

Reinforcement Learning on Reconfigurable Hardware: Overcoming Material Variability in Laser Material Processing

Giulio Masinelli^{1,2}, Chang Rajani¹, Patrik Hoffmann¹, Kilian Wasmer¹, and David Atienza²

Abstract—Ensuring consistent processing quality is challenging in laser processes due to varying material properties and surface conditions. Although some approaches have shown promise in solving this problem via automation, they often rely on predetermined targets or are limited to simulated environments. To address these shortcomings, we propose a novel real-time reinforcement learning approach for laser process control, implemented on a Field Programmable Gate Array to achieve real-time execution. Our experimental results from laser welding tests on stainless steel samples with a range of surface roughnesses validated the method’s ability to adapt autonomously, without relying on reward engineering or prior setup information. Specifically, the algorithm learned the optimal power profile for each unique surface characteristic, demonstrating significant improvements over hand-engineered optimal constant power strategies — up to 23% better performance on rougher surfaces and 7% on mixed surfaces. This approach represents a significant advancement in automating and optimizing laser processes, with potential applications across multiple industries.

I. INTRODUCTION

Laser material processing, including applications such as welding, cutting, and additive manufacturing, is a critical technology widely employed in various industrial sectors, such as automotive manufacturing [1], aerospace engineering [2], and electronics assembly [3]. These processes — valued for their precision, speed, minimal mechanical interaction, and ability to produce high-quality results — involve focusing a high-power laser beam onto the material’s surface, creating localized effects such as melting, vaporization, or chemical reactions [4]. However, ensuring consistent processing quality across different materials and conditions poses significant challenges. For example, variations in material properties — such as surface roughness, composition, and thickness — can affect the process outcomes, necessitating real-time adjustments to the laser parameters.

Traditionally, these adjustments are manually optimized by engineers, a process that is time-consuming and prone to errors, since even minor changes in material properties can require extensive reprogramming of the laser system [5]. This manual approach is increasingly inadequate for meeting

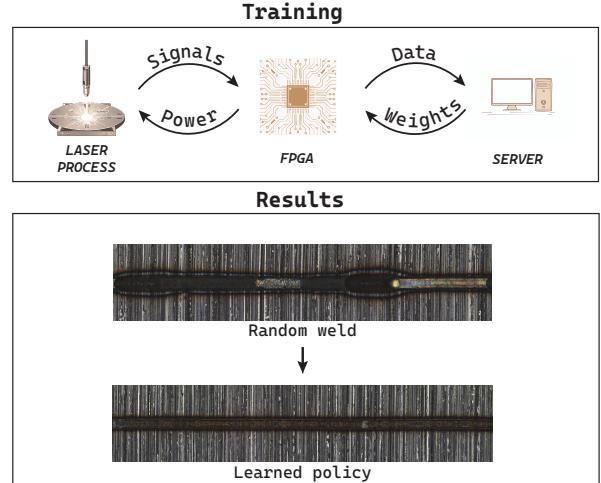


Fig. 1. Illustration of the proposed method. **Top:** The FPGA receives optical signals from the process zone and uses its onboard policy network to determine the laser power in real-time. Between processing runs, the collected data is sent to a server where RL is used to train the policy. **Bottom:** The policy initially starts with random actions and learns to optimize the process outcome, achieving the best possible results while avoiding defects such as keyhole formation.

the demands of modern manufacturing, necessitating the development of advanced control strategies. Among these, Reinforcement Learning (RL) has emerged as a particularly promising technique, demonstrating success in complex real-world control tasks such as robotics [6], nuclear fusion [7], and even laser welding [8], [9].

Specifically, RL methods learn control policies by interacting with the environment and receiving feedback, making them particularly suited for dynamic and complex processes like laser material processing. Nevertheless, although RL has shown considerable success in various industrial applications, its application to laser processes remains relatively unexplored. In fact, the fluctuating dynamics of these processes, combined with difficult-to-interpret sensor signals and challenging performance evaluations, present unique challenges for RL methods.

Despite these obstacles, several notable attempts have been made to apply RL to laser process control. For example, Günther et al. (2016) [10] pioneered this approach in laser welding, using weld width as input and a sigmoid-transformed depth error as reward. Although innovative, their method relied on a predetermined reference depth and was primarily evaluated in a simplified simulation, potentially limiting its real-world applicability. Advancing this concept, Masinelli et al. (2020) [8] introduced a more sophisticated RL approach, incorporating multiple input signals, including optical and acoustic emissions, to control laser power. How-

¹Laboratory for Advanced Materials Processing (LAMP), Swiss Federal Laboratories for Materials Science and Technology (Empa), Thun, Switzerland. Correspondence to giulio.masinelli@empa.ch

²Embedded Systems Laboratory (ESL), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

© 2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

ever, their method faced generalizability challenges due to a hand-engineered reward function based on a Machine Learning (ML) classifier, and its implementation on a commercial PC introduced latency and non-deterministic execution — limiting its possibility to adapt to surface changes. More recently, Kaneko et al. (2023) [9] proposed an RL control method for laser welding that minimizes bead width error by adjusting both laser power and scan speed. While promising, this approach required a predetermined target bead width and — similar to the work of Günther et al. — was only validated in a simulated environment.

