



Coupling of an analytical rolling model and reinforcement learning to design pass schedules: towards properties controlled hot rolling

C. Idzik¹ · A. Krämer¹ · G. Hirt¹ · J. Lohmar¹

Received: 31 December 2021 / Accepted: 15 March 2023 / Published online: 19 April 2023
© The Author(s) 2023

Abstract

Rolling is a well-established forming process employed in many industrial sectors. Although highly optimized, process disruptions can still lead to undesired final mechanical properties. This paper demonstrates advances in pass schedule design based on reinforcement learning and analytical rolling models to guarantee sound product quality. Integrating an established physical strengthening model into an analytical rolling model allows tracking the microstructure evolution throughout the process, and furthermore the prediction of the yield strength and ultimate tensile strength of the rolled sheet. The trained reinforcement learning algorithm Deep Deterministic Policy Gradient (DDPG) automatically proposes pass schedules by drawing upon established scheduling rules combined with novel rule sets to maximize the final mechanical properties. The designed pass schedule is trialed using a laboratory rolling mill while the predicted properties are confirmed using micrographs and materials testing. Due to its fast calculation time, prospectively this technique can be extended to also account for significant process disruptions such as longer inter-pass times by adapting the pass schedule online to still reach the desired mechanical properties and avoid scrapping of the material.

Keywords Hot rolling · Pass schedule design · Reinforcement learning · Fast rolling models · Properties control

Introduction

Rolling is a widely-used and established forming process employed in process chains for different metallic components and several industrial sectors, e.g., the construction and automotive industry. Allwood et al. (2012) showed that about 95% of steel products and over 60% of aluminum products undergo at least one rolling process during their production. Therefore, even relatively small optimizations of the already highly optimized rolling process have a significant effect on a global scale regarding energy and material consumption.

Nowadays, mostly two factors affect the process efficiency in the (hot) rolling process. One of these factors is the process design, more specifically the pass schedule generation (Spuzic et al., 2017). The pass schedule defines the height reduction or draft in each pass, the rolling velocity and the

inter-pass time. The design of the pass schedule has to consider several constraints, e.g., maximum allowable rolling force and torque, economic aspects, e.g., the overall process duration, and ecological aspects, e.g., the energy consumption. At the same time, the schedule has to guarantee product properties, e.g., ultimate tensile strength, within customer specified tolerances (Li et al., 2012; Liu et al., 2019).

Typically, pass schedules are generated by specialized heuristics designed by experts based on their knowledge, experiences and with the support of fast analytical rolling models or complex finite element method (FEM) simulations (Özgür et al., 2021; Pandey et al., 2020). Optimization procedures are less prevalent, as in reverse rolling the number of passes required is unknown and additionally different objectives with complex interdependencies need to be satisfied. Simultaneously, novel product specifications cause more restrictive process windows (Peng et al., 2021; Schmidtchen & Kawalla, 2016; Shen et al., 2022), leading to even higher demands on the pass schedules design. Novel approaches to finding optimal schedules that improve on established heuristics are being sought here.

✉ C. Idzik
christian.idzik@ibf.rwth-aachen.de

¹ Institute of Metal Forming, RWTH Aachen University,
Aachen, Germany

The second relevant factor that might lead to inefficiency is irregularities, such as deviations in material properties and process disruptions like elongated inter-pass times due to scheduling problems. These irregularities can lead to properties outside the specified tolerances. This is why rolled material is sometimes scrapped or downgraded in industry (Allwood et al., 2012). Therefore, the aim is to design pass schedules that on the one hand satisfy all above-mentioned requirements and on the other hand are robust with respect to common process disruptions and deviations in material properties.

A novel approach to this is to combine methods of artificial intelligence (AI) with physical models and rolling mill process data (Rath et al., 2019). An early example presented by Larkiola et al. (1998), shows that the data-driven calibration of a neural network based roll force model for a tandem cold rolling mill can improve efficiency by 1.8% due to higher rolling speeds enabled by the gained precision in roll force. Recently, Scheiderer et al. (2020) demonstrated that reinforcement learning (a subset of AI), can use stored model data to generate pass schedules while accounting for multiple objectives. Furthermore, combining AI and big data might help to explore unconventional pass schedules, revealing hidden relationships in the process and thus extend knowledge. However, a direct coupling between a process model and a reinforcement learning algorithm in order to ensure product specifications has not been trialed, yet.

Therefore, the goal of this paper is to couple reinforcement learning (RL) with a fast analytical rolling model (FRM) to design pass schedules that yield properties within specified tolerances and simultaneously satisfy the conventional objectives, e.g., maximize productivity as well as consider typical limitations, e.g., allowable rolling forces. FRM are chosen since they provide the required results within seconds compared to minutes when using FEM. In contrast to using stored data as done by Scheiderer et al. (2020), FRM can also provide results for process parameters that would exceed machine limits thus enabling the algorithm to learn these limits quicker. In the future, transfer learning could be used to quickly create not just a single but several pass schedules for different conditions using a model that only needs to be calibrated once.

In this paper first, an overview of analytical FRMs, pass schedule generation and machine learning applications in metal forming is given in the state of the art. Afterwards, an existing rolling model is extended to predict yield strength (YS) and ultimate tensile strength (UTS). Next, the extended rolling model and a reinforcement learning algorithm are coupled, where the reward function adheres to specified tolerances (height, grain size), final mechanical properties (UTS), machine limits and total energy consumption. Thereafter, the results of training the reinforcement learning algorithm are presented to demonstrate the suitability of the coupling for

optimized pass schedules. Finally, experimental validation trials on a laboratory scale rolling mill with consecutive tensile tests to determine the YS and UTS are presented.

State of the art

In this section FRM to calculate pass schedules and there most relevant features are introduced first. Next, established approaches to generate pass schedules and the relevant objectives are presented. Afterwards, an introduction to artificial intelligence and reinforcement learning in particular is given. Finally, the state of the art is assessed with regards to opportunity for improved pass schedule generation.

Fast rolling models

FRMs are typically based on simplified mechanical assumptions also known as slab method combined with mostly semi-empirical material equations. In dependence of specific use cases, several FRMs were developed and are still used today—especially in industry.

The development of these FRMs goes back to the 1920 s of the last century, where Kármán (1925) and Siebel (1925) used basic mechanics to describe and analyze the rolling process. Based on their fundamental findings, known as the slab method, Sims (1954) published simplified equations to predict roll forces and roll torques. The description of the microstructure and product properties requires a basic understanding of the physical metallurgy, which was pioneered by, Johnson and Mehl (1939), Avrami (1939) and Kolmogorov (1937). The authors developed an equation, today known as the JMAK-equation, to describe the kinetics of crystallization. Based on this fundamental knowledge the understanding of dynamic and static recrystallization and its effect on microstructure during hot deformation processes were extended for instance by Sellars and Tegart (1966) and Jonas et al. (1969) in the 1960 s. Later a concept to capture the microstructural changes also during partial recrystallization was put forward by Beynon and Sellars (1992). To calibrate these semi-empirical models, material-dependent parameters are required. Sellars (1979) demonstrated how to obtain these parameters for C-Mn and low-alloy steels from laboratory experiments while Hodgson and Gibbs (1992) focused on micro-alloyed steels.

Based on this knowledge, numerous FRMs were developed. One well-known model named SLIMMER was developed by Beynon and Sellars (1992). The authors coupled a thermal and a microstructure model with Sims' rolling theory in order to predict roll forces and torques and describe the microstructure evolution during multi-pass hot rolling. For the thermal model an explicit finite difference technique was used, which subdivided a transverse section of

the rolled material into quadrilateral elements. Seuren et al. (2014), showed that such a model can be extended to incorporate shear strain determined in FE simulations and in turn improved the microstructure prediction over roll stock thickness. Furthermore, Lohmar et al. (2014a) combined inverse modeling and fast rolling models to determine semi-empirical material model parameters based on industrial rolling data.

The extension of FRMs, most significant for the present study, is the prediction of mechanical properties of the rolled product like YS and UTS. There is a plentitude of models to describe mechanical properties associated with different steel grades, like bainitic steels (Edmonds & Cochrane, 1990) or dual-phase steels (Lanzilloto & Pickering, 2013). The following survey is restricted to the ferritic-pearlitic microstructure of C-Mn steels—relevant to the S355 steel considered in this study. A comprehensive literature review about different modelling approaches for final ferrite grain size, YS and UTS was given by Lenard et al. (1999).

An important early contribution related to mechanical properties was made by Hall (1951) and Petch (1953) in the 1950s who discovered the inverse correlation between the increase of YS and the ferrite grain size. Today, this phenomenon is known as Hall–Petch effect. Gladman et al. (1972) extended the Hall–Petch effect for ferritic-pearlitic microstructures by considering the influence of the volume fraction of pearlite and adding a solid solution term to the relationship which considers the effects of Si and N on the mechanical properties. In a similar fashion, Choquet et al. (1990) modelled the final ferrite grain size based on the austenite grain size, retained strain, cooling rate and chemical composition and proposed a YS and UTS prediction equation for a range of C-Mn steels ($0.01 < C < 0.3$, $0.1 < Mn < 1.6$). In a related study Hodgson and Gibbs (1992) extended a modelling approach for the ferrite grain size first proposed by Sellars and Beynon (1985) to incorporate the influence of carbon and manganese. To enable a simpler prediction of YS and UTS compared to Choquet et al. (1990), the pearlite fraction influence was omitted while a precipitation hardening term was added. Still today, model improvements to predict YS and UTS are put forward. A recent example is work by Singh et al. (2013) that adds a dislocation strengthening term based on a dislocation density model presented by Wang and Tseng (1996).

Finally, in recent years, industrial data has been increasingly used for data-driven modelling of final properties. The complexity of the models ranges from relatively simple examples, e.g., by Saravanakumar et al. (2012) where five inputs like coiling temperature and carbon equivalent are used for the YS and UTS prediction to more complex approaches that enable a predictions for different steel grades. In a model developed by Xie et al. (2021) for example a deep

neural network trained using 27 inputs is used to predict inter alia YS and UTS for four steel grades.

Pass schedule generation

The aim of pass schedule generation first and foremost is to lay out a sequence of rolling passes that transfers the initial slab geometry into the final strip geometry. To optimize productivity, the draft in each pass should be maximized in accordance with the relevant constraints, i.e., maximum allowable draft, roll torque and force, bite condition, strip shape and flatness. Historically pass schedules were laid out by iterative approaches where maximum allowable draft passes were stringed together until the desired thickness was reached. Then the draft was reduced to comply with the aforementioned constraints adding passes in this process if necessary and also calculating the slab temperature evolution to ensure a suitable delivery temperature. When developing one of the first such iterative scheduling procedure Fujii and Saito (1975) noted that roll schedules for reversing mills are restricted by maximum design draft or torque in the first passes, followed by force constraints once the material hardens. The draft in the finishing passes is limited by the necessity to control strip crown and shape.

Maintaining a constant relative crown, i.e., the thickness difference between strip edge and center with respect to the current strip thickness requires less roll bending and thus linearly reducing roll force with decreasing strip thickness as already pointed out by Shohet and Townsend (1968) and Okamoto et al. (1975). Building on this insight Jonsson and Mäntylä (1985) proposed a backwards calculation scheme for plate schedules that derived the required roll force from the final desired plate crown and the bending characteristics of the mill. In a second step, the height reduction leading to this roll force is determined and the process is repeated. In subsequent research Nakajima et al. (1984) were able to improve on the constant relative crown principle by devising the shape vector method for crown control in hot tandem rolling. In this method a shape vector is determined that balances longitudinal and normal material flow differences between edge and center to prevent shape defects. The shape vector method was adopted for plate mill pass schedules by Mäntylä et al. (1989). Nakajima et al. (1985) instead used experimental data collected during rolling to propose regression models for crown control in plate rolling schedules. Apart for the advances in crown control, the aforementioned scheduling procedures are largely comparable to those of Fujii and Saito (1975).

A strategy to obtain different draft distributions throughout all passes of a schedule was proposed by Szerenyi (1984). For a given number of passes n , a known initial and final cross-section or height h_0 and h_n as well as a factor q the

resulting relative height λ is based on a geometric series and defined as

$$\lambda_i = \lambda_{i-1} \cdot q \text{ with } \lambda_1 = \sqrt[n]{\frac{h_n}{h_0}} / q^{\frac{n}{2} \cdot (n-1)} \quad (1)$$

The factor q controls if the draft distribution is descending ($q > 1$), constant ($q = 1$) or ascending ($q < 1$). In a more recent publication, Moon and Lee (2009) proposed a concept for plate rolling where first the target thickness is intentionally undercut by scheduling maximum allowable draft passes and adhering to common constraint. The draft of all passes is then reduced and redistributed in either of two specific ways to hit the target: The thickness ratio correction (TRC) approach aims to maintain a uniform microstructure over thickness by applying large drafts while the reduction ratio correction (RRC) approach applies an even distribution of drafts to reduce surface defects.

With the advent of computers with greater capabilities research focused on determining the optimal solution for all passes of a schedule at once. Pietrzyk et al. (1990) proposed a system of equations that fully facilitates the available torque of the mill motors via setting the rolling velocity and devised an iteration scheme to satisfy the most relevant constraints in rolling: maximum allowable reduction, maximum roll torque and force. The solution scheme linearizes the governing equations and uses the Gauss method to solve them. If constraints are violated the rolling velocity or load factor are reduced and the calculation is repeated multiple times if need be. While the number of rolling passes is a priori known in tandem rolling, in reverse rolling Pietrzyk's approached requires an additional procedure to determine the passes beforehand, as they need to be known to construct the system of equations. Moreover, Pietrzyk et al. (1990) proposed that in cold rolling, necessary reheating should be taken into account in the roll schedule generation in order to increase formability if required.

