清華大學
电机工程与应用电子技术系
Department of Electrical Engineering
Tsinghua University

Centre for Ocean Energy Research

Maynooth University
National University of Ireland Maynooth

MPC-based real-time energy management of hybrid energy storage systems for wave energy converters

Xuanyi Zhu[1], Zechuan Lin[2], Xuanrui Huang[1], Kemeng Chen[1], Yifei Han[1] and Xi Xiao[1]

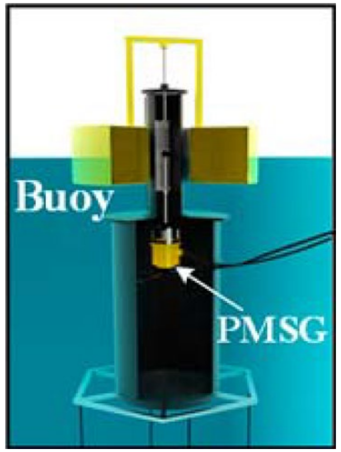
[1] Dept. of Electrical Engineering, Tsinghua University, China
[2] Centre for Ocean Energy Research, Maynooth University, Ireland

Content

Background

Grid integration of wave energy

- Wave energy converters produce **oscillatory (usually bidirectional) power**.
- On the other hand, the grid wants **unidirectional and stable** power input.
- Power-smoothing needs to be achieved using **energy storage systems**.
- Additionally, energy storage is essential for enhancing the resilience of offshore grids.



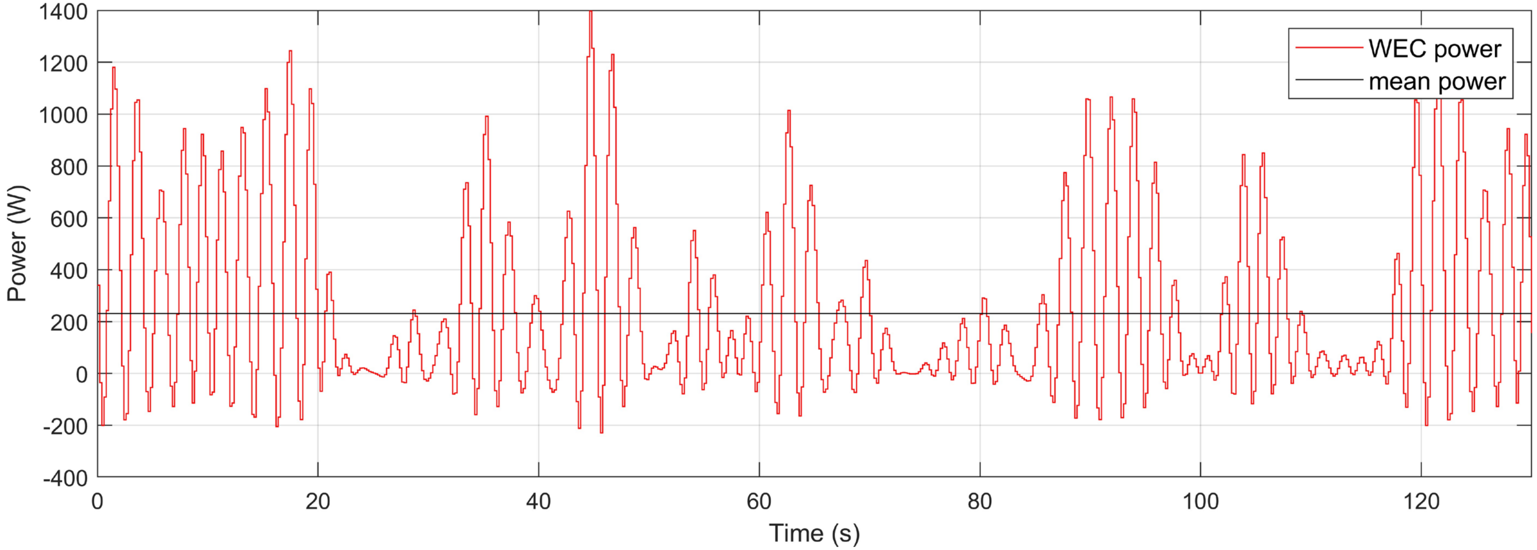
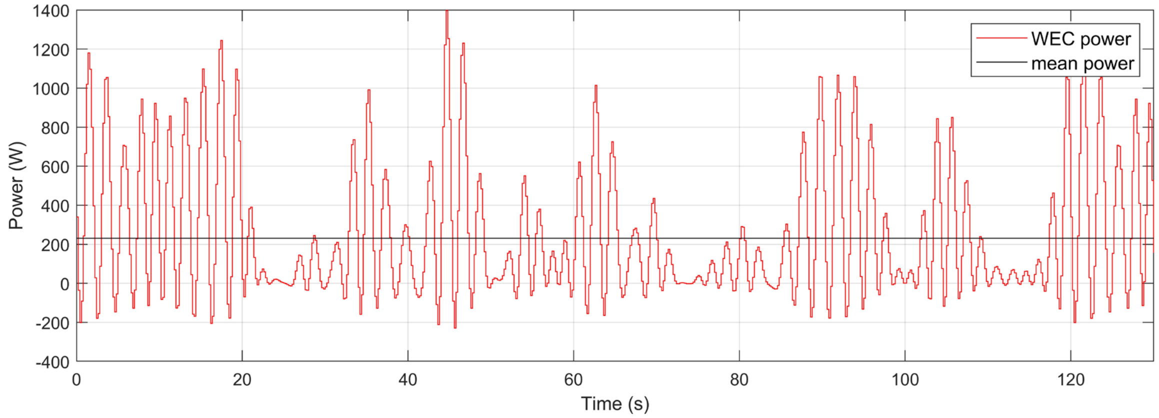


Figure: Output power profile of a point-absorber wave energy converter under reactive control.

