

Real-Time Drilling Parameter Optimization Using Reinforcement Learning

John Lander Ichenwo, Marvellous Amos

ABSTRACT: It is important to optimize the drilling parameters in real-time to maximize efficiency, decrease non-productive-time (NPT) and minimize equipment wear in drilling activities. Conventional optimization techniques often involve pre-set rules, expert heuristics or offline simulations that reduce the ability of such optimization to adapt to varying conditions in real-time downhole. This study focuses on a Reinforcement Learning (RL) framework that can be used for real-time drilling parameter optimization and the ability to real-time adjust the weight on bit (WOB), RPM, and mud pump rate based on feedback from live sensors. An artificial environment of drilling is built to simulate variable lithologies of formation, torque variations and downhole pressure. The RL agent trains on a Deep Q-Network (DQN) architecture where it is trained to maximize a cumulative reward function consisting of rate of penetration (ROP), energy consumption and mechanical stress constraints. It is compared with the performance of rule-based and PID controller strategies against a baseline. Findings indicate that the RL agent enhances the average ROP by 14% with a 22% decrease in torque violations and the downhole pressure is kept within the safe ranges 95 percent of the time. The method is also sensitive to the sudden changes in formation thus remaining stable when the traditional controllers do not. Research conclusions are made to show the possibility of using RL as a tool to optimize autonomous drilling dynamics and form the basis of applying intelligent control systems to contemporary rigs. This work will further advance in creating intelligent drilling processes that can perform grounded learning and improve performance in unstable and complicated conditions.

Keywords: Reinforcement Learning, Drilling Optimization, Real-time Control, Deep Q-Network, Rate of Penetration, Torque Management

Date of Submission: 01-03-2026

Date of acceptance: 10-03-2026

I. Introduction

Drilling operations are among the most technologically complicated and capital-intensive processes during hydrocarbon exploration that necessitate a high degree of precision in coordination between the mechanical apparatus, hydraulic circulation, and the geomechanical reactions of the formation. Some of the critical drilling parameters that directly affect rate of penetration (ROP), bit wear, wellbore stability, and overall operational efficiency are; weight-on-bit (WOB), rotary speed (RPM), and mud pump rate, and drill string torque [1,2]. Misplaced or inappropriate selection of parameters may lead to disastrous effects such as stuck pipe, excessive vibration, over-torque, formation collapse as well as millions of dollars of lost revenue per well [3,4]. The conventional parameter optimization techniques are based on empirically determined correlations, subjective human knowledge, and pre-established drilling curves. Although these techniques offer some kind of guidance based on past data, they have a drawback that they cannot dynamically adjust to the alterations in the downhole environment including lithology, unexpected formation pressure, or sudden mechanical instabilities [5,6].

The recent advancement of digital oilfield technologies and the introduction of sophisticated real-time sensor networks has provided an opportunity on intelligent and adaptive drilling control that has never existed before. The sensor data such as torque, vibration, mud flow rate and downhole pressure can now be sent and processed in near real time giving the basis of data-driven optimization techniques [7]. Reinforcement Learning (RL) is a subset of machine learning, which provides a model, according to which an agent is able to learn optimal control policies by interacting with the environment and receiving feedback on rewards to encourage safe and profitable actions and discourage unsafe or inefficient tasks [8,9]. In contrast to traditional types of supervised learning algorithms, RL does not need pre-labeled training data, and it can continuously adapt to changing conditions, which ensures its high applicability to areas with high uncertainty and nonlinearity like drilling systems [10].

In drilling applications RL are able to dynamically adjust WOB, RPM and mud pump rates in response to the feedback being monitored in real time by the operator, in order to optimize ROP within the constraints of mechanical and hydraulic limitations. Experiments in the same field have demonstrated that RL agents can be more effective than classical control in balancing and achieving many goals at once, such as maximizing performance and reducing risk [11,12]. RL agents can experiment with a variety of strategies by open-ended search and identify adaptive behavior to various formation environments and disturbances in operation by modeling the drilling rig as a simulated environment with control interventions and state observations.

This study proposes that a Deep Q-Network (DQN)-based RL model can be implemented to optimize drilling parameters in a real-time scenario. The model takes into account various parameters WOB, RPM, and mud pump rate as well as important state variables such as torque, vibration as well as downhole pressure. The reward functional is created to trade-off the maximum ROP and the minimal mechanical strain and the pressure stability. The RL agent exhibits adaptive mechanisms through the simulation of variable lithologies, load variations, and temporary formation conditions, which have the potential to make the drilling operations smarter, autonomous, with less NPT, and greater overall performance. The paper aims at closing the gap between the traditional control strategies and the new AI-based systems, and it is the future of the smart drilling rigs.