These studies highlight the potential of RL in laser process control while underscoring persistent challenges in the field:

- The need for a more generalizable approach rooted in physics-based understanding, rather than relying on situation-specific reward functions or predetermined targets.
- Reliance on simulated environments for validation, potentially overlooking real-world complexities.
- Implementation challenges that introduce latency or limit the system's ability to respond to rapid changes in processing conditions.

To address these limitations, we present a novel real-time method for laser material processing, utilizing reinforcement learning to adjust laser parameters based on optical signals from the process zone (PZ). Our system combines an FPGA-implemented policy for rapid execution with a server-based training component, leveraging the FPGA's high-speed data processing capabilities for real-time control while utilizing the server's computational power to continuously update the model's parameters.

Although our approach is general and applicable to various laser processes, we demonstrate its effectiveness in the context of laser welding. Specifically, our work advances the field through the following key contributions:

- **Real-Time Closed-Loop Control on FPGA:** We develop a closed-loop control system implemented on an FPGA, capable of adjusting laser power in real time at microsecond scales, ensuring rapid response to dynamic conditions.
- **Autonomous Adaptation Without Prior Knowledge:** Our RL method adapts to different surface conditions and material properties, without requiring assumptions or manual tuning.
- **Experimental Validation on Real-World Surfaces:** We validate our system in laser welding, conducting extensive experiments on 316L stainless steel samples with varying surface roughness, demonstrating significant improvements over optimal constant power strategies.

II. EXPERIMENTAL METHODOLOGY

The core components of the setup include a laser source, an optical laser head, a stage holder mounted on a moving stage, and optical photodiodes.

The laser source is a fiber laser system LFS150 (Coherent Switzerland AG, Switzerland), with a maximum output capacity of 250 W at a wavelength of 1070 nm. The diameter of the laser spot measures 34 μm (within $1/e^2$) in the focal plane. Operating in continuous-wave (CW) mode, the source allows power modulation via an external voltage source (0–5 V). For this paper, we configured the system to operate within a power range of 25–100 W, which corresponds to the full range of the external voltage control.

All experiments were conducted in air at atmospheric pressure. To prevent weld oxidation and minimize plume absorption and scattering effects on optical signals, an argon flow was directed to the PZ blowing at a constant rate of 2 L/min.

Line welds were achieved by mounting samples on a linear Zaber LRT0250DL stage (Zaber Technologies, Canada), moving at a constant velocity of 50 mm/s. The movement of the sample was synchronized with the laser source to ensure that irradiation began only after the stage reached the set velocity.

A. Sensor System

The laser head incorporates a customized optical system that directs the on-axis optical radiation from the PZ to two photodiodes, enabling simultaneous measurement of optical reflection and emission. For optical reflection (OR) measurements, we employed a silicon (Si) free-space amplified photodetector (PDA100A2, Thorlabs) with a spectral range of 320–1100 nm, set at a gain of 30 dB. Reflection data provide information on surface conditions and transitions between melting regimes, such as from conduction mode (where heat is transferred into the material with minimal vaporization) to keyhole mode (where intense laser power creates a vapor-filled cavity or ‘keyhole’) [11]. For optical emission (OE) measurements, we used an Indium Gallium Arsenide (InGaAs) free-space amplified photodetector (PDA10CS2, Thorlabs) with a spectral range of 900–1700 nm, set at 50 dB gain. This detector was paired with an NF1064-44 notch filter (Thorlabs, CWL: 1064 nm, FWHM: 44 nm). The emission data complement the reflection information, providing a measure of the thermal state of the PZ and allowing an indirect assessment of the temperature of the heated and therefore emitting surface [11]. Both sensors had a measured field of view (FoV) of 3 mm in diameter, exceeding the typical melt pool size (100–300 μm), thus allowing monitoring of both the immediate melt area and its Heat-Affected Zone (HAZ).

B. FPGA Board and Data Acquisition

The novel aspect of our setup lies in its data acquisition and processing system, built around the Digilent Eclypse Z7 board with a Xilinx Zynq-7020 SoC. This system integrates a 667 MHz dual-core Cortex-A9 processor with an FPGA operating at 100 MHz. Two SYZYGY-compliant interfaces connect expansion modules: one with an AD9648 ADC (dual-channel, 14-bit, up to 125 MS/s) for recording photodiode signals at 100 kS/s, and another with an AD9717 DAC

(dual-channel, 14-bit, 125 MS/s) for laser power modulation. The 100 kS/s sampling rate was chosen based on preliminary studies showing negligible frequency components above 50 kHz.

