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

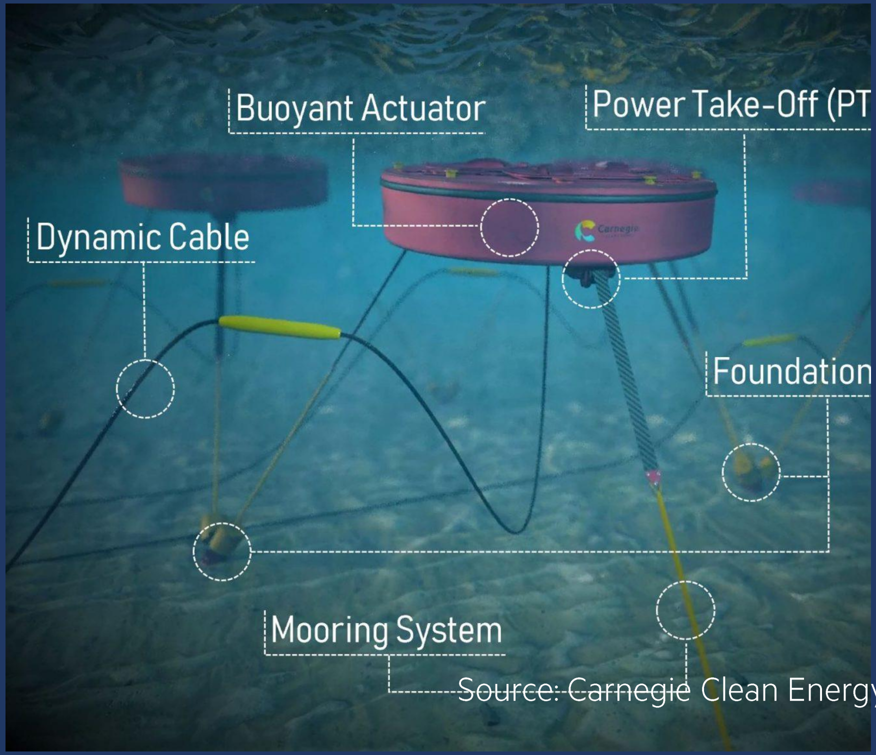
NeurIPS 2021 : Tackling Climate Change with Machine Learning Workshop

Soumyendu Sarkar[1], Vineet Gundecha[1], Alexander Shmakov[1], Sahand Ghorbanpour[1], Ashwin Ramesh Babu[1], Paolo Faraboschi[1],

Mathieu Cocho[2] , Alexandre Pichard[2] , Jonathan Fievez[2]

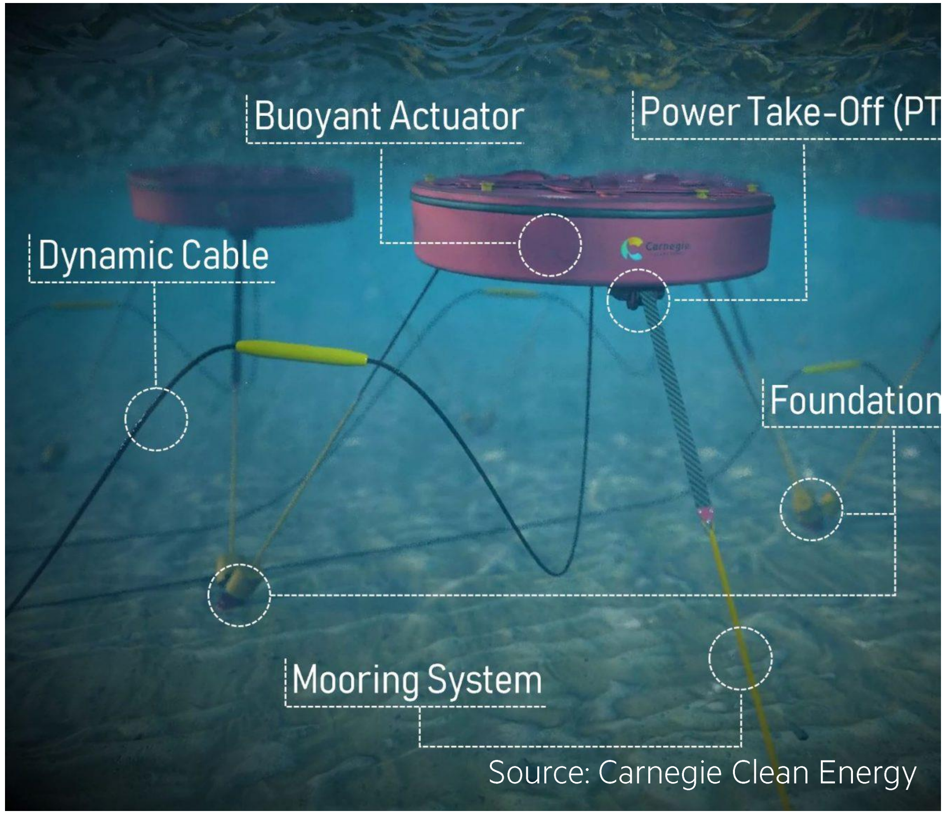
[1] Hewlett Packard Enterprise (Hewlett Packard Labs)