Apart from the challenges of crown control discussed already by Nakajima et al. (1984), generating pass schedules for tandem rolling mills imposes additional requirements due to the tight coupling between successive roll stands. According to Buchholz (1976) the range of possible draft distributions in between the stands is identified by a combined forwards and backwards analysis based on maximum and minimum design draft per stand. Combining the forwards and backwards analysis gives a "tube" of thicknesses attributed to each stand that is widest for the intermediate stands. The actual schedule then needs to be chosen within this "tube" typically by determining the schedule that maximizes the throughput of the mill while not violating any constraint.

When scheduling for tandem mills the number of passes is always known a priori as it coincides with the number of roll stands. Wang et al. (2000) used this knowledge about the problems dimensions to devise an heuristic procedure for a tandem cold rolling mills where the search is initialized using an empirical rolling schedule as start point. The objective function considers equal power distribution between stands, interstand tensions and strip shape. The procedure used genetic algorithm, a metaheuristic method to find and optimize solutions in high dimensional problems. Genetic algorithms are inspired by the biological process of natural selection applying concepts of mutation, crossover and selection. Drawing upon the same basic idea, Qi et al. (2012) and Li et al. (2012) developed optimization procedures to obtain pass schedules for seven-stand hot tandem mills. Their works differ with regards to the optimization procedure and the objective function. While Qi et al. (2012) used a conventional procedure to obtain the maximum of an objective function that combines equal relative motor power and roll force for shape control via weighting factors, Li et al. (2012) used a proprietary multi-objective optimization procedure that can account for power distribution, roll force distribution and strip crown and shape simultaneously. Recently, evolutionary algorithms such as the genetic algorithm have been increasingly used for pass schedule design and optimization. Wu et al. (2018) trained ANNs to predict inter alia YS and UTS and coupled the ANNs with genetic algorithm to design hot rolling processes for Q345B steel. Another recent example of pass schedule generation for a six-stand hot tandem mill via genetic algorithms was presented by Hernandez et al. (2019). The authors defined multiple objectives, i.e., reduced overall rolling time as well as roll bending, crown and wear while considering typical limitations such as the maximum allowable force. Besides that, nonlinear programming is also used to optimize the process parameters, as the work by Ozsoy et al. (2013) shows. The authors used nonlinear programming and a hill-climb algorithm to optimize pass schedule in order to reduce the process time and the energy consumption.

A comprehensive literature review based on 90 hot rolling mill scheduling publications from 1989 to 2020 carried out by Özgür et al. (2021) show the relevance of the hot rolling design optimization all over the world. The authors show that different methods are used for the design and optimization. More than half of the publications (62%) studied use heuristics. However, the vast majority of methods presented only consider single-objective optimization (80%). The authors conclude, among other things, that the different approaches are often specialized for specific use cases and use different metrics and datasets, making comparison between them impossible. This makes a transfer from one use case to the next usually difficult or impossible.

Machine learning and its applications in metal forming

ML as a subset of AI focuses on learning patterns and inferences. One of the early key contributions by Rosenblatt (1958) is a probabilistic model, named perceptron, that aims to replicate the functions of biological neurons in human brains. These artificial neurons transform an input vector into a scalar output by calculating a weighted sum of all the input elements and then feeding them through an activation or transfer function. The output depends on the weights and the used function. Thus, multiple layers of these artificial neurons, called Artificial Neural Networks (ANN), can learn non-linear interrelations. In order to learn these non-linear interrelations different techniques were developed. These techniques can be subdivided into three categories, supervised learning, unsupervised and reinforcement learning. Supervised learning needs a training set of labeled data and is used for classification, prediction and regression (Sutton & Barto, 2018) while the goal of unsupervised learning is to find hidden structures in unlabeled data (Sutton & Barto, 2018) by visualization, clustering, outlier detection and dimension reduction (Larkiola et al., 1998). Finally, in reinforcement learning the algorithm learns by mapping situations to actions in order to maximize a numerical reward (Sutton & Barto, 2018). In the following, some exemplary applications of each machine learning category in the context of metal forming are given.

There are several examples of supervised learning in the realm of metal forming. One of the first applications was presented by Korczak et al. (1998) in the 1990 s. The authors aimed to improve the prediction of the non-linear relationship between chemical composition, microstructure and final mechanical properties for low carbon steels after hot rolling based on an ANN. Their ANN is able to predict Vickers hardness, YS and UTS. A validation using measured properties showed a good agreement. Besides that, ANNs were often used to predict rolling forces either stand-alone as discussed by Lee and Choi (2004) or in combination with analytical models (hybrid model) as demonstrated by Moussaoui et al. (2006). Later, Zhang et al. (2016) improved this hybrid model concept to meet stricter quality requirements, e.g., with regards to final thickness, which requires an even more accurate prediction of the roll force. They combined a conventional rolling force model, based on the equation by Sims and Wright (1963) with an ANN, which was trained based on online process data. Notably, this online learning increased the robustness and simultaneously the roll force prediction. Nevertheless, there is continued work in this area as Shen et al. (2022) showed. The authors trained several ANNs with industrial data to predict the roll force for a complete 7-stand hot strip mill based on the chemical composition and process parameters. In addition to these elaborate ANNs, there is a

growing amount of work that combines ANNs with physical models. For example, Hwang et al. (2020) used ANNs, decision trees and a classical physical model (Sims, 1954) to predict robustly the roll force and temperature taking advantages of classical rolling models and data-driven approaches. In addition to that, ANNs are used successfully to detect anomaly in industrial hot rolling data as Jakubowski et al., (2021).

As shown, supervised learning was commonly used in hot rolling to predict certain properties, sometimes also in combination with unsupervised learning. For instance, Lieber et al. (2013) used unsupervised learning for clustering and dimensionality reduction of different surface features and supervised learning for classification of surface defects. This served to avoid the further processing of material that would be scrapped later. Still, the applications of unsupervised learning in hot rolling process is relatively rare. In general, unsupervised learning methods are mainly used for image processing. Recent examples are microstructure cluster analysis (Kitahara & Holm, 2018), surface defect classification (Di et al., 2019) and defect segmentation of hot-rolled steel strip surface (Youkachen et al., 2019).

In recent years, reinforcement learning achieved prominence by defeating humans in playing games like Go (Silver et al., 2017). However, one of the first approaches to using reinforcement learning for optimization in manufacturing dates back to 1998 by Mahadevan and Theocharous (1998). A detailed overview of the current state of the art of reinforcement learning algorithms was given by Sutton and Barto (2018). In their review Wuest et al. (2016) specifically touched upon current applications of reinforcement learning in the context of manufacturing. According to the authors, in 2016 reinforcement learning was not widely applied in manufacturing. However, a few examples are described in the following: Günter et al. (2014) published a self-learning approach for laser welding to increase the performance by enabling a rapid setup process. The presented approach includes deep neural networks, which extract features and a trained reinforcement learning algorithm that uses these features as an input to control the process in real-time. In the realm of metal forming Dornheim et al. (Dornheim & Link, 2018; Dornheim et al., 2019) coupled a deep reinforcement learning algorithm with an FE simulation model to enable multi-objective optimization of a deep drawing process. The presented approach controls the blank holder force over the process in such a way that the product was manufactured as material-efficiently as possible and with as little residual stresses as possible while ensuring the desired product geometry. Following on from this, Guo and Yu (2019) trialed different reinforcement learning algorithms to improve the blank holder force control. The authors concluded that more advanced reinforcement learning algorithm, e.g., Twin Delayed Deep Deterministic Policy Gradient (TD3) lead to

further improvements. Moreover, there are first activities in forging as well. Zhang et al. (2021) used RL to identify optimal parameter of a mechanism model in real-time without historical data. This allows an accurate online simulation of the forging process. Reinisch et al. (2021) coupled a forging model with a RL algorithm to design pass schedules for open-die forging. The authors coupled a fast process model was coupled with Double Deep Q-Learning to design and optimize the process in terms of final ingot geometry, press force and process duration. The designed processes led to executable processes and show the enormous potential of the methodology.

A first example of reinforcement learning for pass schedule generation was presented by Scheiderer et al. (2020) based on collaborative work involving some authors of the present paper. Therein a simulation as a service architecture for the application of reinforcement learning to protect intellectual properties of different stakeholders, e.g., process experts and AI experts was proposed. To demonstrate the capabilities, pass schedules for reverse hot rolling were generated via a RL-algorithm that was coupled to data pre-calculated using an FRM and provided via a web interface. The results show that such a coupled approach is generally capable of generating pass schedules. However, a direct coupling of RL and FRM could eliminate the need for interpolation of data and provide results that are even more reliable. Moreover, Gamal et al. (2021) demonstrated that RL in combination with process data identifies optimal model parameters and thus improve predictions for bar and wire hot rolling processes.

Literature assessment

When reviewing the relevant literature, it becomes apparent that a plentitude of different approaches to generate and optimize pass schedules have been trialed. Most of them rely on some experience and process knowledge to lay out the underlying heuristics. Approaches that are based on optimization instead, mostly do not unlock the full potential of optimization as they only account for one or two objectives. In addition, genetic algorithms have been applied with some success. Due to the used meta-heuristics these sit somewhere between heuristics and optimization. However, all approaches share the same shortcoming: Results are generally not transferable to a different problem, e.g., a different final geometry. Therefore, the procedures must be run for each rolling schedule individually, which is particularly time-consuming for optimization and genetic algorithms. When extending the training data to cover many different conditions the combination of RL and FRM in contrast might be able to generate several different pass schedules based on the same training.

There are different reinforcement learning algorithms available. However, most of them can only handle discrete spaces, which is not expedient for designing pass schedules. An algorithm compatible with continuous space is the Deep Deterministic Policy Gradient (DDPG), developed by Silver et al. (2014). Besides, the DDPG algorithm is particularly interesting for application in pass schedule design due its successful applications in solving 20 different continuous physical problems (Lillicrap et al., 2016) among others a deep drawing process (Guo & Yu, 2019). The DDPG algorithm uses a specialized concept called the actor-critic principle. This reinforcement learning architecture using two ANNs for optimization was developed by Konda and Tsitsiklis (2001). One ANN, called the actor, represents the agent's policy, while the other ANN, the critic, models the Q function. The Q function describes how rewarding it is to perform a certain action in a defined state. In the training process, the actor on the one hand is trained using the critic's evaluation of the current policy, allowing for updating the actor towards choosing better action. The critic on the other hand is trained using the observed experience in order to model the Q-function accurately.

A pass schedule generation procedure that pursues multi-optimization objectives and takes into account the final product properties has not yet been presented. Therefore, the goal of this paper is to trial a pass schedule generation via a direct coupling of the FRM and a DDPG algorithm leading to an efficient pass schedule and the desired final UTS.

Methodology: coupling fast models and machine learning for pass schedule design

The combination of ML and a FRM can enable the design of optimized pass schedule, regarding multiple objectives and an explicit inclusion of product properties. To achieve this, three steps regarding the FRM and the reinforcement learning algorithm are necessary. As mentioned before, an existing rolling model (Lohmar et al., 2014b) is used for the coupling with a reinforcement learning algorithm. The model is able to calculate inter alia the roll force and the austenite grain size, but it is not capable of predicting final properties after the hot rolling process. Therefore, the first step is to extend the FRM to predict product properties. Here YS and UTS are used as industrially relevant representatives to demonstrate the idea. In the second step, the ML algorithm needs to be set up. Since reinforcement learning algorithms can interact with the FRM without any limits in combining process parameters, they look the most promising and are therefore chosen. Consequently, the second step also includes the definition of the reward function. The third step is to couple the FRM and the reinforcement learning algorithm.

All three steps are detailed in the following in individual subsections.

Material, microstructure and properties modelling

The material considered in this paper is a S355 structural steel. A structural steel is taken because it is used in various areas and has a microstructural behavior that can be well described with existing material models. This makes the structural steel S355 ideal for testing the coupling of RL and FRM for the pass schedule design. The chemical composition is given in Table 1 and was determined using an optical emission spectrometer.

To model hot rolling of this particular steel using a FRM flow stress, static recrystallization kinetics and grain size evolution during and after deformation need to be known. Furthermore, to predict mechanical properties additional equations are needed to infer yield strength (YS) and ultimate tensile strength (UTS) from chemical composition and microstructure.

The semi-empirical equations to describe material behavior are briefly introduced in the following. The flow stress calculation is based on an equation put forward by Hensel

and Spittel (1978),

$$\sigma_f = 3750 * \dot{\varphi}_{eq}^{(-0.11+0.00024*T)} * \exp(-0.003*T) * \varphi_{eq}^{0.28} * \exp(-0.41 * \varphi_{eq}) \tag{2}$$

where φ_{eq} is the equivalent strain, $\dot{\varphi}_{eq}$ is the equivalent strain rate and T is the temperature.

The material dependent constants were determined from isothermal compression tests carried out on a servo-hydraulic testing machine. Figure 1 shows both the experimental flow curves (red) determined from 900 to 1200 °C and for strain rates between 0.1 and 10 s⁻¹ and the flow curve fit (black) based on the previous introduced equation. Apart from the lacking ability to reproduce the dynamic recrystallization (DRX) visible in some curves the flow curves are captured well. DRX does not typically occur during conventional hot rolling of steels and thus is intentionally neglected.