C. Material

The samples used were 2 mm thick plates of 316L stainless steel with a melting temperature of 1400 °C. This material was selected due to its wide industrial applications, including use in chemical and petrochemical equipment [12], food processing [13], and medical devices [14]. Additionally, 316L's HAZ is easily recognizable in cross-sections due to its distinctive textural changes [15], making it ideal for post-processing analysis.

To test the algorithm's performance across a range of surface conditions, we experimented on three distinct sets of samples characterized by their surface roughness, quantified using the arithmetic mean height (S_a). The first set consisted of brushed samples with $S_a = 1.47 \mu\text{m}$; the second set comprised sandblasted samples with $S_a = 1.20 \mu\text{m}$; and the third set included mixed samples with alternating brushed and sandblasted sections — specifically, 10 mm brushed ($S_a = 1.47 \mu\text{m}$), 20 mm sandblasted ($S_a = 1.23 \mu\text{m}$), and 10 mm brushed ($S_a = 1.47 \mu\text{m}$).

This variety of surface conditions allowed us to evaluate the adaptability and robustness of our RL algorithm in different welding scenarios. All surface roughness measurements were performed using an S Neox 3D Optical Profiler (Sensofar Metrology, Spain) operating in white-light interferometry mode.

III. ALGORITHM IMPLEMENTATION

Our experimental setup employs an algorithm that integrates FPGA inference, data streaming to an external server, replay buffer updates, model training with Quantization-Aware Training (QAT), and continuous weight updates. This setup ensures that the FPGA is always operating with the most current model parameters, allowing rapid adaptability to changing conditions in the welding setup.

The process begins with the FPGA performing inference using pre-trained model weights stored in the Programmable Logic (PL) BRAM. During a single-line welding — which we refer to as an episode — data from the photodiodes are acquired through the ADC module at a sampling rate of 100 kS/s and processed by performing a forward pass of the neural network to compute control actions in real-time.

Upon the completion of an episode, the collected data — including sensor signals and applied actions — are transferred from the FPGA to the external server via Ethernet. This data transfer is managed by the ARM microcontroller on the System-on-Chip (SoC) — also known as the processing system (PS). The server maintains a replay buffer, which stores experience tuples (s, a, r, s') representing the state, action, reward, and next state for each time step.

Between welding operations, the server performs model training using QAT. This process employs a digital twin of the FPGA created using the Python library Brevitas (Xilinx,

USA) [16], which simulates the FPGA's behavior, including quantization effects. After each training iteration, the server serializes the updated model weights and transmits them back to the SoC. The ARM microcontroller then uses Direct Memory Access (DMA) to update the PL BRAM with these new weights.

The following pseudocode outlines the procedure of our experimental setup:

```

1: procedure MAINLOOP
2:   Initialize FPGA policy  $\pi$  and digital twin
3:   Stream initial weights from server to PS
4:   Transfer weights from PS to PL BRAM via DMA
5:   while not converged do
6:     Acquire data from ADC on FPGA
7:     if OR signal  $\geq 0.1V$  then
8:       for  $t = 1$  to  $N_{\text{steps}}$  do
9:         Acquire optical data  $s_t$ 
10:        Sample  $a_t \sim \mathcal{N}(\pi_\mu(s_t), \pi_\sigma(s_t))$ 
11:        Apply  $a_t$  to laser
12:        Store  $(s_t, a_t)$  pair in FIFO buffer
13:     end for
14:   end if
15:   Transfer data from FIFO buffer to PS via DMA
16:   Stream data to server via Ethernet
17:   Calculate  $R_t = \text{OR}(s_{t+1}) \forall t$ 
18:   Perform  $N_{\text{steps}}$  gradient steps of SAC learning
19:   Update model weights via QAT
20:   Transfer new weights to PS and to PL
21: end while
22: end procedure
```

A. Reinforcement Learning Setup

We formulate the laser welding process as a Partially Observable Markov Decision Process (POMDP) [17], since the complicated dynamics of molten metal during welding are not directly observable. A POMDP is a discrete-time stochastic process in which at every time step t the state s_t represents the true condition of the weld, including the dynamics of the melting pool and the temperature distribution. An agent interacts with this environment by taking actions a_t , which in our case correspond to voltages controlling the laser power between 25W and 100W. However, due to the partial observability of the process, the agent must rely on observations o_t as input. As noted before, our observations are derived from two optical sensors, providing indirect information about the melt pool dynamics. After each action, the environment transitions to a new state s_{t+1} , and the agent receives a reward $r(s_{t+1}, s_t, a_t)$ based on the quality of the weld.

