment, which is considered in this paper.

4. Compartmental Network Design

The material network considered in this paper is

$$\mathcal{N}_s = \{c_{1,1}^1, c_{2,2}^2, c_{3,3}^3, c_{1,2}^4, c_{2,3}^5, c_{2,3}^6\} \quad (5)$$

and it is shown in Figure 2. This network is valid for solid materials and not for fluids because it involves the disassembly stage ($c_{2,2}^2$).

The mass of non-renewable raw materials leaves the reservoir ($c_{1,1}^1$) at $t = t_{1,out,4} \equiv 0$. Consider that $\mathcal{N}_s$ processes 2 materials, namely, β_1 and β_2 . Let $m_{f,b,i}^0$ be the mass of material β_i extracted at $t = 0$, with $i \in \{1,2\}$. Thus, the weighted mass of finite-time sustainable material extracted at $t = 0$ is

$$\bar{m}_0 = c_{f,b,1} m_{f,b,1}^0 + c_{f,b,2} m_{f,b,2}^0, \quad (6)$$

where $c_{f,b,1} \in (0,1]$ and $c_{f,b,2} \in (0,1]$ are the criticality coefficients of the materials β_1 and β_2 , respectively.

As shown at the top of Figure 2, the disassembler ($c_{2,2}^2$) processes the products made of β_1 and β_2 after their first use (final stage of $c_{1,2}^4$). The disassembler has two output flows indicated at the bottom of Figure 2: one flow is the one in $c_{2,3}^5$ and it involves the mass of parts that are not disassembled, namely, $\bar{m}_u$. These parts are directly sent to the incinerator ($c_{3,3}^3$). The other flow involves the mass of parts that are disassembled successfully, and then, reused, namely, $\bar{m}_r$. From the principle of mass conservation, it holds that

$$\bar{m}_0 = \bar{m}_u + \bar{m}_r. \quad (7)$$

Let $s \in [0,100]$ be the percentage success measuring to what extent the disassembly stage is successful, it follows that

$$\bar{m}_u = \bar{m}_0 \left(1 - \frac{s}{100}\right) \quad (8)$$

and

$$\bar{m}_r = \bar{m}_0 \frac{s}{100}. \quad (9)$$

Since the mass transfer in $\mathcal{N}_s$ (5) is a dynamical system, let us now specify the time instants at which the masses enter or exit the compartments. The time instants are indicated at the bottom of Figure 2, where $t_{i,out,j}$ is the time at which the material exits compartment i for compartment j , while $t_{i,in,j}$ is the time at which the material enters the compartment i from compartment j . Specifically, $\bar{m}_0$ exits the non-renewable reservoir at $t_{1,out,4} \equiv 0$. Then, the weighted mass $\bar{m}_0$ enters the disassembly facility ($c_{2,2}^2$) in $t_{2,in,4}$. The disassembly stage lasts T_d . The disassembled and not-disassembled masses of materials (i.e., $\bar{m}_r$ and $\bar{m}_u$, respectively) exit the disassembler at $t_{2,out}$, and hence, $t_{2,out} = t_{2,in,4} + T_d$. The material mass $\bar{m}_u$ enters the incinerator at $t_{3,in,5}$, while the reused material does so at $t_{3,in,6} > t_{3,in,5}$. The incineration lasts T_i after which the life of the material ends at $t = t_f$. The time necessary to transport $\bar{m}_u$ is $T_t = t_{3,in,5} - t_{2,out}$, while the time for reusing $\bar{m}_r$ is $T_r = t_{3,in,6} - t_{2,out}$.

From Definition 3, it follows that the weighted net mass of finite-time sustainable materials transported in batches in $\mathcal{N}_s$ (i.e., solids) is

$$\bar{m}_{f,b}(t) = \begin{cases} \bar{m}_0, & 0 \leq t < t_{2,in,4} + T_d + T_t \\ \bar{m}_0 + \bar{m}_0 \left(1 - \frac{s}{100}\right), & t_{2,in,4} + T_d + T_t \leq t < t_{2,in,4} + T_d + T_r \\ \bar{m}_0 + \bar{m}_u + \bar{m}_r, & t_{2,in,4} + T_d + T_r \leq t < t_{2,in,4} + T_d + T_r + T_i. \end{cases} \quad (10)$$

Regarding the functionality coefficient $\mu_{f,b}$, we consider that the non-disassembled batch m_u is sent to the incinerator even if not faulty, whereas the reused batch m_r has no functionality when it is sent to incineration. Hence, from (4), it follows that

$$l = 1 \Rightarrow \mu_{f,b} = 2. \quad (11)$$

Thus, the dynamics of circularity (2) in $\mathcal{N}_s$ is

Table 1. Values used for the parameters. The variables s , T_d , $m_{f,b,1}^0$, and $m_{f,b,2}^0$ change as covered in Section 4.1.

Parameter	Value
l	1
$c_{f,b,1}$	0.1
$c_{f,b,2}$	0.95
$t_{1out,4}$	0 s
$t_{2,in,4}$	2,592,000 s (i.e., 1 month)
T_t	3,600 s (i.e., 1 hour)
T_r	2,592,000 s (i.e., 1 month)
T_i	86,400 s (i.e., 1 day)

$$\begin{aligned}
\lambda(\mathcal{N}_s; t, \bar{m}_0, s, T_d) &= -\frac{1}{t_f} \int_0^{t_f} 2 \bar{m}_{f,b}(t) + \Delta \dot{\bar{m}}_{f,c}(t) \mathbf{r}^0 dt \\
&= -\frac{2}{t_f} [\bar{m}_0(t_{2,in,4} + T_d + T_t) \\
&\quad + \left(\bar{m}_0 + \bar{m}_0 \left(1 - \frac{s}{100} \right) \right) (T_r - T_t) + 2 \bar{m}_0 T_i],
\end{aligned} \tag{12}$$

where

$$t_f = t_{2,in,4} + T_d + T_r + T_i. \tag{13}$$

As visible in (12), the disassembler ($c_{2,2}^2$) affects the circularity λ through the percentage success s and the disassembly time T_d . Hence, the next section will cover the design of $c_{2,2}^2$ with the aim of increasing λ . To extract a numerical value of circularity (12), we considered the values summarized in Table 1. Those values are for demonstration purposes and not depicting a particular case study. Yet, they are chosen to be representative of realistic scenarios. Indeed, they assume multiple criticality coefficients (0.1 and 0.95, respectively) and the duration for reusing the material (i.e., 1 month) is set to be much longer than the time needed to transport it (i.e., 1 hour) or to perform pick-and-place with the robot (less than 1 second according to the simulation results reported in Table 3). The duration of the first use of the product is defined by $t_{2,in,4}$, which is set to be 1 month since it is a possible value for a defective product, e.g., a PC desktop.

4.1. Disassembler Compartment Design

We considered four disassembly tasks of increasing complexity to be performed by a Franka Emika Panda robot. Hence, we build on the *panda-gym* library (Gallouédec et al., 2021). The four tasks (i.e., RL environments) are detailed below and their visual simulations are shown in Figure 3. Over the following, we consider that each part to be disassembled weights 1 kg and is made either of material β_1 or β_2 .