[2]Carnegie Clean Energy



Source: Carnegie Clean Energy

Motivation : Lowering the Levelized Cost of Energy (LCOE) for Wave Energy



- Waves are more consistent and reliable renewable energy source than wind and solar
- Lowering the Levelized Cost of Energy (LCOE) for Wave Energy Converters is key
    - Increase in energy capture efficiency, boosting revenue
    - Reducing structural stress to limit maintenance and operating cost
    - Protect the wave energy converter from acute weather events, preserving investments and lowering effective capital cost

Problem: Complexity of Control for the latest multi-generator multi-legged WEC

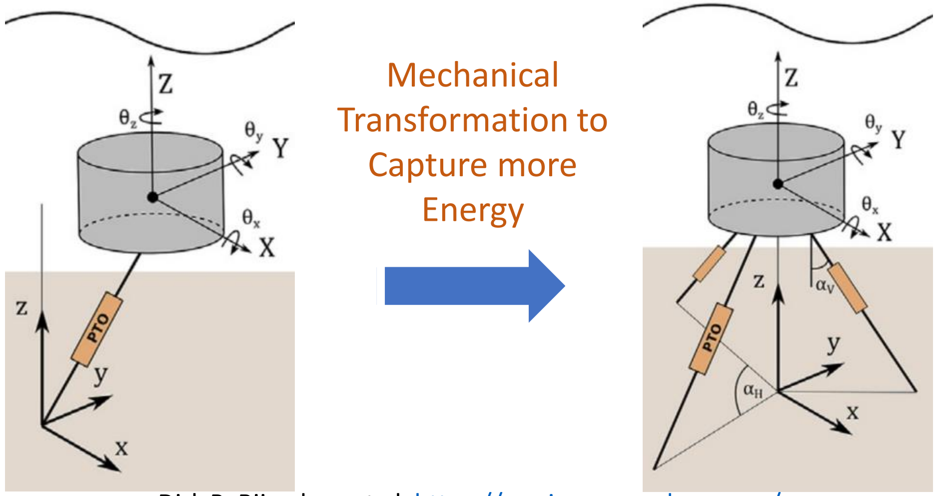
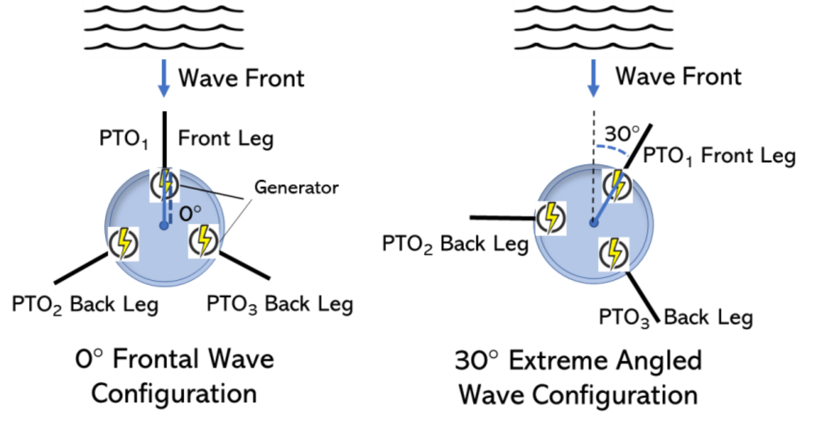