The goal of the RL agent is to maximize the expected discounted return, given by $\mathbb{E}[\sum_{t=0}^T \gamma^t r(s_{t+1}, s_t, a_t)]$, where γ is the discount factor. This approach optimizes for the sum of the given reward quantity for the entirety of the episode — in this case one-line weld. Notice that, while reward engineering is challenging in general [18], choosing a good reward for laser welding is even more difficult, since many useful metrics of weld quality (such as weld depth used by

[10] or weld track width used by [9]) are very challenging to obtain in real-time, or have to be acquired by extensive post-hoc manual labor.

To overcome these limitations, we used the findings of Wittemer et al. [19], who provide valuable information on the relationship between optical signals and weld pool dynamics. Their research reveals that the OR signal reaches a peak value just before the melt pool transitions from conduction mode to keyhole mode, corresponding to the formation of the largest stable weld.

Based on this understanding — while we present a system that can in principle be used with any reward — we choose a simple reward function which encourages large stable welds while staying within conduction mode (avoiding vaporization of material), using the OR signal as the reward:

$$r(s_{t+1}) = \frac{\text{OR}(s_{t+1})}{10}. \quad (1)$$

This simple reward scales between 0 and 1 based on the photodiode sensor values (0-10 V).

Each episode consists of 80 steps, each 10 ms long. For the first 25 episodes, actions are sampled uniformly at random for exploration; subsequently, they are drawn from the learned policy. Every 10 episodes, a test episode is conducted in which actions are derived solely from the mean of the policy distribution $\pi_\mu(s_t)$, without sampling.

As the core RL algorithm we chose Soft Actor-Critic [20] with adaptive entropy tuning [21], since maximum entropy RL algorithms have been found to perform well on continuous control tasks [22]. We used the following hyperparameters for the SAC agent: hidden layer sizes [32, 64], ReLU activation, target entropy of -2, learning rate of 3×10^{-4} , batch size of 100, and a discount factor of 0.99. The agent performs 80 gradient steps at each training iteration, using mini-batches of size 100 sampled from the replay buffer.

B. Agent Implementation in HLS

To achieve the low-latency inference required for the RL agent, we implemented a neural network policy on the FPGA using High-Level Synthesis (HLS) [23], enabling efficient hardware realization directly from high-level algorithm descriptions.

The policy network consists of a Multi Layer Perceptron (MLP) with an input layer that receives data signals from the ADC, two hidden layers that perform weighted sums and activation functions, and an output layer that produces the final inference results. The HLS implementation optimizes the network for efficient parallel processing and pipelining, fundamental for high-speed inference operations. The core clock of the HLS module is set at 100 MHz via the internal clock divider. Specifically, we designed a pipelined architecture in which the processing of the current data window overlaps with the acquisition of the subsequent one. This overlap ensures continuous operation without idle cycles and minimizes latency in computing control actions.

In particular, during data acquisition, each processing window consists of exactly one million clock cycles, corresponding to a 10 ms time frame due to the decimation factor of 1,000 at a 100 kS/s sampling rate. After the acquisition phase, the processing of the data window is completed in just 354 clock cycles (3.54 μ s), significantly shorter than the interval between two consecutive data-points from the ADC (10 μ s). This rapid processing ensures that the computed control action is available almost instantaneously relative to the sampling period.

To balance precision with resource utilization, we employ integer arithmetic for weights and biases, maintaining them at 8-bit precision. However, to prevent overflow during computations, we implement a growing bitwidth strategy for activation values. This approach allows the bitwidth to increase progressively through the network layers, ensuring computational accuracy while optimizing resource usage.

In the output layer, the integer values are scaled and converted to floating points to compute the mean $\pi_\mu(s_t)$ and standard deviation $\pi_\sigma(s_t)$ for action sampling using the reparameterization trick [24]. The scaling factor for this conversion is a parameter provided by the digital twin, ensuring consistency between the FPGA implementation and the training environment:

$$a_t = \pi_\mu(s_t) + \pi_\sigma(s_t) \cdot \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1). \quad (2)$$

To avoid generating normally distributed numbers on the FPGA, ε are sampled on the server and streamed along with the weights, matching the number of steps in the environment.

For the activation functions, we primarily use the Rectified Linear Unit (ReLU) throughout the network due to its efficient implementation on the FPGA. However, for action squashing, we approximate the hyperbolic tangent function (\tanh) using a piecewise polynomial approximation optimized for the FPGA architecture.

The combination of these optimizations results in an efficient and fast MLP implementation on the FPGA, capable of meeting the real-time requirements of the laser welding control system.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of our RL laser welding control system, we performed experiments on three types of 316L stainless steel samples: brushed, sandblasted, and mixed surfaces. For each sample type, we performed welding operations using both our learned policy and an optimal constant power strategy for comparison.

