Experimental Validation and Analysis of Deep Reinforcement Learning Control for Wave Energy Converters

Abishek Subramanian[a], Shangyan Zou[a], Bret Bosma[b], Junhui Lou[b], Courtney Beringer[b], Bryson Robertson[b], Pedro Lomonaco[b]

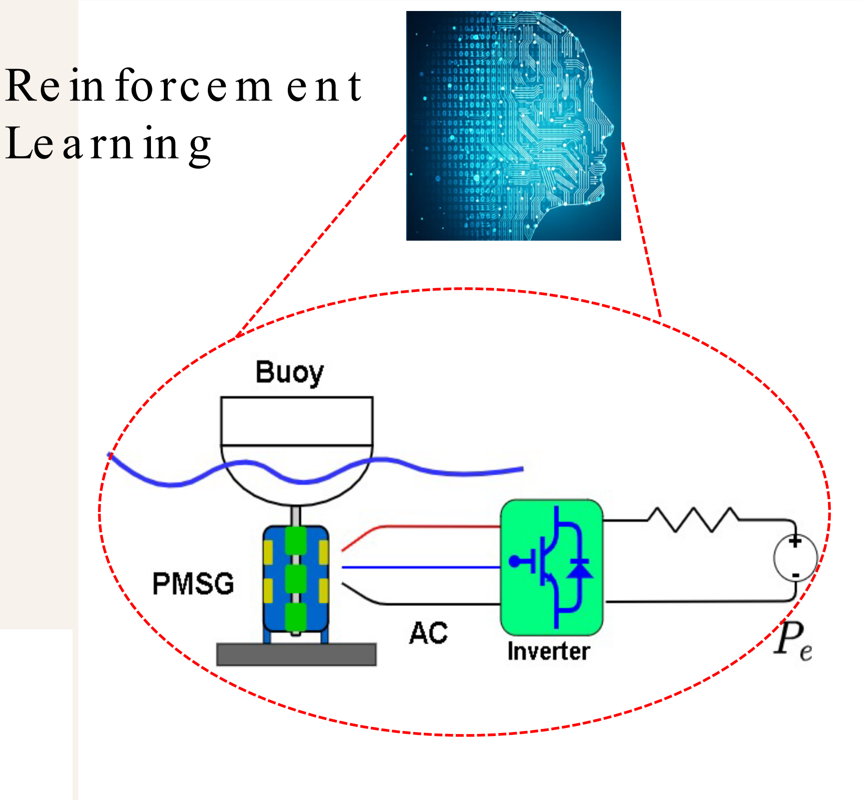
a: Department of Mechanical and Aerospace Engineering, Michigan Technological University

b: School of Civil and Construction Engineering, Oregon State University

umerc + OREC 2025

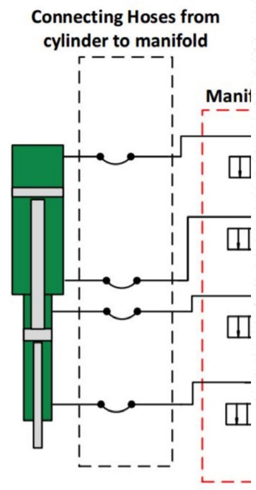
Background

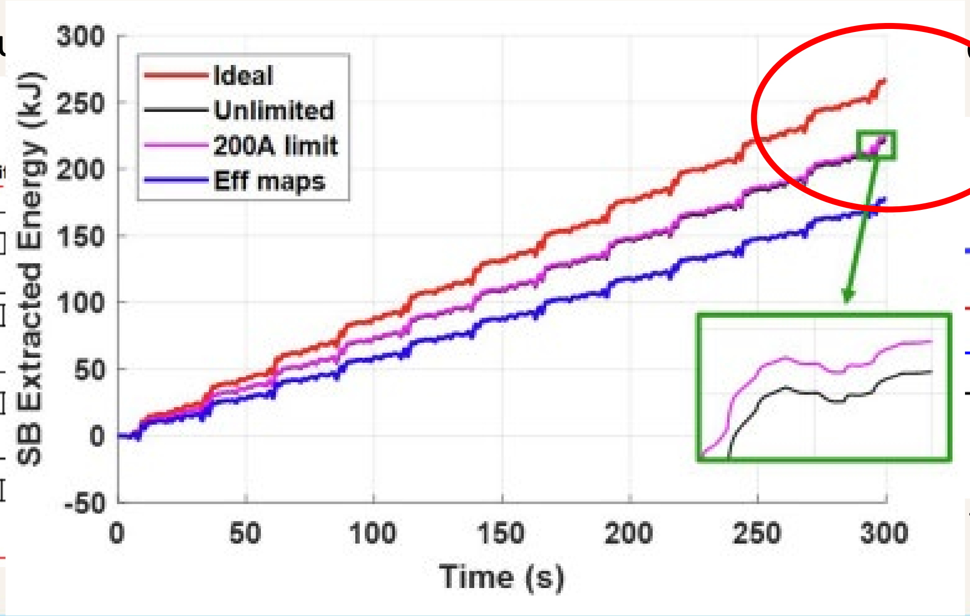
- A major challenge that hinders the economic viability of WECs is the development of **highly efficient Power Take-Off (PTO) control algorithms**.
- On one hand, conventional model-based controls for WECs are typically developed without considering the PTO dynamics (e.g., impedance matching), which may be misleading.
- On the other hand, directly developing a model-based control w.r.t all subsystems from wave to wire is extremely cumbersome.
- Reinforcement learning based method may offer a solution given it is model-free, adaptive, and robust.

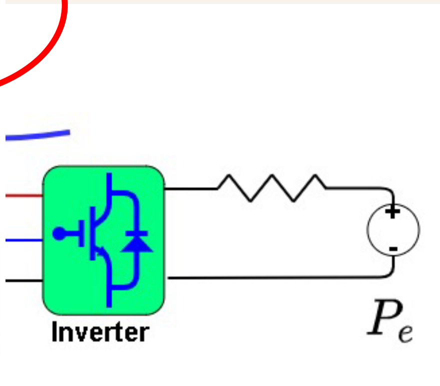


Hydrau

e PTO



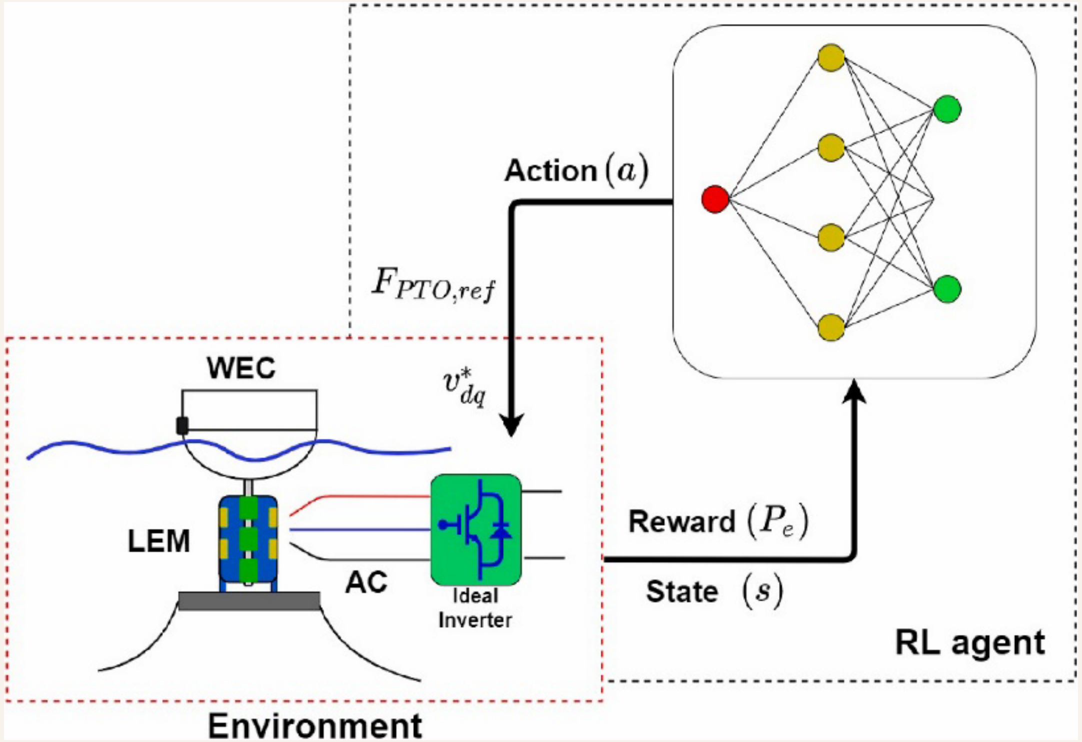




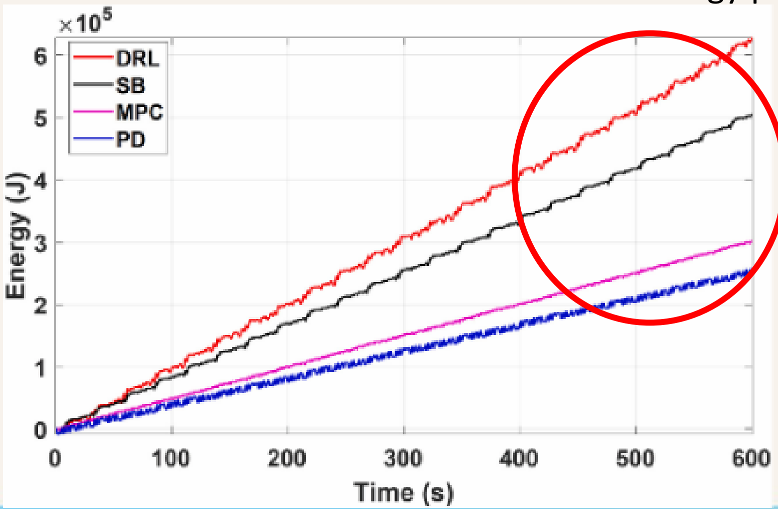
The control performance is significantly degraded when the control is designed without considering the PTO unit

