Function Approximations for Reinforcement Learning Controller for Wave Energy Converters

Soumyendu Sarkar[1*], **Vineet Gundecha**[1], **Sahand Ghorbanpour**[1], **Alexander Shmakov**[1], **Ashwin Ramesh Babu**[1], **Alexandre Pichard**[2], **Mathieu Cocho**[2]
[1] Hewlett Packard Enterprise
soumyendu.sarkar, vineet.gundecha, sahand.ghorbanpour, alexander.shmakov ashwin.ramesh-babu @hpe.com
[2] Carnegie Clean Energy
apichard, mcocho @cce.com

Abstract

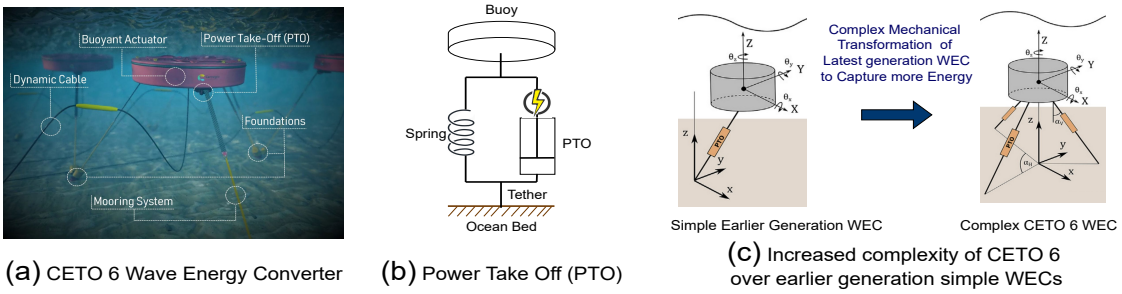
Waves are a more consistent form of clean energy than wind and solar and the latest Wave Energy Converters (WEC) platforms like CETO 6 have evolved into complex multi-generator designs with a high energy capture potential for financial viability. Multi-Agent Reinforcement Learning (MARL) controller can handle these complexities and control the WEC optimally unlike the default engineering controllers like Spring Damper which suffer from lower energy capture and mechanical stress from the spinning yaw motion. In this paper, we look beyond the normal hyper-parameter and MARL agent tuning, and explored the most suitable architecture for the neural network function approximators for the policy and critic networks of MARL which act as its brain. We found that unlike the commonly used fully connected network (FCN) for MARL, the sequential models like transformers and LSTMs can model the WEC system dynamics better. Our novel transformer architecture, Skip Transformer-XL (STrXL), with several gated residual connections in and around the transformer block performed better than the state-of-the-art with faster training convergence. STrXL boosts energy efficiency by an average of 25% to 28% over the existing spring damper (SD) controller for waves at different angles and almost eliminated the mechanical stress from the rotational yaw motion, saving costly maintenance on open seas, and thus reducing the Levelized Cost of wave energy (LCOE).
Demo: https://tinyurl.com/4s4mmb9v

1 Introduction and Motivation

Lowering the Levelized cost of energy for wave energy converters is key to bringing stability to decarbonization of electric energy generation as it is a very reliable form of clean energy. As shown in Figure 1(c), to maximize energy capture from all translational and rotational motion components, the simple earlier generation one generator WEC with one tether(leg) design is transformed to having 3 generators on 3 interdependent legs (tethers) in CETO 6, that this work focuses on. The CETO WECs have been deployed in Australia with planned deployments in Europe. The complexity of this design coupled with the variability of waves in terms of directions, frequency components, and heights, and the mechanical stress from yaw, needed refinements to multi-agent RL so that it can optimally control for these multiple objectives. The main contributions of this paper can be summarized as follows:

[*]Corresponding author

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.