The optimal constant power for each surface type was determined through a grid search over the laser power range from 25W to 100 W. Specifically, we selected the power setting that yielded the highest average OR signal along the weld lines, correlating with the best weld quality in conduction mode achievable with traditional constant power welding.

Figure 2 illustrates the learning progress and performance of our algorithm on the three sample types during both the

training and testing episodes. The figure gives the episode return, which refers to the cumulative reward obtained by the agent over the course of a single episode. It is important to note the distinction between training and testing episodes. During training, the agent explores and exploits, leading to potentially noisy performance metrics. Test episodes, on the other hand, provide a cleaner and more stable measure of the agent's true capabilities by evaluating the learned policy's performance separately from the training process.

For brushed samples, the algorithm showed a gradual improvement in performance. The training curve indicates a slow but steady increase in return, eventually matching the optimal constant-power strategy after approximately 200 episodes. This gradual improvement reflects the challenge of outperforming an already optimized constant power on a uniform surface. Once stabilized, the policy achieved a 6.8% higher test return than the baseline, demonstrating improvement even under seemingly optimal conditions.

In the case of sandblasted samples, our algorithm demonstrated its most impressive performance. The policy quickly learned to outperform the optimal constant power strategy, showing significant improvements within the first 40 episodes. This rapid learning and superior performance can be attributed to two key factors related to the nature of the sandblasted surfaces and the welding process itself. Firstly, sandblasting creates more homogeneous surfaces, which significantly reduces the variability in optical signals. This reduction in signal noise enhances the consistency of action selection and simplifies the learning process, as the policy's actions are derived directly from these optical inputs. Secondly, the welding process on sandblasted surfaces introduces a dynamic element that favors our adaptive approach. In fact, as welding begins, the sandblasting-induced surface roughness disappears. This transition creates a necessity for dynamic laser power adjustment, giving our policy an upper hand over the constant power baseline. The combination of these factors allowed our policy to achieve a substantial 22.6% improvement over traditional optimized constant laser power strategies, highlighting the effectiveness of our approach in handling various welding scenarios.

The mixed sample scenario, which combined both brushed and sandblasted surfaces, presented a more complex challenge. In fact, the learning curve shows greater variability, reflecting the policy's attempts to adapt to changing surface conditions within single welding lines. After approximately 200 episodes, the algorithm successfully learned a policy that outperforms the optimal constant power approach, ultimately achieving a 7.30% improvement in terms of test return. Figure 3 illustrates the relationship between the OR and laser power actions during a test episode on this sample. Notably, the learned policy adopts an ideal power profile: it initiates the weld with a power spike to overcome the initial high reflectivity and start the melting process, then reduces the power to maintain stable welding on the brushed surface. Upon entering the sandblasted region (approximately steps 20 to 60), the policy increases the laser power to compensate for the surface condition change, as evidenced by

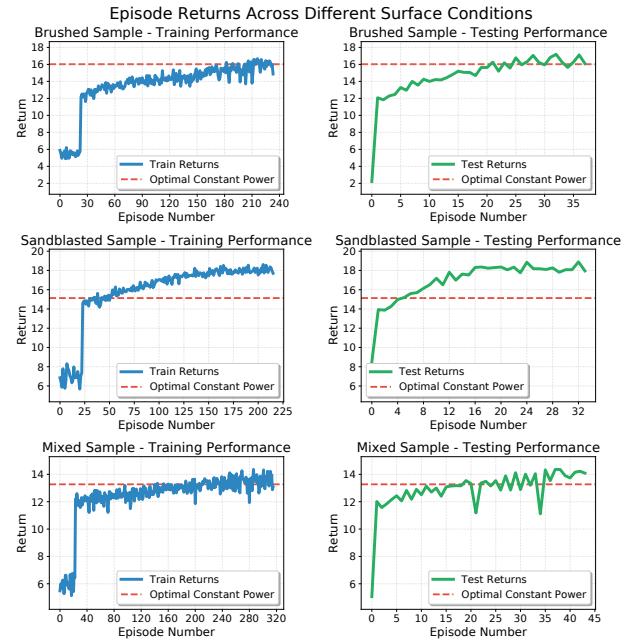


Fig. 2. Comparison of our RL algorithm performance on different sample types during training and testing episodes. Three rows of plots are shown: Brushed, Sandblasted, and Mixed samples. Each row contains two graphs: Train Episode Returns (left) and Test Episode Returns (right). The blue lines represent the episode returns, while the red dashed lines indicate the Optimal Constant Power.