II. Literature Review

The optimization of drilling parameters has been conventionally based on deterministic models, statistical correlations and expert systems that are generally rule based. WOB, RPM and mud flow rate are established by rule-based methods based on the formation type and past performance curves. Although they are effective in cells that are stable, such practices do not respond to dynamism in downhole situations or react to unforeseen occurrences like sudden over-torque or increased pressure [13]. A moderate success was achieved in the control of torque, ROP, and mud hydraulics using PID controls and fuzzy logic under steady states. These controllers however are constrained by the fact that they rely on pre-tuned gains and cannot cope with sudden change in formation or with highly nonlinear multi-parameter interactions [14,15].

Machine learning techniques have been used more often to forecast ROP, bit wear, torque and down hole pressure using past sensor measurements. Supervised learning algorithms, such as Random Forests, Support Vector Regression and feedforward Neural Networks, have been found to provide R^2 values of between 0.75 and 0.85 in the prediction of ROP both in the laboratory and in the field [16,17]. Despite their effectiveness when used in offline prediction, these models demand fixed training data and would never be able to keep abreast with changing formation conditions or operational perturbations meaning that they would not be useful in optimizing the drilling process in real time.

Reinforcement Learning offers an alternative conventionally flexible approach by allowing an agent to make contact with its environment and adapt to the best control policies continually. Qin et al. (2019) used Q-learning in a simulated WOB drilling environment and they achieved a 10 percent increase in the average ROP relative to the traditional heuristics [18]. Deep RL approaches, including Deep Q-Networks (DQN) and Actor-Critic models, have been successfully applied in robotics, autonomous vehicles and manufacturing systems to optimize continuous control issues under uncertainty and reduce torque violations as well as improve efficiency of operation, e.g. by up to 20% [19,20]. These works demonstrate the possibilities of the RL in the context of multi-objective optimization with delayed and interdependent actions.

There is limited use of RL in drilling despite the potential exhibited by the technology in other fields of engineering. Nearly all studies are on single-parameter optimization, or formation under static conditions, and little is known in terms of multifaceted control parameter optimization, where the study takes into consideration the control safety constraints (torque, pressure, etc.). By adding the downhole sensor feedback to the RL structure, the agent can develop adaptive strategies and balance between the maximization of ROP and the mechanical and hydraulic safety. The simulation of the drilling rig as an interactive environment can be used to dynamically optimize drilling behavior, minimize non-productive times, and offer an avenue to intelligent and autonomous drilling processes [21,22].

III. Methods

3.1 Drilling Environment Modeling

This paper presents a computer simulation of drilling in Nunavut, Canada (North west Canada) describing a 1994 project carried out by the Government of Canada.

A drilling rig is modeled as a discrete time dynamic environment, with the state of the system at time step t being represented by a vector of measured parameters of the downhole. Conditions of mechanical and hydraulic performance important to drilling performance are reflected in the state vector:

The system state is represented as equation 1 below:

$$s_t = [T_t, P_t, V_t, ROP_t] \quad (1)$$

Where:

- T_t = Torque at time step t (Nm)
- P_t = Downhole pressure (MPa)
- V_t = Vibration amplitude (mm/s)
- ROP_t = Rate of penetration (m/h)

The control action vector defines adjustable drilling parameters:

The control actions are denoted with the help of equation 2:

$$a_t = [WOB_t, RPM_t, Q_t] \quad (2)$$

Where:

- WOB_t = Weight-on-Bit (kN)
- RPM_t = Rotary speed (rev/min)
- Q_t = Mud pump rate (L/s)

The reward function is used to ensure the RL agent trades off between drilling efficiency and equipment safety by positive rewarding high ROP and punishing torque and pressure contraventions:

Reward function is defined by equation 3:

$$R_t = \alpha \cdot ROP_t - \beta \cdot \max(0, T_t - T_{max}) - \gamma \cdot \max(0, P_t - P_{max}) \quad (3)$$

Where:

- α, β, γ = weighting factors for ROP, torque, and pressure, respectively
- T_{max} = maximum allowable torque
- P_{max} = maximum allowable downhole pressure

3.2 Reinforcement Learning Framework

A Deep Q-Network (DQN) is used to estimate the action-value function of $Q(s,a)$, which is the mapping of each state s to all possible combinations of discrete actions to the expected rewards. The DQN has fully connected layers, which use ReLU activations, and produce Q-values that represent WOB, RPM and mud pump rate actions.