Why hybrid energy storage system (HESS)?

- Wave power exhibits **multi-time-scale** fluctuation:



1. Within each wave cycle (e.g., 0-6 s).
2. Across different individual waves (e.g., 10-60 s).
3. Sea-state changes (e.g., 30 min).

- Different energy storage devices have different strengths.

Supercapacitor (SC): High power density, low energy density. Suitable for absorbing fast power fluctuations.

Battery: High energy density, low power density. Suitable for absorbing slow power fluctuations.

Hence, combination of SCs and batteries (HESS) is suitable for WEC power smoothing.

Energy management strategy (EMS) of HESS for WECs

- The task of EMS is to **optimally share the power** between the SC and battery, given the WEC input power and respecting the grid demand.
- The EMS needs to consider several (conflicting) **performance metrics/constraints:**

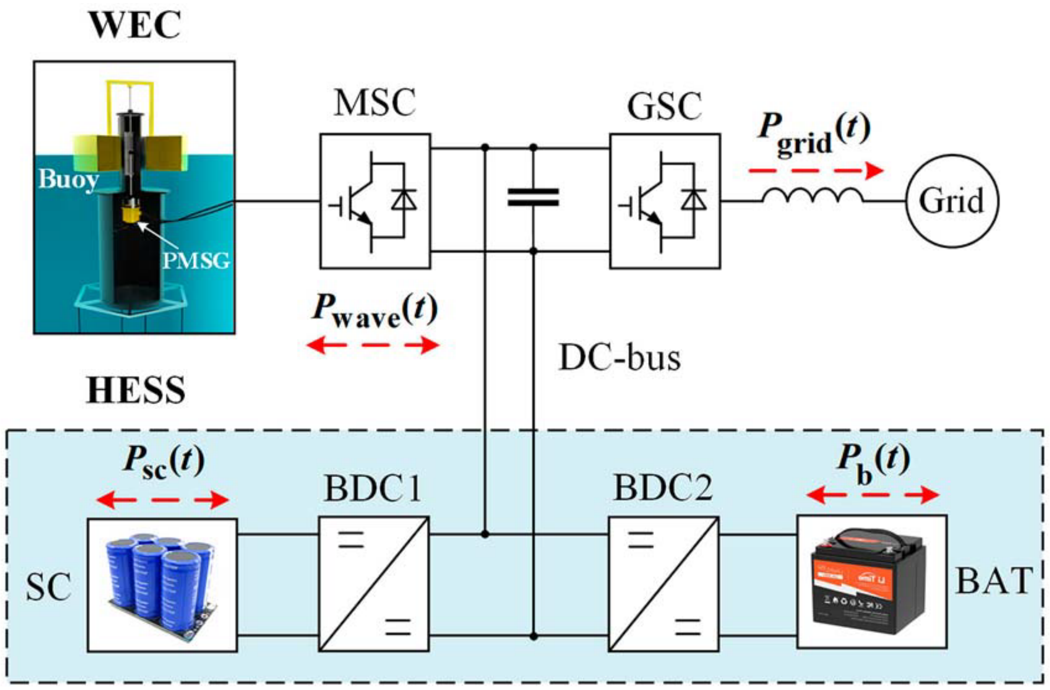


Figure: Diagram of the WEC-HESS system.

1. Track the grid power reference (if given), or appropriate grid power.
2. HESS power conversion losses.
3. Battery lifetime, represented by the Ah throughput, current changing rate, …
4. Management of state-of-charge (SoC) of the battery and SC.
5. Current limits of power converters.
6. …

The low-pass filter approach

- A typical EMS for WECs is based on **low-pass filters (LPF)**[1].
- The grid takes **low-freq.** components. The battery takes **mid-freq.** components. The SC takes **high-freq.** components. Aligning with the HESS principle for WECs.

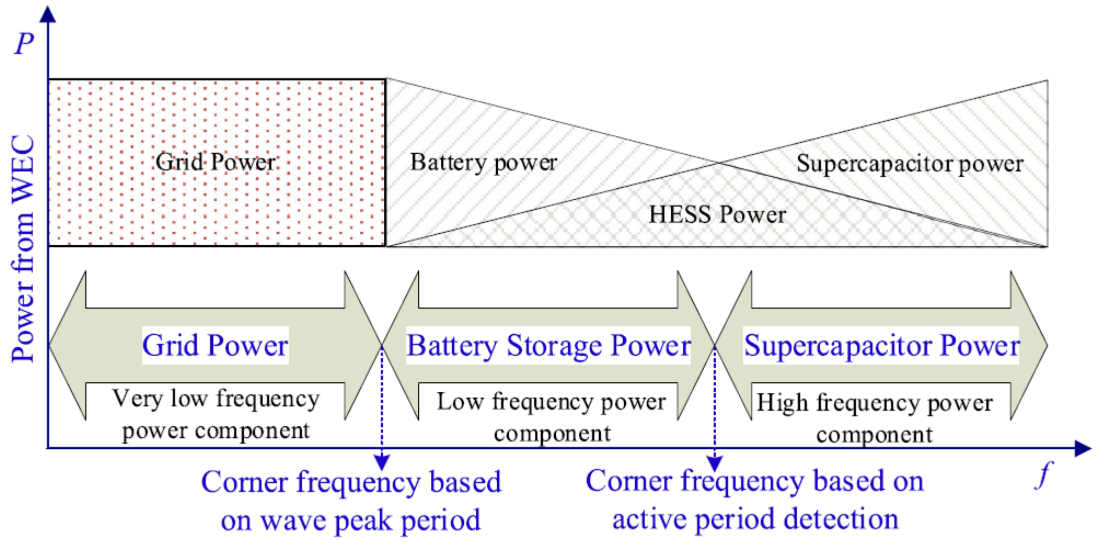


Figure: Diagram of the LPF-based EMS.