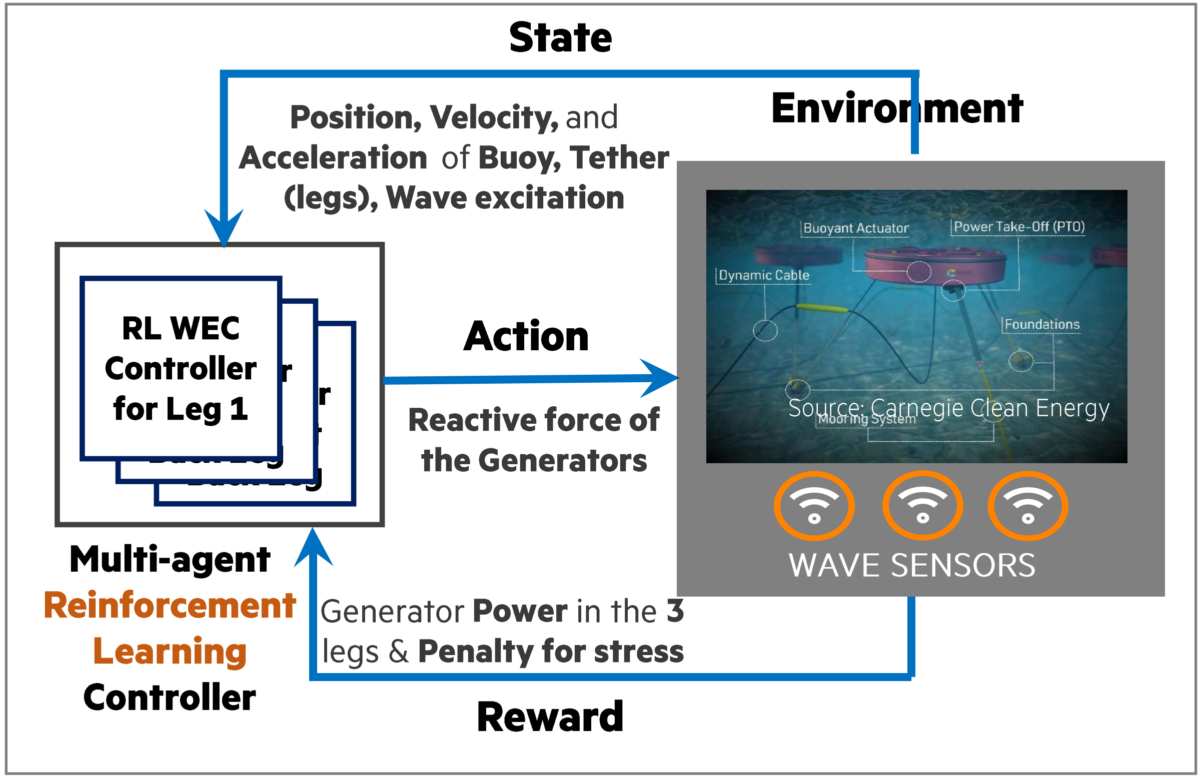
Image source: Dirk P. Rijnsdorp et al. https://marineresearch-gs.com/wp-content/uploads/2019/12/6_Rijnsdorp-DP-et-al.pdf University of Western Australia



- To maximize energy capture the simple earlier generation one generator WEC with one tether(leg) design is transformed to having 3 generators on 3 interdependent legs (tethers) to leverage translational and rotational motions
- Complexity of control has gone up significantly with the state-of-the-art Wave Energy Converters (WEC)
- Variability of the waves, angles of wave fronts and asymmetry of the WEC further complicates the control
- Existing controllers like Spring Damper are unable to leverage the full potential of this complex mechanical structure
- Reinforcement Learning is able to better control the reactive forces of the generators on multiple tethers (legs) of WEC

Multi-agent Reinforcement Learning controller with augmentations

- 3 legs and the generators on each of the legs act differently
- Heterogeneity requires Multi-Agent Reinforcement Learning



Key Challenges

- Iterative agent state design for observability and correlation with control objective
- Sustaining the training convergence for optimum controller
- Simultaneous convergence for all wave time periods