The static recrystallization kinetics is modelled based on the well-known JMAK equation. The Eqs. 3, 4 and the material parameter are adopted from Hodgson and Gibbs (1992).

$$X_{SRX} = 1 - e^{\log(0.5) * (t/t_{50})^1} \text{ with} \tag{3}$$

Table 1 Chemical composition of the used S355 in weight %

	C	Si	Mn	P	S	N	Cu	Al
Weight %	0.1	0.3	1.6	0.012	0.001	–	0.2	0.02

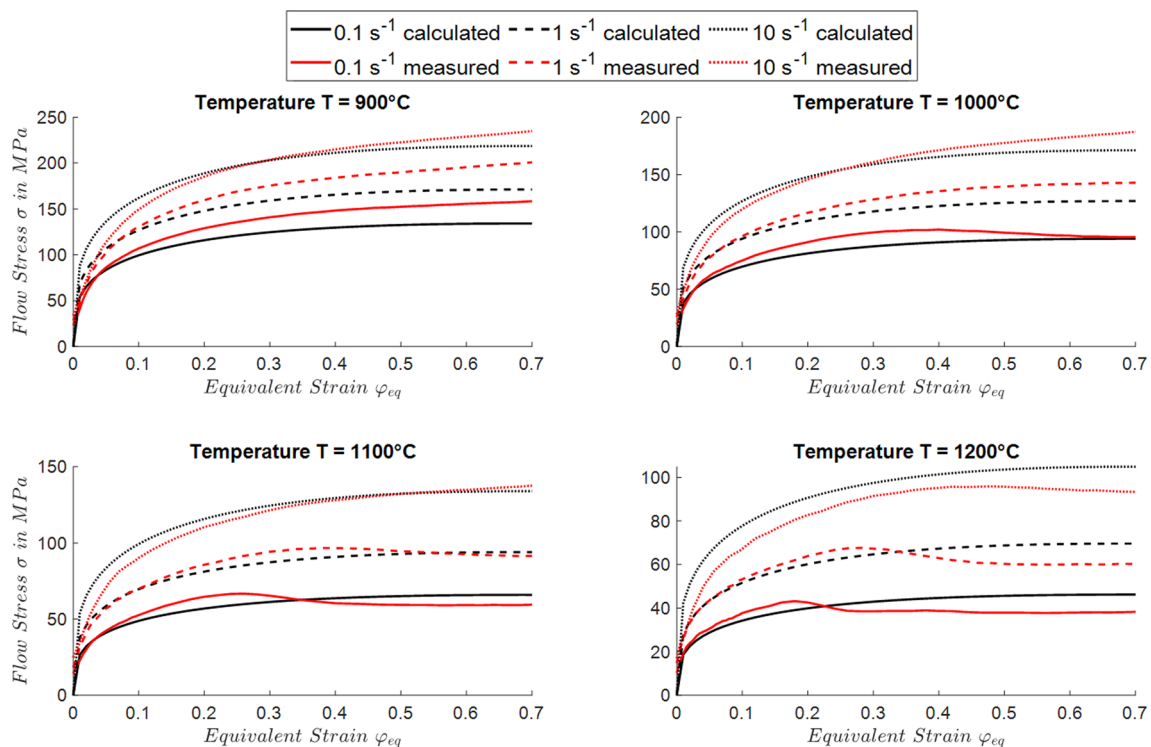


Fig. 1 Flow Stress for S355 determined by compression tests at Institute of Metal Forming (in red the measured and in black the calculated one)

$$t_{50} = 6.31 * 10^{-18} * d_{\gamma,0}^{-2.38} * \exp(2) * \exp\left(\frac{330000}{8.31 * T}\right) \quad (4)$$

where $d_{\gamma,0}$ is the austenite grain size and t is the inter-pass time.

As mentioned before, the used FRM considers static recrystallization and grain growth by simple semi-empirical equations introduced by Beynon and Sellars (1992).

The grain size after full static recrystallization $d_{sr,x}$ is calculated using an Eq. 5 with parameter that were also put forward by Hodgson and Gibbs (1992)

$$d_{sr,x} = 1 * d_{\gamma,0}^{0.4} * \varphi_{acc.}^{-0.5} * \exp\left(\frac{-45000}{8.31 * T_m}\right) \quad (5)$$

where $\varphi_{acc.}$ is the accumulated equivalent strain, and T_m is the average temperature.

In order to describe the final average austenite grain size $d_{\gamma,1}$, a case distinction is necessary. Depending on whether the material is fully recrystallized or not, other formulas must be used to describe $d_{\gamma,1}$. If the material is not fully recrystallized, a law of mixtures based on work by Beynon and Sellars (1992) is used to determine the final austenite grain $d_{\gamma,1}$ size based on the statically recrystallized fraction (X_{SRX}), the grain size after full static recrystallization ($d_{\gamma,srx}$) and the initial grain size ($d_{\gamma,0}$)

$$d_{\gamma,1} = X_{SRX} * d_{\gamma,srx} + (1 - X_{SRX})^2 * d_{\gamma,0} \quad (6)$$

If the material is fully recrystallized, grain growth can occur. Grain growth is modelled based on Eq. 7 and its material parameter were originally suggested by Hodgson and Gibbs (1992)

$$d_{\gamma,1} = \left(d_{\gamma,srx}^{4.5} + 4.1 * 10^{23} * t_{gg} * \exp\left(\frac{-435000}{8.31 * T}\right)\right)^{1/4.5} \quad (7)$$

where t_{gg} is the time available for grain growth. Typically, this time is calculated by subtracting the time for full recrystallization from the inter-pass time.

To enable the prediction of the final mechanical properties, i.e., YS and UTS, two established physical models put forward by Hodgson and Gibbs (1992) and Singh et al. (2013) are implemented. Precipitation strengthening is excluded due to its small contribution.

$$\sigma_{YS} = \sigma_{SSYS} + \sigma_{DISYS} + K_{YS} * d_{\alpha}^{-0.5} \quad (8)$$

$$\sigma_{UTS} = \sigma_{SSUTS} + K_{UTS} * d_{\alpha}^{-0.5} \quad (9)$$

Here, σ_{YS} is the YS, σ_{UTS} is the UTS, σ_{SSYS} , σ_{SSUTS} , σ_{DISYS} and $K * d_{\alpha}^{-0.5}$ are the contributions from solid

solution strengthening, dislocation strengthening and the grain-boundary (Hall–Petch) relationship, respectively. All parameters despite the chemical composition for a structural steel S355 are taken from literature and detailed below. For the calculation of solid solution strengthening σ_{SSYS} and σ_{SSUTS} despite nitrogen content (N) the actual measured chemical composition, see Table 1, was used. Because the nitrogen content (N) was not measured, but has a very important influence on σ_{SSUTS} , see Eq. 11.

$$\sigma_{SSYS} = 15.4 * (3.5 + 2.1 * Mn + 5.4 * Si) = 130.592 MPa \quad (10)$$

$$\sigma_{SSUTS} = 165 + 54 * Mn + 100 * Si + 652 * P + 635 * C + 3339 * N = 386.11 MPa \quad (11)$$

$$\sigma_{DISYS} = 1.44 * 10^{-3} * \sqrt{\rho} \quad (12)$$

$$\rho = \rho_n(1 - X) + \rho_s X \quad (13)$$

$$\rho_n = \frac{C(1 - e^{-B\varepsilon})}{B} + \rho_0 e^{-B\varepsilon} \quad (14)$$

$$\rho_s = 3.82 * 10^9 * \varepsilon^{0.2} * [(1613 - T)/290]^2 \quad (15)$$

$$C = 8.5 * 10^{10} (1 + D_0^{0.5}) \quad (16)$$

$$B = 6227 \varepsilon^{-0.28} * e^{-7500/T} \quad (17)$$

$$d_{\alpha} = d_{\alpha,0} * (1 - 0.45 * \sqrt{\varepsilon_r}) \quad (18)$$

$$d_{\alpha,0} = d_{\gamma,cool} \quad (19)$$

For the calculation of the grain boundary dependency K_{YS} and K_{UTS} are set to 19.7 MPa μm and 11 MPa μm respectively, as proposed by Hodgson and Gibbs. (1992). The solid solution strengthening σ_{SSYS} , σ_{SSUTS} (see Eqs. 10 and 11) is material-dependent and calculated using the chemical composition, shown in Table 1. The dislocation strengthening σ_{DISYS} (see Eq. 12) is calculated as suggested by Singh et al. (2013). In order to calculate the dislocation density ρ (see Eq. 13), first the dislocation densities of the dynamically recovered ρ_n (see Eq. 14) and dynamically recrystallized ρ_s (see Eq. 15) regions are required. For ρ_n the constants C (Eq. 16) and B (Eq. 17) have to be calculated. In order to calculate the strengthening through grain-boundaries, the ferrite grain size d_{α} is calculated based on the initial ferrite grain size $d_{\alpha,0}$ and the retained strain ε_r at the beginning of the transformation using Eq. 18 presented by Beynon and Sellars (1992). The authors presented an equation to calculate the

Table 2 Thermal boundary conditions for hot rolling of S355

$\rho [kg/m^3]$	$\alpha_{conv} [W/(m^2 * K)]$	$\alpha_{roll} [W/(m^2 * K)]$	$\epsilon_{rad} [-]$	$\epsilon_{diss} [-]$
7600	15	35000	0.8	0.9

ρ represents the average density for the temperature range (800–1200 °C); α_{conv} is the heat transfer coefficient to the air during the inter-pass time; α_{roll} is the heat transfer coefficient to the rolls; ϵ_{rad} is radiation coefficient; ϵ_{diss} is the dissipation factor

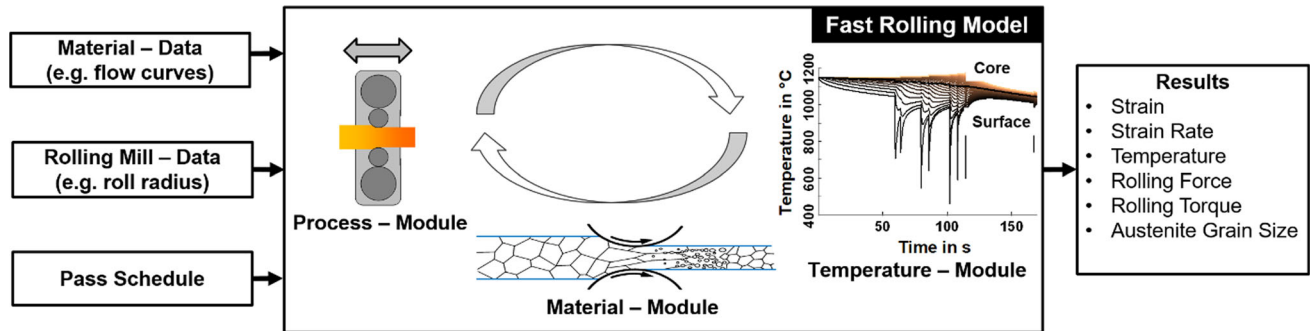


Fig. 2 Schematic structure of the FRM presented in (Lohmar et al., 2014a)

initial ferrite grain size based on the austenite grain size at the transformation ($d_{\gamma,cool}$), carbon equivalent and cooling rate after the last pass and the transformation. Here, it is assumed that the initial average ferrite grain size $d_{\alpha,0}$ is the same as the final austenite grain size ($d_{\gamma,cool}$), before the austenite transforms to ferrite due to the small deviation between the assumption and the more complex equation. The equations presented here allow the calculation of product properties after hot rolling. However, you need an FRM to supply the quantities required here, such as $d_{\gamma,cool}$ and ϵ_r . Next, the used FRM is described in detail.

Integrating properties prediction into the FRM

In this paper, a FRM developed at the Institute of Metal Forming is used (Lohmar et al., 2014a) for convenience. The model consists of several modules, predicting the deformation in the roll gap, the temperature evolution, the material behavior, the microstructure evolution and the resulting rolling force and torque. It is based on the slab method put forward by Siebel (1925) and Karman (1925). Hence simple equations are used to calculate the equivalent strain ϵ_{eq} , see Eq. 20, and the strain rate $\dot{\epsilon}_{eq}$, see Eq. 21.

$$\epsilon_{eq} = \frac{2}{\sqrt{3}} * \ln\left(\frac{h_0}{h_1}\right) \tag{20}$$

$$\dot{\epsilon}_{eq} = \frac{v_{roll}}{l_d} * \epsilon \tag{21}$$

Here, h_0 and h_1 are the entry and the exit height of the rolling stock, respectively and v_{roll} represents the rolling velocity. The evolution of the temperature T during the time t

is calculated by a one-dimensional finite-differences method considering heat conduction inside the rolling stock, radiation and convection on the surface, heat transfer to the rolls and dissipation \dot{Q} due to the applied deformation, see Eq. 22. To capture the thermal behavior of the S355 steel grade, the material density ρ , the specific heat capacity c_p and the thermal conductivity λ discussed in the materials subsection are required.

$$\rho * c_p * \frac{\partial T}{\partial t} = \lambda * \left(\frac{\partial^2 T}{\partial^2 y^2}\right) + \dot{Q} \tag{22}$$

The used thermal boundary conditions are listed in Table 2 and are in accordance with values typically found in technical literature or were determined experimentally earlier.