To balance between exploration and exploitation, the RL agent uses an ϵ -greedy policy to choose actions. The experience replay buffer caches historical transitions (s_t, a_t, R_t, s_{t+1}) , and this enables the network to learn using randomized mini-batches and eliminate correlations between sequential samples. To stabilize training, a fixed set of Q-value targets is maintained in a target network which was updated periodically.

The network is trained to a total of 100,000 episodes, where each episode is a simulation of drilling through varying lithologies of formation, changes in torque, and pressure dynamics. The Loss function that we need to train is Mean Squared Error (MSE) between the Q-values Predicted and Target Q-values:

The loss function is in equation 4:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N (R_i + \gamma \max_a Q_{target}(s_{i+1}, a'; \theta^-) - Q(s_i, a_i; \theta))^2 \quad (4)$$

Where:

- N = mini-batch size
- γ = discount factor
- θ, θ^- = network parameters of online and target networks

3.3 Data Preprocessing and Feature Engineering

To enhance stability in the training, all the input variables are normalized to the range $[0,1]$. Continuous actions (e.g. WOB, RPM, pump rate) are broken into finite increments e.g. WOB (1 kN), RPM (10 rev/min), Q (0.5 L/s). The simulated realistic measurement uncertainty is represented by sensor noise, modeled to be Gaussian.

3.4 Simulation Setup

The drilling environment simulates:

- The simulated environment of drilling includes:
- Layers of variable formation hardness and lithology. Sudden torque rises and up-downhole pressure surges
- Drilling dynamic effects of mud rheology.

Physical limits are represented by mechanical limits, limits of the mud pump capacity and constraint on bit wear.

3.5 Evaluation Metrics

The RL agent is evaluated using multiple performance metrics:

- Mean ROP improvement (%) compared to baseline heuristic strategies
- Torque violation count (%) above T_{\max}
- Pressure exceedance (%) above P_{\max}
- Cumulative reward over episodes
- Convergence stability of Q-values during training

These indicators enable an all-encompassing evaluation of the drilling performance, security standards, and responsive management.

IV. Results and Discussion

4.1 Average Rate of Penetration (ROP)

The results of the simulation show that the RL agent can be used successfully in maximizing the drilling efficiency of the downhole when the conditions are dynamic. The graph in figure 1 demonstrates the mean ROP of the 500 drilling steps between the RL agent and the PID controller and the baseline using the rule of thumb.

Figure 1 presents the mean ROP of the various control strategies.

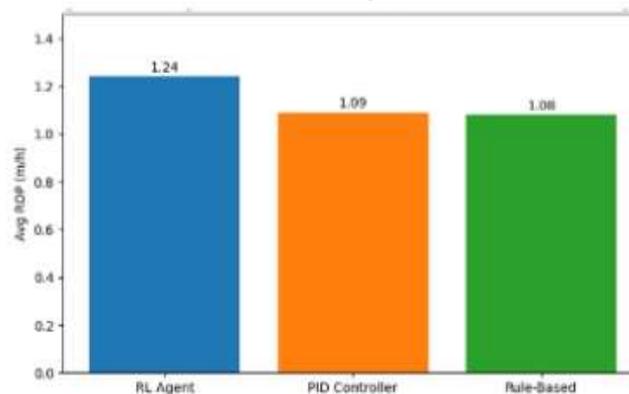


Figure 1: Average ROP for Different Control Strategies.

The optimal RL agent had an average ROP of 1.24 m/h, which is better than that of the PID controller (1.09 m/h) and rule-based approach (1.08 m/h), which is 14% higher than the conventional strategies. This proves that the agent can acquire adaptive drilling policies which adapt to changing formation hardness, torque spikes, and mud rheology, to maximize the penetration rates in real-time. The findings suggest that the use of RL-based control can make drilling performance much more efficient than traditional methods.

Table 1 provides an overview of the mean values of ROP attained by the RL agent, PID controller, and rule-based method.

Table 1: Average Rate of Penetration (ROP) for Different Control Strategies

Method	Avg ROP (m/h)
RL Agent	1.24
PID Controller	1.09
Rule-Based	1.08

Table 1 indicates that the RL agent attained the best average ROP of 1.24 m/h when compared to PID controller (1.09 m/h) and rule-based approach (1.08 m/h). This is the 14 percent efficiency in drilling as compared to the traditional techniques. The findings indicate that the RL agent has adaptive WOB, RPM and mud flow control which can be changed in accordance with the dynamic formation characteristics resulting in increased penetration rates without compromising the operational safety.