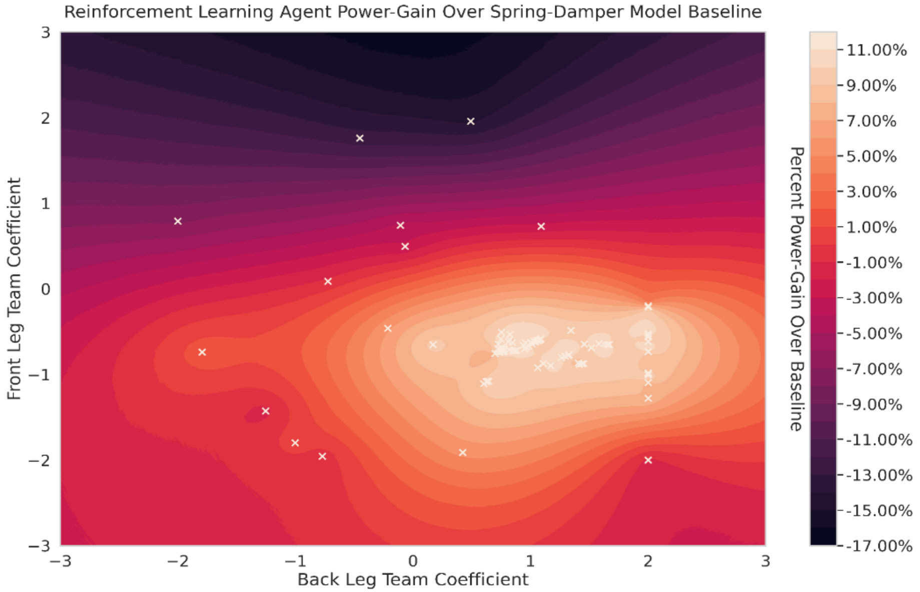
Refinements to PPO for training stability and convergence

- Proximal Policy Optimization (PPO) performed better than other RL algorithms like the DDQN, Soft Actor-Critic, and A3C
- PPO design augmentation and tuning for convergence
- LSTMs to leverage the time-series nature of states and partial observability into the oncoming wave excitation

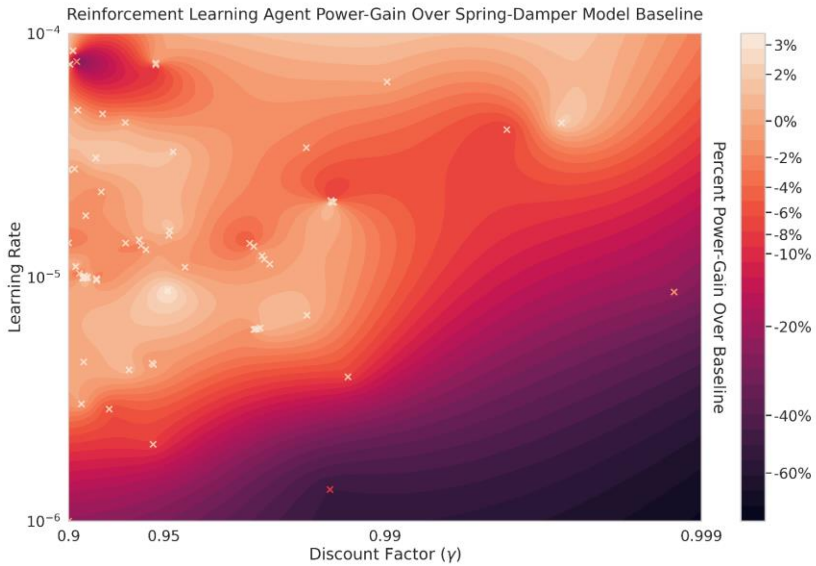
MARL Hyper-parameter optimizations are key for higher power gains

- Multi-dimensional hyper-parameter optimization was key for RL agent performance gains
- Multi-agent Co-opetition: Disparity in power and optimum trade-off by individual legs, to get most combined power in all legs, make the best solution a combination of co-operation and competition for different agents.