(a) CETO 6 Wave Energy Converter (b) Power Take Off (PTO) (c) Increased complexity of CETO 6 over earlier generation simple WECs

Figure 1: Architecture of Multi-Agent RL controlling the WEC

- Explore and evolve the sequential models of Transformers and LSTMs for the function approximations for the policy and critic networks of multi-agent RL to model the system dynamics better, instead of the traditional approach of just RL hyper-parameter tunings and RL agent optimizations with fully connected neural networks(FCN).
- Propose a **novel transformer block architecture Skip Transformer-XL (STrXL)** which performs better and trains faster than state-of-the-art transformers for this problem as shown in figure 3. STrXL will also make it easier to use transformers in other RL applications where training instability, slow speed, and computation budget make it challenging.

2 Background and Related Work

2.1 Wave Energy Converter (WEC)

The CETO 6 WEC is composed of a cylindrical Buoyant Actuator (BA), submerged approximately 2 meters under the ocean's surface as shown in Figure 1. The BA is secured to the seabed through three mooring legs, each of which terminates on one of the three power take-offs (PTOs) located within the BA. The PTO resists the extension of the mooring legs, thereby generating electrical power similar to regenerative braking in cars. Optimal timing of the PTO forces resisting the wave excitation force is key to maximizing WEC performance. The control strategies mostly used are pure damping control, spring damper control, latching control, model predictive control, and so on.

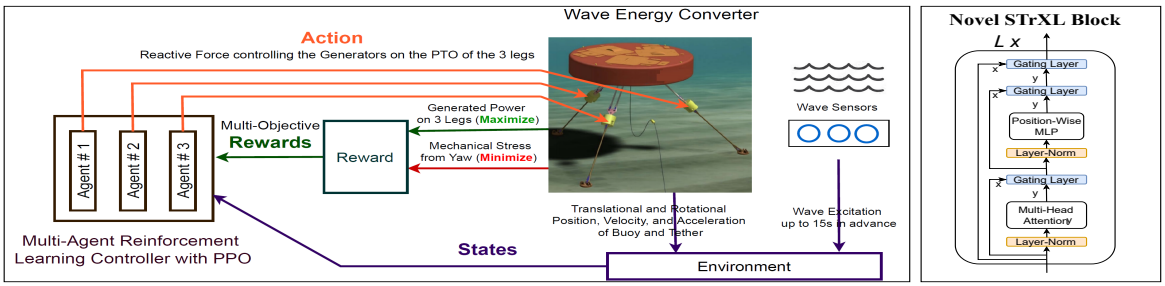
2.2 Related Work

RL has been applied to continuous control tasks for different applications (11), (7) (20). Research for WEC controllers as in (1) (2) ((4)) (5) have been applied to one degree of freedom point absorbers. (20), (24), (25) used RL to control three-legged WECs, but the design had partial success as the RL function approximation was limited to a fully connected neural network (FCN). In this work, we investigate a variety of ML model architectures for the actor and the critic in the PPO, which can better learn the sequential characteristics of the WEC. Our studies show that these models outperform the FCN by a significant margin. Also we investigate different architectural modifications like STrXL that the transformer model needs to effectively train and converge in a MARL design.

3 Reinforcement Learning and Function Approximation

3.1 Multi-agent RL with PPO

After exploring different RL algorithms like Deep Q-Learning (DQN)(17), Soft Actor-Critic (SAC) (8), and Asynchronous Advantage Actor-Critic (A3C) (15), we limited our focus to Proximal Policy Optimization (PPO) ((21)) for this study as PPO outperformed other models. Also as in WECs the generators mounted on the individual legs act quite differently based on the placement in the mechanical structure and mean wavefront, multi-agent RL was chosen as single agent RL failed to control the WEC effectively. The RL states, actions and rewards are explained in Figure 2 (a).



(a) Multi-agent Reinforcement Learning controller for Wave Energy Controller (b) Novel STrXL Transformer

Figure 2: Design of Multi-Agent RL for WEC and our novel STrXL Transformer Block

3.2 Function Approximation for RL

The periodic nature of ocean waves and the spring-type inertial response of the WEC require a system model which can represent and process time series information, unlike the most widely used default feed-forward networks (FCN) for RL. This also enables combining long-term behavior from past observations and future wave states from sensors placed further into the ocean, into the current state.

