Multi-objective Reinforcement Learning Controller for Multi-Generator Industrial Wave Energy Converter

Soumyendu Sarkar [1]* **Vineet Gundecha**[1] **Alexander Shmakov**[1] **Sahand Ghorbanpour**[1]
Ashwin Ramesh Babu[1] **Paolo Faraboschi**[1] **Mathieu Cocho**[2] **Alexander Pichard**[2]
Jonathan Fievez[2]
[1] Hewlett Packard Labs @ Hewlett Packard Enterprise
[2] Carnegie Clean Energy

Abstract

Waves are one of the greatest sources of renewable energy and are a promising resource to tackle climate challenges by decarbonizing energy generation. Lowering the Levelized Cost of Energy (LCOE) for wave energy converters is key to competitiveness with other forms of clean energy like wind and solar. Also, the complexity of control has gone up significantly with the state-of-the-art multi-generator multi-legged industrial Wave Energy Converters (WEC). This paper introduces a Multi-Agent Reinforcement Learning controller (MARL) architecture that can handle these multiple objectives for LCOE, helping the increase in energy capture efficiency, boosting revenue, reducing structural stress to limit maintenance and operating cost, and adaptively and proactively protect the wave energy converter from catastrophic weather events, preserving investments and lowering effective capital cost. We use a MARL implementing proximal policy optimization (PPO) with various optimizations to help sustain the training convergence in the complex hyperplane. The MARL is able to better control the reactive forces of the generators on multiple tethers (legs) of WEC than the commonly deployed spring damper controller. The design for trust is implemented to assure the operation of WEC within a safe zone of mechanical compliance and guarantee mechanical integrity. This is achieved through reward shaping for multiple objectives of energy capture and penalty for harmful motions to minimize stress and lower the cost of maintenance. We achieved double-digit gains in energy capture efficiency across the waves of different principal frequencies over the baseline Spring Damper controller with the proposed MARL controllers.

1 Introduction and Related Work

Waves in the ocean are one of the more consistent and predictable sources of renewable energy, and the exploitable resource of coastal wave energy has been estimated to be over 2 TW, representing about 16% of the world energy consumption (Yusop et al. 2020). Some significant challenges of deploying Wave Energy Converters(WEC) include variability of the wave time period, height, and directionality in offshore locations, leading to the complexity of capturing energy. Also, WEC must be operated to minimize maintenance cost, and withstand rare but extreme wave conditions.

1.1 Wave Energy Controller (WEC)

The industrial WEC considered in this study is composed of a submerged cylindrical Buoyant Actuator (BA) similar in structure as in Figure 1. The BA is secured to the seabed with three mooring legs, each of which terminates on one of the three power take-offs (PTOs) located within the BA.

* Soumyendu Sarkar is the corresponding author {Soumyendu.Sarkar@hpe.com}

Tackling Climate Change with Machine Learning workshop at NeurIPS 2021

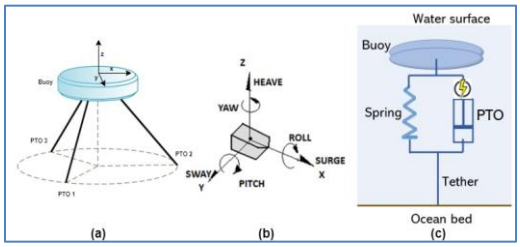


Figure 1: Geometry of the three-tether wave energy converter: (a) 3D view, (b) PTO and motion with 6 degrees of freedom, (c) WEC. (Sergiienko et al. 2020)

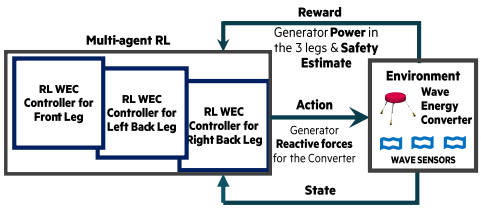


Figure 2: Multi-agent RL with 3 agents controlling the reactive forces of the generators on the 3 legs

The electric generator of the PTO resists the extension of the mooring legs applying varying reactive force controlled by the RL controller, thereby generating electrical power. RL controllers need to optimize the timing and value of the PTO forces in relation to the wave excitation force, which is key to maximizing WEC energy capture and conversion efficiency. The different controllers currently deployed are damping control, spring damper control, latching control, and model predictive control with various degrees of success but fail to leverage multi-generator WECs well.