The roll force F is calculated by a simple equation developed by Sims and Wright (1963), see Eq. 23. F depends on the contact length l_d , the width of the rolling stock b , the mean flow stress σ_m and the geometric factor Q_p that compensates for inaccuracies due to friction and shear as discussed by Seuren et al. (2010) and Lohmar et al. (2014b). For the calculation of flow stress, recrystallization and grain size evolution, the simple semi-empirical equations introduced earlier are used. Figure 2 schematically shows the structure of the FRM together with the relevant input and output.

$$F = l_d * b * \sigma_m * Q_p \tag{23}$$

An extension of the FRM is required to also predict the mechanical properties of the sheets after hot rolling and cooling. In this paper, the YS and UTS of the structural steel S355 is predicted. For this steel grade, mostly conventional air-cooling to room temperature is applied after hot rolling. Thus,

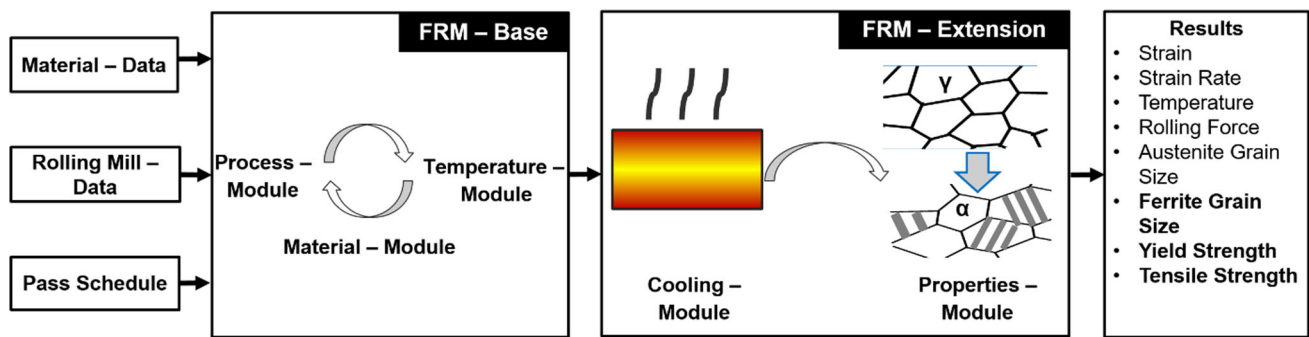


Fig. 3 Schematic structure of the extended FRM developed in this paper

during cooling the microstructure transforms from austenitic to ferritic-pearlitic. Therefore, a cooling and a transformation module are added to the existing FRM. The temperature evolution of the workpiece up until the onset of the austenite to ferrite transformation, e.g., 830 °C, is calculated in the cooling module based on the temperature profile after the last pass. Using this information, the retained strain, recrystallized volume fraction and austenite grain size prior to transformation are calculated in the cooling module. Finally, the transformation module is triggered, calculating the YS and UTS based on the ferrite grain size based on Eqs. 8, 9 and 18 discussed in detail in the materials subsection. The extended FRM with additional outputs is shown in Fig. 3.

Setting up the DDPG algorithm for automated pass schedule design

There are numerous RL algorithms that can be used for process optimization. However, Deep Q-Learning (DQN) (Dornheim et al., 2019; Reinisch et al., 2021) or DDPG (Gamal et al., 2021; Lillicrap et al., 2016) are most commonly used to optimize processes. Both have proven themselves and are already implemented in many frameworks like TensorFlow or Matlab. Therefore, an established algorithm is used here in order to provide a proof-of-concept for hot rolling. The biggest difference between them is that DQN has only one ANN whereas DDPG has two ANNs, which enables it to propose continuous actions or process parameters. The ANNs used are initialized using common initialization methods which are explained later in chapter 4. In the future, physics-based ANNs or other pre-trained ANNs could be used for this purpose.

As mentioned before, the DDPG algorithm is the reinforcement learning approach of choice due to its ability to suggest continuous instead of discrete values. Setting up the DDPG algorithm (Silver et al., 2014) includes the definition of the state, the possible actions and the reward function as well as assignment of the hyperparameters. The set up will

be detailed below following an introduction of the general DDPG algorithm structure.

As mentioned above the DDPG uses a Critic and an Actor Network as well as states, actions and rewards to learn a given task. As Fig. 4 shows, in pass schedule generation the Critic Network uses the state, action and reward information of the previous rolling pass and tries to estimate the expected reward for each possible new state (i.e., the next rolling pass) based on the current state and possible actions. It then provides the state-action pairs and its estimated rewards to the Actor Network, which uses this information to choose the action with maximum reward. In both cases, the backpropagation method is used to train the Critic and Actor Network. The backpropagation method developed by Rumelhart et al. (1986) in 1986, computes the gradient of the loss function, difference between actual ANN output and desirable output, with respect to each weight of the ANN. It computes the gradient one layer at a time, iterating backward from the last layer. This method allows to train ANNs efficiently.

The state, is defined by the rolling stock height, accumulated strain, temperature at the surface and core, the recrystallized volume fraction and average austenite grain size prior to a given rolling pass. A unique state definition is very important for the DDPG algorithm. A non-converging training process is likely, if for example the same current height and grain size leads to several possible grain sizes in the next pass. To circumvent this temperature and accumulated strain are also considered within the state.

The action defines the next pass in terms of height reduction and inter-pass time within the allowed ranges. The range of the height reduction (Δh) and the inter-pass time is dependent on the current state. For example, the maximum Δh is dynamically set to 40% of the rolling stocks height unless this violates the bite condition, in which case the bite condition defines the maximum Δh . This leads to drafts between 15.68 and 5 mm whereas the inter-pass time ranges from 5 s (minimum reversing time) to 15 s. The reward function used to assess the state is problem specific and steers the algorithm by giving numerical feedback on the goodness of the new state.

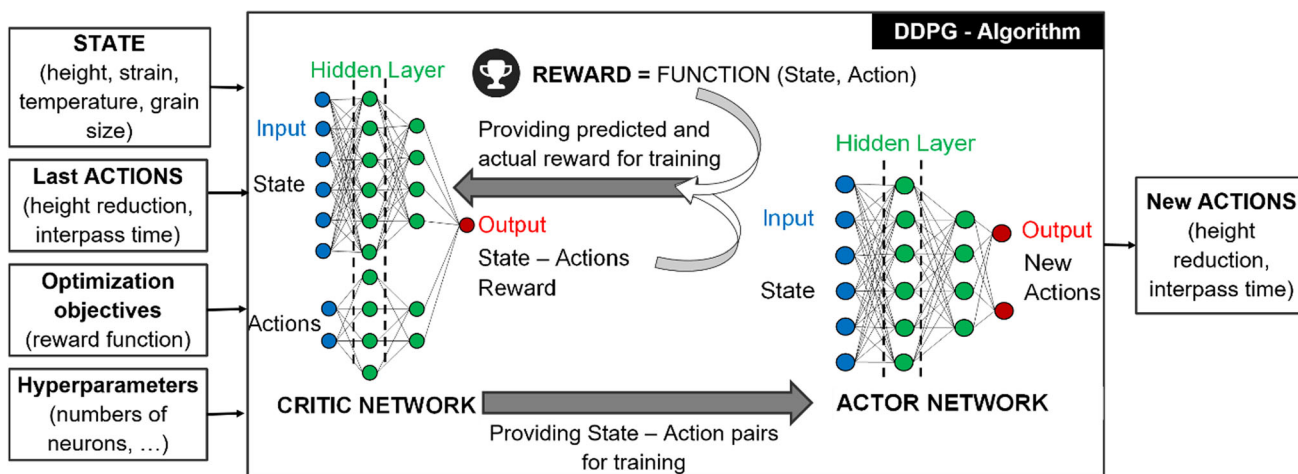


Fig. 4 Schematic structure of the Deep Deterministic Policy Gradient (DDPG) algorithm (Silver et al., 2014)

The reward function used to design an optimal pass schedule must consider several desired optimization objectives.

In this paper, the total reward (r_t) given to the DDPG is the sum of each collected reward r_i during one pass schedule. The reward function is evaluated after each step of the training process that corresponds to a single rolling pass. It can also be considered as an evaluation benchmark for later analysis. The reward for one pass consists of five optimization objectives, see Eq. 24. Four of them are considered for each pass, namely height (r_H), average austenite grain size (r_D), rolling force (r_F) and the thermal and mechanical energy consumption (r_E). The fifth optimization objective the final UTS, (r_M) is only considered in the final pass because the final mechanical properties r_M are offset by the schedule as a whole.

$$r_t = \sum_i^n r_i = r_{H_n} + r_{D_n} + r_{F_n} + r_{E_n} + r_{M_n} + \sum_i^{n-1} r_{H_i} + r_{D_i} + r_{F_i} + r_{E_i} \quad (24)$$

Designing a pass schedule is a twofold problem. The difficulty lies within optimizing each pass, e.g., maximizing the draft, while also optimizing the pass schedule as a whole in terms of reaching all targets as quickly as possible. Those two goals do not always go hand in hand, which is the reason most classical optimization techniques fail. To tackle this challenge, the total reward is divided into several intermediate rewards and an additional completion reward that can substitute the last intermediate reward. The intermediate rewards, cover the optimization of each pass individually, while the completion reward assesses the entire pass schedule.

The intermediate rewards (see Table 3) apply as long as the height after the current pass (h_{prev}) has not reached the target height (h_{target}) and thus the pass schedule cannot be

considered complete. In this case the mechanical properties (r_M) are not assessed. In order to promote reaching the specific targets as fast as possible, the rewards for the height (r_H) and grain size (r_D) are the bigger the closer they are to their target (h_{target} , $d_{y,target}$) and the bigger the change (Δh , Δd_y) compared to the previous pass (h_{prev} , d_{prev}) is. In contrast, high energy consumption (E) and roll force (F) exceeding machine limits are penalized. Once the height is within the target tolerance or below it, the pass schedule is considered complete and thus the intermediate reward in the last pass is replaced by the completion reward.

The completion reward function is split as follows (also see Table 3). If the final height (h_{final}) is achieved within the target tolerance ($h_{target,max}$, $h_{target,min}$), the mechanical properties (r_M), in terms of the UTS, are accounted for in addition to the previously mentioned rewards. The completion reward for height and grain size decreases symmetrically the further away from the target their value is. The completion reward for the UTS rewards higher values. In order to obtain an optimal schedule considering multiple objectives, weights have to be assigned to the individual contributions. Achieving the final height has the highest priority and thus the highest weight of 10. The next priorities are reaching the final austenite grain size ($d_{y,roll}$) within the tolerance (weight of 6) and then maximizing the UTS (weight of 2). If the final height (h_{final}) is below the tolerance, the rewards for grain size and mechanical properties are negated and the reward for the height is the more negative the further away from the target it is.

In a nutshell, the three goals (height, grain size, mechanical properties) should be encouraged and thus provide a positive reward if they meet the target. In contrast, more consumed energy and the force exceeding the machine limits are punished and therefore give a negative reward. This negative reward is important so that the DDPG algorithm does not add

Table 3 Definition of the reward function consisting of the intermediate and the completion components

Reward objective	Intermediate reward $i = 1 \dots n - 1$	Completion reward $i = n$	
	Target height not reached ($h_{prev} > h_{target}$)	Target height within tolerance ($h_{target,max} \geq h_{final} \geq h_{target,min}$)	Target height below the tolerance ($h_{final} < h_{target,min}$)
Height	$r_H = \frac{\Delta h}{h_{prev} - h_{target}}$	$r_H = 10 - (h_{final} - h_{target})^2$	$r_H = -10 - (h_{final} - h_{target})^2$
Grain size	$r_D = \frac{\Delta d_y}{d_{y,prev} - d_{y,target}}$	$r_D = 6 - \frac{4*(d_{y,roll} - d_{y,target})^2}{d_{y,target}^2}$	$r_D = 0$
Mechanical properties	$r_M = 0$	$r_M = 2 + \frac{(\sigma_{UTS} - \sigma_{SSUTS})^2}{\sigma_{SSUTS}^2}$	$r_M = 0$
Energy consumption	$r_E = -0.2 * E$		
Rolling force	$r_F = \begin{cases} -10 - \frac{(maxForce - F)^2}{maxForce^2}, & \text{if } F > max\ Force \\ 0, & \text{else} \end{cases}$		

Table 4 Hyperparameters of the two neural networks used in the DDPG algorithm

Network	Number of hidden layers	Neurons per hidden layer	Learning rate
Critic network	2	256	0.0005
Actor network	1	24	0.0001

additional passes just to increase the reward. Therefore, the intermediate reward is designed in such a way that it generates negative rewards when the applied height reductions are small.

The DDPG uses ANNs that consist of multiple layer of neurons, an input and output layer and hidden (intermediate) layers. The number of hidden layers and neurons per layer increases the complexity that the network can learn, but requires more data for training. The learning rate represents the amount that the weights of the ANN are updated in order to minimize the error of the training. Typically, due to the high complexity of the reward function, the number of hidden layers and neurons per layer in the critic network is bigger than in the actor. Figure 4 shows a simplified structure of the DDPG algorithm. The key hyperparameters of both neural networks are given in Table 4. The determination of the hyperparameters was based on typical published parameters like from Gamal et al. (2021). First, the ANNs were chosen and tested with relatively small number of neurons per hidden layer. After that, the other hyperparameters were determined via trail-and-error. In future works, it is planned to use more advanced methods such as Bayesian optimization to tune the hyperparameters.

Coupling of the FRM and the DDPG algorithm and ANN training

As a third step illustrated in Fig. 5, the FRM and DDPG algorithm are coupled to train the algorithm to design pass schedules. This training is necessary to adjust the weights inside the ANNs in order to obtain proper results. After a successful training, the DDPG algorithm is able to solve the trained task within seconds. Additionally, the trained ANNs can be used to solve similar tasks by applying transfer learning as described by Taylor (2009).