We investigated the WEC controller performance and speed of convergence for FCN, LSTMs, and Transformers of varying depths ((10; 22; 18; 16)) for RL policy and critic function approximators. Transformers with multi-head attention, temporal convolution network, and contextual horizon with relative position encoding, proved to be ideally suited for PPO function approximation for WEC. To mitigate the limited sequence length for canonical transformers unlike LSTMs, we used the Tr-XL architecture which keeps a memory of hidden states corresponding to previous sequences.

3.3 Skip Transformer-XL (STrXL) Architecture

Using canonical transformers as function approximators for RL is challenging as it is very difficult to train and optimize as established by (18), Unlike supervised learning tasks ((22)). So we explored the effect of various gated bypasses for Transformer FA on the stability and speed of training. Inspired by residual network architecture ((9)) and expanding earlier work of (18), we propose a novel transformer block "Skip Transformer-XL" or STrXL (Figure 2(right)). In STrXL an additional bypass connection with a gating layer around the transformer block helps accelerate training convergence when compared to the previous designs like GTrXL. With the layer normalization placed on the input stream of the submodules ((18)) and tucked inside the bypass, an identity map from the transformer's input at the first layer to the transformer's output after the last layer is established, unlike canonical transformers. The state encoding is passed untransformed to the policy and value heads, enabling the agent to learn a Markovian policy at the start of training. For WEC, the reactive behaviors need to be learned before memory-based ones can be effectively utilized. Also, the GRU-style gating mechanisms in place of the residual connections within the transformer block helped stabilize learning and improved performance. This enables better performance and faster convergence during training.

4 Experiments

The CETO 6 wave energy converter (WEC) platform simulator was used to accurately model the mechanical structure, the mechanical response, the electro-mechanical conversion efficiency for generator and motor modes, and the fluid dynamical elements of the wave excitation. Wave data collected from WEC deployment sites at Albany, and Garden Island in Western Australia, Armintza in Spain (Biscay Marine Energy Platform: BiMEP) ((12)), and Wave Hub on the north coast of Cornwall in United Kingdom ((13)) were used along with Jonswap spectrum. For evaluation, we used 1000 episodes for each principal wave period and height. 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.

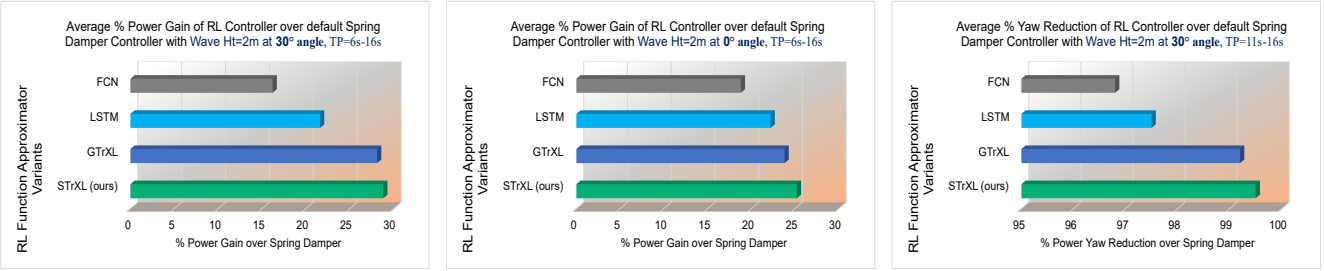


Figure 3: **Left:** % Increase of Energy Capture over SD controller for 0° and 30° waves for ht=2m with different Function Approximations, **Right:** % Yaw Reduction by RL over SD for ht=2m, angle=30°.

Table 1: % reduction of Yaw by RL over SD for ht=7m, angle = 30°

WTPs	11	12	13	14	15	16	Avg
% yaw ↓	98.8	98.8	98.9	99.1	98.9	98.6	98.6

5 Results

The power generated by the baseline spring damper controller tuned to a specific wave time period and height under consideration is used as a reference for evaluation to estimate the gain of energy capture by Reinforcement Learning (RL) controllers as a percentage improvement. A direction of 0° indicates frontal waves with the wavefront aligned with the front leg, and for evaluation, we used the same seed for sampling waves for episodes between RL and SD.