1.2 Related work

There has been recent work on applying RL to control simple one-legged WECs in different academic settings. (Anderlini et al., 2016, 2017, 2018, 2020) uses RL to obtain optimal reactive force for a two-body heaving point absorber with one degree of freedom. To our knowledge, **RL has not been used to control advanced industrial multi-legged and multi- generator WECs** with six degrees of freedom of motion, where the complexity and impact are even greater.

2 Reinforcement learning design for WEC Multi-generator control

The heterogeneity and complexity of WEC require a versatile controller like Multi-Agent Reinforcement Learning (MARL). The three legs and the generators mounted for each of the legs act differently, as they tend to generate different amounts of energy based on the orientation of the mechanical structure and wave directionality. Simpler one agent RL with multiple actions failed to control the WEC effectively. Hence, separate agents of MARL as shown in Figure 2, were used to control the reactive force of the generators on each of the three legs to learn the policy better.

2.1 Environment state, action, and reward design

For training, the state information is provided as a vector represented by s, where "e" represents the buoy position, "g" represents the tether extension, and "z" represents wave excitation. All RL agents share the continuous observation space of position and wave.

$$s = [\ e\ \dot{e}\ \ddot{e}\ g\ \dot{g}\ z\ \dot{z}\ ]^T$$

The continuous **action space** for the individual RL agent is defined by the reactive force $f_{gen(i)}$ for the controlled generator, where "i" represents the index for the agent.
The **reward** is defined as,

$$Reward_i = \alpha.(P_{own(i)} + \eta_i.P_{others}) + (1-\alpha)yaw$$

Where P represents the generated power defined by $-f_{gen} * \dot{e}$. η is the hyperparameter for the team coefficient and α is the hyperparameter for yaw minimization of individual legs.

2.2 Refinements to PPO for training stability and convergence

For WEC, the Proximal Policy Optimization (PPO) for policy optimization performed better than other RL algorithms that we tried, like the DQN, Soft Actor-Critic, and A3C. We mitigated stalling convergence problem during training optimization with PPO design choices, data transformation, and tuning, a methodology that will also help tackle similar control problems for wind power. We used LSTMs to leverage the time-series nature of the states and partial visibility into the oncoming wave excitation from the wave sensors.

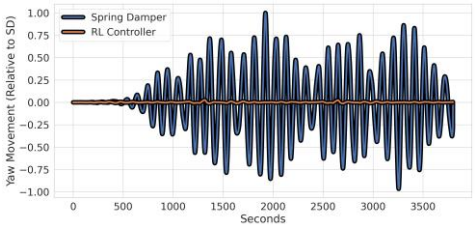


*Figure 3: Comparison of Yaw movement between RL and the spring damper controllers for an episode with **wave of height 2m** and principal wave period of 12s. Values are relative to maximum SD yaw.*

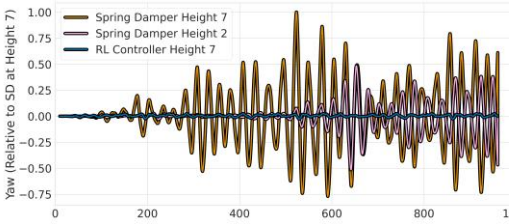


*Figure 4: A comparison of Yaw for the RL and spring damper controllers on an extreme **wave of height 7m** and period 12s. The SD yaw for a wave of height 2m is also included as a reference.*

2.3 Cooperation vs Competition

Though on the surface it looks like a cooperative MARL problem, the disparity in the power generated by individual legs and the complex nature of trade-off by one leg, to get additional power in other legs, makes the optimum solution a combination of co-operation and competition. We added a signed hyperparameter "η" of team coefficient to have an option for both positive and adversarial contribution of the power from other legs in the reward.