The training can be divided into iterations and steps, in each iteration the algorithm starts at an initial state and takes steps until a condition or goal is reached. In our case an iteration represents an entire pass schedule and is completed when the height is within or below the specified tolerance. These iterations are repeated until either the average total reward limit or the number of allowed iterations are exceeded.

Each iteration starts by setting the initial state, which in our case represents the roll stock properties when leaving the furnace in terms of height, strain, temperature and austenite grain size as well as by initializing the iteration complete flag associated with reaching or undercutting the target height (h_{target}). The DDPG algorithm then designs an initial pass schedule in steps on a pass-by-pass basis starting from the initial state. In each step an action (a combination of draft and inter-pass time) that adds a new pass is chosen by the Actor Network based on the network's weights and biases, the current state as well as artificial exploration noise. This noise helps to initialize the training process by introducing additional variance into the data sets and reduces during training to allow for exploitation of experience gained. The concept of exploration noise is detailed in the Appendix. Never the less, as long as no experience was collected by the DDPG

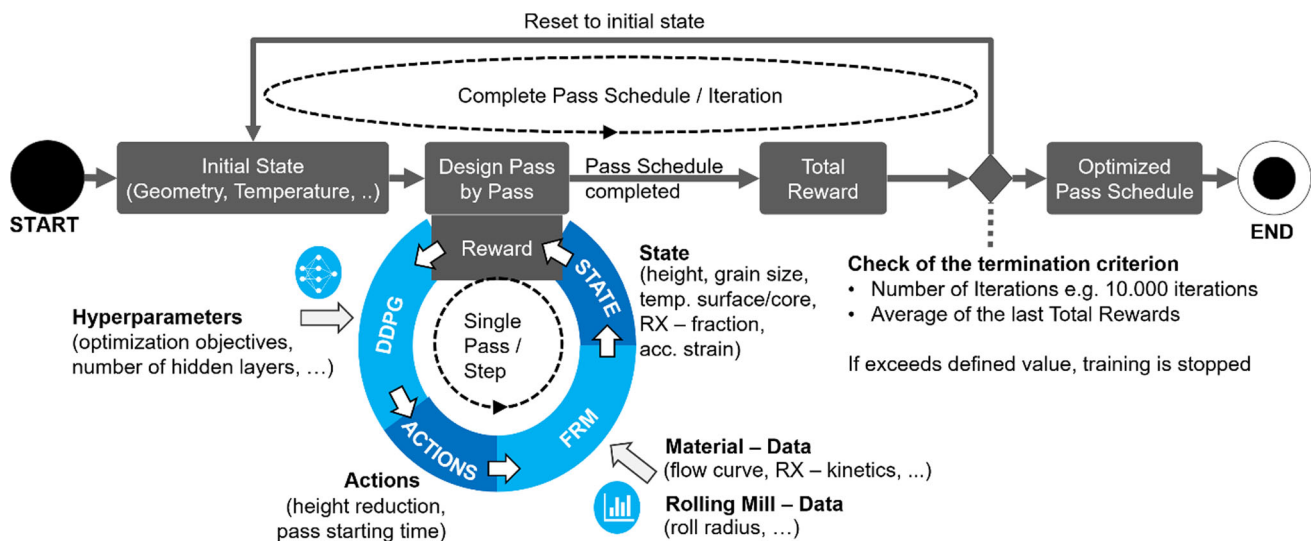


Fig. 5 Flowchart of the coupling between the FRM and the DDPG algorithm. The intermediate reward is given after each pass if the height > target height, otherwise the completion reward is given

algorithm in a so-called replay buffer the initial actions chosen by the Actor Network strongly depend on the weight and bias initialization. After a pass was added the incomplete pass schedule is passed to the FRM, which calculates the roll stock properties after the pass, thus defining a new state. To aid convergence of the DDPG algorithm, intermediate rewards based on the height, and grain size after each pass are provided already for the incomplete pass schedule using the reward function (see Table 3). Next, if the final height is not reached, an additional step is taken and a pass is added until the final height is reached or undercut and the pass schedule is thus complete.

As soon as the final height is reached, the previously mentioned complete flag is set to true and the completion reward (see Eq. (24)) is calculated using the state after the final pass and saved along with the pass schedule instead of the intermediate reward calculated for any other pass. The total reward (r_t) for the iteration is then given as the sum of the intermediate rewards and the completion reward. As a final step, state-action-pairs together with the rewards from the completed iteration and thus the complete pass schedule are stored to the replay buffer. The weights and biases of the two used ANNs, the Actor and Critic Network are then updated based on the data in the replay buffer. Further details on the replay buffer are given in the Appendix. The update of the Actor and Critic Network concludes each iteration.

Prior to starting the next iteration, a termination check is conducted. In this paper, 10.000 iterations were chosen as termination criterion because through many training runs no increases in the total reward (r_t) beyond 10.000 iterations was detected. If the termination criterion is exceeded, the training stops and the weights and biases of the ANNs are

frozen. Otherwise, the state is reset to the initial state and the described iteration starts over.

Results: design of an optimized pass schedule

The goal of this section is to provide a proof of concept for a pass schedule designed via the coupling of a FRM and a DDPG algorithm. For demonstration, the initialization of the DDPG algorithms, the initial and final pass schedule and the corresponding reward evolution are shown.

The overall benchmark for designing an optimized pass schedule is maximizing the reward function for a given initial state. For faster convergence, the variable target values in the reward function and their specified tolerances were chosen wider than typically used in industry: The target height (h_{target}) is $25 \text{ mm} \pm 1 \text{ mm}$ and the target austenite grain size $d_{\gamma, target}$ is $35 \mu\text{m} \pm 5 \mu\text{m}$ after rolling prior to cooling. Additionally, the reward function maximizes the UTS, minimizes the energy consumption and considers the force limitations of the laboratory rolling mill at the IBF to enable rolling the optimized scheduled on this mill later. Therefore, the initial geometry differs from typical industrial application. It is defined by a height of 140 mm, a width of 180 mm and a length of 500 mm. Furthermore, the initial distribution of the austenite grains is assumed to be homogenous and its size is set to $200 \mu\text{m}$. As mentioned before, neither the initial temperature nor rolling velocity are currently considered to reduce complexity and improve convergence. Instead, a homogeneous initial temperature of $1200 \text{ }^\circ\text{C}$ is used and the rolling velocity for all passes was 250 mm/s.

After setting up reward function and initial state (initial height and grain size) an initialization of the Actor and Critic ANNs is required, as their replay buffer is empty at the start of training. Thus, no data regarding the relationship between state, action and reward is available at first. As detailed earlier the replay buffer is filled based on actions (draft and inter-pass time) according to the initialization method leading to an initial complete pass schedule. This initial pass schedule can have a significant impact on the training process with regards to both quality and performance, because it defines the parameter space, in which the DDPG algorithm starts its search based on the data collected in the replay buffer. Most importantly for rolling, the initialization implicitly defines the number of passes in the initial schedule. For example, if the height reduction were minimized during initiation for the conditions defined above, the resulting number of passes would be 23 compared to eight if the reduction were maximized.

To obtain a suitable pass schedule initialization with random, minimum and maximum actions (draft and inter-pass time) was trialed. As mentioned earlier the possible height reduction (Δh) is between 15.68 and 5.0 mm, the possible inter-pass time ranges from 5 to 15 s. The results and required computation times for the training (on an Intel Xenon CPU E3-1270) based on 10,000 iterations are summarized in Table 5. The best final reward was obtained for the training based on the max. initialization method and this training furthermore required the least time to complete. Thus in the following only the results obtained using the max. initialization will be discussed.

Figure 6 summarizes the height and strain (left) as well as the temperature and grain size evolution (right) of the initial (Initial iteration, top) and the final (10,000 iterations, bottom) pass schedules. The figures on the left, a) and c), display accumulated strain in blue (left axis) and the height of the rolling stock in black (right axis) over the process time. Moreover, the final YS and UTS as well as the total energy consumption (E) are shown. The figures on the right, b) and d), visualize the temperature distribution over the roll stock height in red (left axis) and the average austenite grain size in black (right axis). Furthermore, the average temperature ($\bar{\theta}$) and average austenite grain size right after rolling ($d_{\gamma,roll}$) as well as the average austenite grain size after cooling ($d_{\gamma,cool}$) and the final ferrite grain size (d_{α}) are depicted.

A comparison of the initial and final schedules reveals both similarities and differences. Starting with the similarities, both pass schedules consist of eight passes that correspond to the minimum number of passes possible. This is favored by both the initialization of the actor network and the reward function, introduced above (see Table 3). The process duration, i.e., rolling and cooling is about 300 s and almost identical in both cases. Moreover, the austenite grain size ($d_{\gamma,roll}$) after rolling is near identical (40.0 vs. 39.0 μm) and

Table 5 Overview of the outcome of different initialization strategies for the DDPG algorithm

Initiali-zation method	Initial pass schedule				Pass schedule after 10,000 iterations				Training time	
	No. of passes	Final height in mm	$d_{\gamma,roll}$ in μm	Process time in s	Initial R_t	Final height in mm	$d_{\gamma,roll}$ in μm	Process time in s		Final R_t
Max	8	15.0	40.0	188.0	-126.0	24.3	39.0	141.8	10.9	10 h 14 min
Min	23	25.0	10.1	276.2	-8.3	24.0	44.0	185.7	-1.6	52 h 39 min
Random	16	25.3	23.3	248.1	8.1	15.0	49.0	190.6	-127.9	14 h 26 min

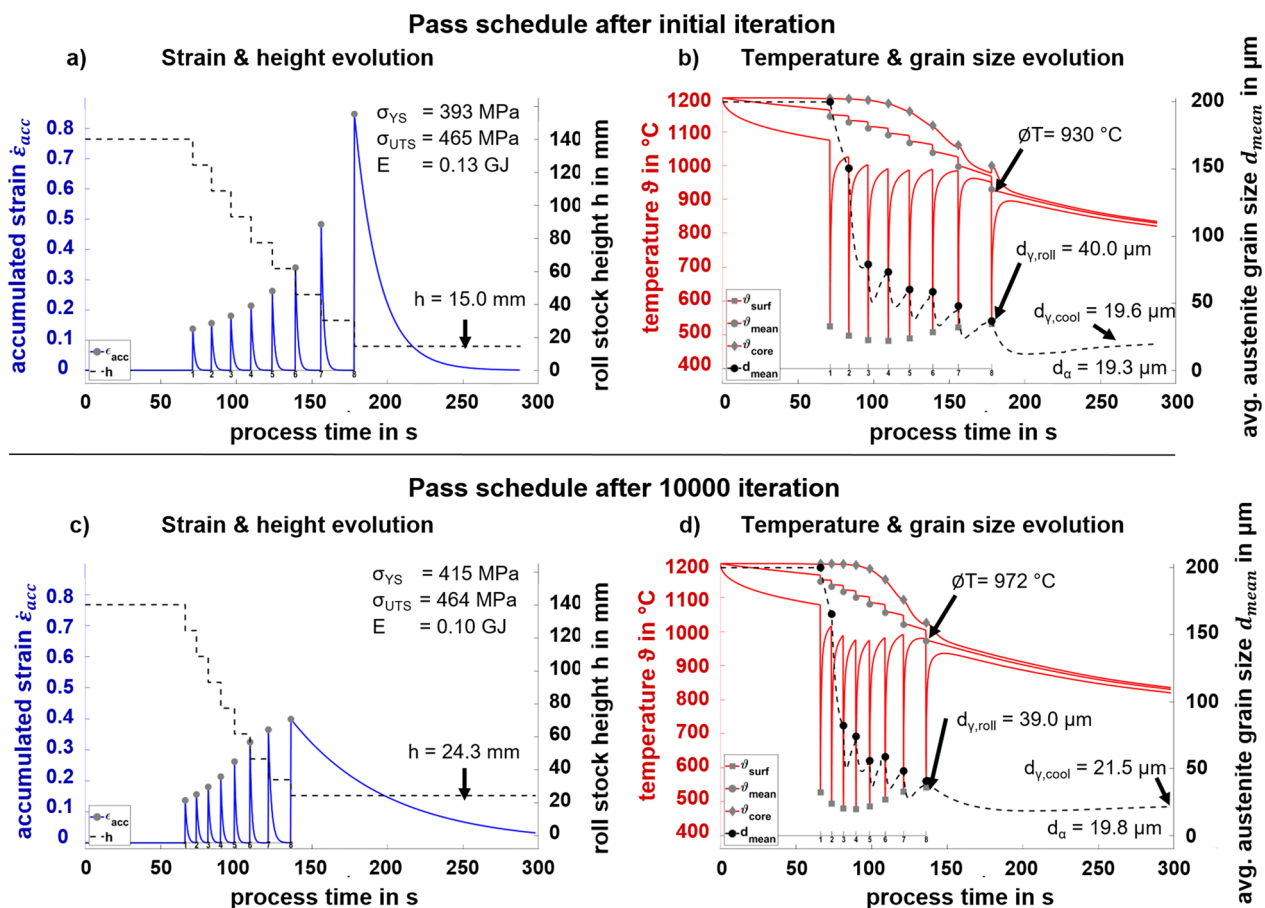


Fig. 6 Training results after initial iteration (top) and 10,000 iterations (bottom). On the left hand side, the strain (left axis) and height (right axis) evolution is shown (a) + c). On the right hand side, the temperature (left axis) and grain size (right axis) evolution are displayed (b) + d)

close to the upper limit of the set tolerance ($d_{\gamma,target} = 35 \pm 5 \text{ } \mu\text{m}$).