Figure 3(left) shows that for **electric energy conversion** at **0° frontal waves**, the MARL with STrXL performs on an average **25.2%** better than the baseline spring damper (SD) controller, while the LSTM performs 22.2% better and FCN performs 18.8% better on an average for the entire range of wave time periods 6s to 16s. For **angled waves of 30°**, the MARL with STrXL (**28.8%**) performs much better than LSTM (21.6%) on average. STrXL is better than state-of-the-art GTrXL, but it trains much faster with high stability as can be seen in the appendix. The STrXL performance peaks at a depth of 3, the LSTM at the depths of 2 and 3, and FCN at a depth of 2.

Figure 3(right) shows that for wave periods of 11s to 16s, where yaw is a significant problem of default SD controller, the **yaw is reduced by more than 99%** with STrXL, significantly reducing mechanical stress with huge maintenance savings. Even with extreme wave height of 7m, Table 1 shows that the PPO with STrXL reduces the yaw by over 98.6% over SD.

6 Conclusions

The proposed MARL controller yields 25%+ gain over the baseline Spring Damper controller (SD) for the entire spectrum of ocean waves, boosting energy production with revenue implications. The MARL also helped reduce mechanical stress significantly, lowering maintenance and operating costs and actively mitigating adverse effects of high waves helping preserve capital investments and lowering LCOE making wave energy more of a reality.

We found that robust RL function approximation sequence models of suitable architectures and depths are key to achieving higher performance for complex real-life use cases like WEC, and RL agent (PPO) refinements alone cannot do that. The proposed novel STxRL architecture with GRU gated bypass inside and around the transformer block help solve the challenging training convergence problem of RL with transformers. As STrXL trains faster and performs better than the state-of-the-art GTrXL, it may help other complex multi-agent RL applications and facilitate greener computation.

References

[1] 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.

[2] 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.

[3] Anderlini, E.; Forehand, D. I.; 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.

[4] Anderlini, E.; Forehand, D. I.; 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.

[5] 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): 845.

[6] CETO Technology. 2020. CETO Technology

[7] Duan, Y.; Chen, X.; Houthooft, R.; Schulman, J.; and Abbeel, P. 2016. Benchmarking deep reinforcement learning for continuous control. In International conference on machine learning, 1329–1338. PMLR.

[8] Haarnoja, T.; Zhou, A.; Abbeel, P.; and Levine, S. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In International conference on machine learning, 1861–1870. PMLR.

[9] He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778.

[10] Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. Neural computation, 9(8): 1735–1780.

[11] Lillicrap, T. P.; Hunt, J. J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; and Wierstra, D. 2015. Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

[12] BiMEP. 2021. BiMEP

[13] Wave Hub. 2021. Wave Hub

[14] Lim, B.; Arık, S. O.; Loeff, N.; and Pfister, T. 2021. Tempo- ¨ ral fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting.

[15] Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Lillicrap, T.; Harley, T.; Silver, D.; and Kavukcuoglu, K. 2016. Asynchronous methods for deep reinforcement learning. In International conference on machine learning, 1928–1937. PMLR.

[16] Lim, B., Arık, S.Ö., Loeff, N. and Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), pp.1748-1764.

[17] Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; and Riedmiller, M. 2013. Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602.

[18] Parisotto, E.; Song, F.; Rae, J.; Pascanu, R.; Gulcehre, C.; Jayakumar, S.; Jaderberg, M.; Kaufman, R. L.; Clark, A.; Noury, S.; et al. 2020. Stabilizing transformers for reinforcement learning. In International Conference on Machine Learning, 7487–7498. PMLR.

[19] Rijnsdorp, D. P.; Hansen, J. E.; and Lowe, R. J. 2018. Simulating the wave-induced response of a submerged wave energy converter using a non-hydrostatic wave-flow model. Coastal Engineering, 140: 189–204.

