Cost of Energy Optimised by Reinforcement Learning

WES Control Systems Stage 1 Public Report

MaxSim

This project has been supported by Wave Energy Scotland

1 Project Introduction

The goal is to apply reinforcement learning (RL) algorithms to learn good control policies for specific wave energy converters (WECs). We plan for our methods to be applicable to all types of WEC. Even though the same RL algorithms could be used for training, the control policies they produce will clearly be specific to each WEC type. A longer-term goal would be to learn individual control policies for individual WECs of the same type. The policies could change as the machine experiences wear, minor faults, biofouling or repairs. Policies would also be specific to the development level. The development level impacts the complexity and accuracy of information available about levelised cost of energy (LCOE), and the commercial focus and targets of the device developer.

The main consortium members for Stage 1 were Paul Stansell (MaxSim), Alexandra Price (Wave Conundrums Consulting), and Alex Hagmüller and Max Ginsburg (AquaHarmonics). Specialist advice was provided by David Forehand (University of Edinburgh), David Pizer, and Max Carcas (Caelulum). Our WES project partners were Mocean and CorPower.

2 Description of Project Technology

2.1 Inputs and Outputs

Market research conducted with our WEC developer partners, CorPower, Mocean and AquaHarmonics allowed us to assess the nature of the inputs and outputs available to the RL system. The inputs to the control system could be all or any of the sensor data typically available on a WEC. The control system could also have access to simple design-fixed parameters. There would be the option to include data from the shore, or from nearby sensors such as wave buoys.

The outputs would be all the controllable aspects of the WEC that have an impact on cost of energy. For all WECs this would include the demand load of each power take-off (PTO) unit. Depending on the type of WEC, the controllable aspects could include a signal to adjust the geometry of the hydrodynamic interface, the mooring extension, physical negative spring, ballast, flywheels, etc.

2.2 Design drivers

This project addresses three challenges for control of WECs. It has the potential to overcome the difficulties of:

- **Model-based control:** model-predictive control is only as good as the model; it needs bespoke development for each device type and each particular machine, and it requires that a human understands the dynamics and mechanisms. The RL methods we propose are not required to perform system identification because they optimise the control policies directly without the need for a model.
- **Sea-state control:** calculations of spectra require 0.5-2 hours of data and are always centred on a time in the past. The assumption of statistical stationarity is often invalid (e.g. due to tidal interaction). Even when this assumption is valid, waves tend to group, and the optimal PTO damping for a group of small waves is likely to be different to that for a group of large waves. The RL method we propose will use as inputs the instantaneous state of the local wave environment so will not need averages of spectra measured over extended periods of time.
- **Making improved power capture the main purpose of the control system:** investors require both performance and reliability targets to be met. Focussing only on maximising power capture could result in poor reliability, and detracts from the opportunities to use control to reduce LCOE by controlling both

performance and reliability. If, for example, capturing the last possible increase in energy yield came at the cost of a disproportionate increase in wear and fatigue, this would likely not reduce the LCOE.

Often the economic viability of a WEC is thought of in terms of how well it performs on a number of competing metrics, such as capture width and peak loads. WES lists about eighteen such metrics, many of which are tightly coupled, meaning that a control that improves one is likely to worsen another. For example, an increase in peak performance (which reduces LCOE) is likely to be accompanied by an increase in peak loads (which increases LCOE). This trade-off between desirable metrics is illustrated schematically in Figure 1.

In this project, we focus on the primary goal of reducing the LCOE. It is therefore self-evident that LCOE should be the primary metric on which to judge success of achieving the goal. As RL attempts to find a control policy that minimises a long-term reward function, it is clear that the long-term reward function for RL should be LCOE.