On the other hand, some differences are noticeable. Ending at 15 mm the initial pass schedule does not fulfill the set height tolerance ($25 \pm 1 \text{ mm}$) whereas the final pass schedule ends at 24.3 mm and thus does. As both schedules have identical initial height, this requires a lower total height reduction and consequently a smaller average draft per pass in the final schedule. When comparing the mechanical properties, it is noticeable that the difference in YS is slightly bigger (393 vs. 415 MPa) than the UTS difference (465 vs. 464 MPa) due to the dependence of YS on dislocation density, see Eq. 8. Furthermore although $d_{\gamma,roll}$ is almost identical in both cases (40.0 vs. 39.0 μm), the difference in the austenite grain size after cooling ($d_{\gamma,cool}$) is greater (19.6 vs. 21.5 μm) due to slower static recrystallization in the final pass schedule. Thus, a higher retained strain at the beginning of the transformation results in a smaller ferrite grain size (d_{α}), see Eq. 18. This results in almost the same d_{α} (19.3 vs. 19.6 μm). While the initial temperature is fixed to 1200 $^\circ\text{C}$ in both schedules, the inter-pass times and thus the total rolling time (190.51 vs.

141.82 s) is much shorter in the final pass schedule than in the initial one. This results in an average temperature difference after rolling (ΔT) of greater than 40 $^\circ\text{C}$ (930 vs 972 $^\circ\text{C}$). Consequently, the cooling time up until transformation is a bit longer for the final pass schedule. Besides that, the temperature difference between the two pass schedules is too small for it to have a significant effect on the microstructure evolution.

To sum up, during training the DDPG algorithm learned to adjust the drafts and the inter-pass times in such a way that the desired final height and grain size are reached. Especially, the reduction of the inter-pass time results in an increased average temperature during rolling and in combination with the reduced total and average height reduction the rolling forces and the energy consumption are reduced, too. The differences in YS and UTS can be explained by the fact that only the YS calculation considers dislocation density, see Eqs. 8 and 9. The dislocation density is correlated with the retained strain ϵ_r prior to the transformation from austenite to ferrite (see Eq. 18). The retained strain is greater for the

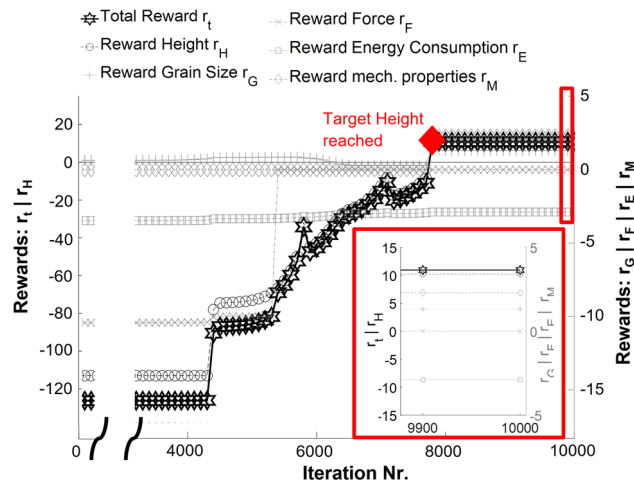


Fig. 7 Evolution of the reward constituents after each iteration; the left axis displays the total and height reward, the right axis displays the grain size, force, energy consumption and mechanical properties reward. The red marker at 7700 iterations highlights that the final height reached the tolerance for the first time. In a red rectangle the reward distribution of the last two stored reward distributions (9900 and 10.000 iterations) is detailed

final pass schedule and therefore the dislocation density and the YS increase.

The reward evolution is shown in Fig. 7. As presented before, the reward r_i is evaluated for each pass and consists of five optimization objectives, see Table 3, height r_H (left axis), average austenite grain size r_G , final UTS, r_M , rolling force r_F and the thermal and mechanical energy consumption r_E (all right axis). The total reward r_t is the sum of the rewards r_i in all passes and thus considered as evaluation metric. Remember, the intermediate reward only considers r_H, r_G, r_F, r_E as the final mechanical properties r_M are considered only in the last pass and thus in the completion reward. The total reward r_t (left axis, also see Eq. 24) is the sum of the intermediate rewards of pass $1 \dots n - 1$ and the completion rewards of the final pass n .

Three stages of development can be identified from the reward evolution, the first one due to DDPG's need of training samples (0–4500 iterations), in the second stage learning occurs resulting in an optimization (4600–7700 iterations) and in the final stage (7800–10.000 iterations) a (local) maximum is reached. As there are two stages—the first and the final—with constant rewards, only the evolution of the rewards in the second stage is described in more detail:

In the second stage between 4600 and 7700 iterations, all the active rewards increase. The mechanical properties reward r_M is still inactive (as the target height is not reached) and thus zero. The active rewards show a distinctly different increase behavior. While the height reward r_H gradually increases over time, the reward for the force r_F abruptly jumps from -10 to zero at about 5300 iterations when the

machine limits are no longer exceeded. At the same time the reward for the energy consumption r_E increases constantly but much slower. Finally, the grain size reward increase slightly, then decreases to shortly below zero and finally increases again whilst reaching its absolute maximum. In consequence also the total reward r_t increases in the second state from -126 to 11 . The second stage is completed as soon as the target height is reached after 7700 iterations, recognizable by a positive height reward r_H . This is due to the activation of the reward for the mechanical properties r_M , see reward definition in Table 3.

A detailed distribution of the rewards in the final stage is shown in the red rectangle in Fig. 7. However, after achieving the target height, no further improvement is achieved within the next 2300 iterations, which is probably due to a local minimum. This local minimum is most likely caused by the current reward function that is discontinuous (see Table 3) and thus can cause erratic gradients. Apart from improving the reward function, more iterations and/or a higher weight on the exploration behavior of the algorithm might help mitigate this problem. The exploitation versus exploration problem is detailed in the Appendix.

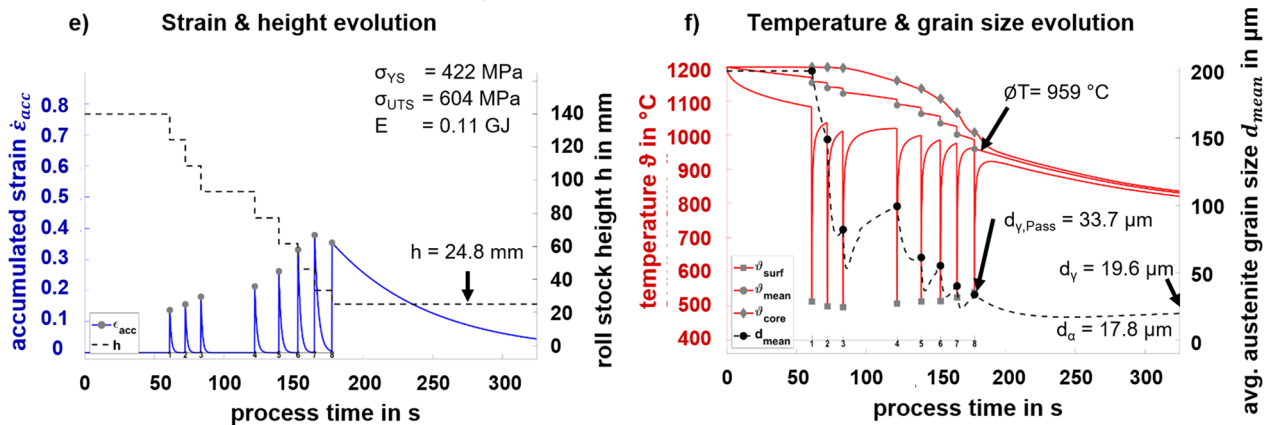
In summary, a FRM and a DDPG algorithm where successfully coupled and the DDPG algorithm's ANNs were trained based on an automatically generated, non-ideal pass schedule whose overall height reduction was too large. After training the DDPG is able to design a pass schedule that accounts for five different optimization objectives simultaneously in ~ 3 s (using an Intel Xenon CPU E3-1270). While a further optimization of the reward function which is the evaluation metric here and hyperparameters is required to better prevent local minima and associated training stalls, the presented concept clearly is worth further investigation. An important focus here will be the influence of the different reward components on the training process and final pass schedules. For this, a comprehensive study with different use cases is planned.

Validation and discussion of the exemplary designed pass schedule

Finally, the resulting properties of the designed pass schedule, i.e., the austenite grain size after hot rolling $d_{\gamma, roll}$, the final ferrite grain size d_{α} , as well as YS and UTS are compared to a rolling experiment. This shows firstly if the designed pass schedule can guarantee properties within specified tolerances and secondly if the predictive accuracy of the FRM is high enough. The validation rolling trial carried out based on the designed pass schedule is described first and after that, the experimentally determined properties are shown, compared to the FRM predictions and analyzed.

Table 6 Process parameters of the designed and experimental pass schedule

Pass No	Height after pass in mm	Δh in%	Designed start time in s	Exp. start time in s
1	124.3	11.2	65.0	60.0
2	108.6	12.6	72.3	71.5
3	93.0	14.4	79.8	82.4
4	77.3	16.9	87.8	121.0
5	61.6	20.3	96.5	138.3
6	46.2	25.0	106.0	152.0
7	33.3	27.9	117.0	163.9
8	24.9	25.3	130.3	176.5

Actual rolled pass schedule calculated with FRM**Fig. 8** FRM simulation of the experimental pass schedule shown in Table 6. On the left, the strain (left axis) and height (right axis) evolution is shown (a). On the right, the temperature (left axis) and grain size (right axis) evolution are displayed (b)**Validation rolling trial based on the designed pass schedule**

The pass schedule was hot rolled on a laboratory rolling mill at the IBF. Due to the manual operation of the rolling mill, the desired starting times were not met for each pass. A significant process disturbance occurs after the third pass, which results in an extended inter-pass time of about 30 s. Thus, the actual rolled pass schedule was recalculated using the FRM. Calculations for both schedules the originally designed and the actual rolled one are compared first to analyze in how far the pass schedule designed via DDPG is robust with respect to process disturbances and retains similar final properties. In Table 6 the process parameters of the designed and the experimental pass schedule are summarized; n.b. apart from the start times of the individual passes also the final thickness differs slightly.

In Fig. 8 the FRM results of the experimental pass schedule are presented. The figure on the left, (a), displays accumulated strain in blue (left axis) and the height of the rolling stock in black (right axis) over the process time. Moreover, the final YS and UTS as well as the total energy consumption (E) are shown. The figure on the right, (b), visualizes

the temperature distribution over the roll stock height in red (left axis) and the average austenite grain size in black (right axis). Furthermore, the average temperature ($\bar{\Delta T}$) and average austenite grain size right after rolling ($d_{\gamma,roll}$) as well as after cooling ($d_{\gamma,cool}$) and the final ferrite grain size (d_{α}) are depicted.

Looking at Fig. 6 bottom and Fig. 8 a comparison between the designed and the experimental pass schedule can be drawn: As mentioned above, the process time of the experimental pass schedule is larger due to the large inter-pass time in between passes three and four. This deviation causes a temperature difference of 11 °C (970 vs. 959 °C) at the end of the rolling process that in turn has an influence on the microstructure evolution during rolling. The larger inter-pass time lead to significant grain growth and a different austenite grain size d_{γ} after pass three. However, the grain size difference mostly vanishes throughout the rest of the schedule leading to a final austenite grain size $d_{\gamma,roll}$ of 33.7 μm in the experiment vs. a $d_{\gamma,roll}$ of 39 μm in the designed schedule. Both grain sizes are within the required tolerance of $35 \pm 5 \mu\text{m}$ after rolling. Similarly, the final ferrite grain sizes d_{α} differ only by a few μm (17.8 vs. 19.8 μm) and thus lead to very similar YS (422 vs. 415 MPa) and UTS (470 vs. 464 MPa). The slightly

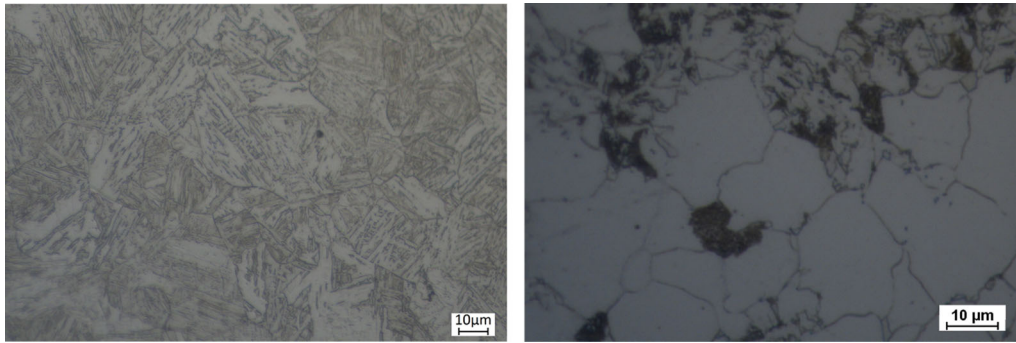
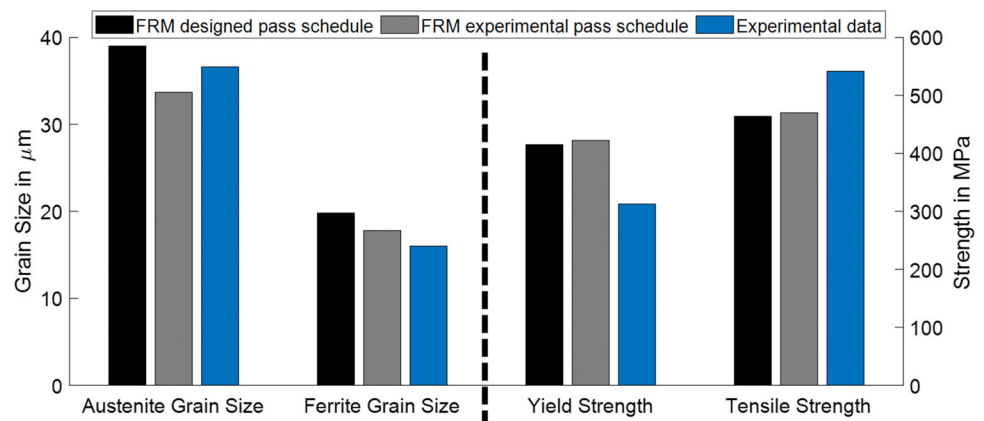


Fig. 9 Microstructure in the center of the hot-rolled material once quenched (left) and once normalized after quenching (right)

Fig. 10 Comparison between the measurements and FRM-predicted values



smaller grain size of the experimental pass schedule leads to a slightly bigger YS and UTS purely due to the Hall–Petch effect as the solid solution strengthening is identical in both schedules.