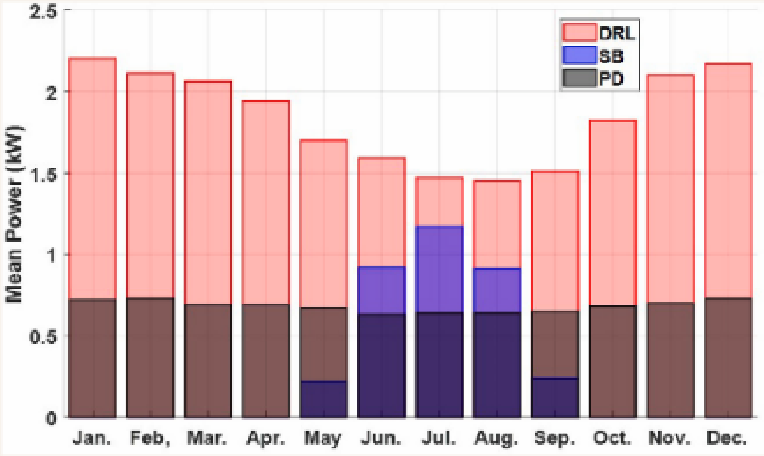
Background

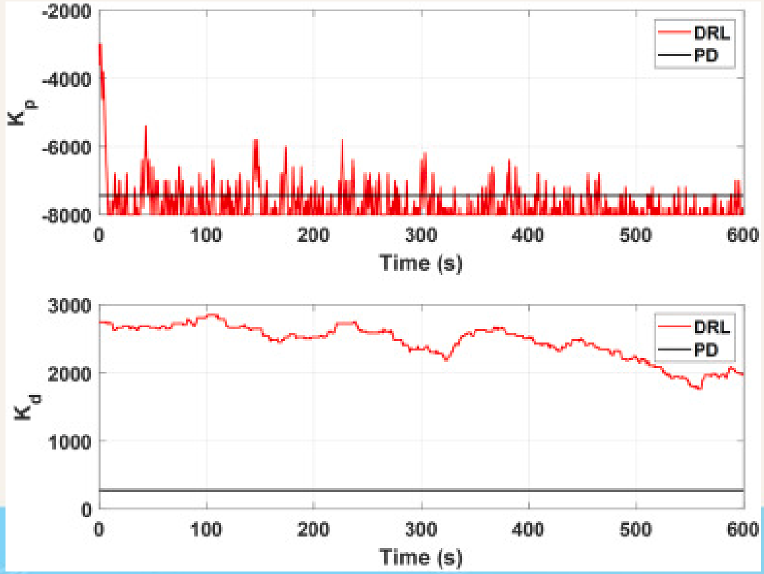
- We have previously developed a Deep Reinforcement Learning (DRL) based control that is able to **optimize the performance of the WEC in a global manner** via direct interaction with the environment [1].
- The developed control collects real-time reward (electrical power output), and states to train the action-value function and determine the next action by maximizing the target.
- On paper, we have proved **promising performance** of the DRL control in terms of both power production and power quality.
- However, the practical performance of this control ***remains unknown***.



On paper, we have shown that the proposed control is able to outperform some state-of-the-art controls in energy production.







[1] Zou, S., Zhou, X., Weaver, W. and Abdelkhalik, O., 2022. Deep reinforcement learning control of wave energy converters. *IFAC-PapersOnLine*, *55*(27), pp.305-310.

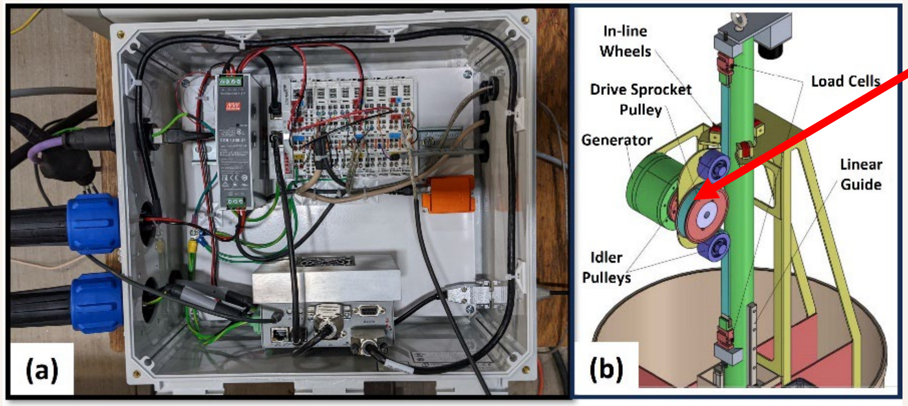
Background

- Therefore, the research team at MTU collaborated with OSU on a TEAMER project to experimentally validate and analyze the practical performance of the DRL control.
- The main objective is to answer the following **research questions:**

(1) *What is the practical performance of DRL control* in terms of power production, adaptivity, robustness, computational speed, and losses?

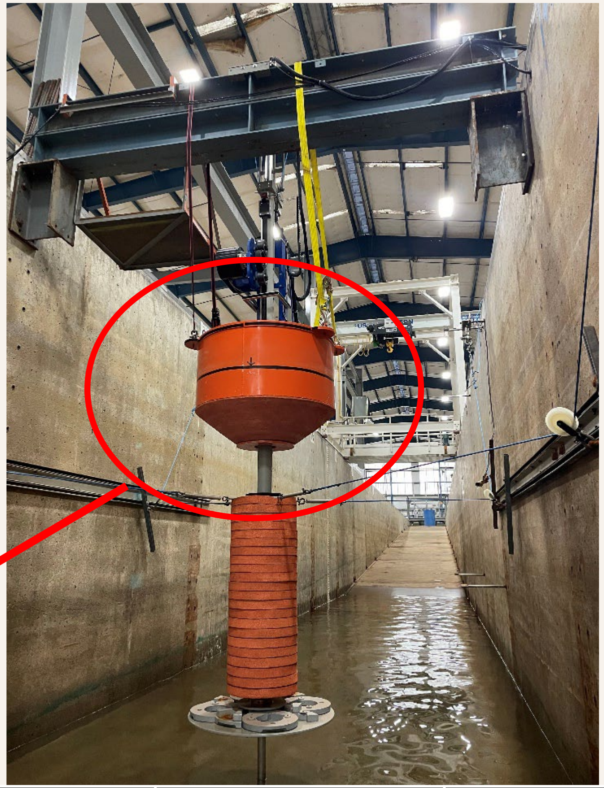
(2) *What are the challenges and limitations of practically implementing DRL control*?

LUPA power electronics and PTO unit



Symbol	Quantity	Value	Unit
τ_{cn}	Continuous Torque	46	N.m
τ_{pk}	Peak Torque	137.9	N.m
K_t	Torque Constant	8.51	N.m/Arms
Ω_{max}	Max Speed	150	rpm

LUPA single body heaving configuration and the actual hardware in the wave flume.



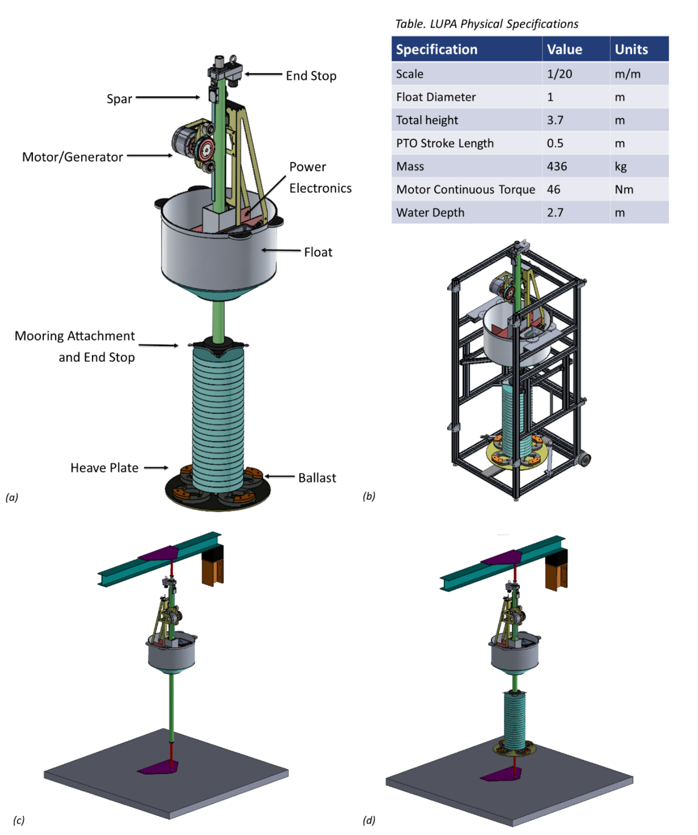


Table. LUPA Physical Specifications

Specification	Value	Units
Scale	1/20	m/m
Float Diameter	1	m
Total height	3.7	m
PTO Stroke Length	0.5	m
Mass	436	kg
Motor Continuous Torque	46	Nm
Water Depth	2.7	m

Component	Brand	Part Number	Specification
Generator	Akribis	ADR220-B175-S/P-J/K-3.0-RA-26B-P25-Z75	Continuous Torque: 50Nm
Encoder	Renishaw	L-9517-9448-05-B	Resolution: 26 bit
Drive	ElmoMC	G-OBOE 13A/230VAC ECTSW ENC SRC FIN+FAN	Peak Current: 26A
Load Cells	Futek	FSH00971	Capacity: 4448N
VRU	Xsens	MTI-20-2A8G4	Roll/Pitch: 0.5^{o}
Draw Wire	Micro-Epsilon	WDS-1000-P60-SR-SSI	Resolution: 0.012mm

Methodology: LUPA numerical model

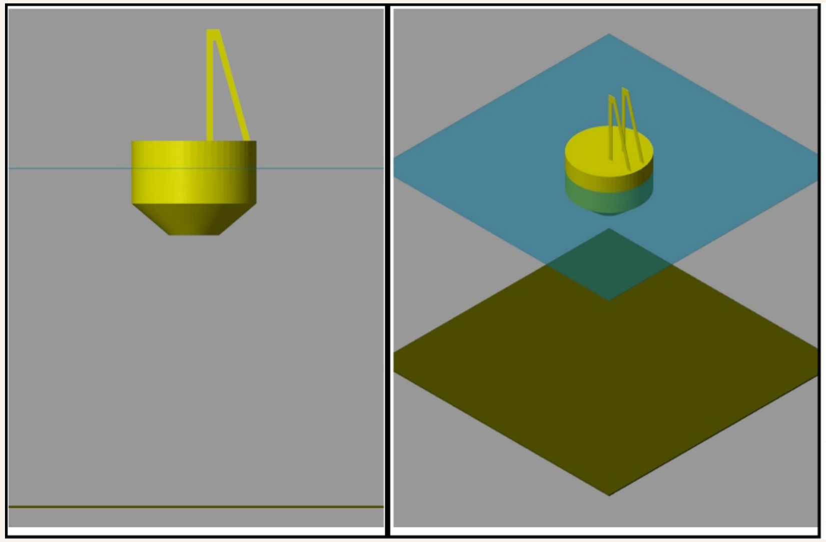
- An experimentally calibrated LUPA numerical model (for single-body heaving configuration) is used for control training [2].
- The hydrodynamics is modelled as:

$$m\ddot{z} = F_{ex} + F_r + F_{hs} + F_{drag} + F_{pto} + F_{loss}$$

It is noted that in addition to the drag force, the force F_{loss} represents physical system losses that include the mechanical friction and electromagnetic damping/inertia in the motor.