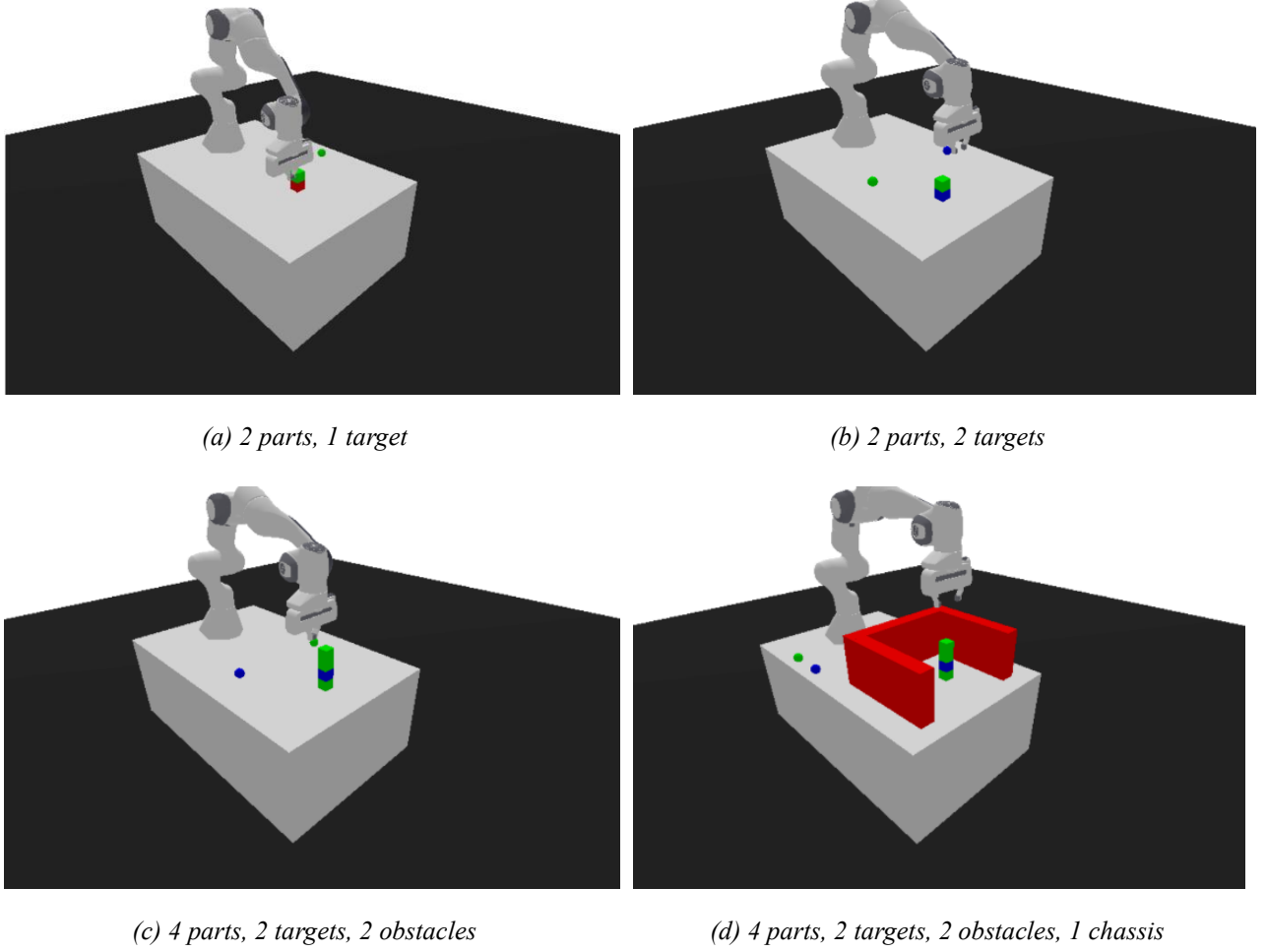
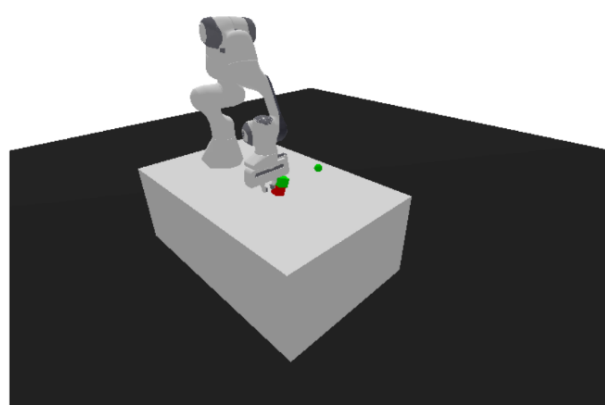


Figure 3. Visual simulation of disassembler compartment performance after training ($c_{2,2}^2$ in Figure 2).

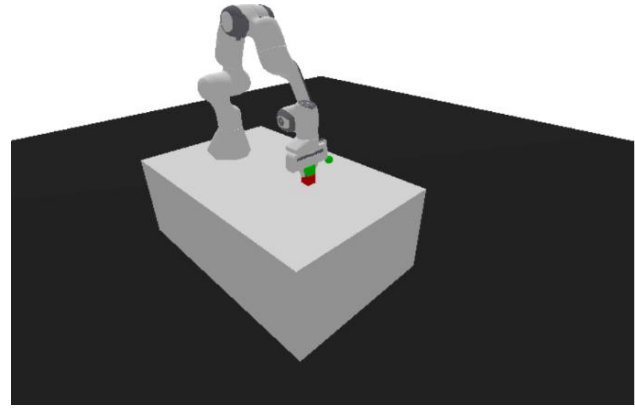
4.1.1. 2 Parts and 1 Target The first disassembly task consists of separating two stacked parts modeled as boxes. The robot must pick and place the box at the top in a target location indicated by a sphere in Figure 3(a). One part is made of β_1 and the other part is made of β_2 . Since $m_{f,b,1}^0 = m_{f,b,2}^0 = 1$ kg, it follows that $\bar{m}_0 = 0.1 + 0.95 = 1.05$ kg. The training settings and performance of the controllers in one training for this task are reported in Table 2, Table 3, and Figures 5(a)-5(c), where

$$\zeta = r_e - r_s, \quad (14)$$

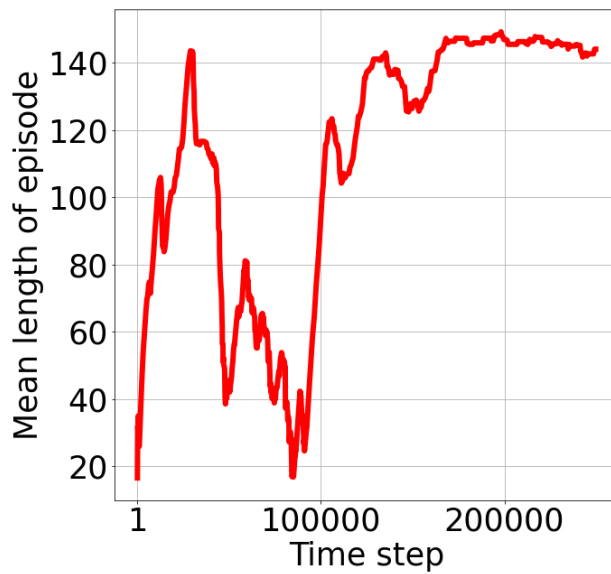
with r_e and r_s denoting the mean reward over 100 episodes at the end and at the start of the training, respectively. The best performing algorithms, namely, SAC and TQC, increase their reward during training and, during testing, yield a percentage success s of 100% and 80%, respectively. The disassembly time T_d is estimated considering that, according to (Gallouédec et al., 2021), one simulated time step of the RL controller is equivalent to approximately 40 milliseconds of the real time. Equation (2) yields that the circularity with SAC and TQC is -2.1 and -2.3, respectively, while it reduces to -3.1 with TD3 because it has $s = 0\%$ as a consequence of its learning failure in this single training. Table 4 reports the mean and standard deviation of T_d , s , and λ over 5 trainings and tests. SAC-HER shows the best performance even over multiple runs. Figures 4(a) and 4(b) show that SAC and TQC perform the task in different ways. Specifically, the former pushes the part at the top towards the target (Figure 4(a)), whereas the latter picks and drags both parts (Figure 4(b)). As a consequence, T_d of the SAC controller is lower than that of the TQC controller (0.4 sec vs. 0.8 sec).