Co-opetition between multiple RL agents and Hyper-parameter optimizations



Discount Factor (γ) vs Learning Rate

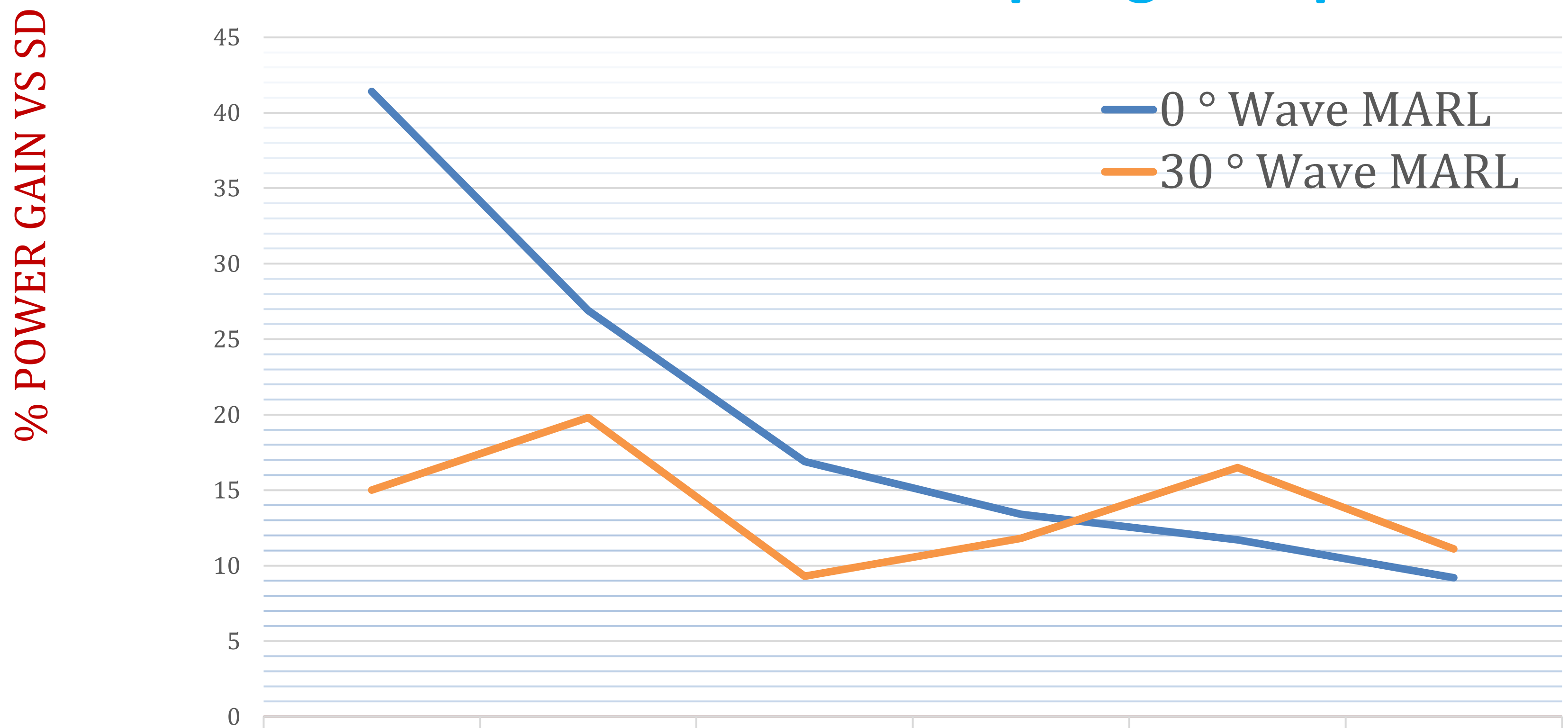


$$Reward = P_{own} + \eta \, . P_{others}$$

where, η = team coefficient, P_{own} is the power of the generator being controlled and P_{other} is power from other generators

Results : Double digit % Power Gain over state-of-the-art Spring Damper controller