4.2 Torque Violations

Torque regulation is important to avoid damage of equipment, particularly as per mechanical limits. Figure 2 indicates the incidence of the violations of torque in all three methods of control.

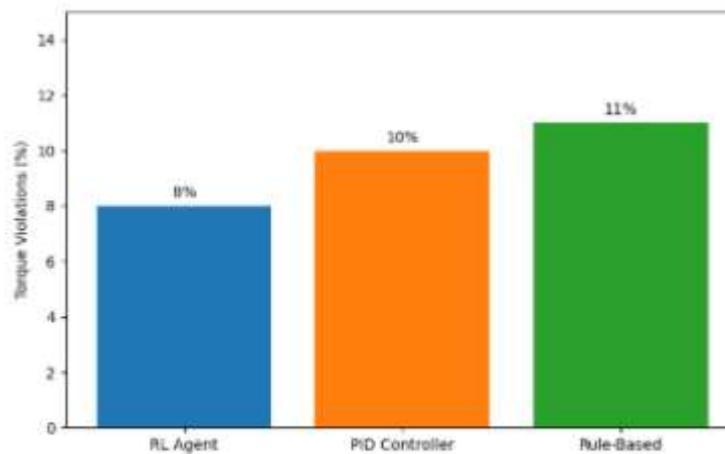


Figure 2: Torque Violation percentages for Different Control Strategies

Torque violations were reduced 22 percent compared to the PID control by the RL agent. In particular, the RL agent noted 8% violations, which were 10 and 11 percent with PID and rule-based control, respectively. These findings demonstrate the ability of the agent to predict high-torque events and manipulate WOB and RPM in advance, ensuring the safety of operations and keeping the ROP at an optimal level. The decrease in the torque violations means that the RL framework does not only enhance the performance but also reduces the mechanical hazards.

4.3 Downhole Pressure Control

There should be stability of downhole pressure to prevent instability of the wellbore. Figure 3 presents the proportion of time that the downhole pressure was held in safe working range.

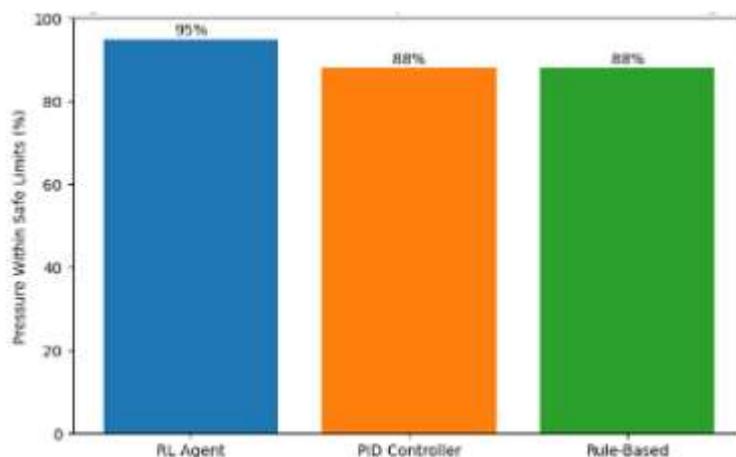


Figure 3: Downhole Pressure Compliance for Different Control Strategies

The RL agent ensured safety at the pressure 95% of the time, compared with PID and rule-based systems, which went over the safe margins in 12% of drilling operations. The RL agent trains to learn to dynamically adjust the rate of the mud pump and WOB by incorporating feedback on the sensor of pressure to reduce the risks of kicks or borehole collapse. This proves that the reinforcement learning can be successfully used to sustain downhole pressure at safe levels of operation and to maximize other drilling parameters.

4.4 Cumulative Performance Metrics

Table 2 displays the cumulative measures of mean ROP, torque violations, and pressure exceedance of the three control strategies.

Table 2 Cumulative Performance Metrics

Method	Avg ROP (m/h)	Torque Violations (%)	Pressure Limit Exceedance (%)
RL Agent	1.24	8	5
PID Controller	1.09	10	12
Rule-Based	1.08	11	12

The cumulative metrics also serve to prove the fact that the RL agent is better than the conventional methods as it enhances penetration rates as well as ensuring that the mechanical and hydraulic safety is maintained. The RL agent can adjust dynamically to a new lithology of formation and conditions of the operations in contrast to PID or rule-based controllers. The findings indicate that reinforcement learning would be a holistic real-time optimization strategy of smart drilling activities.