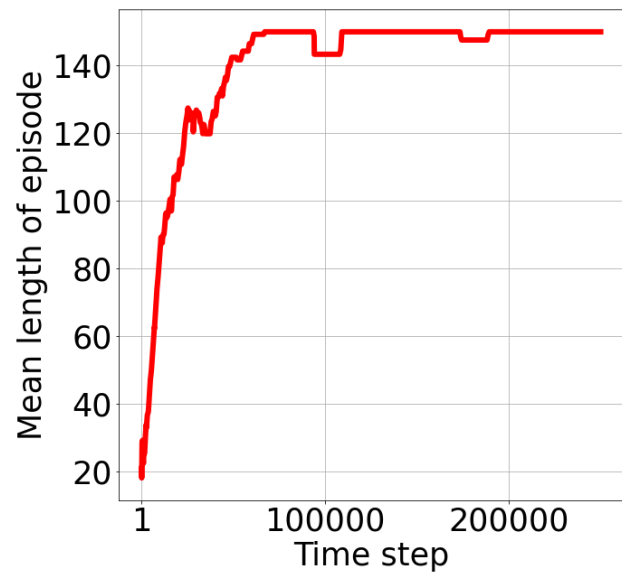
(a) Pushing of SAC-HER controller



(b) Picking of TQC-HER controller



(c) SAC-HER stops to collide



(d) TQC-HER stops to collide

Figure 4. Further analysis of the robot behaviors after training (Figures 4(a) and 4(b)) and during training (Figures 4(c) and 4(d)). Specifically, Figures 4(a) and 4(b) regard the scenario with 2 parts and 1 target, while Figures 4(c) and 4(d) regard the scenario with the chassis.

Table 2. Training tuning for the case of disassembly of 2 parts with 1 target.

	Max episode length	Time steps
SAC-HER	50	4.5×10^5
TQC-HER	50	4.5×10^5
TD3-HER	50	4.5×10^5

4.1.2. 2 Parts and 2 Targets In this disassembly task, the robot is asked to bring one part to a target location and the other part to another target. The situation is shown in Figure 3(b), where the targets are indicated with spheres. Similarly to the previous task, one part is made of β_1 and the other part is made of β_2 , and hence, $\bar{m}_0 = 1.05$ kg. The training tuning and testing results are reported in Table 5, Table 6, and Figures 5(d)-5(f). Although all the algorithms increase their reward during training, the selected number of time steps (1.5×10^5) is not sufficient to perform successfully the task. Thus, $s = 0$ for all of them, and hence, $\lambda = -3.1$.

Table 3. Performance of the first training and test of different algorithms in the case of disassembly of 2 parts with 1 target. Trainings were executed on a Colab NVIDIA L4. In bold the resulting circularity (12).

	Train. time (h:min:sec)	r_s	r_e	ζ	$(c_{f,b,1}, c_{f,b,2})$	$m_{f,b,1}^0$ (kg)	$m_{f,b,2}^0$ (kg)	T_d (Sec)	s	λ
SAC-HER	2:41:46	-50	-14.3	35.7	(0.1, 0.95)	1	1	0.4	100.0	-2.1
TQC-HER	2:44:28	-50	-12.2	37.8	(0.1, 0.95)	1	1	0.8	80.0	-2.3
TD3-HER	1:38:58	-50	-50	0.0	(0.1, 0.95)	1	1	86400*	0.0	-3.1

* When disassembly fails (i.e., when $s = 0$), $c_{2,2}^2$ works as a waste collection center where material is stored for $T_d = 86400$ seconds (i.e., 1 day).

Table 4. Mean(standard deviation) of T_d , s , and λ over 5 trainings for the case of disassembly of 2 parts with 1 target, i.e., the task covered in Table 3.

	T_d	s	λ
SAC-HER	0.51(0.08)	98.0(4.0)	-2.15(0.04)
TQC-HER	17280(34560)	70.0(35.8)	-2.44(0.36)
TD3-HER	69120(34559)	4.0(8.0)	-3.11(0.08)

Table 5. Training tuning for the case of disassembly of 2 parts with 2 targets.

	Max episode length	Time steps
SAC-HER	100	1.5×10^5
TQC-HER	100	1.5×10^5
TD3-HER	100	1.5×10^5

4.1.3. 4 Parts, 2 Targets, and 2 Obstacles In this disassembly task, there are 4 stacked parts as shown in Figure 3(c). The robot is asked to pick the part in blue, which is stacked under 2 green parts. Hence, the 2 parts at the top are obstacles to be moved to a target location (the green sphere). In contrast, the part in blue should be moved to the blue sphere. In this scenario with 4 parts, 2 parts are made of β_1 and 2 parts are made of β_2 . Hence, $m_{f,b,1}^0 = m_{f,b,2}^0 = 2$ kg, and hence, $\bar{m}_0 = 2.1$ kg. The training tuning is reported in Table 7, the learning curves in Figure 5(g)-5(i), and the test results in Table 8. Although both SAC and TQC increase the reward, the number of training time steps is not sufficient to reach the reward saturation. Thus, all the algorithms yield $s = 0$ during testing, and hence, $\lambda = -6.3$. Note that the circularity in this task is lower than in the previous disassembly task (-3.1 vs. -6.3) because now there is more material to be disassembled, and hence, an incomplete disassembly ($s = 0$) is more detrimental to circularity than in the previous task. In other words, *the impact on circularity of the performance of the RL controller has a positive correlation with the mass and the criticality of the materials to be disassembled*. This finding will be discussed further in Section 4.2.

Table 6. Performance of training and test of different algorithms in the case of disassembly of 2 parts with 2 targets. Trainings were executed on a Colab NVIDIA Tesla T4. In bold the resulting circularity (12).

	Train. time (h:min:sec)	r_s	r_e	ζ	$(c_{f,b,1}, c_{f,b,2})$	$m_{f,b,1}^0$ (kg)	$m_{f,b,2}^0$ (kg)	T_d (Sec)	s	λ
SAC-HER	1:03:19	-40.0	-30.3	9.7	(0.1, 0.95)	1	1	86400	0.0	-3.1
TQC-HER	1:02:27	-40.2	-32.6	7.6	(0.1, 0.95)	1	1	86400	0.0	-3.1
TD3-HER	0:40:00	-42.9	-41.3	1.6	(0.1, 0.95)	1	1	86400	0.0	-3.1

* When disassembly fails (i.e., when $s = 0$), $c_{2,2}^2$ works as a waste collection center where material is stored for $T_d = 86400$ seconds (i.e., 1 day).

Table 7. Training tuning for the case of disassembly of 4 parts with 2 targets and 2 obstacles.