$$Reward = P_{own} + \eta \, . P_{others}$$

3 Design for Trust

3.1 Maintenance mitigation and Yaw minimization with RL

The rotational motion of the voluminous buoy (yaw) causes the tether connections to wear out faster and has potential maintenance implications. The yaw motion is most significant in extreme cases of angled waves of 30 degrees. The penalty for the yaw movement is accounted in the reward shaping with a hyperparameter α for the three individual agents:

$$Reward = (\alpha)\, power + (1 - \alpha)\, yaw$$

where α is a tunable yaw penalty hyperparameter, lesser the α, stronger the penalty. This led to significant improvements in yaw reduction resulting in much less displacement than the currently deployed spring damper (SD) controller, as seen in Figure 6. Also, adding the penalty for yaw to the reward improved power generation, as seen in Figure 7, likely because yaw minimization is simpler for RL to implement which resulted in more directed power in the PTO. This combined reward serves the dual purpose of energy capture maximization and stress minimization on the WEC to avoid costly maintenance in the open sea with submerged structures.

3.2 Assured ML and enforcing preferred zone of operation

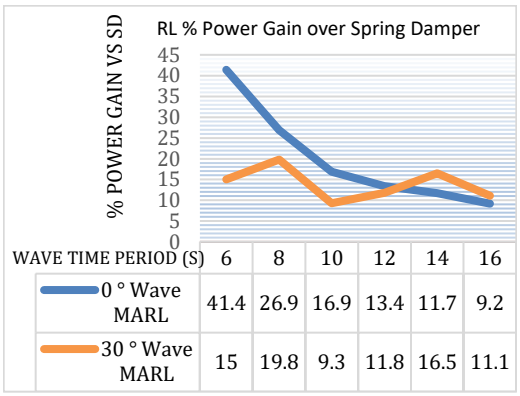
In addition to reward shaping, we performed clipping on the RL action of generator reactive force to adhere to maximum and minimum tension in the spring extensions and the maximum reactive forces on the generator ensuring the preferred zone of operation maintaining integrity.

3.3 WEC control with RL for survival condition

For extreme and dangerous conditions of 7m high waves at an angle of 30°, the high yaw motion is mitigated with yaw penalty coupled with the LSTM model of the policy and the critic which can track to minimize yaw with long episode horizons.

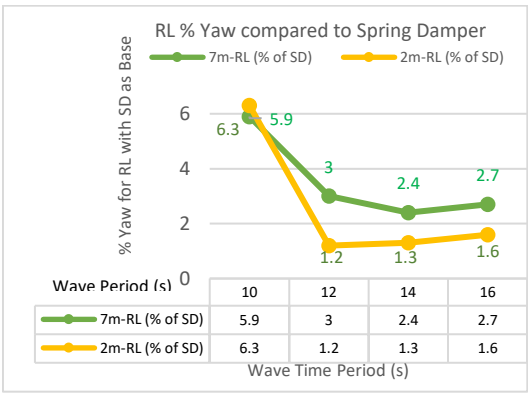
4 Results

As the waves follow a characteristic spectrum, while evaluating power generation performance for each wave frequency, multiple waves from the Jonswap spectrum have been sampled for episodes lasting several minutes each. All results are based on a simulator, which replicates the CETO 6



WAVE TIME PERIOD (S)	6	8	10	12	14	16
0 ° Wave MARL	41.4	26.9	16.9	13.4	11.7	9.2
30 ° Wave MARL	15	19.8	9.3	11.8	16.5	11.1

*Figure 5: **% Power gain for of RL** controller over Spring Damper for 2m high waves of different time periods*



Wave Period (s)	10	12	14	16
7m-RL (% of SD)	5.9	3	2.4	2.7
2m-RL (% of SD)	6.3	1.2	1.3	1.6

*Figure 6: **Yaw for RL as a % of Yaw for Spring Damper (SD)** for normal (2m median height) and survival condition (7m height) waves at 30°*

industrial wave energy converter. For regular operation, we show results of median wave height of 2m for the entire wave frequency spectrum spanning time periods of 6s to 16s.