Comparison of predicted and experimental properties

In order to determine the average former austenite grain size, the hot-rolled material was quenched in water via a cooling line. Afterwards, samples were taken from the quenched roll stock to analyze the microstructure. However, for determining the YS and UTS the microstructure has to be ferritic-perlitic for this steel grade to comply with the typical delivery conditions to customers in industry. Therefore, the samples were annealed for 40 min at 900 °C. In Fig. 9 the microstructure of the quenched and annealed samples is presented. While the microstructure of the quenched sample is either bainitic or martensitic with an average hardness of 314 HV 10, the annealed material sample shows a distinct ferritic-perlitic microstructure with an average hardness of 190 HV 10. The hardness tests were carried out in the center of the specimen according to ISO 6507 using a Zwick ZHV10 hardness tester.

In Fig. 10 the average former austenite $d_{\gamma,roll}$ and ferrite grain size d_{α} from the experiment are compared to the FRM prediction for the experimental pass schedule and for the pass schedule originally designed via RL. The average measured $d_{\gamma,roll}$ is 36.6 μm while the FRM predicts 33.7 μm for the experimental schedule, demonstrating that $d_{\gamma,roll}$ can be predicted within a few μm. The measured average d_{α} of the annealed sample is 16.0 μm also showing good agreement with the FRM prediction for the experimental schedule of 17.8 μm.

Finally, tensile tests of the annealed samples according to ISO 6895 are carried out using a Zwick Z100 testing machine. The annealed material has a YS of 313 MPa and a UTS of 542 MPa. In contrast, the FRM predicts a YS of 422 MPa and UTS of 470 MPa. Thus, the predicted mechanical properties are 35% higher for YS and 13% lower for UTS. The noticeable difference in YS might be related to the simplified modelling of the dislocation density that is not currently based on an internal state variable model. Furthermore, a dedicated calibration of the microstructure, YS and UTS parameters for S355 structural steel should improve the accuracy of the FRM. Never the less, the microstructure, YS and UTS results show that the schedule laid out by the DDPG algorithm is quite robust in regards to extended inter-pass times, at least in the rolled geometry range considered.

Rolling trials with intentional irregularities have to be carried out to fully assess the robustness of the pass schedules generated via the DDPG algorithm.

6 Summary and outlook

In this paper, a novel concept was presented that shows the feasibility of coupling a FRM and DDPG algorithm for pass schedule generation. First, an existing FRM was extended to predict YS and UTS. Next, a complex reward function was developed for the DDPG algorithm in order to design pass schedules. By coupling the FRM and the DDPG algorithm a pass schedule that satisfied the target height and grain size while also maximizing the mechanical properties of the sheet, minimizing the energy consumption and adhering to the machine limits was laid out. When hot rolling this pass schedule on a laboratory scale rolling mill the predicted austenite and ferrite grain sizes were met within a few μm . Adversely, the resulting YS was predicted with an accuracy of 35% while the deviation in UTS was only about 13%. The deviations, especially in YS, likely stem from using common material-dependent parameters from literature and the uncertainties in the cooling rate after hot rolling. In the future carefully calibrating the material and property model parameters should improve the accuracy.

In summary, the presented coupling of FRM and RL can, for the first time, provide pass schedules that consider multiple objectives including final mechanical properties. It can therefore assist process experts from academia and industry in designing better pass schedules. Especially where current, mostly manual approaches have reached their limits. While further improvements to the reward function and hyperparameters are needed to increase efficiency, this proof of concept successfully demonstrates the potential of coupling conventional engineering models and reinforcement learning methods to tackle complex problems in metal forming. In this respect, this work could lead to further developments and applications of RL in engineering fields. Prospectively the presented approach could enable the real-time adaption of pass schedules to unforeseen process disturbances also in the context of properties controlled hot rolling.

Acknowledgements We thank Jürgen Nietsch for performing Vickers hardness tests and microstructural analysis.

Funding Open Access funding enabled and organized by Projekt DEAL. This study is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2023 Internet of Production—390621612.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

This appendix gives further details on the exploration behavior and the replay buffer of the DDPG algorithm.

Exploration behavior

In order to find an optimal solution for a given task, the DDPG algorithm needs to weight exploitation versus exploration behavior. Exploitation means choosing actions that are best according to current knowledge, while exploration means choosing actions that will increase knowledge and potentially enable higher rewards in the future. In other words, while exploration is necessary for the DDPG algorithm in order to learn, exploitation is necessary for the agent's performance and for convergence. Several implementations of different exploration behavior are available, one well-known example is the ϵ -greedy method. Here, the algorithm selects the currently best action with probability $1 - \epsilon$ and a random action with probability $0 \leq \epsilon \leq 1$ instead. Later Lillicrap et al. (2016) presented an alternative noise-base exploration which results in a more efficient exploration process. For sampling noise, the authors used the Ornstein–Uhlenbeck process, developed by Uhlenbeck and Ornstein (1930) and afterwards added the noise directly to the actions. Since this method gave good results in physical control problems, it is also used in the present study.

Replay buffer

The concept of a replay or experience buffer was introduced in 1992 by Lin (1992). One prominent application of a replay buffer was presented by Minh et al. (2013) who used a Deep Q-Learning algorithm to successfully play Atari games. In

the replay buffer, the experience or more specifically the state action pairs and the corresponding reward observed during the training, collected by the DDPG algorithm is stored. This is necessary as using ANNs in the concept of Reinforcement Learning causes two problems: The first problem is that the data needed to train the Actor and Critic Network is highly dynamic, i.e., the data changes during the training as the algorithm learns new relationships and behavior. For proper training of ANNs on the other hand, it is necessary that data is static and independent. Here, the second problem occurs. Data used to train the ANNs are typically highly correlated, i.e., the current state usually depends on the previous state action pair (Sutton & Barto, 2018). For example, the height within each pass schedule is strictly monotonously decreasing due to the pass by pass design but is reset to its initial value for the next pass schedule and the pattern repeats. Such correlated data might result in bad performance of the ANN because it will simply memorize the pattern during learning. The replay buffer serves to overcome the problem by changing the data distribution and breaking up the correlation in the training data. For this, the replay buffer has a fixed size and for each new state action pair, the oldest pair is replaced (Lillicrap et al., 2016). To train the ANNs batches from the replay buffer drawn at random are used. This ensures that, the training data available to the ANNs is as independent as possible. In addition to that, the replay buffer enables multiple replays of the same interaction, so the gained experience can be used in several updates of the Actor and Critic Network and is not lost. Furthermore, this greatly improves the data efficiency.

References

- Allwood, J. M., Cullen, J. M., & Carruth, M. A. (2012). *Sustainable materials: With both eyes open; [future buildings, vehicles, products and equipment - made efficiently and made with less new material]*. UIT Cambridge.
- Avrami, M. (1939). Kinetics of phase change: General theory. *The Journal of Chemical Physics*, 7, 1103–1112. <https://doi.org/10.1063/1.1750380>
- Beynon, J. H., & Sellars, C. M. (1992). Modelling microstructure and its effects during multipass hot rolling. *The Iron and Steel Institute of Japan*, 32(3), 359–367.
- Buchholz, F.-G. (1976). *Berechnung und Optimierung von Stichplänen für den stationären Betrieb kontinuierlicher Kalt- und Warmwalzstraßen*. Dissertation. Technische Hochschule, München.
- Choquet, P., Fabrègue, P., Giusti, J., Chamont, B., Pezant, J. N., & Blanchet, F. (1990). Modelling of forces, structure and final properties during the hot rolling process on the hot strip mill. In S. Yue (Ed.), *Hamilton, Ontario, Canada, 26–29 August* (pp. 34–43). Montreal, Canada: Canadian Institute of Mining and Metallurgy.
- Di, H., Ke, X., Peng, Z., & Dongdong, Z. (2019). Surface defect classification of steels with a new semi-supervised learning method. *Optics and Lasers in Engineering*, 117, 40–48. <https://doi.org/10.1016/j.optlaseng.2019.01.011>
- Dornheim, J., & Link, N. (2018). *Multiobjective reinforcement learning for reconfigurable adaptive optimal control of manufacturing processes* (pp. 1–5). IEEE.
- Dornheim, J., Link, N., & Gumbsch, P. (2019). Model-free adaptive optimal control of episodic fixed-horizon manufacturing processes using reinforcement learning. *International Journal of Control, Automation and Systems*, 18, 1593–1604. <https://doi.org/10.1007/s12555-019-0120-7>
- Edmonds, D. V., & Cochrane, R. C. (1990). Structure-property relationships in Bainitic steels. *Metallurgical Transactions A*, 21A, 1527–1540.
- Fujii, S., & Saito, M. (1975). *A new mathematical model for plate mill control: Construction department*. Nippon Kokan K.K.
- Gamal, O., Mohamed, M. I. P., Patel, C. G., & Roth, H. (2021). Data-driven model-free intelligent roll gap control of bar and wire hot rolling process using reinforcement learning. *International Journal of Mechanical Engineering and Robotics Research*. <https://doi.org/10.18178/ijmerr.10.7.349-356>
- Gladman, T., McIvor, I. D., & Pickering, F. B. (1972). Some aspects of the structure-property relationships in high-carbon ferrite-pearlite steels. *Journal of the Iron and Steel Institute*, 210, 916–930.
- Günther, J., Pilarski, P. M., Helfrich, G., Shen, H., & Diepold, K. (2014). First steps towards an intelligent laser welding architecture using deep neural networks and reinforcement learning. *Procedia Technology*, 15, 474–483. <https://doi.org/10.1016/j.protcy.2014.09.007>
- Guo, P., & Yu, J. (2019). Optimal control of blank holder force based on deep reinforcement learning. In *2019 IEEE international conference on industrial engineering and engineering management (IEEM), Macao, Macao, 15.12.2019–18.12.2019* (pp. 1466–1470). IEEE. <https://doi.org/10.1109/IEEM44572.2019.8978743>.
- Hall, E. O. (1951). The deformation and ageing of mild steel: III discussion of results. *Proceedings of the Physical Society: Section B*, 64, 747–753. <https://doi.org/10.1088/0370-1301/64/9/303>
- Hensel, A., & Spittel, T. (1978). *Kraft- und Arbeitsbedarf bildsamer Formgebungsverfahren*. VEB Deutscher Verlag für Grundstoffindustrie.
- Hernández Carreón, C. A., Mancilla Tolama, J. E., Castilla Valdez, G., & Hernández González, I. (2019). Multi-objective optimization of the hot rolling scheduling of steel using a genetic algorithm. *MRS Advances*, 4, 3373–3380. <https://doi.org/10.1557/adv.2019.436>
- Hodgson, P. D., & Gibbs, R. K. (1992). A mathematical model to predict the mechanical properties of hot rolled C-Mn and microalloyed steels. *ISIJ International*, 32, 1329–1338. <https://doi.org/10.2355/isijinternational.32.1329>
- Hwang, R., Jo, H., Kim, K. S., & Hwang, H. J. (2020). Hybrid model of mathematical and neural network formulations for rolling force and temperature prediction in hot rolling processes. *IEEE Access*, 8, 153123–153133. <https://doi.org/10.1109/ACCESS.2020.3016725>
- Jakubowski, J., Stanis, P., Bobek, S., & Nalepa, G. J. (2021–2021). Explainable anomaly detection for hot-rolling industrial process. In *2021 IEEE 8th international conference on data science and advanced analytics (DSAA), Porto, Portugal, 06.10.2021–09.10.2021* (pp. 1–10). IEEE. <https://doi.org/10.1109/DSAA53316.2021.9564228>.
- Johnson, W. A., & Mehl, R. F. (1939). Reaction kinetics in processes of nucleation and growth. *Transactions of the Metallurgical Society of AIME*, 135(8), 396–415.
- Jonas, J. J., Sellars, C. M., & Tegart, W. J. (1969). Strength and structure under hot-working conditions: Review 130. *Metallurgical Reviews*, 14(1), 1–24.
- Jonsson, N.-G., & Mäntylä, P. (1985). On-line control system for profile, shape and temperature in 4-high mills. *Proceedings of the 27th mechanical working & steel processings conference*, 129–136.