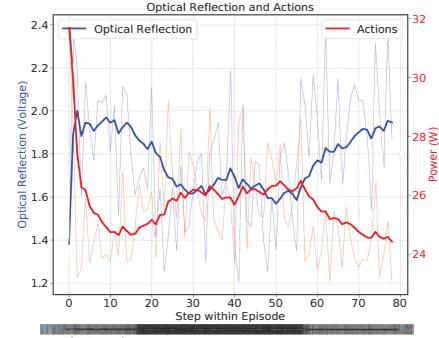


Fig. 3. Comparison of OR and laser power actions during a test episode. OR (blue) and laser power actions (red). A power spike at the beginning initiates the melting process, while the increase in the middle of the episode corresponds to adjustments made for the sandblasted region. The microscope image at the bottom illustrates the corresponding processed line.

the decrease in the OR signal. After exiting the sandblasted region, the policy reduces the power again when returning to the brushed surface. This dynamic adjustment demonstrates the policy's ability to respond effectively to varying surface conditions in real time.

To further validate our learned policy, we performed post-fabrication analysis of the welded samples. Figure 4 compares representative melt-pool cross-sections obtained from the mixed sample, after grinding, polishing, and chemical etching with Aqua Regia, for both the optimal constant power strategy (left column) and our learned policy (right column) for both brushed (top row) and sandblasted (bottom row) surfaces. In each condition, the learned policy consistently produced wider and deeper melt pools while avoiding key-hole formation (that often leads to porosity [25]). These enhanced melt pools indicate better penetration and fusion, which are critical for weld quality in conduction mode [26], [27]. In this mode, porosity is minimal [28], leading to welds

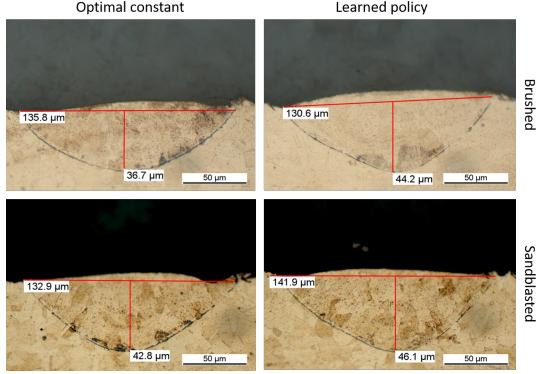


Fig. 4. Representative melt-pool cross-sections taken from the mixed sample. **Left column:** Optimal constant power strategy. **Right column:** Learned policy. **Top row:** Brushed surface. **Bottom row:** Sandblasted surface.

that typically exhibit more reliable mechanical properties. Indeed, tensile strength is strongly influenced by microstructural features and fusion zone characteristics [29], both of which correlate with melt-pool geometry. Accordingly, the larger melt pools observed here suggest improved tensile performance in addition to the apparent gains in penetration and fusion. While direct mechanical testing (e.g., tensile or fatigue tests) could offer further quantitative comparisons of weld performance, these cross-sectional results already underscore the policy's adaptability and effective process optimization across diverse surface conditions.

V. DISCUSSION

In this paper, we presented a methodology for generic laser material processing using RL, focusing on laser welding to showcase its feasibility. Our system demonstrates remarkable efficiency in both speed and resource utilization, successfully maintaining conduction-mode welding across various steel surfaces and identifying optimal power profiles with minimal training data. This sample- and time-effectiveness underlines its adaptability for rapid deployment in diverse manufacturing scenarios.

Nonetheless, real manufacturing conditions often involve factors not fully captured in our current setup. For instance, ambient temperature fluctuations or material additives can alter optical signals in ways that may require retraining or more robust state representations. Similarly, although basing the reward on the mean OR signal proved effective in avoiding keyhole formation, alternative reward functions encompassing a broader range of weld quality metrics — such as porosity, tensile strength, or weld depth — could further refine control.

Moreover, it is important to note that our current observation space, while sufficient for the presented experiments, may face challenges in more complex scenarios. Multi-material welding, for example, might generate similar sensor readings yet demand distinct power adjustments — a phenomenon known as perceptual aliasing [30]. Additionally, at very high speeds, the strongest reflection may occur off-axis, rendering the coaxial signal maximization less effective. Future work could address such challenges

by integrating advanced sensing technologies — like Optical Emission Spectroscopy or off-axis measurements — and by employing domain randomization [31] or more diverse training data to accommodate environmental variability. In this way, the system would gain a richer understanding of the weld pool and be more robust to changes in both material and processing conditions.

Overall, the present results demonstrate how RL-driven laser welding can adaptively optimize process parameters in real time, capitalizing on FPGA-based low-latency inference to keep up with high-speed manufacturing demands. By refining the reward function and observation space, and by accounting for environmental factors, this approach has the potential to generalize to a wide range of laser applications — ultimately offering a scalable, automated solution to ensure consistently high-quality outputs.