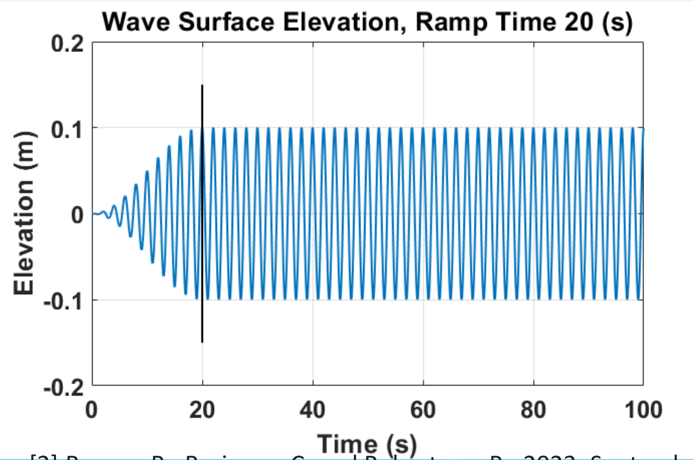
$$F_{loss} = -C_{ld}\dot{z}$$

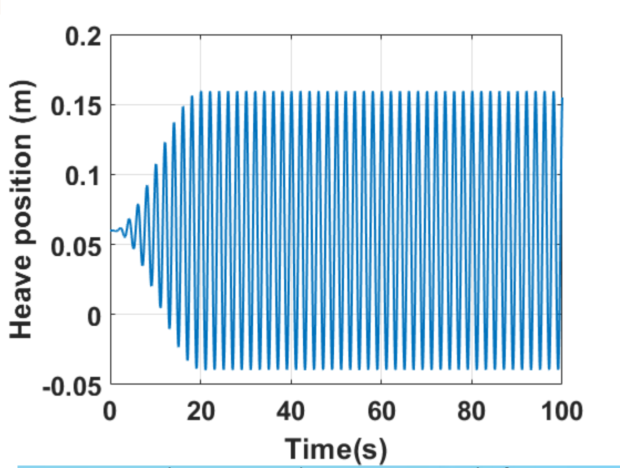
where this force initially is assumed to be **zero** in our pretraining, but later found to be necessary to add (due to model mismatch) and calibrated according to testing data.

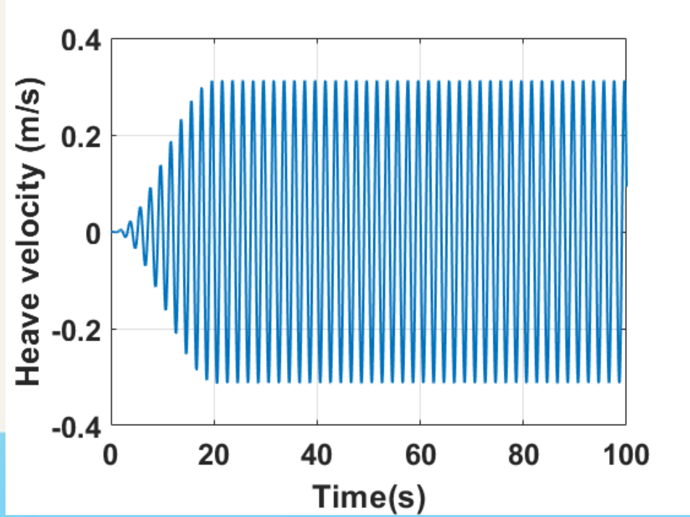


Specifications	Value	Unit
Mass	436	kg
PTO stroke	0.5	m
Moment of Inertia (X)	66.17	Kg.m^2
Moment of Inertia (Y)	65.33	Kg.m^2
Moment of Inertia (Z)	17.16	Kg.m^2
Quadratic Drag coefficient	0.15	-

LUPA response under a regular wave with a height of 0.2m and a period of 2s (wave rider mode).







During the actual experiments, the research team did observe significant mechanical losses in the drivetrain.

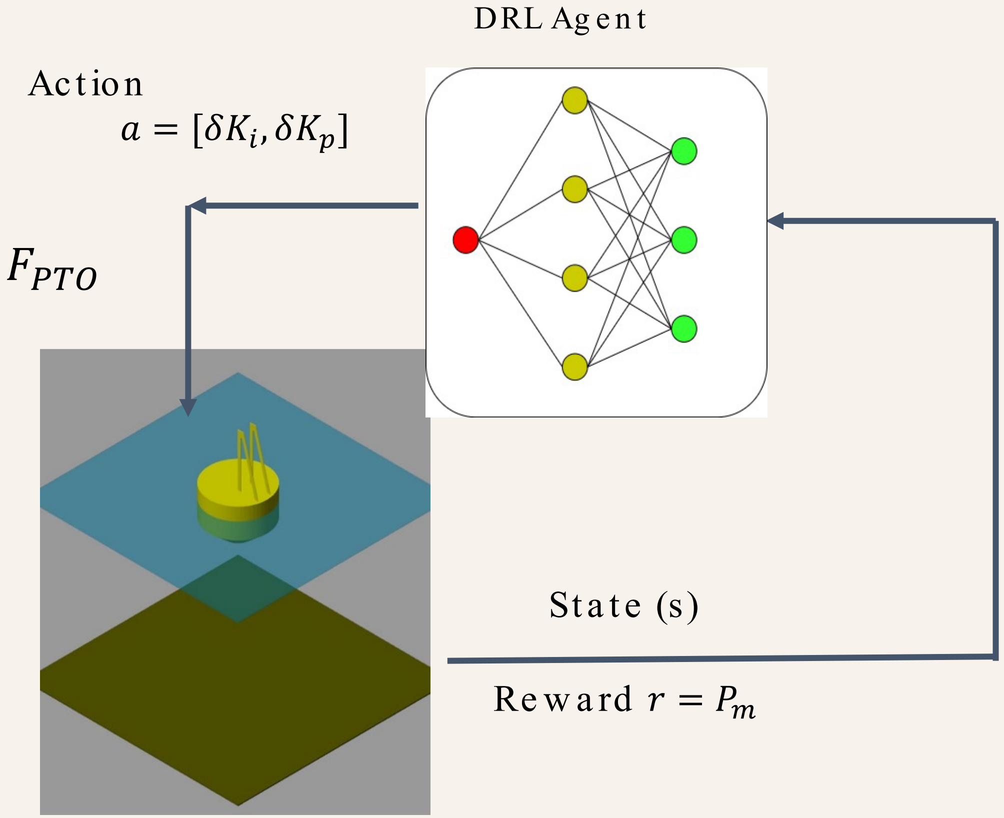
[2] Bosma, B., Beringer, C. and Robertson, B., 2023, September. Experimental passive and reactive control of a Laboratory Scale WEC Point Absorber. In *Proceedings of the European Wave and Tidal Energy Conference* (Vol. 15).

Methodology: DRL control

- Next, the DRL control is developed for LUPA to maximize the wave power production. A **time-varying PI control law** is applied:
  $F_{PTO} = -K_i(t)z(t) - K_p(t)\,\dot{z}(t)$
  where the PI gains are adapted based on discrete actions:
  $A = \{a|(\delta K_i, 0), (0, \delta K_p), (0,0), (-\delta K_i, 0), (0, -\delta K_p)\}$

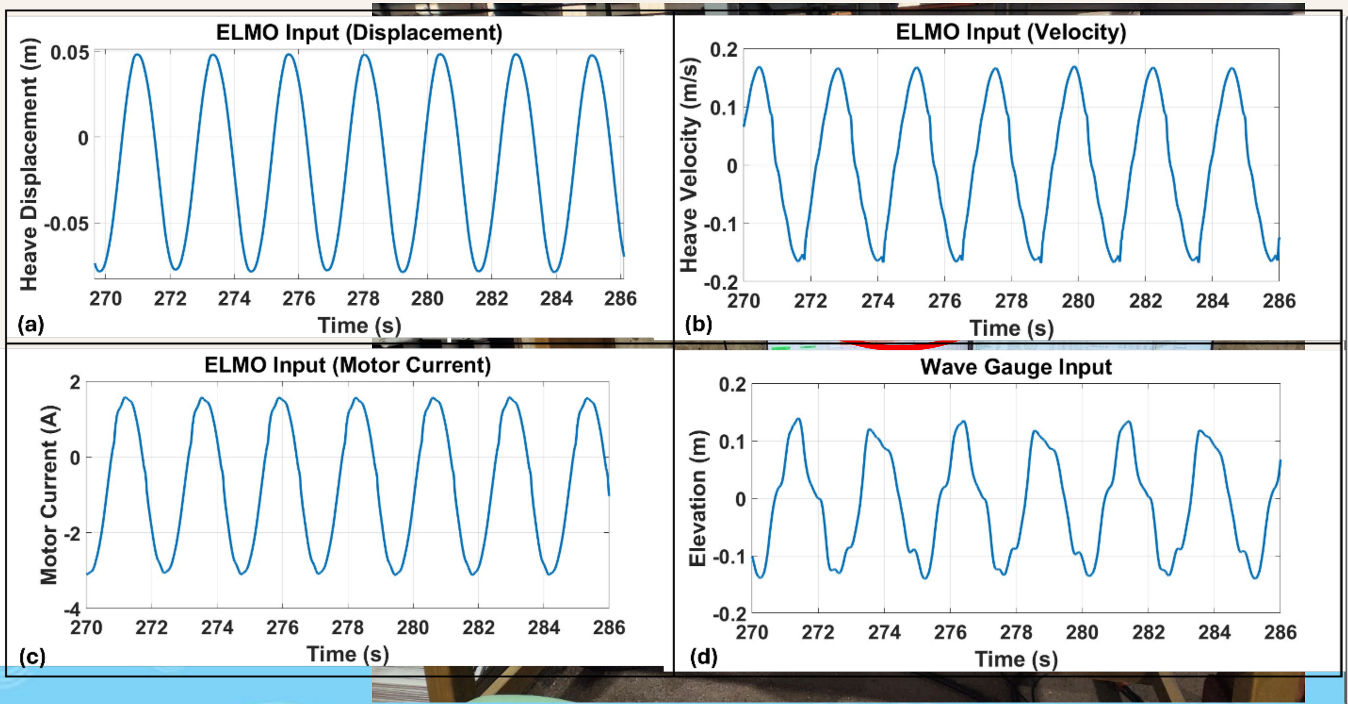
- The optimal action of the DRL agent is determined by maximizing the target network:
  $\tilde{Q} = r + \gamma \max_{a'} Q(s', a'; \theta_i^-)$ *Target network*
  $Q = Q(s, a; \theta_i)$ *Prediction network*
  where s and r represents the state and reward collected from the environment:
  $s = [z, \dot{z}]$
  $r = P_m = -F_{PTO}\dot{z}$