[20] Sarkar, S.; Gundecha, V.; Shmakov, A.; Ghorbanpour, S.; Babu, A. R.; Faraboschi, P.; Cocho, M.; Pichard, A.; and Fievez, J. 2022. Multi-Agent Reinforcement Learning Controller to Maximize Energy Efficiency for Multi-Generator Industrial Wave Energy Converter. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, 12135–12144

[21] Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

[22] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In Advances in neural information processing systems, 5998–6008.

[23] 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.

[24] Sarkar, S.; Gundecha, V.; Shmakov, A.; Ghorbanpour, S.; Babu, A. R.; Pichard, A.; and Cocho, M. 2022. Skip Training for Multi-Agent Reinforcement Learning Controller for Industrial Wave Energy Converters. In 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE), 212–219

[25] Sarkar, S.; Gundecha, V.; Shmakov, A.; Ghorbanpour, S.; Babu, A. R.; Faraboschi, P.; Cocho, M.; Pichard, A.; and Fievez, J. 2021. Multi-objective Reinforcement Learning Controller for Multi-Generator Industrial Wave Energy Converter. In NeurIPs Tackling Climate Change with Machine Learning Workshop

Appendix A Additional Results

A.1 Gain in Energy Capture and Yaw reduction for different function approximations for RL agent

Table 2: Energy Capture Gain by the RL controller over spring damper controller for different PPO function approximators

RL % Gain of Energy Capture over default Spring Damper (SD controller)												
% Gain for Wave Height = 2m, and Wave Angle = 0 degrees												
Wave Time Period(s)	6	7	8	9	10	11	12	13	14	15	16	Avg
FCN	38.4	35.4	23.0	19.7	15.1	14.1	13.5	11.9	12.9	11.7	11.4	18.8
LSTM	41.3	35.5	27.8	24.1	18.6	15.4	15.9	17	17.7	15.9	15.3	22.2
GTrXL	40.2	36.1	28.2	23.9	19.3	14.9	23.2	17.9	18.9	18.3.2	15.8	23.8
STrXL (ours)	40.1	38.9	32.2	25.2	21.4	22.3	24.1	18.5	19.0	18.1	17.1	**25.2**
% Gain for Wave Height = 2m, and Wave Angle = 30 degrees												
Wave Time Period(s)	6	7	8	9	10	11	12	13	14	15	16	Avg
FCN	33.4	32.9	20.1	9.6	5.3	7.6	10	14.6	15.8	16.1	12.3	16.2
LSTM	34.6	33.3	26.7	20.3	14.5	14.3	16.3	20.8	17.9	20.1	18.7	21.6
GTrXL	39.2	35.2	27.6	21.8	17.1	17.8	21.3	29.4	36.9	30.6	31.9	28.1
STrXL (ours)	39.7	34.7	28.7	22.7	17.4	18.1	22.6	31.5	38.7	31.2	31.3	**28.8**

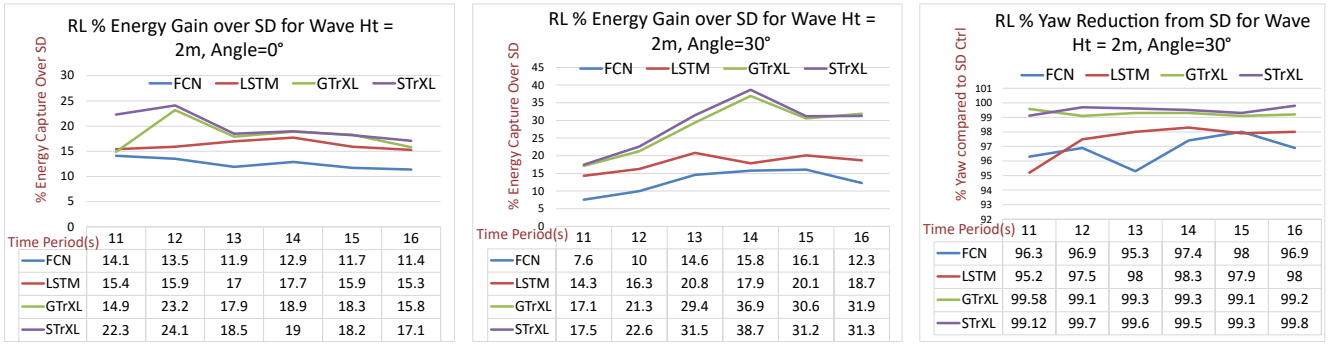


Figure 4: **Left:** % Increase of Energy Capture over SD controller for 0° and 30° waves for Height=2m with different Function Approximations, **Right:** % Yaw Reduction by RL over SD for ht=2m, angle=30°.