The power generated by the baseline spring damper (SD) controller with resonant spring constant and damping constant is used as a reference for evaluation to estimate the gain of energy capture by RL controllers as a percentage improvement. A direction of 0° indicates frontal waves with the wave-front aligned with the front leg. For evaluation, we used the same seed for sampling waves for multiple episodes between RL and SD.

Table 2 shows a significant improvement in captured power with RL controller over baseline spring damper (SD) controller for the entire frequency spectrum of ocean waves. For frontal waves (0°), the MARL performs on an average of 19.9% better than the spring damper over the entire frequency range of the waves, while for 30° angled waves, MARL controller performs 13.9% better than SD on an average. This shows that MARL is versatile for non-frontal angled waves. The variation of gains by the RL controller with wave time periods is because the spring damper is more resonantly tuned to the mechanical structure of the WEC for a certain frequency band.

Table 3 shows that 3-agent MARL almost eliminated the yaw, which causes mechanical stress, while still making significant energy capture gains over baseline spring damper, as shown in Table

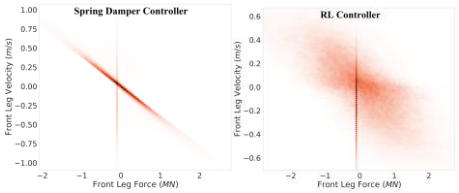


Figure 7: Generator Reactive force vs Velocity of tether: Front leg for frontal waves.

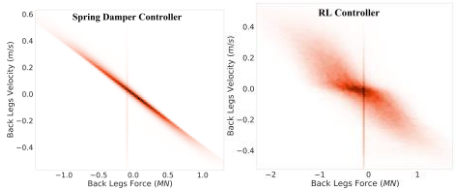


Figure 8: Generator Reactive force vs Velocity of tether: Back legs for frontal waves

2. Table 3 also shows that for natural disasters with surging waves of 7m height, the 3-agent MARL can almost eliminate yaw, just like it did for waves of normal height.

The intuition is that the reactive forces for the generator on the legs will be proportional to the velocity of the tether as energy is captured working against this motion. However, the RL controller is fuzzy about it, implying that it takes a more long-term view and compromises short-term objectives for greater gains on energy capture at the more opportune segments of the wave cycles.

5 Conclusions

The proposed MARL controller yielded double-digit gains over the entire spectrum of waves boosting revenue opportunities with higher energy production. At the same time, it helped reduce mechanical stress, which impacts maintenance and operating costs, and actively mitigated adverse effects of high waves characteristic of disaster events, helping to preserve capital investment. This MARL architecture with the mentioned objectives is applicable to other clean energy problems like

wind energy. The PPO refinements to stabilize training for global optima can be used in many other complex control applications with multiple entities to control.

6 References

Anderlini, E.; Forehand, D.; Bannon, E.; and Abusara, M. 2017a. Reactive control of a wave energy converter using artificial neural networks. International Journal of Marine Energy, 19: 207–220.

Anderlini, E.; Forehand, D.; Bannon, E.; Xiao, Q.; and Abusara, M. 2018. Reactive control of a two-body point absorber using reinforcement learning. Ocean Engineering, 148: 650–658.

Anderlini, E.; Forehand, D. I. M.; Bannon, E.; and Abusara, M. 2017b. Control of a Realistic Wave Energy Converter Model Using Least-Squares Policy Iteration. IEEE Transactions on Sustainable Energy, 8(4): 1618–1628

Anderlini, E.; Forehand, D. I. M.; Stansell, P.; Xiao, Q.; and Abusara, M. 2016. Control of a Point Absorber Using Reinforcement Learning. IEEE Transactions on Sustainable Energy, 7(4): 1681–1690.

Anderlini, E.; Husain, S.; Parker, G. G.; Abusara, M.; and Thomas, G. 2020. Towards Real-Time Reinforcement Learning Control of a Wave Energy Converter. Journal of Marine Science and Engineering, 8(11).

Bouville, M. 2008. Crime and punishment in scientific research. arXiv:0803.4058.