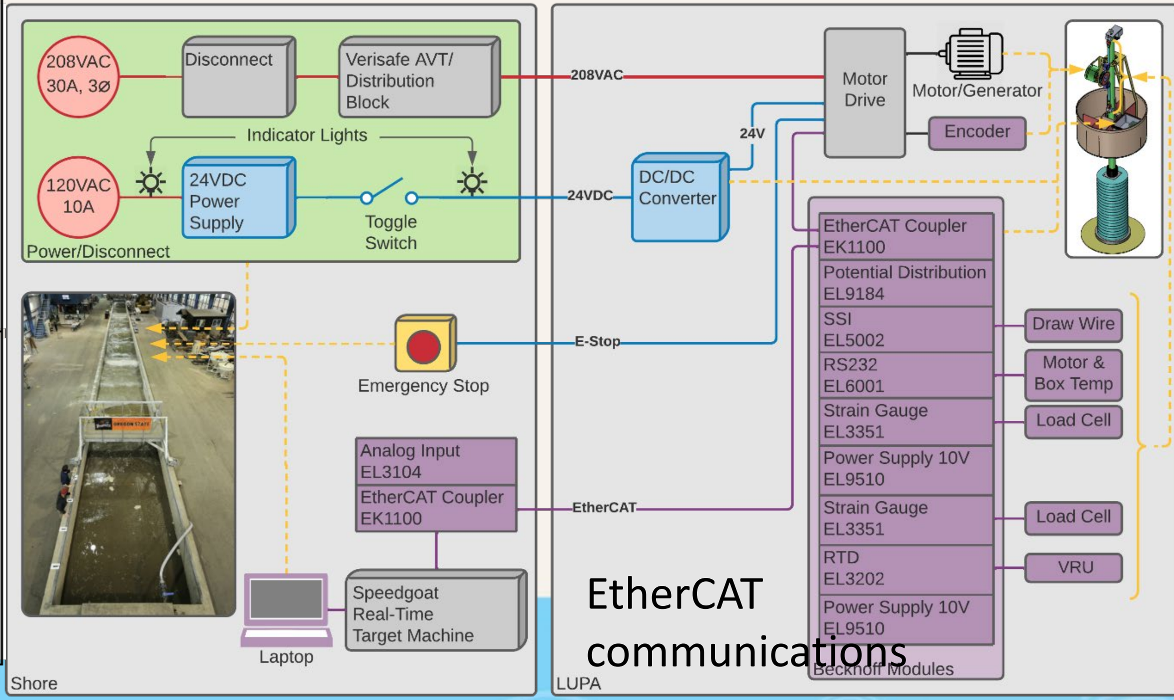
- The network parameters θ_i will be trained such that the error between the target network and prediction network will be minimized:
  $\nabla_{\theta_i} L(\theta_i) = E_{s,a,r,s'}[(\tilde{Q} - Q)\nabla_{\theta_i} Q(s, a,; \theta_i)]$



Method: Physical implementation

- The experimental setup of LUPA is presented in the figures. More specifically, LUPA uses EtherCAT connections to communicate with varied sensors and motor power electronics via a SpeedGOAT machine.
- The pre-trained control (a Neural Network) is directly deployed to the SpeedGOAT machine via a MATLAB/SIMULINK block “generatePolicyBlock” (very straightforward!).
- During the real-time tests, the processing time of the DRL control is very small, which is around 0.008ms << 1ms as the sampling rate.
- In addition, no significant noise has been found for key measurements.





EtherCAT communications

Results: Testing of the DRL control

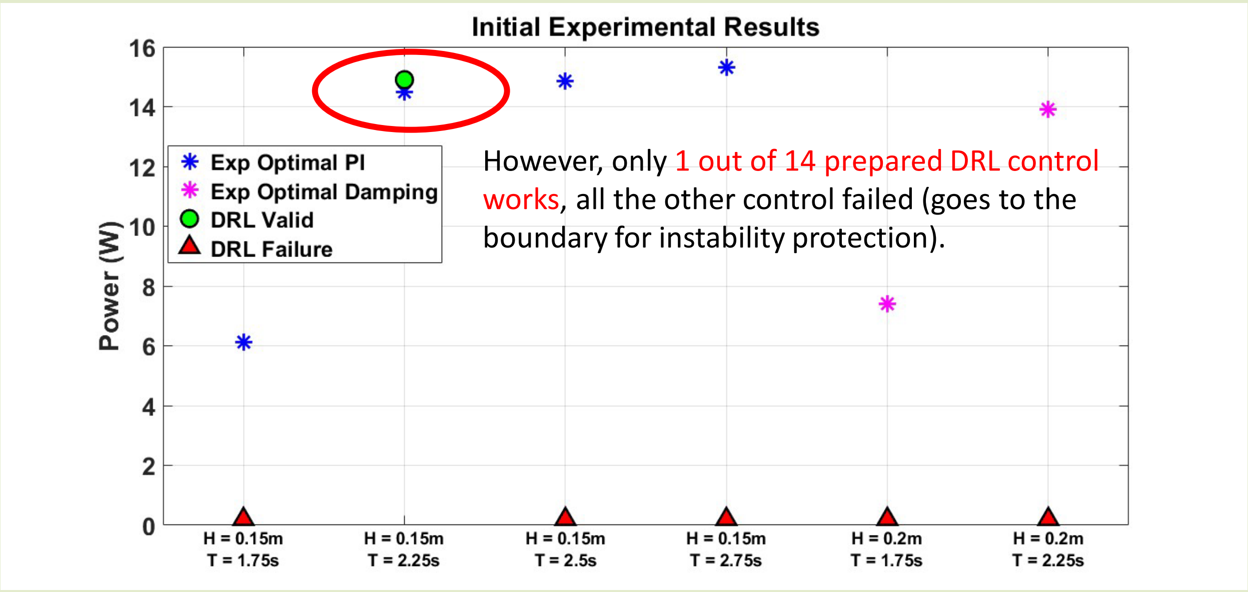
- Despite the easy process of control real-time implementation, the **control performance is very poor in practice initially**.
- An optimal feedback control (OFC) is also tested to benchmark the performance of the DRL control [2].
- It is found that the control failed due to the following reasons:

  (1) Model mismatch

  (2) Random initial conditions

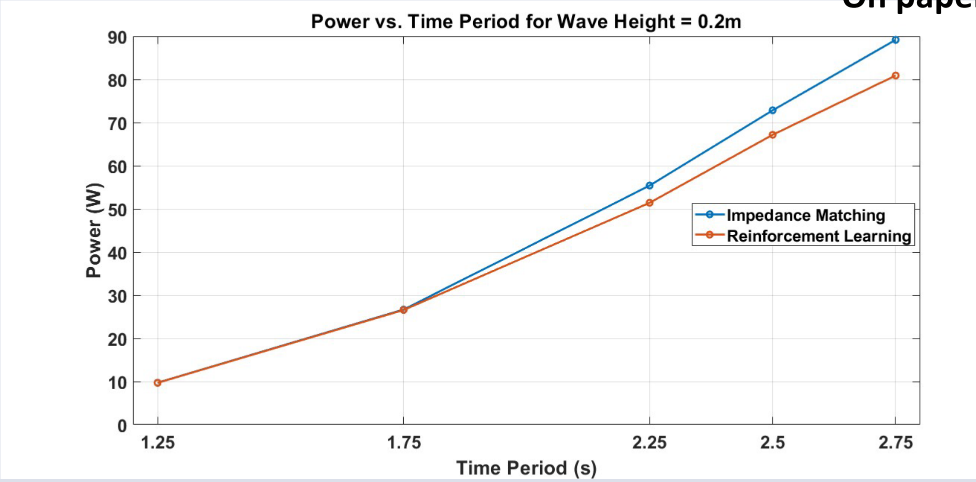
  (3) Nonlinear events

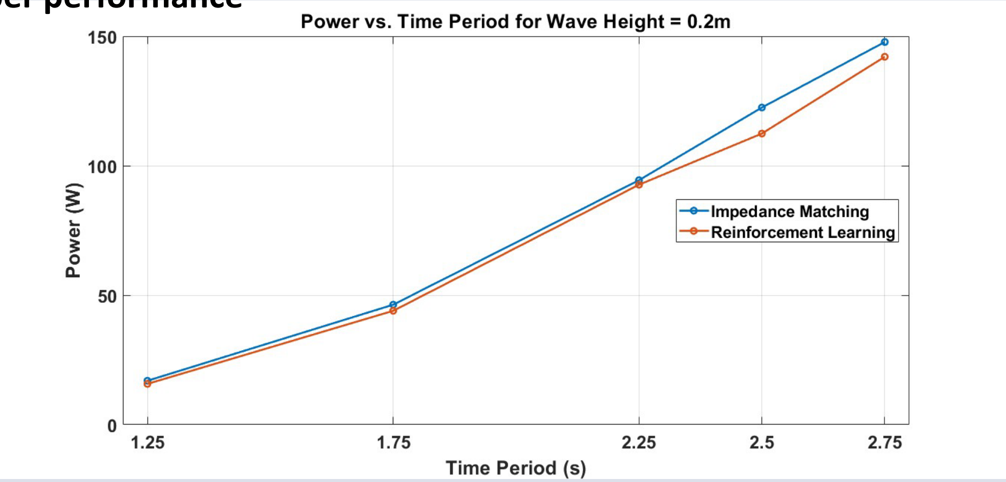
  (4) Process noises



Practical performance

On paper performance





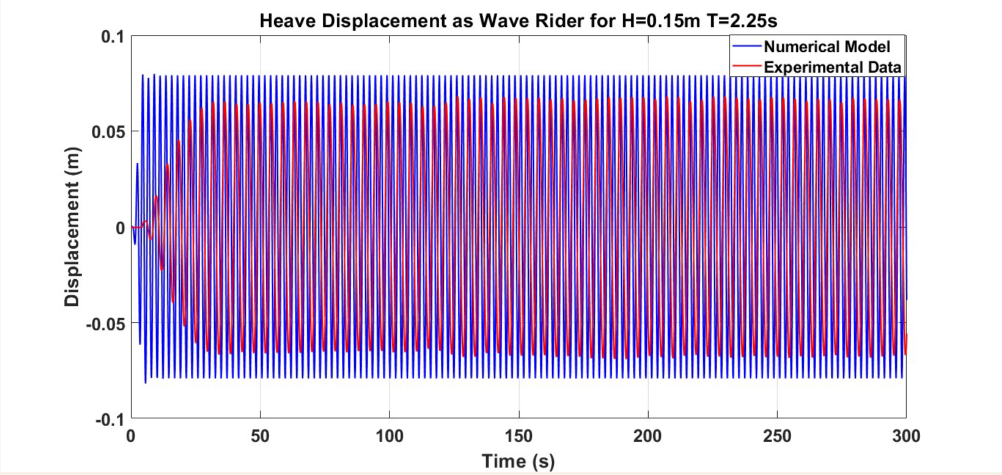
The DRL shows a very promising performance in the numerical environment, and absorbs nearly the same power as an impedance-matching control (considered as the theoretical maximum).