VI. CONCLUSION

Laser material processing and in particular laser welding is a critical technology in industries such as automotive manufacturing, aerospace engineering, and electronics assembly, prized for its precision, speed, and ability to produce high-quality welds. However, ensuring consistent weld quality in varying material properties remains a significant challenge, often requiring painstaking manual optimization.

In this paper, we have proposed a novel control strategy that addresses these limitations. Our approach employs a reinforcement learning algorithm implemented on an FPGA, achieving extremely low latency. This rapid processing allows the system to send control actions to the laser power modulation almost instantaneously, even before receiving the next signal data-point, ensuring real-time responsiveness. Furthermore, our method operates effectively without requiring extensive tuning of the reward function, making it highly adaptable to different welding scenarios.

Our experimental results demonstrated significant improvements over static power strategies in various surface conditions, including brushed, sandblasted, and mixed steel samples. The system showed remarkable adaptability, learning to outperform optimal constant power strategies by up to 22.58% on sandblasted surfaces and 7.30% on mixed surfaces in terms of the reward function. Although the reward serves as a useful proxy, we recognize that the ultimate quality of the weld is determined by physical metrics. To this end, we conducted post-process analyses of the welded samples, including examining cross-sectional images. These cross sections revealed that our learned policy consistently produced larger and deeper melt pools compared to the constant power strategy, indicating improved weld penetration and overall quality.

In conclusion, our real-time, adaptive laser welding control system represents a significant step forward in automating and optimizing laser welding processes, and its generality makes it able to target many other laser applications with potential benefits across a wide range of industries.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to Thomas Rytz, Toni Ivas, Peter Ramseier, and Pinku Yadav for their invaluable support and contributions to this research. We also extend our thanks to the Strategic Focus Area (SFA) Advanced Manufacturing program for financing this project.