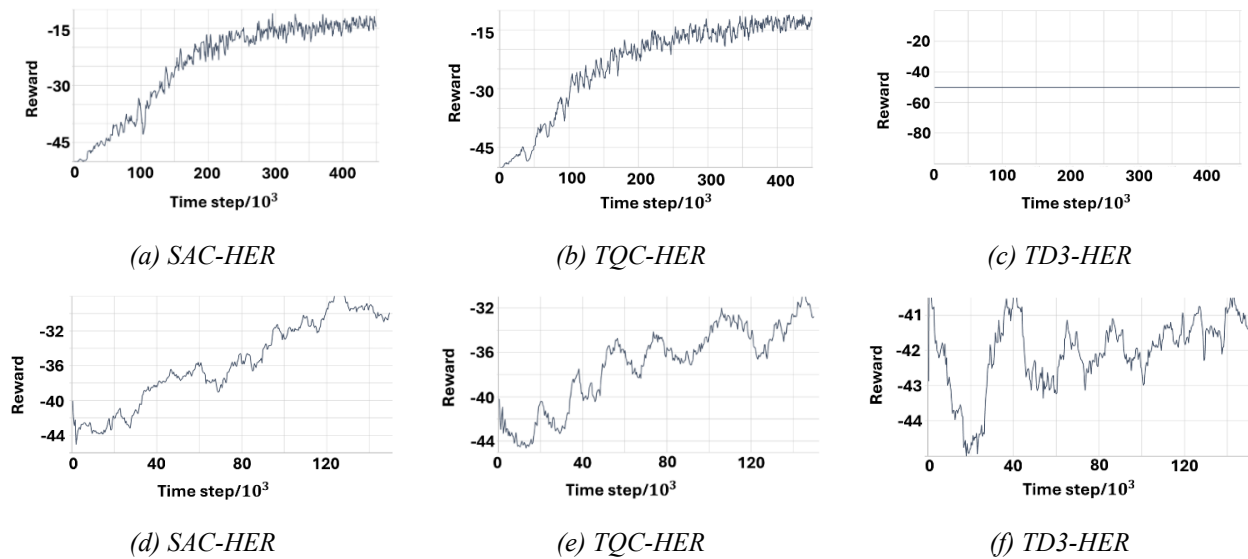
	Max episode length	Time steps
SAC-HER	100	2.0×10^5
TQC-HER	100	2.0×10^5
TD3-HER	100	2.0×10^5

Table 8. Performance of training and test of different algorithms in the case of disassembly of 4 parts with 2 targets and 2 obstacles. Trainings were executed on a Colab NVIDIA Tesla T4. In bold the resulting circularity (12).

	Train. time (h:min:sec)	r_s	r_e	ζ	$(c_{f,b,1}, c_{f,b,2})$	$m_{f,b,1}^0$ (kg)	$m_{f,b,2}^0$ (kg)	T_d (Sec)	s	λ
SAC-HER	1:21:07	-50.1	-39.8	10.3	(0.1, 0.95)	2	2	86400	0.0	-6.3
TQC-HER	1:26:05	-89.1	-39.9	49.2	(0.1, 0.95)	2	2	86400	0.0	-6.3
TD3-HER	1:04:01	-58.9	-51.0	7.9	(0.1, 0.95)	2	2	86400	0.0	-6.3

* When disassembly fails (i.e., when $s = 0$), $c_{2,2}^2$ works as a waste collection center where material is stored for $T_d = 86400$ seconds (i.e., 1 day).

4.1.4. 4 Parts, 2 Targets, 2 Obstacles, and a Chassis Similarly to the previous disassembly task, in this task there are 4 stacked parts with 2 obstacles. In addition, now there is a chassis indicated in red in Figure 3(d). The robot is asked to not collide with the chassis during disassembly (implemented as a penalty and a truncation of the episode). In this task, 2 parts are made of β_1 and 2 parts are made of β_2 , while the chassis is made of β_1 (i.e., the material with low criticality since, usually, a chassis is made of aluminum or plastic). Thus, $m_{f,b,1}^0 = 2$ kg and $m_{f,b,2}^0 = 5$ kg, and hence, $\bar{m}_0 = 2.4$ kg. The training tuning and performance for this task are reported in Table 9, Table 10, and Figures 5(j)-5(l). In the selected training time, none of the algorithms reached a reward that resulted in $s > 0$. However, they learned to not collide with the chassis as visible in Figures 4(c) and 4(d). Specifically, the two figures show the mean length of an episode during training for one training with SAC and TQC, respectively. The episode length is short at the beginning, i.e., less than 30 time steps, because of the frequent collisions between the robot and the chassis, which result in frequent episode truncations. Subsequently, the episode length increases and saturates at 150, which is the maximum value chosen before training (Table 9). The saturation to 150 indicates that the episodes are no longer truncated by collisions, but by the episode length limit.

**Figure 5.** Learning curves corresponding to the following tables: (a)-(c) to Table 3; (d)-(f) to Table 6; (g)-(i) to Table 8; and (j)-(l) to Table 10.

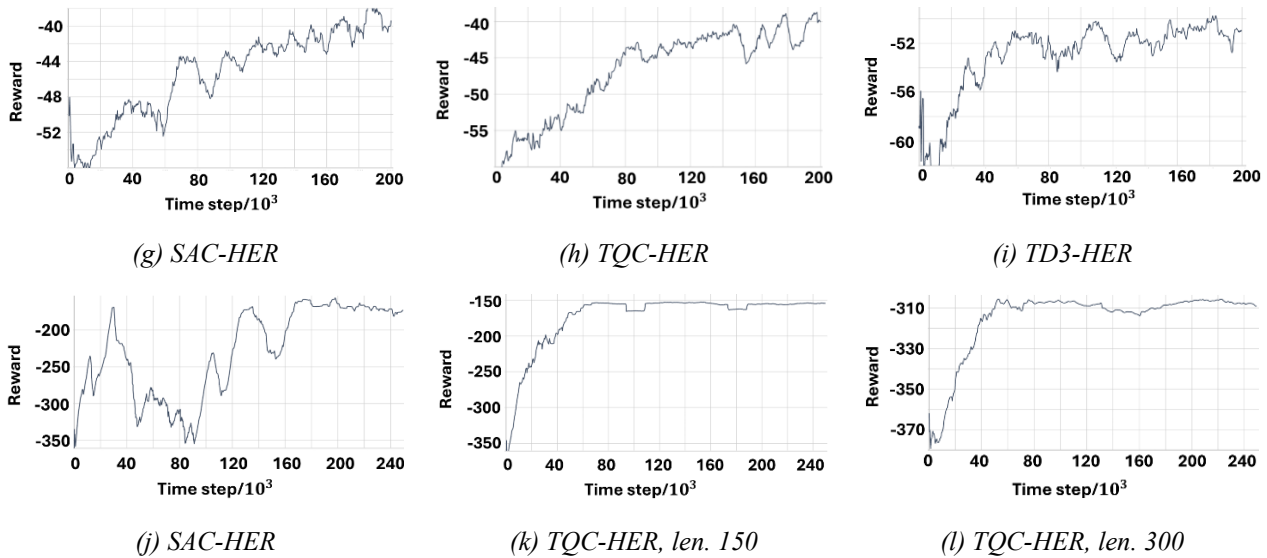


Figure 5 (Cont.). Learning curves corresponding to the following tables: (a)-(c) to Table 3; (d)-(f) to Table 6; (g)-(i) to Table 8; and (j)-(l) to Table 10.

Table 9. Training tuning for the case of disassembly of 4 parts and 1 chassis with 2 targets and 2 obstacles.

	Max episode length	Time steps
SAC-HER	150	2.5×10^5
TQC-HER	150	2.5×10^5
TQC-HER	300	2.5×10^5

In this task, the circularity is $\lambda = -7.2$ with all controllers, which is lower than in all previous disassembly tasks because here $\bar{m}_0$ has the highest value (2.4 kg). This confirms what we observed in the previous task, which is that $s = 0$ is more detrimental to circularity if the mass of materials to be disassembled is larger.

Table 10. Performance of training and test of different algorithms in the case of disassembly of 4 parts and 1 chassis with 2 targets and 2 obstacles. Trainings were executed on a Colab NVIDIA Tesla T4. In bold the resulting circularity (12).

	Train. time (h:min:sec)	r_s	r_e	ζ	$(c_{f,b,1}, c_{f,b,2})$	$m_{f,b,1}^0$ (kg)	$m_{f,b,2}^0$ (kg)	T_d (Sec)	s	λ
SAC-HER	2:15:07	-367.9	-174.0	193.9	(0.1, 0.95)	5	2	86400*	0.0	-7.2
TQC-HER	2:02:02	-370.8	-154.4	216.4	(0.1, 0.95)	5	2	86400*	0.0	-7.2
TQC-HER	2:02:57	-363.9	-308.5	55.4	(0.1, 0.95)	5	2	86400*	0.0	-7.2

* When disassembly fails (i.e., when $s = 0$), $c_{2,2}^2$ works as a waste collection center where material is stored for $T_d = 86400$ seconds (i.e., 1 day).