Results: Model mismatch

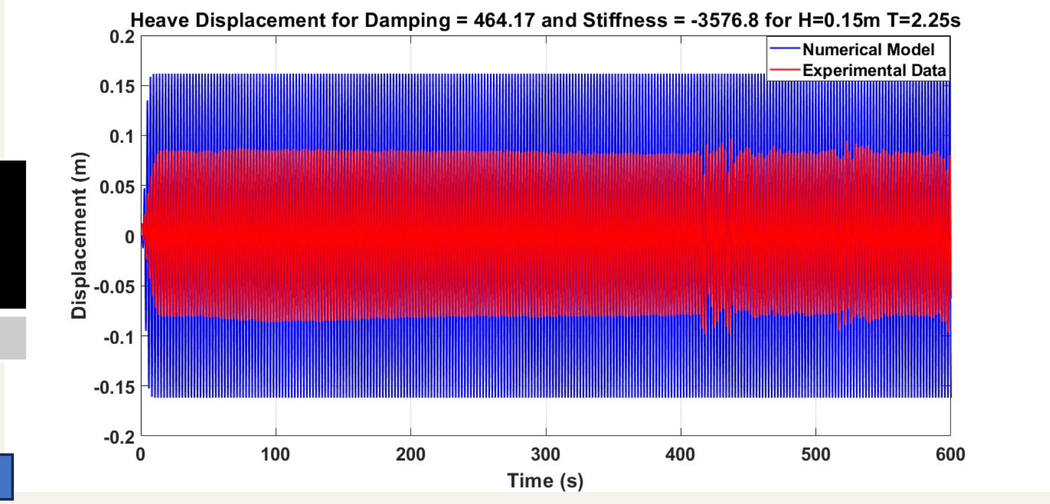
- A significant **model mismatch** is observed during the testing, mainly due to the nonlinear PTO loss.
- Motion of LUPA is **overpredicted** under both controlled and uncontrolled conditions. Physically, this is due to the significant mechanical loss in the drivetrain (belt connection with the motor).
- This additional PTO loss is dependent on the operating conditions of LUPA.
- To address this issue, additionally PTO damping is now added to the model as a linear damping with the damping coefficient being a **random number** in:

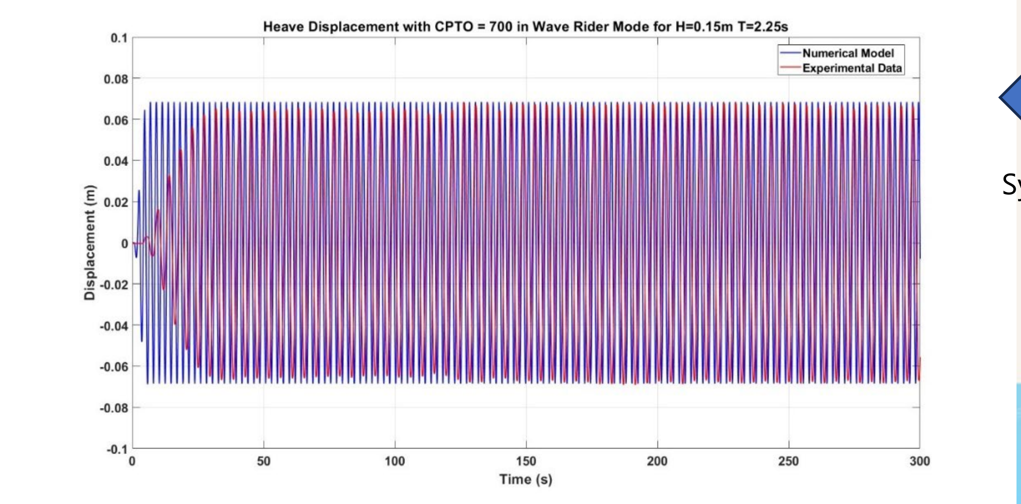
$$PTO_{loss} \in [c_{pto} - 300, c_{pto} + 100]$$

where c_{pto} is identified by matching the system responses under varied conditions.



Wave Conditions	H = 0.15 m T = 1.75s	H = 0.15 m T = 2.25s	H = 0.15 m T = 2.5s	H = 0.15 m T = 2.75s	H = 0.2m T = 1.75s	H = 0.2m T = 2.25s
c_{pto} (N.s/m)	500	700	800	900	1200	800