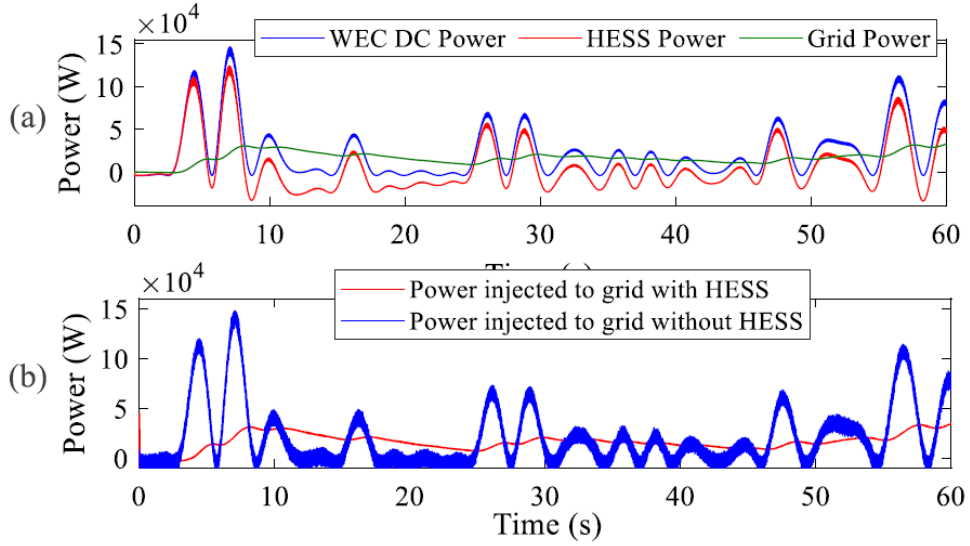


Figure: Control waveform of the LPF method.

- The LPF approach is intuitive but only has limited degrees of control freedom. Cannot **optimally** considers (balances) all control objectives.

---

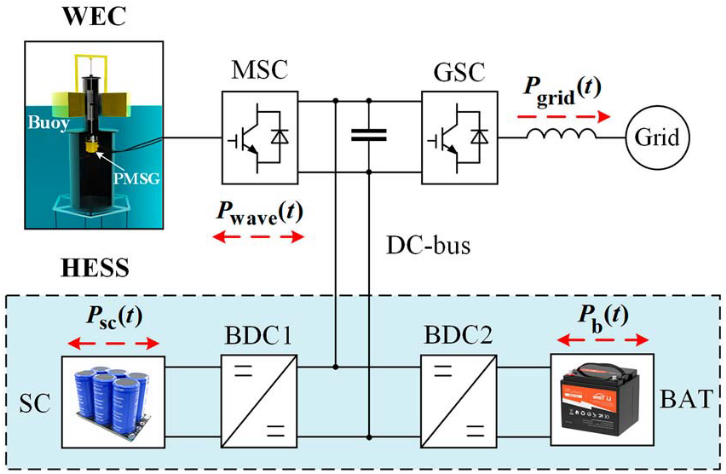
[1] S. Rasool, K. M. Muttaqi and D. Sutanto, "A Multi-Filter Based Dynamic Power Sharing Control for a Hybrid Energy Storage System Integrated to a Wave Energy Converter for Output Power Smoothing," in *IEEE Transactions on Sustainable Energy*, vol. 13, no. 3, pp. 1693-1706, July 2022

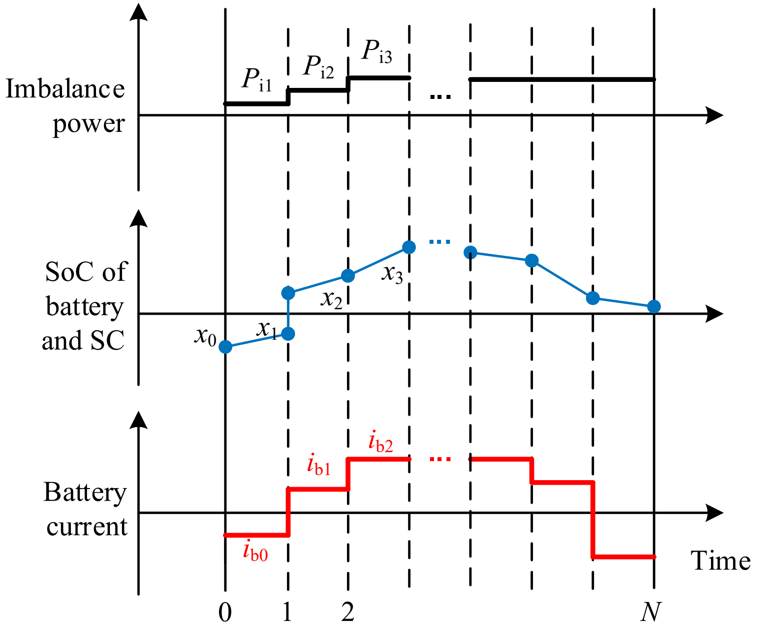
MPC-based EMS

EMS based on model predictive control (MPC)