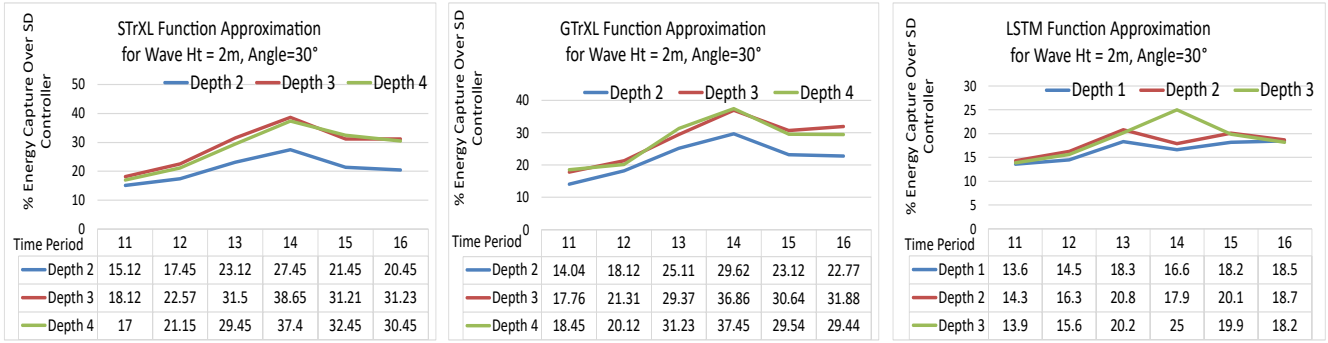


Figure 5: % increase of Energy Capture over SD for Ht=2m Angle=30° with different depths for Function Approximation models.

A.2 Training speed for different function approximations for RL agent

As shown in Figure 6 the STrXL trains faster than the state-of-the-art GTrXL and TrXL-I transformer variants with augmented trainability features.

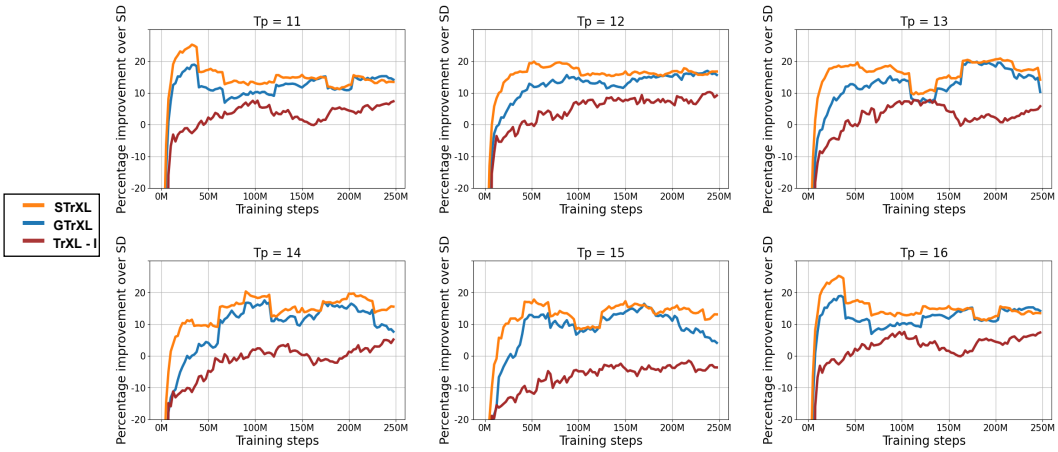


Figure 6: RL Training progression for ht=2m angle=30° for STrXL, GTrXL, and TrXL-I function approximators.