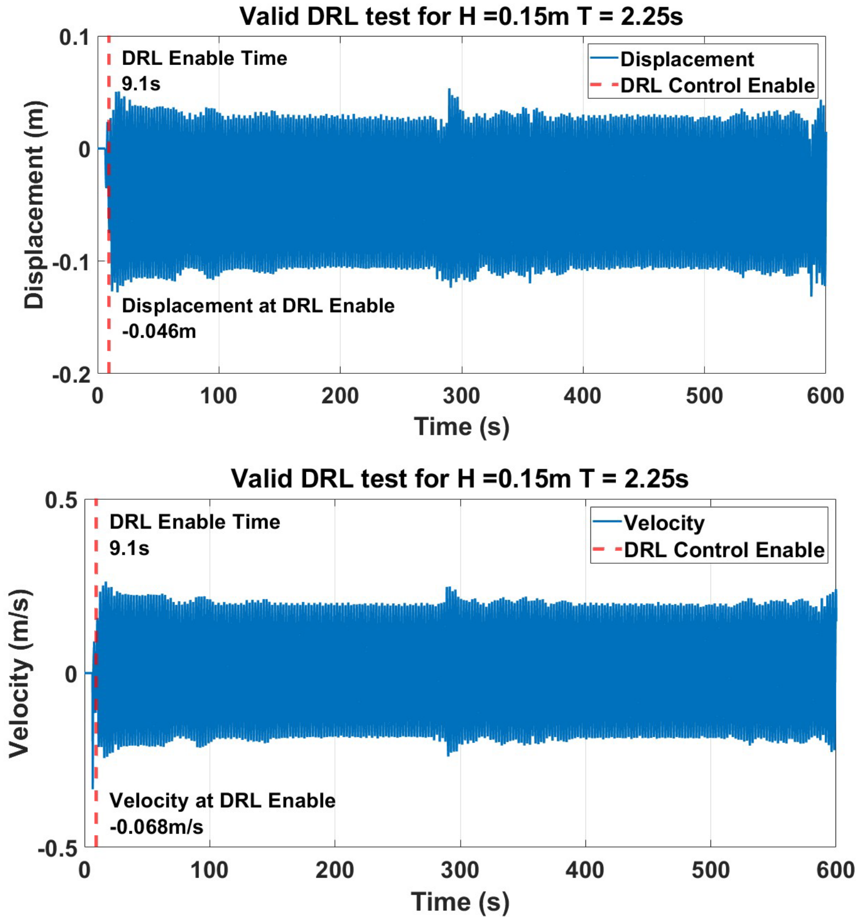


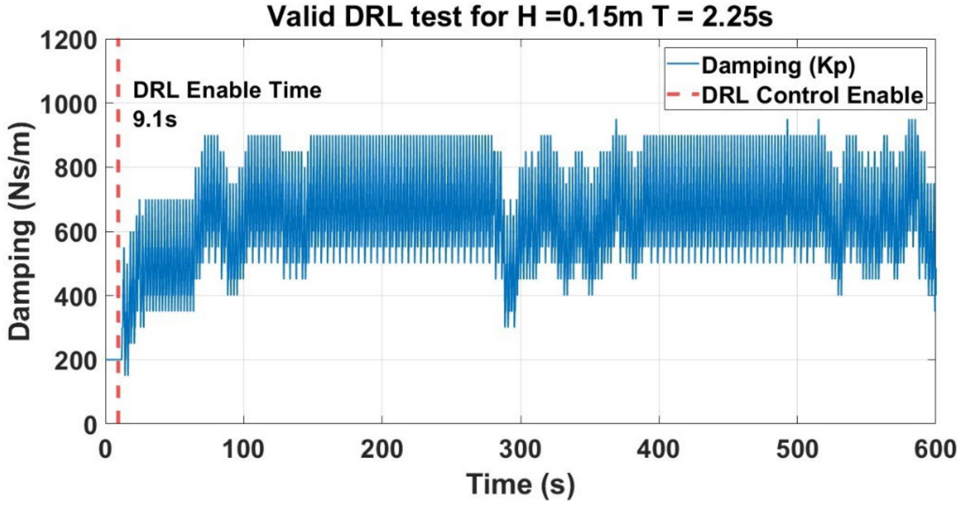


System response of LUPA after fixing this PTO loss

Results: Random initial conditions

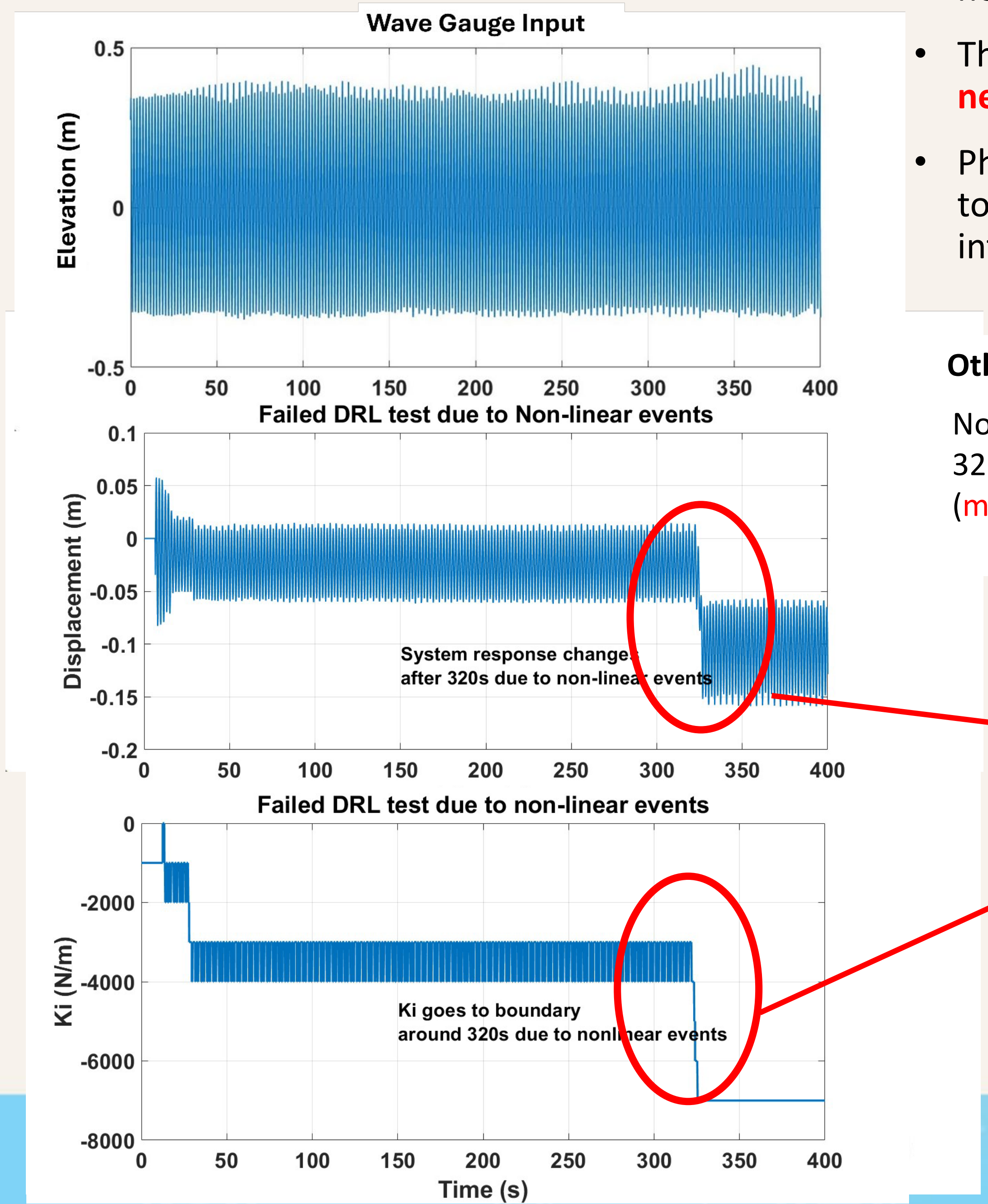
- In numerical simulations, control typically starts from **fixed initial conditions** (e.g., zero displacement and velocity). However, in tests, the control is enabled manually after a few waves pass the WEC (due to the tank setup).
- It is found that the control **is highly sensitive to initial conditions**.
- The following example shows that the control performance can be significantly different under the same wave condition, but only with a different control enable time.





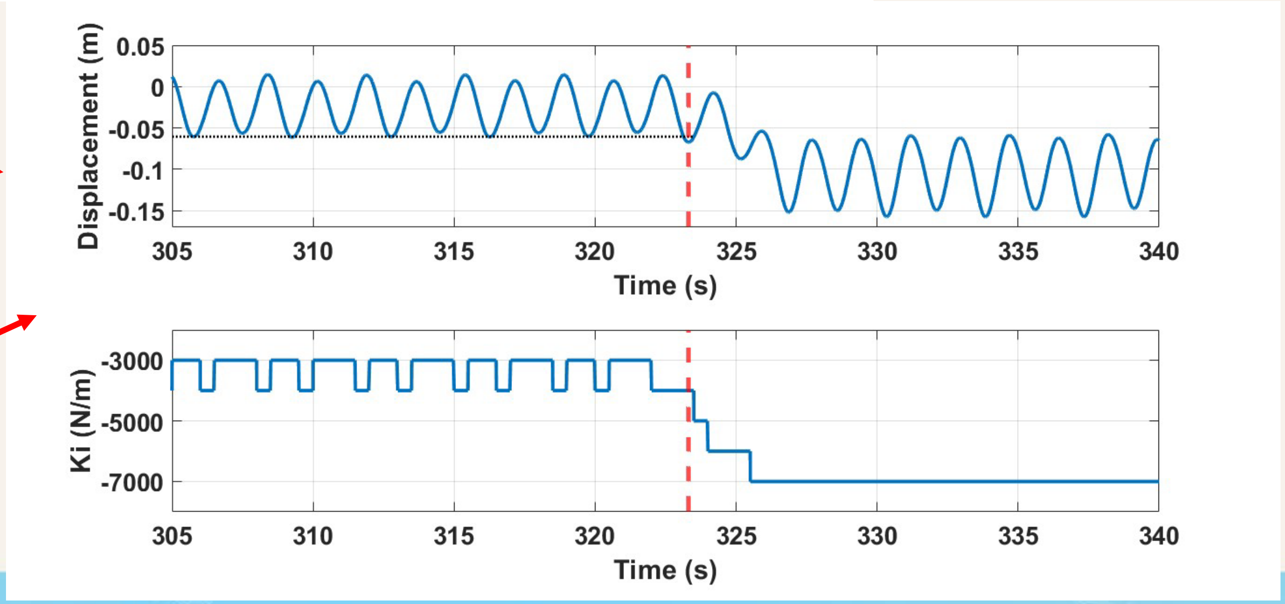
Results: Nonlinear events

- It is found that the control performance can be significantly impacted by nonlinear events.
- The so-called nonlinear events are characterized by **unexpected positive or negative impulses** in motion amplitude, which can easily disrupt the control.
- Physically, these events may be caused by nonlinear wave behavior (e.g., due to the wave paddle) or nonlinearities in mechanical drivetrain/wave-structure interaction.



Other nonlinear events

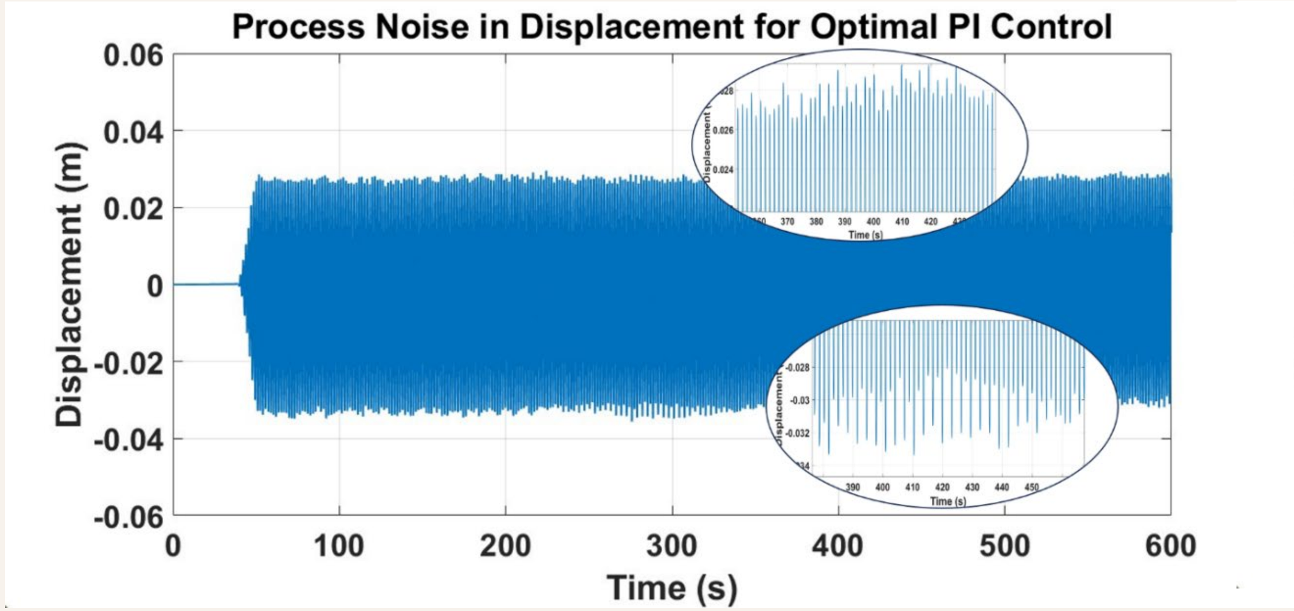
No apparent wave events, but the control goes unstable around 323s, caused by a dramatic low peak of the system response (magnitude change of 21.38%).

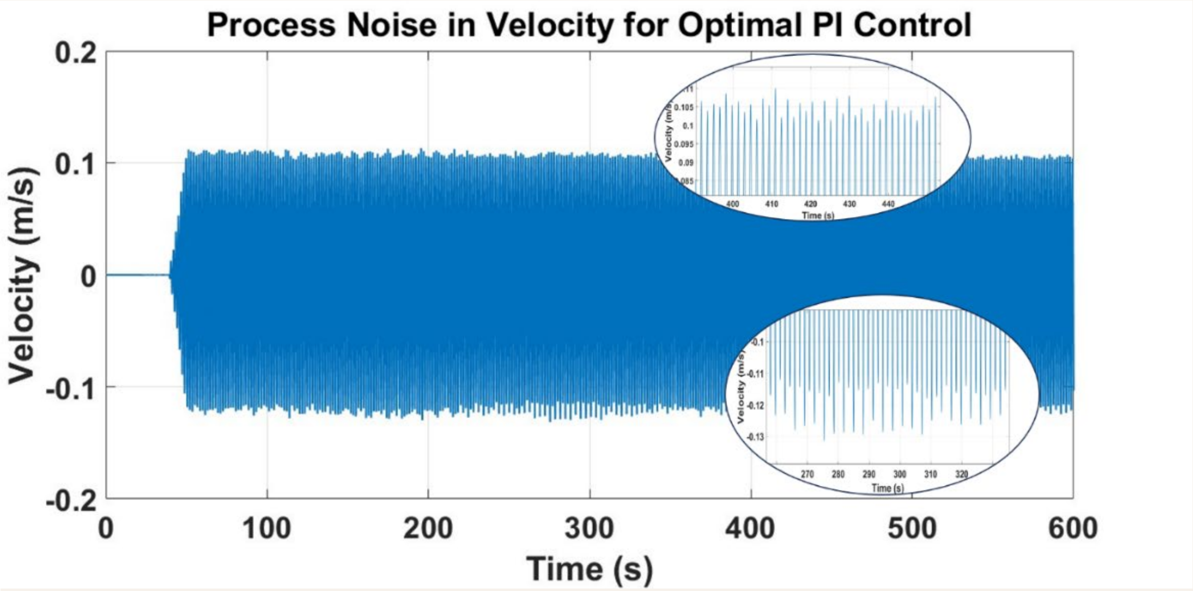


Results: Process noise

- Significant process noise is also observed during the tank tests. Although the process noise has a smaller impact on the system responses (e.g., compared to nonlinear events), it still may impact the control performance.
- Physically, this process noise may result from varied sources such as fluctuations of the waves, uncertainties, and unmodelled nonlinearities in the drivetrain and wave-buoy interaction.

Clear variation in the magnitude of the system response of LUPA even under regular wave conditions.





Results: Improvement

- The mentioned uncertainties and nonlinearities that significantly impact the control performance are now included in the training.