- **Optimisation variable:** Battery current $i_b(k)$.
- **Known external input**: Imbalance power $P_g(k) - P_w(k)$
- **Objective functions:**
- $J_1 = \sum_k E_{loss}(i_b(k))$, total HESS loss
- $J_2 = \sum_k |i_b(k)|$, battery throughput
- $J_3 = \sum_k \left(i_b(k+1) - i_b(k)\right)^2$, current changing rate
- $J_4 = \sum_k \left(u_{sc}(k) - u_{sc}^{ref}\right)^2$, penalty on SC SoC drifting
- **Total objective** $J = w_1 J_1 + w_2 J_2 + w_3 J_3 + w_4 J_4$

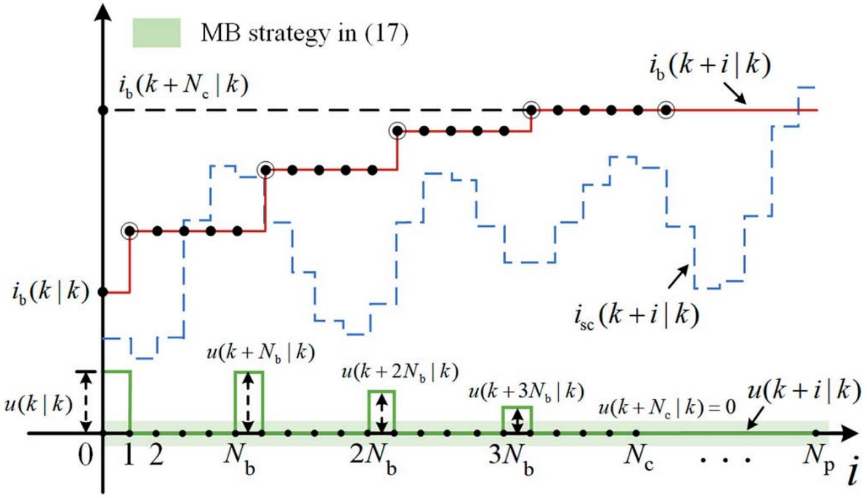
  where w_{1-4} are **weighting factors**.
- **Constraints:** $i_b^{min} \leq i_b(k) \leq i_b^{max}$, $u_{sc}^{min} \leq u_{sc}(k) \leq u_{sc}^{max}$



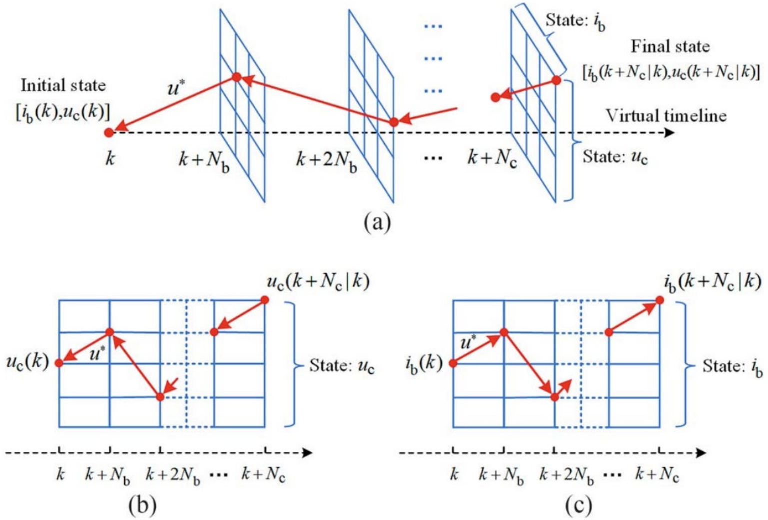


Fast real-time computation of MPC

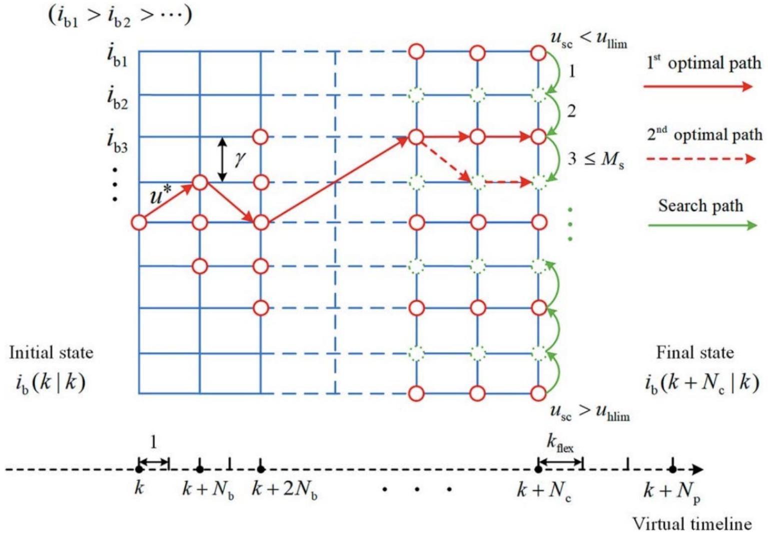
- With $i_b(k)$ the control variable, the system model, objective function, and constraints are **highly nonlinear**, challenging for real-time optimization (NMPC).
- A fast control solution algorithm is developed, integrating several techniques: i) **Move-blocking**, ii) **forward dynamic programming (FDP)**, and iii) **fast searching method**.



i) **Move-blocking** technique, reducing optimisation dimension.



ii) **FDP** is more advantageous than BDP with known initial state and uncertain final state.



iii) **Fast searching method**: flexible(state, time) resolution, heuristic search direction.