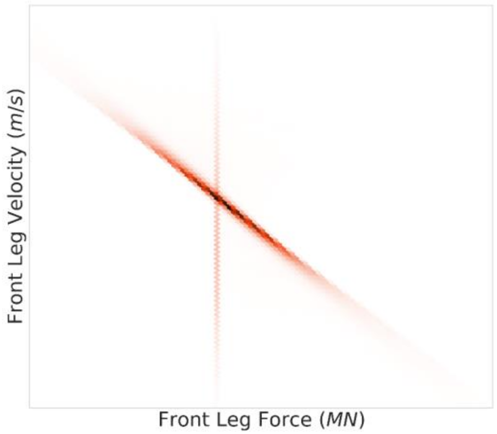
RL % Power Gain over Spring Damper



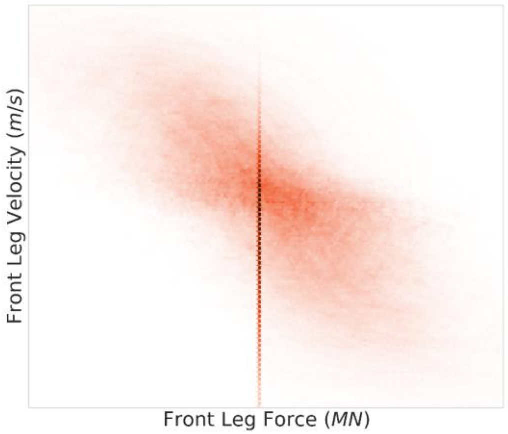
WAVE TIME PERIOD (S)	6	8	10	12	14	16
0 ° Wave MARL	41.4	26.9	16.9	13.4	11.7	9.2
30 ° Wave MARL	15	19.8	9.3	11.8	16.5	11.1

Intuition behind Reinforcement Learning controller performance

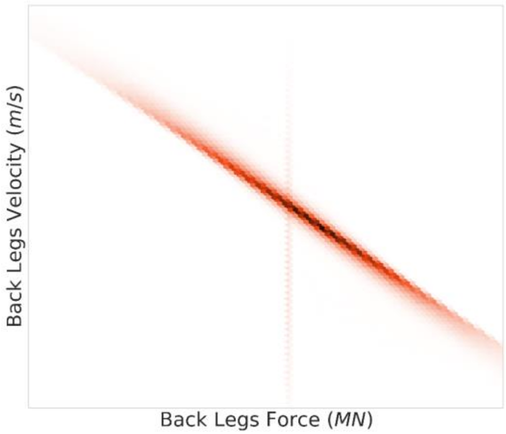
- Spring Damper is more greedy and reactive forces for the generators on the legs are almost proportional to the instantaneous velocity of the tether as energy is captured working against this motion.
- RL controller is fuzzy about the proportionality of reactive force and tether (leg) velocity, as it compromises short-term objectives for greater gains on energy capture at the more opportune segments of the wave cycles with discounted returns.