  More specifically:

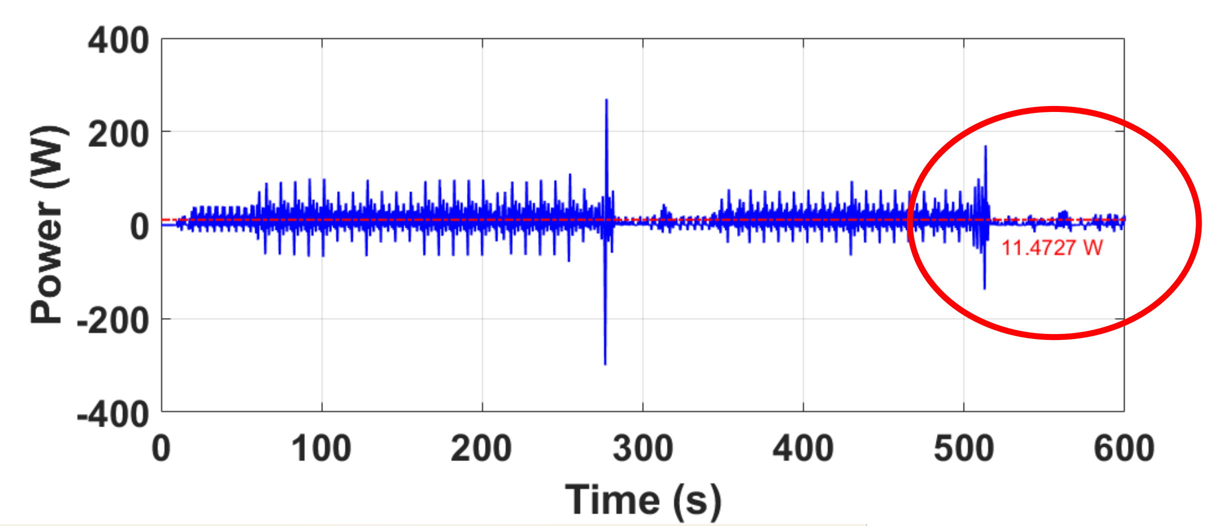
  (a) random control start time such that the initial conditions for the control are random

  (b) directly modify the system response (70% to 130%) within a random time window

  (c) normally distributed white noise

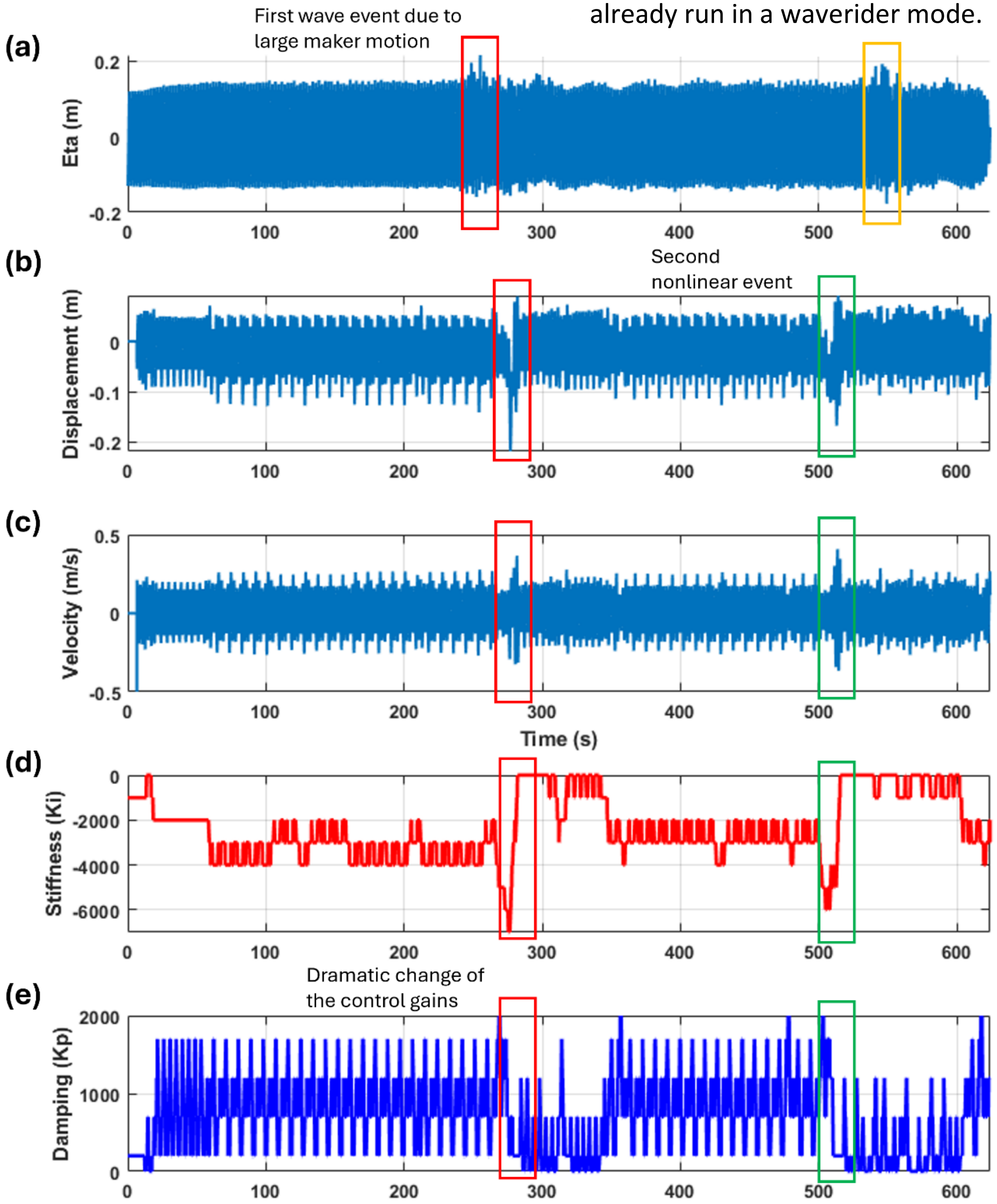
  (d) linear PTO loss with a random damping coefficient selected from a certain range.

- The control **performance is significantly improved**.



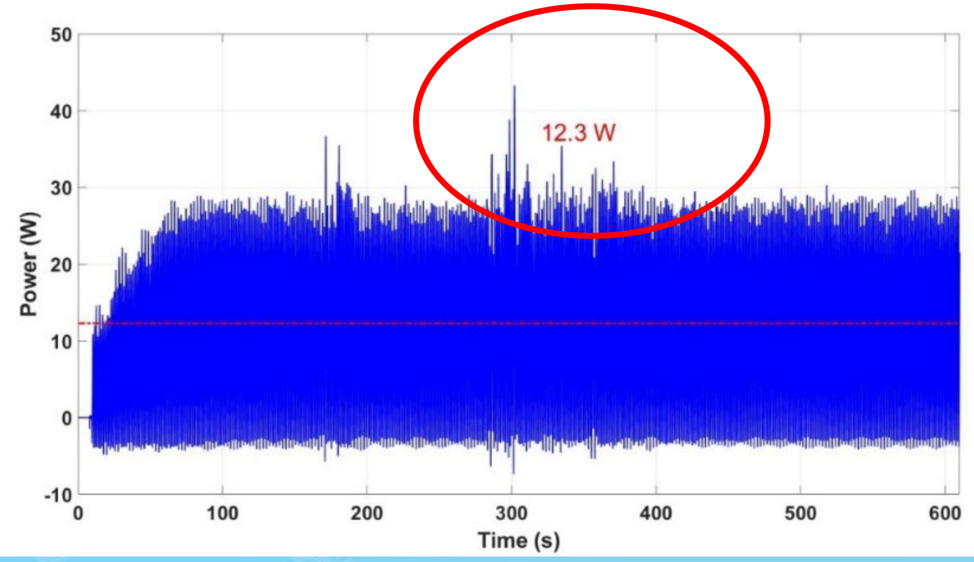
H = 0.15m and T = 2.25s

Third event does not cause noticeable motion change may because the device already run in a waverider mode.

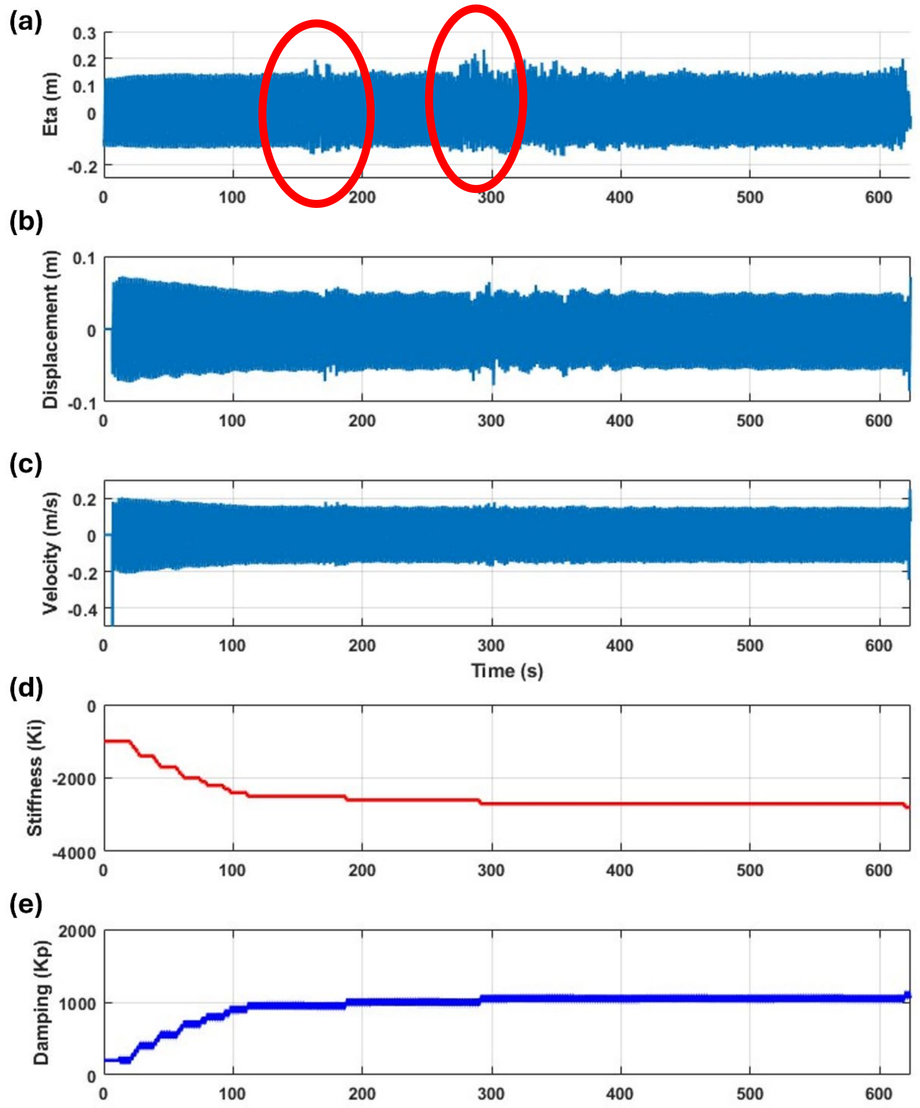


Results: Improvement with different states

- In addition to the improvement on the training environment, we have also investigated the impact of different **selections of observation states** on the control.
- Representative observation states used in the literature are tested.
- From this table, it is clear that S1 and S6 are the optimal selections.
- The detailed performance of the DRL control with S6 is further tested.