Experimental validation

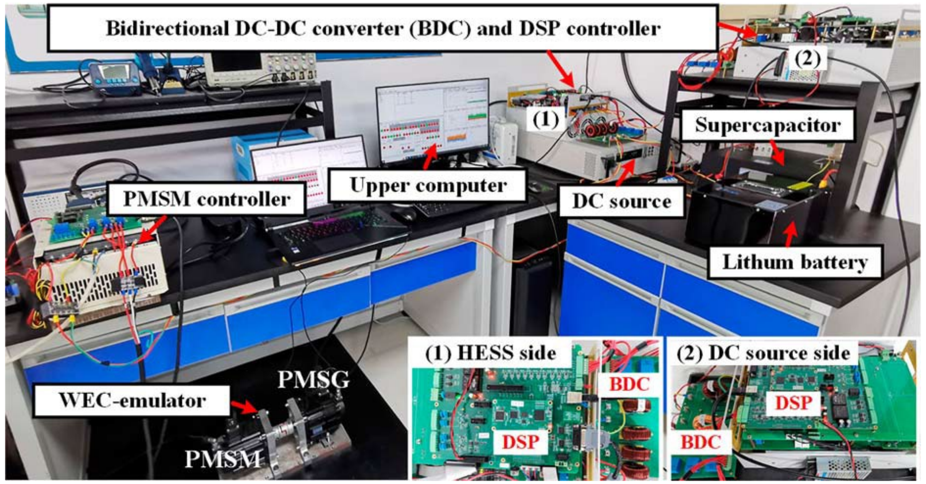
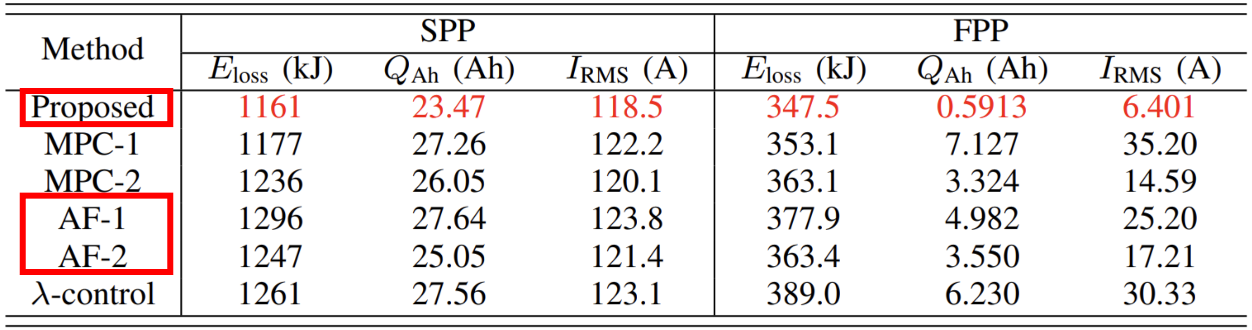


Figure: Experimental platform at THU



Method	SPP			FPP		
	E_{loss} (kJ)	Q_{Ah} (Ah)	I_{RMS} (A)	E_{loss} (kJ)	Q_{Ah} (Ah)	I_{RMS} (A)
Proposed	1161	23.47	118.5	347.5	0.5913	6.401
MPC-1	1177	27.26	122.2	353.1	7.127	35.20
MPC-2	1236	26.05	120.1	363.1	3.324	14.59
AF-1	1296	27.64	123.8	377.9	4.982	25.20
AF-2	1247	25.05	121.4	363.4	3.550	17.21
λ-control	1261	27.56	123.1	389.0	6.230	30.33

Control outperforms existing methods in main objectives. SPP: stepped power profile. FPP: filtered power profile. AF: filter-based methods.