The four disassembly tasks discussed above can be assigned to a complexity measure expressed by an integer value between 1 (simplest) to 4 (most complex). By introducing such a complexity measure, it is possible to plot the value of λ against the task complexity, thus yielding Figure 1(a). The figure indicates that circularity decreases with the increase of task complexity. This shows a correlation between λ and s . In practice, this correlation yields that an higher task complexity makes the RL algorithm less successful, and hence, the material losses increase. In addition, Figure 1(b) shows that circularity increases when the final mean reward of the RL algorithm is higher, i.e., when the trained RL agent is more successful at performing the disassembly task. Figure 1(b) relates the primary performance indicator of RL systems, i.e., the final reward, with the system-level indicator λ , whose maximization is the central goal in circular economy. Since λ is affected by the material masses being disassembled along with the material criticality coefficient $c_{f,b}$ (recall equation (2)), Figure 1(b) also relates, implicitly, these factors to the RL performance. *We argue that*

systematically relating low-level RL performance metrics, e.g., the reward, with circularity is fundamental for developing machine learning systems aiming to implement a circular use of natural resources. Not relating them may undermine the full potential of learning systems in accelerating a circular transition because the link with circularity would be unclear and difficult to quantify.

4.2. Sensitivity of λ and Incinerator Stock to RL Performance

In this section, we analyze the variation of λ to the performance of the RL controller, i.e., to the variables s and T_d .

In practice, $T_d, T_i, T_t \ll t_{2,in,4}, T_r$, and hence,

$$t_f \approx t_{2,in,4} + T_r \quad (15)$$

(consider, for example, the case of car parts, smartphones, desktops, and laptops, that could be used for one or more years before disassembly or disposal). Inserting this in (2), it yields

$$\lambda \approx -\frac{2\bar{m}_0}{t_{2,in,4} + T_r} \left[t_{2,in,4} + \left(2 - \frac{s}{100} \right) T_r \right]. \quad (16)$$

From (16) it follows that the variation of λ with the disassembly time is negligible since T_d is of the order of hours at the most, whereas T_r and $t_{2,in,4}$ are of the order of months or years.

Let

$$\alpha(s) = \frac{t_{2,in,4} + \left(2 - \frac{s}{100} \right) T_r}{t_{2,in,4} + T_r}, \quad (17)$$

with $\alpha(s) \in [1,2]$ since $s \in [0,1]$ and

$$\lim_{T_r \rightarrow \infty} \alpha(0) = 2. \quad (18)$$

Specifically,

$$\alpha(s) = \begin{cases} \frac{t_{2,in,4} + 2T_r}{t_{2,in,4} + T_r} & \text{if } s = 0 \text{ (min } s, \text{ max } \alpha) \\ 1 & \text{if } s = 100 \text{ (max } s, \text{ min } \alpha), \end{cases} \quad (19)$$

and hence,

$$\lambda \approx \begin{cases} -2\bar{m}_0 \frac{t_{2,in,4} + 2T_r}{t_{2,in,4} + T_r} & \text{if } s = 0 \text{ (min } s, \text{ min } \lambda) \\ -2\bar{m}_0 & \text{if } s = 100 \text{ (max } s, \text{ max } \lambda). \end{cases} \quad (20)$$

Equations (16) and (20) show that: first, an increase of disassembly success s yields an increase in circularity if the other equation parameters are fixed; second, since α is independent of $\bar{m}_0$, the sensitivity of λ to s has a positive correlation with $\bar{m}_0$, which confirms that the importance of a performing RL controller increases when there is an increase of the masses and/or criticality of the materials to be disassembled; third, circularity is approximately independent of time in the case of perfect disassembly (i.e., $s = 100$); fourth, since λ has a *negative* correlation with $\bar{m}_0$, circularity increases when there is a reduction of extracted masses (i.e., $m_{f,b,i}^0$) and/or material criticality (i.e., $c_{f,b,i}$) while the other equation parameters are fixed; and fifth, assuming to keep s constant and to set T_r as a variable, from (16) we have that

$$\lim_{T_r \rightarrow \infty} \lambda(T_r) \approx -2\bar{m}_0 \left(2 - \frac{s}{100}\right) \in [-4\bar{m}_0, -2\bar{m}_0], \quad (21)$$

meaning that λ scales proportionally to the weighted mass $\bar{m}_0$ for a large reuse time T_r , i.e., the mass becomes the dominant parameter. This behavior of λ is consistent with an intuitive behavior of circularity since circularity is expected to decrease with an increase of $\bar{m}_0$. However, an *increase* of T_r in equation (16) can yield a *decrease* in λ , which does not align with the definition of circularity of (Ellen MacArthur Foundation, 2020). Indeed, an increase of the reuse time should yield an increase in circularity when the other parameters are fixed. To highlight this phenomenon, we refer to it as an *inverted effect* of the reuse time.

Now, let us see how RL affects the stock inside the incinerator. From equation (6), it follows that the total mass moving in the system is $m_0 = m_{f,b,1}^0 + m_{f,b,2}^0$, which is split, at $t_{3,in,5}$, into the reused stock $m_r = m_0 \frac{s}{100}$ and the waste stock $m_u = m_0 - m_r$ (see Figure 2). Eventually, all the mass enters the incinerator since the reusability of materials and products is not infinite. This is visible for $t > t_{3,in,6}$. Thus, the mass inside the incinerator is affected by s as shown in Figure 6 for the disassembly task corresponding to Table 10. As visible, the incinerator stock behaves as a *step function* whose steps are affected by s . This occurs also in the other disassembly tasks.