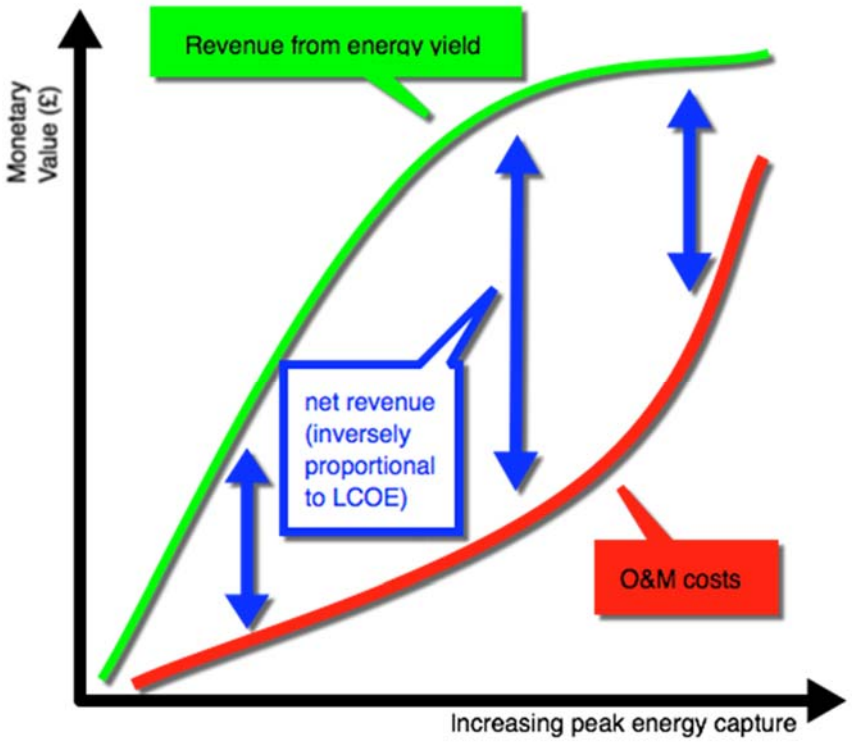

Figure 1 Illustrative graph depicting how RL seeks to maximise the return (blue) to minimise LCOE

However, a difficulty with this approach is that at the current stage of technology development of WECs, it is not clear how the various metrics listed by WES contribute to the LCOE for the many different types of WECs available. When applying RL it will be crucial to have as accurate a representation of LCOE as possible given the technology readiness level (TRL) of the WEC. For WECs at low TRLs a simple proxy for LCOE could be used, such as mean power divided by peak load. For WECs at higher TRLs more advanced proxies could be used, for example, ones that include the replacement cost and likely reduced availability resulting from the failure of a component.

Control actions should be made on a time-scale that optimises the performance-reliability trade-off, rather than choosing fixed coefficients for a given sea-state. We describe our approach as wave-by-wave control. However, the time-scales for varying damping will be less than a wave period to allow for limits to be

implemented when thresholds are exceeded. The control system could also adjust mechanisms that have a cost overhead to change, and hence should be changed on a longer time scale.

2.3 Reinforcement Learning

An extensive literature review summarised how reinforcement learning worked in general and how it could be applied to the problem of WEC control. RL continually updates the control policy that is being applied to the system to maximise the sum of rewards over the long term. A reward is incurred with every action taken, and can include negative rewards to represent penalties. Neural networks can be used to learn how an action taken in a given state is likely to affect future rewards.

RL has been successfully applied in many areas, including dynamic control, and has been shown to outperform human-designed solutions. There are many open-source resources available for development, and every year the research community improves existing algorithms. There is a wide range of methods which can be used to address particular problems. For example, there are risk adverse training methods which can be used to avoid risky exploratory actions. Other methods are sample-efficient, enabling fast learning. Work has also been done on the transfer of policies trained in one environment to a slightly different environment. This suggests that RL can be pre-trained on a simulation, and the policy can then be honed on a real device.

2.4 Potential benefits

Stage 1 investigations, in particular the feasibility study, gave us confidence that this approach was well suited to the problem at hand. We foresee the following benefits of the system once it is mature:

- The reward uses the best available proxy for LCOE at a given technology readiness level. It represents the trade-off between volumetric efficiency and reliability.
- RL policies would be updated throughout the WEC's life. Reasons for making updates include more information becoming available about the reliability of components or the impact of reliability on LCOE, and changing behaviour due to wear, faults, biofouling or repairs.
- RL has been demonstrated to produce excellent results where a mathematical model of the underlying process is not known. It is used to optimise the performance of simple robots by acting directly on the device without the use of simulations.
- RL policies can be tailored to specific WEC devices. This can help the integration of separately developed sub-systems into one unit.
- Pre-training of policies on simulations can make the training on the actual device faster and safer.
- This approach has the potential to not only reduce LCOE, but also risk. Both need to be targeted to attract investors back to wave energy.
- Once it has been demonstrated that the controller can keep loads, stroke and power within design limits in real seas, future design iterations of WECs can include this controller in the optimisation. This will result in a second step-reduction in LCOE. Mechanisms include CAPEX reduction, upscaling, or choosing appropriate relative sub-system ratings.