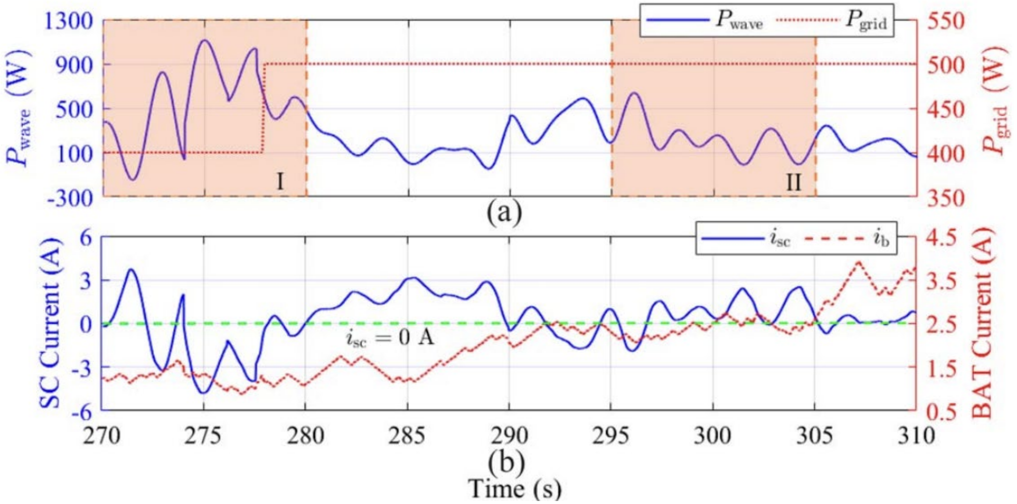
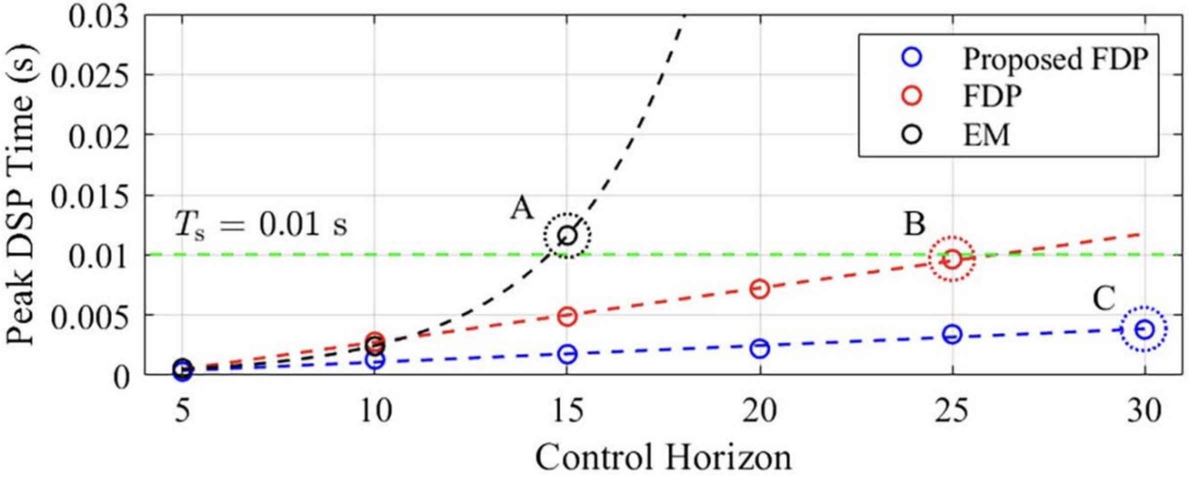


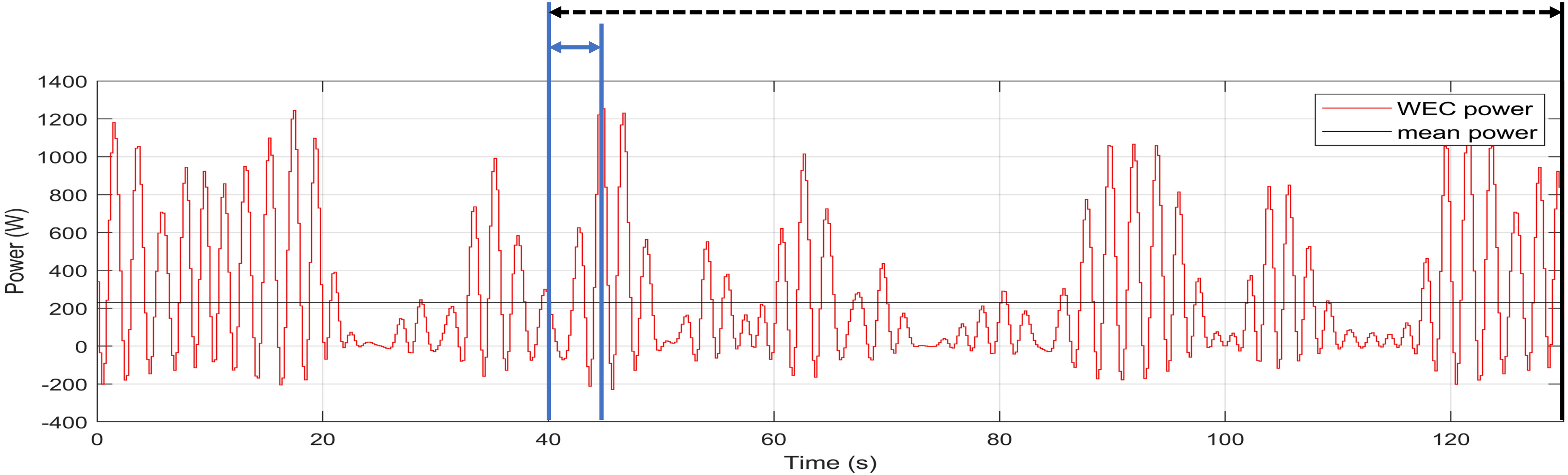
Figure: Control trajectories given a stepped grid power profile.



Control computation much faster than other alternatives. EM: the exhaustive method.