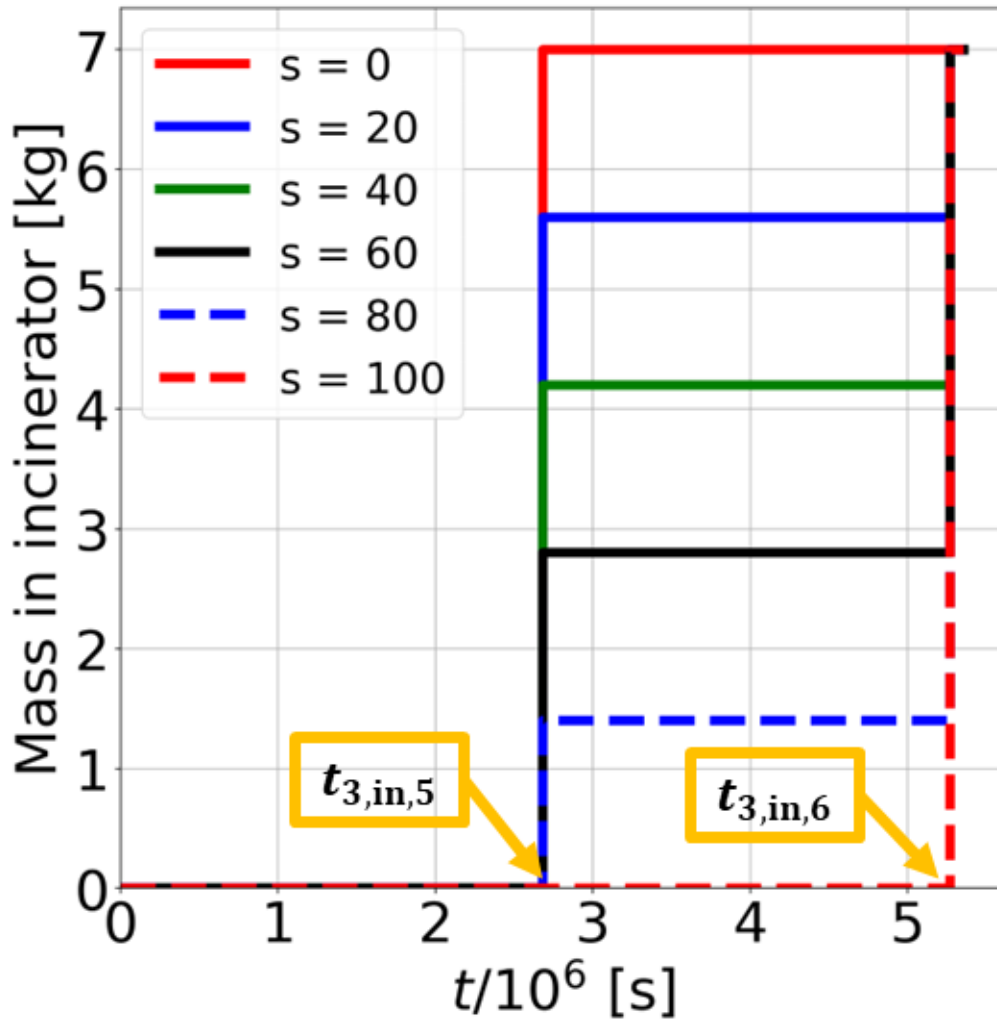


Figure 6. Stock inside the incinerator vs. time and for different values of s . Eventually, all the mass enters the incinerator as in real scenarios, where the reusability of materials is not infinite. Increasing s yields a delay of the time at which the material becomes waste and enters the incinerator.

5. Conclusions and Future Work

This paper developed a network of thermodynamic compartments with RL-controlled robotic disassembly to increase the circularity of the whole system. More complex disassembly tasks are more challenging for RL yielding a lower percentage success, and hence, due to the sensitivity of circularity to the performance of the RL controller, result in lower circularity. In addition, the impact of the performance of the RL-controlled disassembler increases when the criticality and/or the quantity of the materials to be disassembled increases.

In summary, there are three main actions that increase circularity: increasing the performance of the robotic disassembler when dealing with multi-component products (action for the disassembly compartment), reducing the extraction of non-renewable materials (action for the manufacturer compartment), and reducing the criticality of the materials used to make products (action for the manufacturer compartment). In addition, the suggested RL algorithms are SAC-HER and TQC-HER as they showed the best performance, while the criticality of materials has a strong impact on circularity since the criticality coefficient $c_{f,b}$ enters as a multiplying factor of the material mass, e.g., the most intense reduction of circularity is caused by extracting and incinerating a large mass of a highly critical material, e.g., lithium.

This work has implications at different levels. Specifically, at a managerial level, it demonstrates the importance that robotic disassembly has in improving the circularity of resources; operational cost of robotic disassembly is a viable and justifiable option only if large volumes of waste items are reprocessed and recovered. At a policy level, this work highlights the importance of regulations in promoting the extension of material and product life cycles to achieve an economy that is less dependent on finite natural resources. At a theoretical level, our physics-inspired definition of circularity combined with thermodynamical networks provides a framework for a design approach to supply-recovery chains that can be integrated with deep reinforcement learning. While such an approach is more mathematically intensive than pure data-analysis methods, e.g., MFA, it can enhance the accuracy of data-based models in the frequent situation of a lack of data. In addition, TMNs can depict fast dynamics occurring in less than 1 second, whereas MFA studies usually depict only slow dynamics, e.g., variations from year n to year $n + 1$ as in (Houssini et al., 2025), in (Lipp and Lederer, 2025), and in (Grund et al., 2025).

At a network level, future work will investigate the system circularity in the case of extractions of multiple batches of material performed at different times along with addressing the undesired inverted effect of the reuse time, namely, T_r . At a compartment level, future work will further develop the disassembly simulator and the RL controller, and then, implement the RL controller on a real robot for disassembly tasks. A challenge of developing TMNs consists of scalability since the one proposed in this paper, namely, CIRO7.2, comprises only 6 compartments and processes 2 materials, and hence, it is a proof-of-concept. Compartments and materials need to be added to design a network able to accurately depict a real system. Scaling-up requires the time for writing down the dynamical equations and the computational resources to simulate different scenarios corresponding to different network parameters. This paper opens the door for integrating machine learning with physical models of material flows, enabling a new generation of circularity-optimized supply chains.

Acknowledgments The authors thank Giuseppe Saviano at University of Siena for the advice on the panda-gym library and on RL implementation aspects.

Author contributions Federico Zocco: conceptualization, methodology, software, validation, formal analysis, investigation, resources, writing - original draft, writing - review & editing, visualization. Monica Malvezzi: conceptualization, methodology, validation, formal analysis, resources, writing - review & editing, supervision.

Funding The authors declare that no funding was received for this research.

Data availability No data were used for this research.

Declarations

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons License, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons License and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abu-Bakar, Halidu, and Fiona Charnley. 2024. "Developing a Strategic Methodology for Circular Economy Roadmapping: A Theoretical Framework." *Sustainability* 16 (15): 6682. <https://doi.org/10.3390/su16156682>.
- Andrychowicz, Marcin, Filip Wolski, Alex Ray, et al. 2017. "Hindsight Experience Replay." *Advances in Neural Information Processing Systems* 30.
- Baranowski, Mariusz, Piotr Jabkowski, and Daniel M Kammen. 2025. "From the Russian invasion of Ukraine to the battlefield of the future: The geopolitical fight for Ukraine's mineral wealth." *Energy Research & Social Science* 123: 104043. <https://doi.org/10.1016/j.erss.2025.104043>.
- Barteková, Eva, and Peter Börkey. 2022. *Digitalisation for the Transition to a Resource Efficient and Circular Economy*.
- Beghi, Matteo, Francesco Braghin, and Loris Roveda. 2023. "Enhancing disassembly practices for electric vehicle battery packs: A narrative comprehensive review." *Designs* 7 (5): 109. <https://doi.org/10.3390/designs7050109>.
- Bondy, John Adrian, and Uppaluri Siva Ramachandra Murty. 1976. *Graph Theory with Applications*. Vol. 290. Macmillan London.
- Brunner, Paul H, and Helmut Rechberger. 2016. *Handbook of Material Flow Analysis: For Environmental, Resource, and Waste Engineers*. CRC press.
- Cai, WeiBiao, Xiwang Guo, Jiacun Wang, Shujin Qin, Jian Zhao, and Yuanyuan Tan. 2022. "An Improved Advantage Actor-Critic Algorithm for Disassembly Line Balancing Problems Considering Tools Deterioration." *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3336–41.
- Calzolari, Tommaso, Andrea Genovese, and Andrew Brint. 2022. "Circular Economy indicators for supply chains: A systematic literature review." *Environmental and Sustainability Indicators* 13: 100160. <https://doi.org/10.1016/j.indic.2021.100160>.
- CE-HubTeam. 2022. *Demystifying Series: Business and a Circular Economy*.
- Chen, Xinxing, Huaiyang Huang, Yuxuan Liu, Jiqing Li, and Ming Liu. 2022. "Robot for Automatic Waste Sorting on Construction Sites." *Automation in Construction* 141: 104387. <https://doi.org/10.1016/j.autcon.2022.104387>.
- CRM Alliance. 2024. *What Are Critical Raw Materials?*
- Cullen, Jonathan M, and Daniel R Cooper. 2022. "Material Flows and Efficiency." *Annual Review of Materials Research* 52 (1): 525–59. <https://doi.org/10.1146/annurev-matsci-070218-125903>.
- Ellen MacArthur Foundation. 2020. *What Is the Meaning of a Circular Economy and What Are the Main Principles?*
- Ellen MacArthur Foundation. 2022a. *Circulate Products and Materials*.
- Ellen MacArthur Foundation. 2022b. *Introduction to an Adaptive Strategy for Circular Design*.