3 Scope of Work

Stage 1 of the CEORL project consisted of the following activities:

- **Market research:** understand the technical and commercial requirements of developers with different design strategies, at different TRLs; build a list of input and actuator characteristics; explore how control could increase performance and reliability.

- **Literature review:** identify methods and techniques suitable for application to simulated WECs, as well as the real WECs explored in the market research.
- **RL methodology review:** outline the mathematical underpinnings of the method, demonstrate how the equations would apply to a simplified expression of the AquaHarmonics control problem, anticipate risks in the development process, demonstrate feasibility.
- **Identification of modelling requirements and development platforms:** outline WEC simulation method necessary to implement RL and to demonstrate its applicability and advantages, identify RL development tools and explore options for integration with WEC models, demonstrate the feasibility of the development process.
- **Metrics:** identify the information needed to calculate LCOE and the TRL required to have access to that information, review literature about proxies for LCOE, choose metrics and objective functions suitable for the CEORL project, describe anticipated improvements to WES Target Metrics.
- **Interfaces:** consider hardware and software requirements for the CEORL project and beyond, identify which real-life aspects of sensors and actuators are essential to model, demonstrate practical feasibility.
- **Planning Stage 2:** design the R&D process, identify risks and resource requirements, build in contingencies, apply for additional funding, engage collaboration partners.

4 Project Achievements

The information generated by Stage 1 was valuable to the planning of Stage 2. It helped pinpoint challenges and constraints, and it identified the need to recruit specialist software development expertise. Many of our initial technical assumptions were challenged. For example, at the start of Stage 1, we were anticipating producing a standalone control system with discrete state and action spaces. Stage 1 also gave us the opportunity to discuss our ideas with the wave energy community, strengthen our arguments for the novel approach we are taking, and gain confidence that this approach has the potential to make a step change in the prospects for wave energy.

Our market research highlighted that different WEC designs have different types of design limits, and differences in dealing with load shedding and storm protection. The one requirement that was the same for all three partners was for a software component that could be incorporated into an existing system. The methodology review identified specific RL methods to try first in Stage 2. Work with our partners, AquaHarmonics, identified rewards and metrics that were appropriate and complementary to their development process.

The record keeping and electronic collaboration system worked well. However, many administrative tasks fell by the wayside. Our initial estimates of the time required for these admin tasks, as well as internal and external communications, were over-optimistic. In fact, all our estimates of task durations were insufficient, leading to overwork, late milestones, and delays in administrative tasks. This has been accounted for in the schedule of Stage 2, which has had more input from the specialists responsible for various tasks. The project also started over two weeks late as it was discovered that the employment contracts of consortium members were potentially in conflict with WES's IP requirements. We are now aware of this risk and can take appropriate action in Stage 2 to prevent this delay in start date. Max Carcas has joined the consortium and will assist the team to keep to schedule.

5 Recommendations for Further Work

The key steps for commercialisation of this idea consist of methodology development using simulations, implementation in tank tests, and implementation in sea trials. The objectives of the simulations, which will be conducted in Stage 2, are as follows:

- Demonstrate that RL algorithms applied to WECs can converge to good control policies that optimise a proxy for LCOE.
- Demonstrate that RL can improve on sea-state control even when it is not given access to information about the spectrum.
- Demonstrate that an RL algorithm can discover a good policy while operating safely and stably, with access to information that is functionally representative of the real-life situation, in terms of number of samples, sensor uncertainty etc.
- Demonstrate that a policy trained on a particular WEC model can be transferred to a slightly different, or perturbed, WEC model.

Furthermore, Stage 2 will provide evidence that it is feasible to use control actions to improve reliability and that this is as valuable to improving LCOE as is increasing performance.

The purpose of the tank tests is to provide firm evidence of the practicality and advantages of an RL approach to control, as well as to discover and address practical challenges. Improvements will be made to the objective function and metrics that reflect the practical limitations of tank tests and the commercial requirements of developers at lower TRLs.

The purpose of the field trials is to discover and address the practical challenges of operating at sea. The waves are not repeatable, sea states are less stationary, tidal flows and head changes are present, failures are costlier, and communications to shore cannot be guaranteed.

6 Communications and Publicity Activity

The CEORL project presented a poster at the WES annual conference.