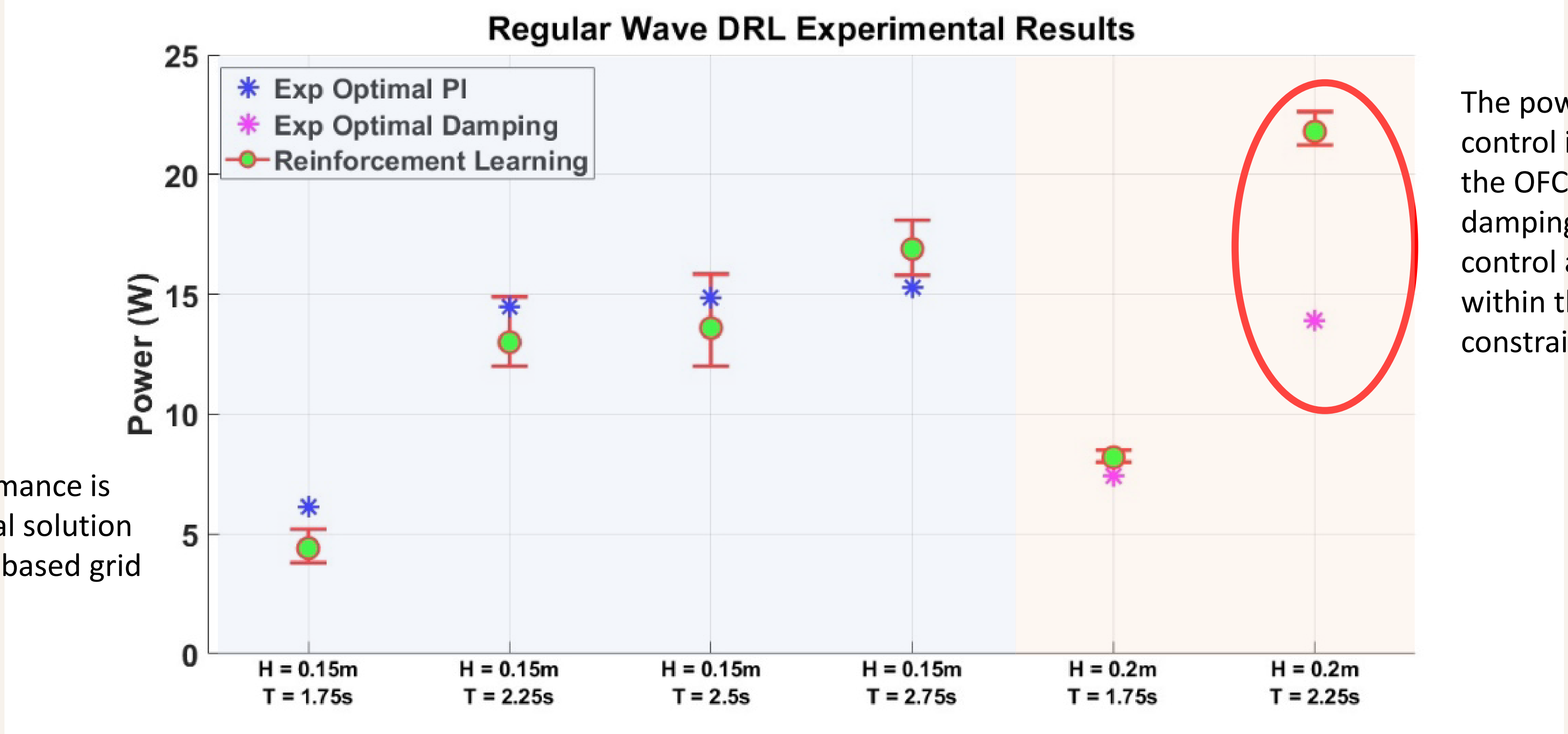
H = 0.2m and T = 1.75s



Two significant wave events occurred at 171 and 283s, the control tends to maintain the PI gains with only slight disturbances (riding through the events in contrast to being adaptive when S1 is used).

Results: Regular waves

- Comparing the performance of the DRL control with S1 and S6 as the observation states, we can tell that when S6 is selected, the control **exhibits a more robust** performance, while is **less adaptive**.
- This feature is well-suited to be applied to regular waves. Therefore, S6 is selected for all regular wave tests.
- After all the improvements, the DRL control now shows a promising performance in practice.

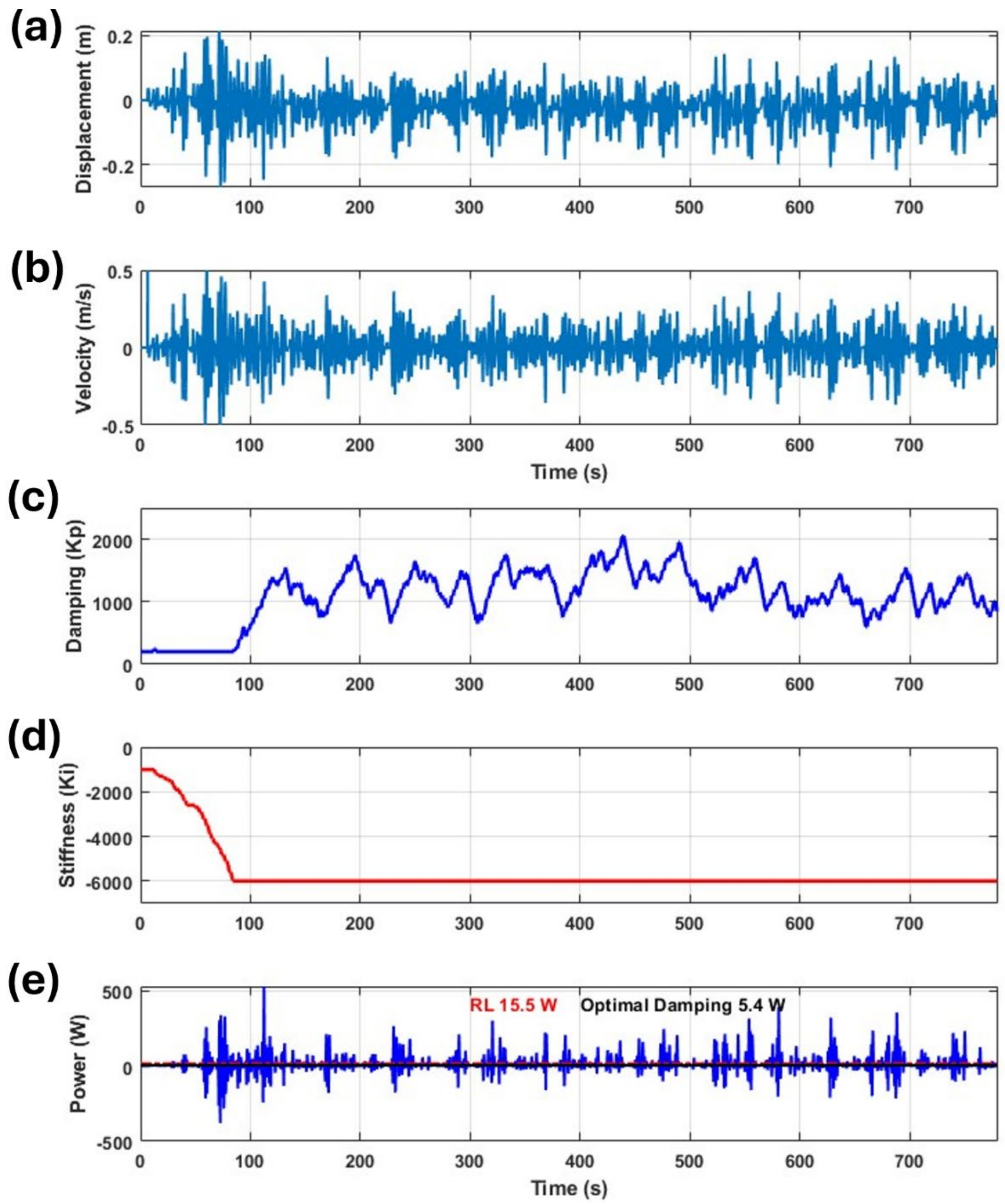


The control performance is close to the optimal solution (obtained via tank-based grid search).

The power produced by DRL control is 56.8% higher since the OFC control only applies damping, while the DRL control allows spring effect within the same motion constraint.

Results: Irregular waves

- The performance of the DRL control is further tested with the irregular wave, which has a significant height of 0.21m and a peak period of 3.09s.
- For irregular waves, the observation state **S1** (displacement and velocity) is selected instead of S6 since S6 is found to be too stiff.
- The control gain shows a strong adaptivity subject to the system responses.
- The power harvested by the DRL control is 11.6W, 14W, and 15.5W under the same wave condition with three different random seeds. **This is significantly better than the OFC control (5.4W).**



Conclusions

- The DRL control is trained using the MATLAB/Simulink Deep Learning Toolbox, which is found to integrate straightforwardly with the real-time system (SPEEDGOAT).
- The processing time of DRL control in real time is found to be very small (around 0.008 ms) and is significantly smaller than the sampling rate (1 ms).
- The robustness of the control in practice is critical, and it is important to introduce significant randomness during training (such as nonlinear events, random initial conditions, nonlinear drivetrain losses, process noises, etc.).
- The performance of the control in terms of power production and robustness is significantly impacted by the selected observation states. It is found when the displacement and velocity are selected, the control is optimal and adaptive but slightly less robust (so recommended for irregular waves). In contrast, when displacement, velocity, significant wave height, period, and prior action are the states, the control is more robust but less adaptive (so recommended for regular waves).
- The DRL control shows a promising performance in terms of power production. For regular waves, its performance is close to the practical maximum for small waves, and significantly better for larger waves. For irregular waves, the DRL control is also significantly better.

More content about this test can be found at https://mhkdr.openei.org/submissions/628

Acknowledgement:
The authors gratefully acknowledge the US Department of Energy, the Pacific Ocean Energy Trust (POET), and the TEAMER program for supporting and funding this work. In addition, the authors would like to thank Dr. Gordon Parker for his valuable discussions and for providing access to high-performance computing resources used for DRL training. Appreciation is also extended to Dr. Tim Maddux for his insightful discussions and on-site support with processing the wave gauge data.