- Emami, Nargessadat, Quoc Anh Nguyen, Alessio Miatto, et al. 2026. "Material Flow Analysis of Plastics in Australia: Towards a Circular Economy of Polymers." *Resources, Conservation and Recycling* 227: 108718. <https://doi.org/10.1016/j.resconrec.2025.108718>.
- Eschmann, Jonas. 2021. "Reward Function Design in Reinforcement Learning." *Reinforcement Learning Algorithms: Analysis and Applications*, 25–33.
- European Commission. 2023. *Critical Raw Materials*.
- Eurostat. 2024. *Waste Statistics*.
- Federal Register. 2023. *Notice of Final Determination on 2023 DOE Critical Materials List*.
- Fet, Annik Magerholm, and Paritosh C Deshpande. 2023. "Closing the Loop: Industrial Ecology, Circular Economy and Material Flow Analysis." In *Business transitions: A path to sustainability: The CapSEM model*. Springer.
- Fujimoto, Scott, Herke Hoof, and David Meger. 2018. "Addressing Function Approximation Error in Actor-Critic Methods." *International Conference on Machine Learning*, 1587–96.
- Gallouédec, Quentin, Nicolas Cazin, Emmanuel Dellandréa, and Liming Chen. 2021. "panda-gym: Open-source goal-conditioned environments for robotic learning." *4th Robot Learning Workshop: Self-Supervised and Lifelong Learning@ NeurIPS 2021*.
- Ghignone, Edoardo, Nicolas Baumann, and Michele Magno. 2023. "TC-Driver: A trajectory conditioned reinforcement learning approach to zero-shot autonomous racing." *Field Robotics* 3: 637–51. <https://doi.org/10.1109/TFR.2024.3499912>.
- Graedel, Thomas E, and Matthew J Eckelman. 2023. *Industrial Ecology and Sustainability*. World Scientific.
- Grund, Sabina, Eric van Genderen, and Martin van Leeuwen. 2025. "Using Material Flow Analysis to Identifying Global and Regional Opportunities for Zinc Circularity." *Journal of Advanced Manufacturing and Processing* 7 (3): e70020.
- Guo, Xiwang, Zhiliang Bi, Jiacun Wang, Shujin Qin, Shixin Liu, and Liang Qi. 2023. "Reinforcement Learning for Disassembly System Optimization Problems: A survey." *International Journal of Network Dynamics and Intelligence* 2 (1): 1–14. <https://doi.org/10.53941/ijndi0201001>.
- Haarnoja, Tuomas, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor." *International Conference on Machine Learning*, 1861–70.
- Haddad, Wassim M. 2017. "Thermodynamics: The Unique Universal Science." *Entropy* 19 (11): 621. <https://doi.org/10.3390/e19110621>.
- Haddad, Wassim M. 2019. *A Dynamical Systems Theory of Thermodynamics*. Princeton University Press.
- Hagelüken, Christian, and Daniel Goldmann. 2022. "Recycling and circular economy—Towards a closed loop for metals in emerging clean technologies." *Mineral Economics* 35 (3): 539–62. <https://doi.org/10.1007/s13563-022-00319-1>.
- Henckens, MLCM, PPJ Driessen, and E Worrell. 2014. "Metal Scarcity and Sustainability, Analyzing the Necessity to Reduce the Extraction of Scarce Metals." *Resources, Conservation and Recycling* 93: 1–8. <https://doi.org/10.1016/j.resconrec.2014.09.012>.
- Henckens, Theo. 2021. "Scarce Mineral Resources: Extraction, Consumption and Limits of Sustainability." *Resources, Conservation and Recycling* 169: 105511. <https://doi.org/10.1016/j.resconrec.2021.105511>.
- Houssini, Khaoula, Jinhui Li, and Quanyin Tan. 2025. "Complexities of the Global Plastics Supply Chain Revealed in a Trade-Linked Material Flow Analysis." *Communications Earth & Environment* 6 (1): 257.
- Ibarz, Julian, Jie Tan, Chelsea Finn, Mrinal Kalakrishnan, Peter Pastor, and Sergey Levine. 2021. "How to Train Your Robot with Deep Reinforcement Learning: Lessons We Have Learned." *The International Journal of Robotics Research* 40 (4-5): 698–721. <https://doi.org/10.1177/0278364920987859>.

- Jia, Yuanyuan, and Shugen Ma. 2021. "A Coach-Based Bayesian Reinforcement Learning Method for Snake Robot Control." *IEEE Robotics and Automation Letters* 6 (2): 2319–26. <https://doi.org/10.1109/LRA.2021.3061372>.
- Ju, Hao, Rongshun Juan, Randy Gomez, Keisuke Nakamura, and Guangliang Li. 2022. "Transferring Policy of Deep Reinforcement Learning from Simulation to Reality for Robotics." *Nature Machine Intelligence* 4 (12): 1077–87. <https://doi.org/10.1038/s42256-022-00573-6>.
- Kirchherr, Julian, Nan-Hua Nadjia Yang, Frederik Schulze-Spüntrup, Maarten J Heerink, and Kris Hartley. 2023. "Conceptualizing the circular economy (revisited): An analysis of 221 definitions." *Resources, Conservation and Recycling* 194: 107001. <https://doi.org/10.1016/j.resconrec.2023.107001>.
- Kiyokawa, Takuya, Jun Takamatsu, and Shigeki Koyanaka. 2022. "Challenges for future robotic sorters of mixed industrial waste: A survey." *IEEE Transactions on Automation Science and Engineering* 21 (1): 1023–40. <https://doi.org/10.1109/TASE.2022.3221969>.
- Koskinopoulou, Maria, Fredy Raptopoulos, George Papadopoulos, Nikitas Mavrakis, and Michail Maniadakis. 2021. "Robotic Waste Sorting Technology: Toward a Vision-Based Categorization System for the Industrial Robotic Separation of Recyclable Waste." *IEEE Robotics & Automation Magazine* 28 (2): 50–60. <https://doi.org/10.1109/MRA.2021.3066040>.
- Kristensen, Christoffer B, Frederik A Sørensen, Hjalte B Nielsen, Martin S Andersen, Søren P Bendtsen, and Simon Bøgh. 2019. "Towards a Robot Simulation Framework for e-Waste Disassembly Using Reinforcement Learning." *Procedia Manufacturing* 38: 225–32. <https://doi.org/10.1016/j.promfg.2020.01.030>.
- Kuznetsov, Arsenii, Pavel Shvechikov, Alexander Grishin, and Dmitry Vetrov. 2020. "Controlling Overestimation Bias with Truncated Mixture of Continuous Distributional Quantile Critics." *International Conference on Machine Learning*, 5556–66.
- Laili, Yuanjun, Xiang Li, Yongjing Wang, Lei Ren, and Xiaokang Wang. 2021. "Robotic Disassembly Sequence Planning with Backup Actions." *IEEE Transactions on Automation Science and Engineering* 19 (3): 2095–107. <https://doi.org/10.1109/TASE.2021.3072663>.
- Laili, Yuanjun, Yongjing Wang, Yilin Fang, and Duc Truong Pham. 2022. *Optimisation of Robotic Disassembly for Remanufacturing*. Springer.
- Lapan, Maxim. 2018. *Deep Reinforcement Learning Hands-on*. Packt Publishing Birmingham.
- Lee, Meng-Lun, Sara Behdad, Xiao Liang, and Minghui Zheng. 2022. "Task Allocation and Planning for Product Disassembly with Human–Robot Collaboration." *Robotics and Computer-Integrated Manufacturing* 76: 102306. <https://doi.org/10.1016/j.rcim.2021.102306>.
- Li, J., M. Barwood, and S. Rahimifard. 2018. "Robotic Disassembly for Increased Recovery of Strategically Important Materials from Electrical Vehicles." *Robotics and Computer-Integrated Manufacturing* 50: 203–12. <https://doi.org/10.1016/j.rcim.2017.09.013>.
- Liao, Jiankan, Sidi Deng, Xun Huan, and Daniel Cooper. 2025. "Bayesian Model Selection for Network Discrimination and Risk-Informed Decision-Making in Material Flow Analysis." *Journal of Industrial Ecology* 29 (4): 1060–76. <https://doi.org/10.1111/jiec.70034>.
- Lieder, Michael, and Amir Rashid. 2016. "Towards circular economy implementation: A comprehensive review in context of manufacturing industry." *Journal of Cleaner Production* 115: 36–51. <https://doi.org/10.1016/j.jclepro.2015.12.042>.
- Lillicrap, Timothy P, Jonathan J Hunt, Alexander Pritzel, et al. 2015. "Continuous Control with Deep Reinforcement Learning." *arXiv Preprint arXiv:1509.02971*. <https://doi.org/10.48550/arXiv.1509.02971>.
- Lipp, Anna-Maria, and Jakob Lederer. 2025. "The Circular Economy of Packaging Waste in Austria: An Evaluation Based on Statistical Entropy and Material Flow Analysis." *Resources, Conservation and Recycling* 217: 108193.
- Mason, Adam R, Tom Bide, Junyang Wang, et al. 2025. "Bayesian material flow analysis of the construction aggregate cycle in England." *Resources, Conservation and Recycling* 215: 108135. <https://doi.org/10.1016/j.resconrec.2025.108135>.