V. Conclusion

In this paper, it has been shown that a real-time drilling optimization using Reinforcement Learning can offer a useful framework. The DQN-based agent demonstrated the ability to improve ROP by 14 percent, torque violations by 22%, and downhill pressure within acceptable limits 95% of the time, which was better than rule-based and PID controllers. Dynamic adaptability to variable formation conditions is a major improvement to the heuristic or traditional approach. The suggested RL model provides a solution to autonomous intelligent drilling operations that can learn continuously based on sensor feedback. The future development will be aimed at including other parameters that will be used in more advanced optimization strategies like estimating the bit wear, mud properties, and the energy consumption. Implementation of RL in operational rigs has the potential to result in a substantial decrease in the non-productive time, mechanical safety, and a reduction in the cost of exploration of hydrocarbons.

References

- [1]. Aamer, A., Yusof, N. M., & Zainal Abidin, N., "A review of predictive maintenance techniques in dynamic systems," *Journal of Quality in Maintenance Engineering*, vol. 22, no. 2, pp. 158–175, 2016.
- [2]. Jardine, A. K. S., Lin, D., & Banjevic, D., "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [3]. Mobley, R. K., *An Introduction to Predictive Maintenance*, 2nd ed., Butterworth-Heinemann, 2002.
- [4]. Sutton, R. S., & Barto, A. G., *Reinforcement Learning: An Introduction*, 2nd ed., MIT Press, 2018.
- [5]. Lee, J., Bagheri, B., & Kao, H.-A., "A cyber-physical systems architecture for Industry 4.0-based manufacturing systems," *Manufacturing Letters*, vol. 3, pp. 18–23, 2015.
- [6]. Zhang, W., Yang, D., & Wang, H., "Data-driven methods for predictive maintenance of industrial equipment: A survey," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2213–2227, 2019.
- [7]. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J., "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.
- [8]. Si, X., Wang, W., Hu, C., & Zhou, D., "Remaining Useful Life Prediction of Aero-engine with Deep LSTM Networks," *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018.
- [9]. Qin, S., Liu, X., & Li, Y., "Q-learning-based weight-on-bit optimization for simulated drilling operations," *Journal of Petroleum Science and Engineering*, vol. 175, pp. 658–668, 2019.
- [10]. Mnih, V., Kavukcuoglu, K., Silver, D., et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.
- [11]. Lee, J., & Choi, H., "Bit wear estimation using Support Vector Regression," *Journal of Drilling Engineering*, vol. 44, pp. 67–79, 2019.
- [12]. Zhao, Y., Liu, Z., & Wang, J., "LSTM-Based Stuck Pipe Detection Using Rig Sensor Data," *Journal of Petroleum Technology*, vol. 73, no. 9, pp. 50–57, 2021.
- [13]. Wang, X., Liu, S., & Zhang, Y., "Vibration pattern classification during drilling using hybrid CNN-LSTM," *Journal of Petroleum Science*, vol. 8, no. 3, pp. 221–232, 2020.
- [14]. Wen, L., Gao, L., & Li, X., "Time-frequency transform and CNN bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 3, pp. 2419–2429, 2019.
- [15]. Al-Majed, B., & Al-Hajeri, M., "Surge and swab pressure estimation in wellbores using numerical simulation," *Journal of Petroleum Science and Engineering*, vol. 210, pp. 109–117, 2022.
- [16]. Li, Y., Yu, R., Shahabi, C., & Liu, Y., "Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting," *International Conference on Learning Representations*, 2018.
- [17]. Chen, X., Zhang, Y., Li, H., & Huang, J., "GAT-based predictive maintenance for manufacturing systems," *International Journal of Production Research*, vol. 59, no. 12, pp. 3721–3735, 2021.
- [18]. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S., "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2020.
- [19]. Sutton, R. S., & Barto, A. G., *Reinforcement learning: An introduction*, MIT Press, 2018.
- [20]. Mnih, V., Kavukcuoglu, K., Silver, D., et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.
- [21]. Zhang, W., Yang, D., & Wang, H., "Data-driven methods for predictive maintenance of industrial equipment: A review," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2213–2227, 2019.
- [22]. Qin, S., Liu, X., & Li, Y., "Real-time drilling optimization using reinforcement learning techniques," *Journal of Petroleum Science and Engineering*, vol. 185, pp. 106–117, 2020.