REFERENCES

- [1] M. Spöttl and H. Mohrbacher, "Laser-based manufacturing concepts for efficient production of tailor welded sheet metals," *Advances in Manufacturing*, vol. 2, no. 3, pp. 193–202, sep 2014. [Online]. Available: <https://link.springer.com/article/10.1007/s40436-014-0088-8>
- [2] I. Serrano-Munoz, J.-Y. Buffiere, R. Mokso, C. Verdu, and Y. Nadot, "Laser powder bed fusion of 316l stainless steel: Industrial relevance and process optimization," *Advanced Engineering Materials*, vol. 22, no. 3, p. 1900617, 2020.
- [3] S. Liu, Y. Chen, H. Guo, X. Tian, H. Lu, J. Zhao, and G. Zhang, "A review on laser processing in electronic and mems packaging," *Journal of Electronic Packaging*, vol. 139, no. 3, p. 030801, 9 2017. [Online]. Available: <https://doi.org/10.1115/1.4036605>
- [4] W. M. Steen and J. Mazumder, *Laser Material Processing*, 4th ed. Springer London, 2010. [Online]. Available: <https://doi.org/10.1007/978-1-84996-062-5>
- [5] M. Schmidt, A. Otto, A. Grimm, and C. Kägeler, "Fault-tolerant laser welding in automotive industries," *ICALEO 2006 - 25th International Congress on Applications of Laser and Electro-Optics, Congress Proceedings*, oct 2006. [Online]. Available: <https://lia.liacp/article/doi/10.2351/1.5060833/834817/Fault-tolerant-laser-welding-in-automotive>
- [6] S. Gu, E. Holly, T. Lillicrap, and S. Levine, "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 3389–3396.
- [7] J. Degraeve, F. Felici, J. Buchli, M. Neunert, B. Tracey, F. Carpanese, T. Ewalds, R. Hafner, A. Abdolmaleki, D. de Las Casas, et al., "Magnetic control of tokamak plasmas through deep reinforcement learning," *Nature*, vol. 602, no. 7897, pp. 414–419, 2022.
- [8] G. Masinelli, T. Le-Quang, S. Zanoli, K. Wasmer, and S. A. Shevchik, "Adaptive laser welding control: A reinforcement learning approach," *Ieee Access*, vol. 8, pp. 103 803–103 814, 2020.
- [9] T. Kaneko, G. Minamoto, Y. Hirose, and T. Sakai, "Reinforcement learning for laser welding speed control minimizing bead width error," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 275–12 281.
- [10] J. Günther, P. M. Pilarski, G. Helfrich, H. Shen, and K. Diepold, "Intelligent laser welding through representation, prediction, and control learning: An architecture with deep neural networks and reinforcement learning," *Mechatronics*, vol. 34, pp. 1–11, 2016.
- [11] K. Taherkhani, E. Sheydaeiyan, C. Eischer, M. Otto, and E. Toyserkani, "Development of a defect-detection platform using photodiode signals collected from the melt pool of laser powder-bed fusion," *Additive Manufacturing*, vol. 46, p. 102152, 10 2021.
- [12] M. A. Fajobi, R. T. Loto, and O. O. Oluwole, "Steel corrosion behaviour in acidic solution for application in petrochemical distillation systems," *IOP Conference Series: Materials Science and Engineering*, vol. 811, no. 1, p. 012030, apr 2020. [Online]. Available: <https://dx.doi.org/10.1088/1757-899X/811/1/012030>
- [13] J. A. Barish and J. M. Goddard, "Anti-fouling surface modified stainless steel for food processing," *Food and Bioproducts Processing*, vol. 91, no. 4, pp. 352–361, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0960308513000047>
- [14] M. Javidi, S. Javadpour, M. Bahrololoom, and J. Ma, "Electrophoretic deposition of natural hydroxyapatite on medical grade 316l stainless steel," *Materials Science and Engineering: C*, vol. 28, no. 8, pp. 1509–1515, 2008. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0928493108000799>
- [15] Y. Rong, Y. Huang, G. Zhang, G. Mi, and W. Shao, "Laser beam welding of 316l t-joint: microstructure, microhardness, distortion, and residual stress," *The International Journal of Advanced Manufacturing Technology*, vol. 90, no. 5, pp. 2263–2270, May 2017. [Online]. Available: <https://doi.org/10.1007/s00170-016-9501-8>
- [16] A. Pappalardo, "Xilinx/brevitas," 2023. [Online]. Available: <https://doi.org/10.5281/zenodo.3333552>
- [17] R. S. Sutton, "Reinforcement learning: An introduction," *A Bradford Book*, 2018.
- [18] A. Gupta, A. Pacchiano, Y. Zhai, S. Kakade, and S. Levine, "Unpacking reward shaping: Understanding the benefits of reward engineering on sample complexity," *Advances in Neural Information Processing Systems*, vol. 35, pp. 15 281–15 295, 2022.
- [19] M. Wittemer, J. Grünewald, and K. Wudy, "Absorbance measurement for in situ process regime identification in laser processing," *The International Journal of Advanced Manufacturing Technology*, vol. 126, no. 1, pp. 103–115, May 2023. [Online]. Available: <https://doi.org/10.1007/s00170-023-11041-9>
- [20] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*. PMLR, 2018, pp. 1861–1870.
- [21] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, et al., "Soft actor-critic algorithms and applications," *arXiv preprint arXiv:1812.05905*, 2018.
- [22] B. Eysenbach and S. Levine, "Maximum entropy rl (provably) solves some robust rl problems," *arXiv preprint arXiv:2103.06257*, 2021.
- [23] Xilinx, *Introduction to FPGA Design with Vivado High-Level Synthesis*, Xilinx Inc., 2019, uG998 (v1.2). [Online]. Available: <https://docs.xilinx.com/v/u/en-US/ug998-vivado-intro-fpga-design-hls>
- [24] D. P. Kingma, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [25] Y. Huang, T. G. Fleming, S. J. Clark, S. Marussi, K. Fezzaa, J. Thiyyagalingam, C. L. A. Leung, and P. D. Lee, "Keyhole fluctuation and pore formation mechanisms during laser powder bed fusion additive manufacturing," *Nature Communications*, vol. 13, no. 1, p. 1170, 2022. [Online]. Available: <https://doi.org/10.1038/s41467-022-28694-x>
- [26] B. Fotovati, S. F. Wayne, G. Lewis, and E. Asadi, "A review on melt-pool characteristics in laser welding of metals," *Advances in Materials Science and Engineering*, vol. 2018, no. 1, p. 4920718, 2018. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2018/4920718>
- [27] W. Suder, X. Chen, D. R. Sierra, G. Chen, J. Wainwright, K. Rajamudili, G. R. Pardal, and S. Williams, "Control of meltpool shape in laser welding," *Welding in the World*, vol. 68, no. 6, pp. 1485–1495, June 2024. [Online]. Available: <https://doi.org/10.1007/s40194-024-01719-3>
- [28] K. Barat, C. N. Kumar, G. Patil, K. Panbarasu, and K. Venkateswarlu, "Defectometry, distortion analysis and metallurgical properties in the keyhole and conduction mode of laser beam welding of al-mg-sc alloy butt joints," *Optics & Laser Technology*, vol. 162, p. 109248, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S003039922300141X>
- [29] A. E. Odermatt, V. Ventzke, F. Dorn, R. Dinsé, P. Merhof, and N. Kashaev, "Effect of laser beam welding on microstructure, tensile strength and fatigue behaviour of duplex stainless steel 2205," *Journal of Manufacturing Processes*, vol. 72, pp. 148–158, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1526612521007428>
- [30] L. Chrisman, "Reinforcement learning with perceptual aliasing: The perceptual distinctions approach," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 1992. [Online]. Available: <https://api.semanticscholar.org/CorpusID:1963904>
- [31] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 23–30.