- Moran, Michael J, Howard N Shapiro, Daisie D Boettner, and Margaret B Bailey. 2010. *Fundamentals of Engineering Thermodynamics*. John Wiley & Sons.
- Munonye, Williams Chibueze. 2025. "Towards circular economy metrics: A systematic review." *Circular Economy and Sustainability* 5 (5): 4093–135. <https://doi.org/10.1007/s43615-025-00604-5>.
- Nature. 2024. *Physics Nobel Scooped by Machine-Learning Pioneers*.
- Nygaard, Arne. 2023. "The geopolitical risk and strategic uncertainty of green growth after the Ukraine invasion: How the circular economy can decrease the market power of and resource dependency on critical minerals." *Circular Economy and Sustainability* 3 (2): 1099–126. <https://doi.org/10.1007/s43615-022-00181-x>.
- OpenAI. 2018a. *Deep Deterministic Policy Gradient*.
- OpenAI. 2018b. *Twin Delayed DDPG*.
- Paletti, Ligeia, Dionysios Markatos, Spiros Pantelakis, and Angelos Filippatos. 2026. "Assessing Circularity performance in aircraft component design using ISO 59004:2024." *Journal of Circular Economy* 4 (1): 1–38. <https://doi.org/10.55845/joce-2026-4139>.
- Potting, José, Marko P Hekkert, Ernst Worrell, and Aldert Hanemaaijer. 2017. "Circular Economy: Measuring Innovation in the Product Chain." In *Planbureau Voor de Leefomgeving*. PBL Publishers.
- Qu, Mo, Yongjing Wang, and Duc Truong Pham. 2023. "Robotic Disassembly Task Training and Skill Transfer Using Reinforcement Learning." *IEEE Transactions on Industrial Informatics* 19 (11): 10934–43. <https://doi.org/10.1109/TII.2023.3242831>.
- Raffin, Antonin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. 2021. "Stable-baselines3: Reliable reinforcement learning implementations." *Journal of Machine Learning Research* 22 (268): 1–8. <http://jmlr.org/papers/v22/20-1364.html>.
- Rapetswa, Kagiso, and Ling Cheng. 2023. "Towards a Multi-Agent Reinforcement Learning Approach for Joint Sensing and Sharing in Cognitive Radio Networks." *Intelligent and Converged Networks* 4 (1): 50–75. <https://doi.org/10.23919/ICN.2023.0005>.
- Rashid, Amir. 2026. "How circular is the linear economy? Analysing circularity, resource flows and their relation to GDP." *Journal of Circular Economy* 4 (1): 442–53. <https://doi.org/10.55845/joce-2026-4179>.
- Singh, Bharat, Rajesh Kumar, and Vinay Pratap Singh. 2022. "Reinforcement learning in robotic applications: A comprehensive survey." *Artificial Intelligence Review* 55 (2): 945–90. <https://doi.org/10.1007/s10462-021-09997-9>.
- Smith, Rachel L, and Sara Behdad. 2025. "From Present to Future: A Review of e-Waste Recycling Processes." *Waste Management* 204: 114863. <https://doi.org/10.1016/j.wasman.2025.114863>.
- Stable-Baselines3. 2026. *TQC*.
- Sutton, Richard S, and Andrew G Barto. 1998. *Reinforcement Learning: An Introduction*. Vol. 1. MIT press Cambridge.
- United Nations. n.d. *Sustainability*.
- Yau, Kok-Lim Alvin, Yung-Wey Chong, Xiumei Fan, Celimuge Wu, Yasir Saleem, and Phei-Ching Lim. 2023. "Reinforcement Learning Models and Algorithms for Diabetes Management." *IEEE Access* 11: 28391–415. <https://doi.org/10.1109/ACCESS.2023.3259425>.
- Yuan, Gang, Xiaojun Liu, Xiaoli Qiu, Pai Zheng, Duc Truong Pham, and Ming Su. 2025. "Human-Robot Collaborative Disassembly in Industry 5.0: A Systematic Literature Review and Future Research Agenda." *Journal of Manufacturing Systems* 79: 199–216. <https://doi.org/10.1016/j.jmsy.2025.01.009>.
- Zhou, Jiayi, and André Månberger. 2024. "Critical minerals and great power competition: An overview." *Stockholm International Peace Research Institute*, 1–50.

- Zocco, Federico. 2026. "Circularity of Thermodynamical Material Networks: Indicators, Examples, and Algorithms." *Circular Economy and Sustainability* 6 (2): 94. <https://doi.org/10.1007/s43615-026-00860-z>.
- Zocco, Federico, Wassim M Haddad, Andrea Corti, and Monica Malvezzi. 2024. "A Unification Between Deep-Learning Vision, Compartmental Dynamical Thermodynamics, and Robotic Manipulation for a Circular Economy." *IEEE Access* 12: 173502–16. <https://doi.org/10.1109/ACCESS.2024.3493755>.
- Zocco, Federico, Daniel R Lake, Seán McLoone, and Shahin Rahimifard. 2025. "Synchronized Object Detection for Autonomous Sorting, Mapping, and Quantification of Materials in Circular Healthcare." *IEEE Transactions on Instrumentation and Measurement* 74: 5505110. <https://doi.org/10.1109/TIM.2025.3574911>.
- Zocco, Federico, and Monica Malvezzi. 2025. "Circular Economy Design Through System Dynamics Modeling." *International Workshop IFToMM for Sustainable Development Goals*, 530–38.
- Zocco, Federico, Denis Sleath, and Shahin Rahimifard. 2025. "Towards a Thermodynamical Deep-Learning-Vision-Based Flexible Robotic Cell for Circular Healthcare." *Circular Economy and Sustainability*, 1–23. <https://doi.org/10.1007/s43615-025-00532-4>.
- Zocco, Federico, Pantelis Sopasakis, Beatrice Smyth, and Wassim M Haddad. 2023. "Thermodynamical Material Networks for Modeling, Planning, and Control of Circular Material Flows." *International Journal of Sustainable Engineering* 16 (1): 1–14. <https://doi.org/10.1080/19397038.2023.2209582>.
- Zorpas, Antonis A. 2024. "The hidden concept and the beauty of multiple 'R' in the Framework of Waste Strategies Development Reflecting to Circular Economy Principles." *Science of the Total Environment* 952: 175508. <https://doi.org/10.1016/j.scitotenv.2024.175508>.