A main limitation: short sightedness

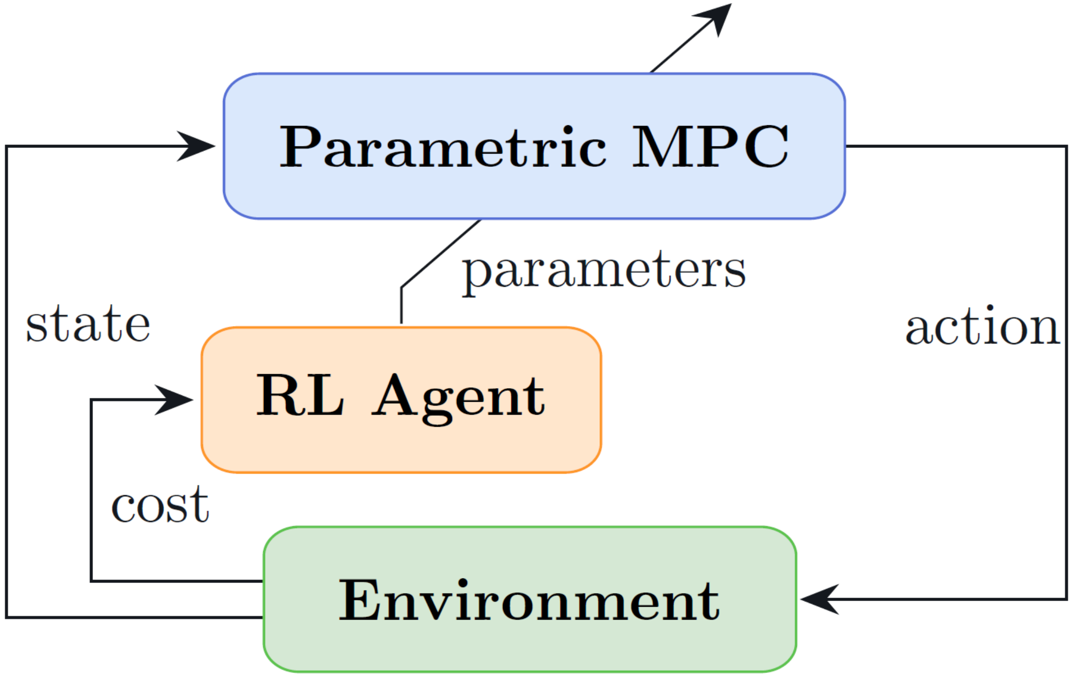
- MPC-based EMS can only optimize over a short horizon (e.g., 5 s) **with fixed weighting factors** – short sightedness.
- **However, the priority of the (conflicting) objectives should be different for different wave conditions, grid power commands, and SoC levels**, to achieve **long-term optimal** performance. (For example, when SoC is low, the penalty for its deviation needs to be larger.)



Learning-augmented MPC-based EMS

Learning-augmented MPC architecture

- We propose a **two-level** control structure.
- The **lower level** is still the short-horizon MPC – ensuring (locally) high-quality control action and constraint enforcement – with weighting factors.
- However, these weighting factors are determined **online** by **an upper-level reinforcement learning (RL) controller (agent)**, as an augmentation for MPC.