Clancey, W. J. 1979. Transfer of Rule-Based Expertise through a Tutorial Dialogue. Ph.D. diss., Dept. of Computer Science, Stanford Univ., Stanford, Calif.

Clancey, W. J. 1983. Communication, Simulation, and Intelligent Agents: Implications of Personal Intelligent Machines for Medical Education. In Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI83), 556–560. Menlo Park, Calif: IJCAI Organization.

Clancey, W. J. 1984. Classification Problem Solving. In Proceedings of the Fourth National Conference on Artificial Intelligence, 45–54. Menlo Park, Calif.: AAAI Press.

Clancey, W. J. 2021. The Engineering of Qualitative Models. Forthcoming.

Duan, Y.; Chen, X.; Houthooft, R.; Schulman, J.; and Abbeel, P. 2016. Benchmarking Deep Reinforcement Learning for Continuous Control. arXiv:1604.06778.

Engelmore, R.; and Morgan, A., eds. 1986. Blackboard Systems. Reading, Mass.: Addison-Wesley.

Hasling, D. W.; Clancey, W. J.; and Rennels, G. 1984. Strategic explanations for a diagnostic consultation system. International Journal of Man-Machine Studies, 20(1): 3–19. Lillicrap, T. P.; Hunt, J. J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; and Wierstra, D. 2019. Continuous control with deep reinforcement learning. arXiv:1509.02971.

NASA. 2015. Pluto: The' Other' Red Planet. https://www. nasa.gov/nh/pluto-the-other-red-planet. Accessed: 2018- 12-06.

Rice, J. 1986. Poligon: A System for Parallel Problem Solving. Technical Report KSL-86-19, Dept. of Computer Science, Stanford Univ.

Robinson, A. L. 1980. New Ways to Make Microcircuits Smaller. Science, 208(4447): 1019–1022.

Yusop, Z. M.; Ibrahim, M. Z.; Jusoh, M. A.; Albani, A.; and Rahman, S. J. A. 2020. Wave-Activated-Body Energy Converters Technologies: A Review. Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 76(1): 76–104.

Yu, C.; Velu, A.; Vinitsky, E.; Wang, Y.; Bayen, A.; and Wu,Y. 2021. The surprising effectiveness of mappo in cooperative, multi-agent games. arXiv preprint arXiv:2103.01955.
Schulman, J.; Wolski, F.; Dhaliwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms.arXiv:1707.06347.

Vinyals, O.; Ewalds, T.; Bartunov, S.; Georgiev, P.; Vezhn-evets, A. S.; Yeo, M.; Makhzani, A.; K¨uttler, H.; Agapiou, J.; Schrittwieser, J.; et al. 2017. Starcraft ii: A new challenge for reinforcement learning. arXiv preprint arXiv:1708.04782.
Brockman, G.; Cheung, V.; Pettersson, L.; Schneider, J.;Schulman, J.; Tang, J.; and Zaremba, W. 2016. Openai gym. arXiv preprint arXiv:1606.01540.
Tan, M. 1993. Multi-Agent Reinforcement Learning: Inde-pendent vs. Cooperative Agents. In Proceedings of the Tenth International Conference on Machine Learning, 330–337. Morgan Kaufmann.
Sergiienko, N. Y.; Neshat, M.; da Silva, L. S.; Alexan-der, B.; and Wagner, M. 2020. Design optimisation of amulti-mode wave energy converter. InInternational Con-ference on Offshore Mechanics and Arctic Engineering, vol-ume 84416, V009T09A039. American Society of Mechani-cal Engineers.
Hochreiter, S.; and Schmidhuber, J. 1997. Long short-termmemory.Neural computation, 9(8): 1735–1780.
Andrychowicz, M.; Raichuk, A.; Sta´nczyk, P.; Orsini, M.;Girgin, S.; Marinier, R.; Hussenot, L.; Geist, M.; Pietquin,O.; Michalski, M.; et al. 2020. What matters in on-policyreinforcement learning? a large-scale empirical study.arXivpreprint arXiv:2006.05990.