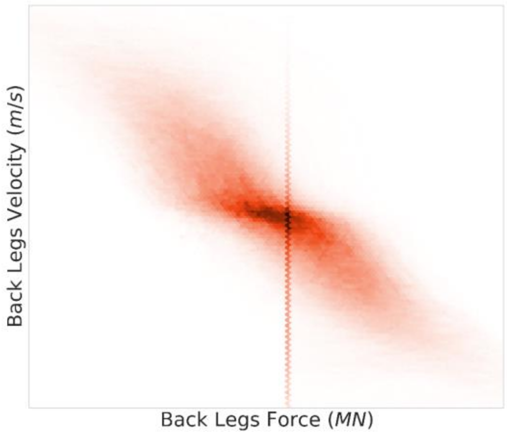
Spring Damper
Front Leg



Reinforcement Learning
Front Leg



Spring Damper
Back Leg

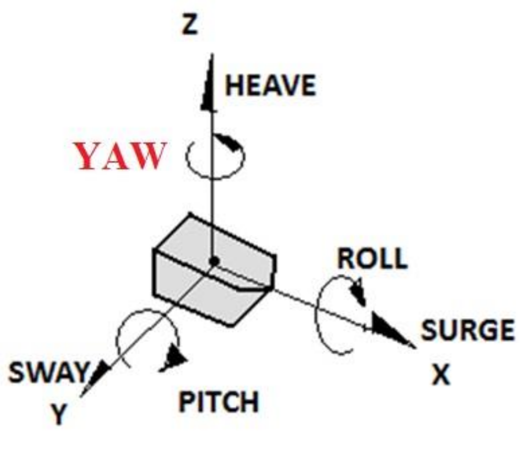


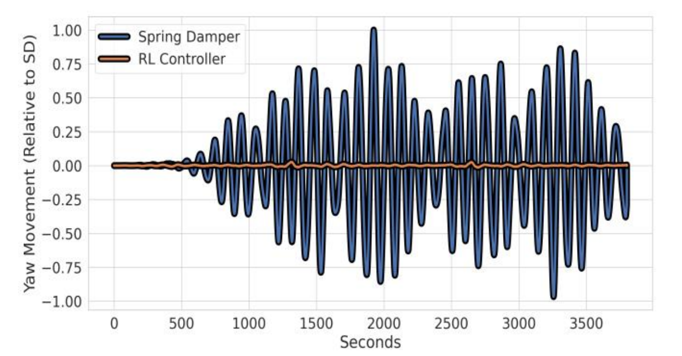
Reinforcement Learning
Back Leg

- Better co-ordination between the multiple generators and legs with varying waves and 6 degrees of motion which the existing state of the art controllers fail to do

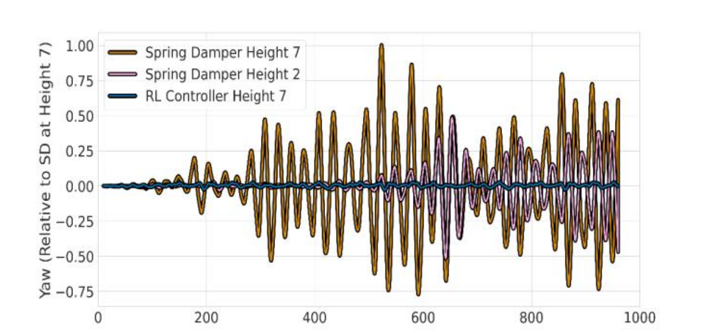
Mechanical stress and Maintenance mitigation with Design for Trust

- Yaw rotational motion of the voluminous buoy causes damaging mechanical stress and is higher for angled waves.
- Penalty for yaw: $\boldsymbol{Reward} = (\boldsymbol{\alpha})\,\boldsymbol{power} + (\boldsymbol{1} - \boldsymbol{\alpha})\,\boldsymbol{yaw}$,
  where α is tunable, lesser α => stronger penalty.
- Result: Yaw is almost eliminated resulting in less maintenance compared to spring damper (SD).
- Reward shaping with yaw minimization improved power generation, as more power directed to the generator
- Combined reward maximizes energy capture maximization and minimizes stress.

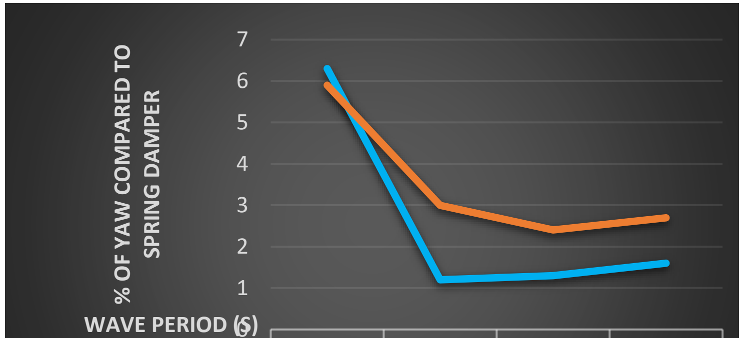




YAW for RL vs SD for normal heights



YAW for RL vs SD for survival conditions



WAVE PERIOD (S)	10	12	14	16
2m- RL (% of SD)	6.3	1.2	1.3	1.6
7m- RL (% of SD)	5.9	3	2.4	2.7

Impact of this work on Wave Energy Converters and beyond

- Double-digit power gains boosting revenue opportunities
- Reduced mechanical stress, which impacts maintenance and operating costs
- Actively mitigated survival conditions, helping to preserve capital investment
- This MARL architecture is applicable to other clean energy problems like wind energy, both for individual wind power generators and wind farms
- PPO refinements to stabilize training for global optima applicable to other applications

Thank You

Soumyendu Sarkar
Soumyendu.Sarkar@hpe.com
Hewlett Packard Labs @ Hewlett Packard Enterprise