- Kitahara, A. R., & Holm, E. A. (2018). Microstructure cluster analysis with transfer learning and unsupervised learning. *Integrating Materials and Manufacturing Innovation*, 7, 148–156. <https://doi.org/10.1007/s40192-018-0116-9>
- Kolmogorov, V. L. (1937). On the statistical theory of the crystallization of metals. *Bulletin of the Russian Academy of Sciences*, 1, 355–359.
- Konda, V., & Tsitsiklis, J. (2001). Actor-critic algorithms. *Society for Industrial and Applied Mathematics*, 42.
- Korczak, P., Dyja, H., & Łabuda, E. (1998). Using neural network models for predicting mechanical properties after hot plate rolling processes. *Journal of Materials Processing Technology*, 80–81, 481–486. [https://doi.org/10.1016/S0924-0136\(98\)00151-4](https://doi.org/10.1016/S0924-0136(98)00151-4)
- Lanzillo, C. A. N., & Pickering, F. B. (2013). Structure–property relationships in dual-phase steels. *Metal Science*, 16, 371–382. <https://doi.org/10.1179/030634582790427433>
- Larkiola, J., Myllykoski, P., Korhonen, A. S., & Cser, L. (1998). The role of neural networks in the optimisation of rolling processes. *Journal of Materials Processing Technology*, 80–81, 16–23. [https://doi.org/10.1016/S0924-0136\(98\)00206-4](https://doi.org/10.1016/S0924-0136(98)00206-4)
- Lee, D. M., & Choi, S. (2004). Application of on-line adaptable neural network for the rolling force set-up of a plate mill. *Engineering Applications of Artificial Intelligence*, 17, 557–565. <https://doi.org/10.1016/j.engappai.2004.03.008>
- Lenard, J. G., Pietrzyk, M., & Cser, L. (1999). *Mathematical and physical simulation of the properties of hot rolled products*. Elsevier.
- Li, W., Liu, X., & Guo, Z. (2012). Multi-objective optimization for draft scheduling of hot strip mill. *Journal of Central South University*, 19, 3069–3078. <https://doi.org/10.1007/s11771-012-1380-z>
- Lieber, D., Stolpe, M., Konrad, B., Deuse, J., & Morik, K. (2013). Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning. *Procedia CIRP*, 7, 193–198. <https://doi.org/10.1016/j.procir.2013.05.033>
- Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y. et al. (2016). *Continuous control with deep reinforcement learning*. <http://arxiv.org/pdf/1509.02971v6>
- Lin, L.-J. (1992). Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, 8, 293–321. <https://doi.org/10.1007/BF00992699>
- Liu, L.-L., Wan, X., Gao, Z., Li, X., & Feng, B. (2019). Research on modelling and optimization of hot rolling scheduling. *Journal of Ambient Intelligence and Humanized Computing*, 10, 1201–1216. <https://doi.org/10.1007/s12652-018-0944-7>
- Lohmar, J., Bambach, M., Hirt, G., Kiefer, T., Kotliba, D., Jochum, M., et al. (2014a). Fast and accurate force prediction for high quality heavy plates by a state of the art rolling model calibrated from mill data via inverse techniques. In P. Darmayan & C. Lerouge (Eds.), *ESTAD2014a*, Paris, France, 07–08 April.
- Lohmar, J., Seuren, S., Bambach, M., & Hirt, G. (2014b). Design and application of an advanced fast rolling model with through thickness resolution for heavy plate rolling. In J. Guzzoni & M. Manning (Eds.), *Ingot casting, rolling & forging, Milan, Italy, 07–09 May*.
- Mahadevan, S., & Theocharous, G. (1998). Optimizing production manufacturing using reinforcement learning. In (Vol. 372, p. 377).
- Mäntylä, P., Myllykoski, L., & Jonsson, N.-G. (1989). Rolling wide thin plates using the profile and shape vector method. *Iron and Steel Engineer* (November), 48–54.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., et al. (2013). *Playing atari with deep reinforcement learning*. <http://arxiv.org/pdf/1312.5602v1>
- Moon, C. H., & Lee, Y. (2009). Methodology for draft schedule design of plate rolling process with peening effect considered. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 223, 1159–1169. <https://doi.org/10.1243/09544054JEM1417>
- Moussaoui, A., Selaimia, Y., & Abbassi, H. A. (2006). Hybrid hot strip rolling force prediction using a Bayesian trained artificial neural network and analytical models. *American Journal of Applied Sciences*, 3, 1885–1889. <https://doi.org/10.3844/ajassp.2006.1885.1889>
- Nakajima, K., Asamura, T., Kikuma, T., Matsumoto, H., Awazuhara, H., Kimura, T., et al. (1984). Hot strip crown control by six-high mill. *Transactions ISIJ*, 24, 284–291.
- Nakajima, K., Kokai, K., Koike, M., Kikuma, T., Ataka, M., & Kako, Y. (1985). New plate mill draft scheduling system for crown and flatness control. *Transactions ISIJ*, 25, 212–218.
- Okamoto, T., Misaka, Y., Yokoi, T., Kise, K., & Fujimaki, I. (1975). The new advanced system of plate mill computer control. *Proceedings World Congress of the International Federation of Automatic Control* 6th, 6, 422–429.
- Özgür, A., Uygun, Y., & Hütt, M.-T. (2021). A review of planning and scheduling methods for hot rolling mills in steel production. *Computers & Industrial Engineering*, 151, 106606. <https://doi.org/10.1016/j.cie.2020.106606>
- Ozsoy, I. C., Ruddle, G. E., & Crawley, A. F. (2013). Optimum scheduling of a hot rolling process by nonlinear programming. *Canadian Metallurgical Quarterly*, 31, 217–224. <https://doi.org/10.1179/cm.1992.31.3.217>
- Pandey, V., Rao, P. S., Singh, S., & Pandey, M. (2020). A calculation procedure and optimization for pass scheduling in rolling process: A review, 126–130.
- Peng, G., Huang, K., & Wang, H. (2021). Dynamic multimode process monitoring using recursive GMM and KPCA in a hot rolling mill process. *Systems Science & Control Engineering*, 9, 592–601. <https://doi.org/10.1080/21642583.2021.1967220>
- Petch, N. J. (1953). The cleavage strength of polycrystals. *J. Iron Steel Inst.*, 174, 25–28.
- Pietrzyk, M., Kusiak, J., & Glowacki, M. (1990). Some aspects of development of models for automatic control of rolling mills. *Steel Research International*, 61(8), 359–364.
- Qi, X., Wang, T., & Xiao, H. (2012). Optimization of pass schedule in hot strip rolling. *Journal of Iron and Steel Research International*, 19, 25–28. [https://doi.org/10.1016/S1006-706X\(12\)60135-6](https://doi.org/10.1016/S1006-706X(12)60135-6)
- Rath, S., Thakur, S. K., Mohapatra, S., & Karmakar, D. (2019). *Application of machine learning in rolling mills: Case studies*.
- Reinisch, N., Rudolph, F., Günther, S., Bailly, D., & Hirt, G. (2021). Successful pass schedule design in open-die forging using double deep Q-learning. *Processes*, 9, 1084. <https://doi.org/10.3390/pr9071084>
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408. <https://doi.org/10.1037/h0042519>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536. <https://doi.org/10.1038/323533a0>
- Saravanakumar, P., Jothamani, V., Sureshbabu, L., Ayyappan, S., Noorullah, D., & Venkatakrishnan, P. G. (2012). Prediction of mechanical properties of low carbon steel in hot rolling process using artificial neural network model. *Procedia Engineering*, 38, 3418–3425. <https://doi.org/10.1016/j.proeng.2012.06.395>
- Scheiderer, C., Thun, T., Idzik, C., Posada-Moreno, A. F., Krämer, A., Lohmar, J., et al. (2020). Simulation-as-a-service for reinforcement learning applications by example of heavy plate rolling processes. *Procedia Manufacturing*, 51, 897–903. <https://doi.org/10.1016/j.promfg.2020.10.126>
- Schmidtchen, M., & Kawalla, R. (2016). Fast Numerical simulation of symmetric flat rolling processes for inhomogeneous materials using a layer model—part I: Basic theory. *Steel Research International*, 87, 1065–1081. <https://doi.org/10.1002/srin.201600047>

- Sellars, C. M. (1979). The physical metallurgy of hot working. In C. M. Sellars & G. J. Davies (Eds.), *Sheffield, England, 17–20 July* (pp. 3–15). London: The Society.
- Sellars, C. M., & Beynon, J. (1985). *Conference on high strength low alloy steels*, 142.
- Sellars, C. M., & Tegart, W. J. (1966). On the mechanisms of hot deformation. *Acta Metallurgica*, 14(9), 1136–1138.
- Seuren, S., Bambach, M., Hirt, G., Heeg, R., & Philipp, M. (2010). Geometric factors for fast calculation of roll force in plate rolling. In Z.-J.-Xuehui (Ed.), *Peking, 15–17 September*. Beijing: Metallurgical Industry Press.
- Seuren, S., Seitz, J., Kraemer, A. M., Bambach, M., & Hirt, G. (2014). Accounting for shear deformation in fast models for plate rolling. *Production Engineering*, 1, 17–24. <https://doi.org/10.1007/s11740-013-0500-4>
- Shen, S., Guye, D., Ma, X., Yue, S., & Armanfard, N. (2022). Multistep networks for roll force prediction in hot strip rolling mill. *Machine Learning with Applications*, 7, 100245. <https://doi.org/10.1016/j.mlwa.2021.100245>
- Shohet, K. N., & Townsend, N. A. (1968). Roll bending methods of crown control in four-high plate mills. *Journal of the Iron and Steel Institute*, 1088–1098.
- Siebel, E. (1925). Kräfte und Materialfluss bei der bildsamen Formgebung. *Stahl Und Eisen*, 45(37), 1563–1566.
- Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., & Riedmiller, M. (2014). Deterministic policy gradient algorithms. In *Proceedings of the 31st international conference on machine learning* (32nd ed., pp. 387–395). Beijing, China.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of Go without human knowledge. *Nature*, 550, 354–359. <https://doi.org/10.1038/nature24270>
- Sims, R. B. (1954). The calculation of roll force and torque in hot rolling mills. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 168(1), 191–200.
- Sims, R. B., & Wright, H. (1963). Roll force and torque in hot rolling mills. *Journal of the Iron and Steel Institute*, 3(5), 261–269.
- Singh, A. P., Sengupta, D., Jha, S., Yallasiri, M. P., & Mishra, N. S. (2013). Predicting microstructural evolution and yield strength of microalloyed hot rolled steel plate. *Materials Science and Technology*, 20, 1317–1325. <https://doi.org/10.1179/026708304225022296>
- Spuzic, S., Narayanan, R., Kovacic, Z., Hapu Arachchige, D., & Abhary, K. (2017). Roll pass design optimisation. *The International Journal of Advanced Manufacturing Technology*, 91, 999–1005. <https://doi.org/10.1007/s00170-016-9424-4>
- Sutton, R. S., & Barto, A. (2018). *Reinforcement learning: An introduction (Adaptive computation and machine learning)*. The MIT Press.
- Szerenyi, I. (1984). Schedule pass planning with dialogic computer program for reversing hot strip rolls. *Arch. Eisenhüttenwesen*, 55(7), 313–320.
- Taylor, M. E., & Stone, P. (2009). Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10, 1633–1685.
- Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical Review*, 36, 823–841. <https://doi.org/10.1103/PhysRev.36.823>
- von Kármán, T. (1925). Beitrag zur Theorie des Walzvorganges. *Zeitschrift Für Angewandte Mathematik Und Mechanik*, 5(2), 139–141.
- Wang, D. D., Tieu, A. K., de Boer, F. G., Ma, B., & Yuen, W. (2000). Toward a heuristic optimum design of rolling schedules for tandem cold rolling mills. *Engineering Applications of Artificial Intelligence*, 13, 397–406.
- Wang, S. R., & Tseng, A. A. (1996). ISS Mech. *Work Steel Processing Conference Proceedings*, 33, 805–818.
- Wu, S., Zhou, X., Ren, J., Cao, G., Liu, Z., & Shi, N. (2018). Optimal design of hot rolling process for C-Mn steel by combining industrial data-driven model and multi-objective optimization algorithm. *Journal of Iron and Steel Research International*, 25, 700–705. <https://doi.org/10.1007/s42243-018-0101-8>
- Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. *Production & Manufacturing Research*, 4, 23–45. <https://doi.org/10.1080/21693277.2016.1192517>
- Xie, Q., Suvama, M., Li, J., Zhu, X., Cai, J., & Wang, X. (2021). Online prediction of mechanical properties of hot rolled steel plate using machine learning. *Materials & Design*, 197, 109201. <https://doi.org/10.1016/j.matdes.2020.109201>
- Youkachin, S., Ruchanurucks, M., Phatrapomnant, T., & Kaneko, H. (2019 - 2019). Defect segmentation of hot-rolled steel strip surface by using convolutional auto-encoder and conventional image processing. In *2019 10th international conference of information and communication technology for embedded systems (IC-ICTES), Bangkok, Thailand, 25.03.2019–27.03.2019* (pp. 1–5). IEEE. doi:<https://doi.org/10.1109/ICTEmSys.2019.8695928>.
- Zhang, D., Du, L., & Gao, Z. (2021). Real-time parameter identification for forging machine using reinforcement learning. *Processes*, 9, 1848. <https://doi.org/10.3390/pr9101848>
- Zhang, F., Zhao, Y., & Shao, J. (2016). Rolling force prediction in heavy plate rolling based on uniform differential neural network. *Journal of Control Science and Engineering*, 2016, 1–9. <https://doi.org/10.1155/2016/6473137>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.