Appendix B Wave Energy Conveters

B.1 Wave Energy Converter (WEC)

The CETO 6 WEC is composed of a cylindrical Buoyant Actuator (BA), submerged approximately 2 meters under the ocean's surface as shown in Figure **??**. The BA is secured to the seabed through three mooring legs, each of which terminates on one of the three power take-offs (PTOs) located within the BA. The PTOs act like winches - they can pay in and out to allow the mooring legs to vary in length and thus converting the chaotic motions of the BA into linear motions. The PTO also resists the extension of the mooring legs, thereby generating electrical power similar to regenerative braking in cars. The high-level structure of the WEC is represented in Figure 7. Optimal timing of the PTO forces resisting the wave excitation force is key to maximizing WEC performance. Various control strategies exist, attempting to get as close as possible to the optimal force function with various degrees of success. These include pure damping control, spring damper control, latching control, model predictive control, and so on.

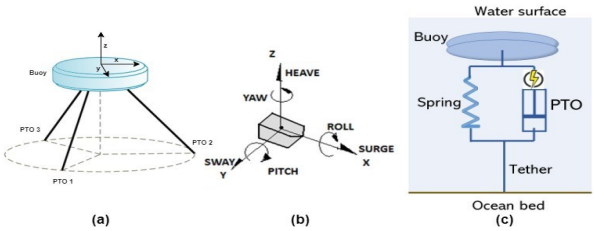


Figure 7: (a)3D view of WEC, (b) PTO motion with 6 degrees of freedom, (c)

B.2 Spring Damper Benchmark Controller (WEC)

The PTO is composed of a mechanical spring and an electrical generator, as represented in Figure 7(c). The damping component is akin to a reactive braking torque against the input shaft, driven by the wave energy source. The captured energy equals the braking mechanical work done by the generator minus losses.

Appendix C Skip Transformer XL (STrXL)

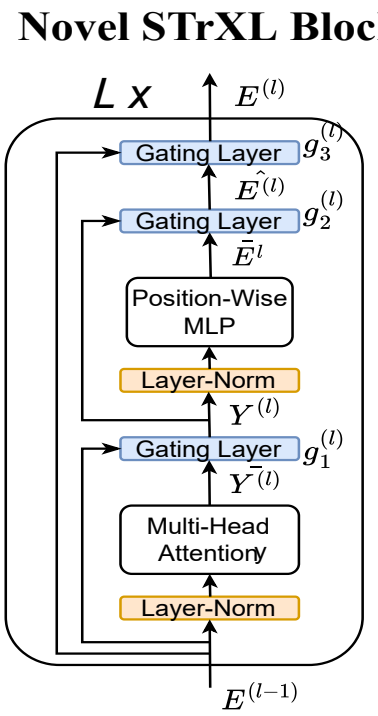


Figure 8: STxRL architecture with GRU gated bypass inside and around the transformer block

The STrXL block can be represented with the following equations. Referring to Figure 8, the input to the transformer block is an embedding from the previous layer $E^{(l-1)}$ and the output of the transformer block is $E^{(l)}$, where $l \in [0, L]$ is the layer index. The Gated Recurrent Unit (GRU) type gating is g_i as shown in Figure **??**. Then the multi-head attention block gated output:

$$\bar{Y^{(l)}} = MultiHeadAttention(LayerNorm(M^{(l-1)}, E^{(l-1)}))$$
$$Y^{(l)} = g_1^{(l)}(E^{(l-1)}, ReLU(\bar{Y}^{(l)}))$$

The MLP block gated output:

$$\bar{E}^{(l)} = f^{(l)}(LayerNorm(Y^{(l)})),$$
$$\hat{E}^{(l)} = g_2^{(l)}(Y^{(l)}, ReLU(\bar{E}^{(l)}))$$

The STrXL gated output:

$$\hat{E}^{(l)} = g_3^{(l)}(E^{(l-1)}, ReLU(\hat{E}^{(l)}))$$