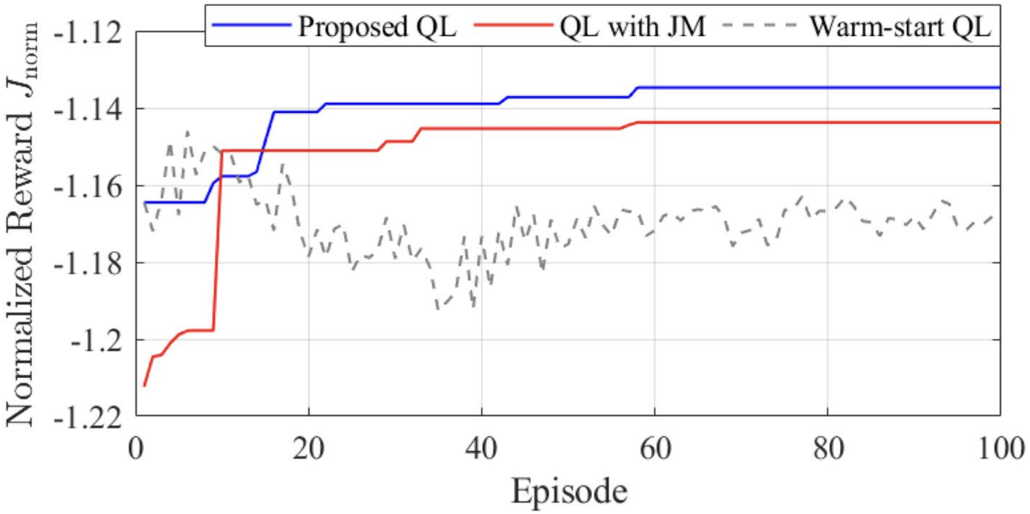
Reinforcement learning formulation

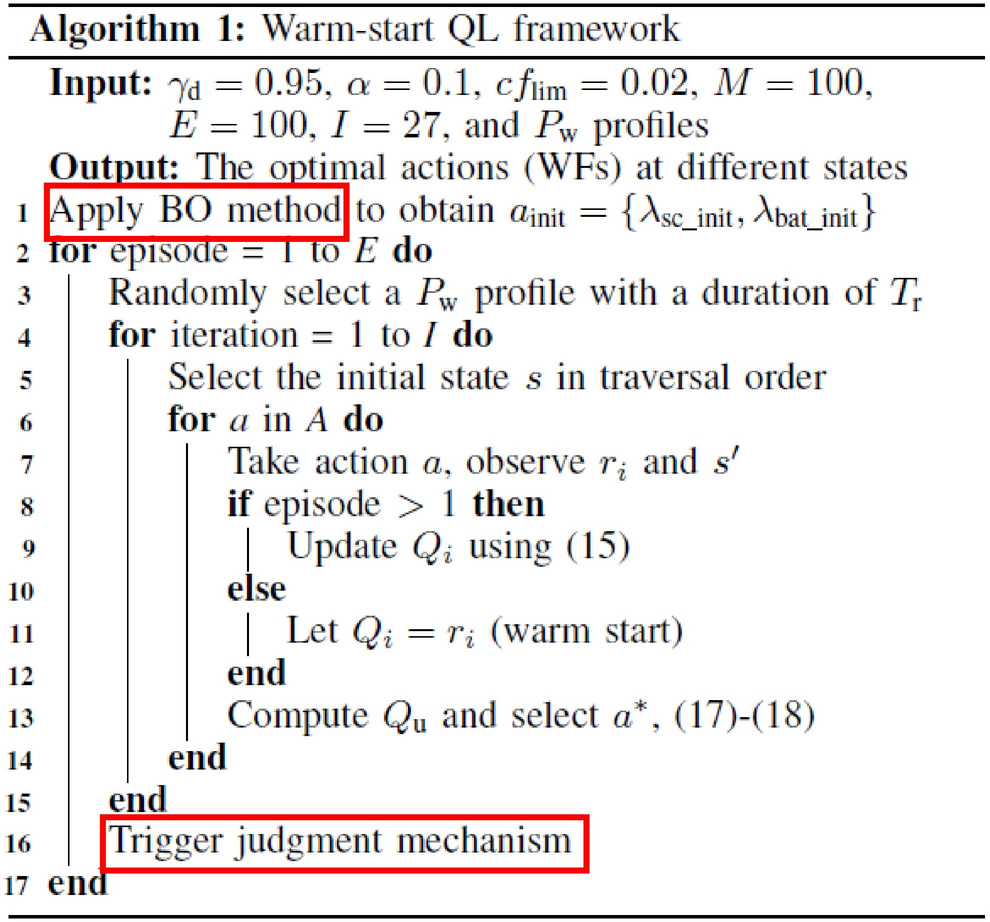
- **Lower-level (short-term) MPC objective**:
$$J = \sum_k P_{loss}(k) + \boldsymbol{w_b} |i_b(k)| + \boldsymbol{w_{sc}} \left(u_{sc} - u_{sc}^{ref}\right)^2$$
where $\boldsymbol{w_b}$ and $\boldsymbol{w_{sc}}$ are the (short-term) weighting factors.
- **RL state space**: SC voltage u_{sc}, grid power demand P_g.
- **RL (discrete) action space**: $\boldsymbol{w_b}$, $\boldsymbol{w_{sc}}$, selected from {0.9, 1, 1.1} times reference values.
- **RL (long-term) reward function**:
$$r_1 = -\int_{t_0}^{t_0+T} P_{loss}(t)\mathrm{d}t, \quad r_2 = -\int_{t_0}^{t_0+T} |i_b(t)|\mathrm{d}t, \quad r_3 = \sqrt{\frac{1}{T}\int_{t_0}^{t_0+T} i_b^2(t)\mathrm{d}t}$$
- **The total reward function** is $r = \boldsymbol{l_1} r_1 + \boldsymbol{l_2} r_2 + \boldsymbol{l_3} r_3$, where $\boldsymbol{l_1}$, $\boldsymbol{l_2}$, $\boldsymbol{l_3}$ are **predesigned long-term weighting factors**.
- **Constraint**: Define $e = \frac{1}{T}\int_{t_0}^{t_0+T} f_{sc}(t)\mathrm{d}t$ as the SC voltage exceedance level, where $f_{sc}(t)$=0 if $u_{sc}(t)$ is within constraint and =1 otherwise. The constraint is $e \leq e_{max}$.

Reinforcement learning algorithm design

- A **Q-learning (QL)** algorithm is designed, integrating a number of techniques:

1. Bayesian optimisation is used calculate an initial actions set to **warm-start** the QL.
2. A **trajectory judgement mechanism (JM)** is designed to avoid erroneous exploration.
3. A **neural-network-based predictor** is applied to mitigate the sim2real gap.





Algorithm 1: Warm-start QL framework

```
Input: $\gamma_d = 0.95$, $\alpha = 0.1$, $cf_{\text{lim}} = 0.02$, $M = 100$,
       $E = 100$, $I = 27$, and $P_w$ profiles
Output: The optimal actions (WFs) at different states
Apply BO method to obtain $a_{\text{init}} = \{\lambda_{\text{sc\_init}}, \lambda_{\text{bat\_init}}\}$
for episode = 1 to $E$ do
    Randomly select a $P_w$ profile with a duration of $T_r$
    for iteration = 1 to $I$ do
        Select the initial state $s$ in traversal order
        for $a$ in $A$ do
            Take action $a$, observe $r_i$ and $s'$
            if episode > 1 then
                Update $Q_i$ using (15)
            else
                Let $Q_i = r_i$ (warm start)
            end
            Compute $Q_u$ and select $a^*$, (17)-(18)
        end
    end
    Trigger judgment mechanism
end
```

Experimental validation

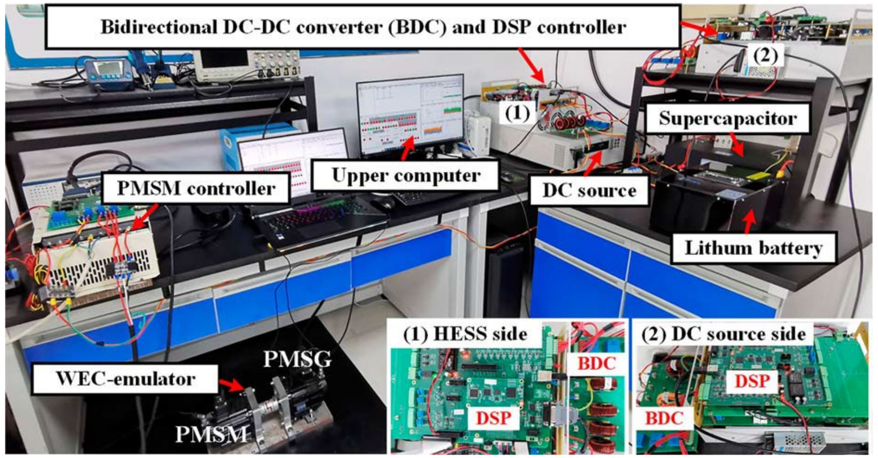


Figure: Experimental platform at THU

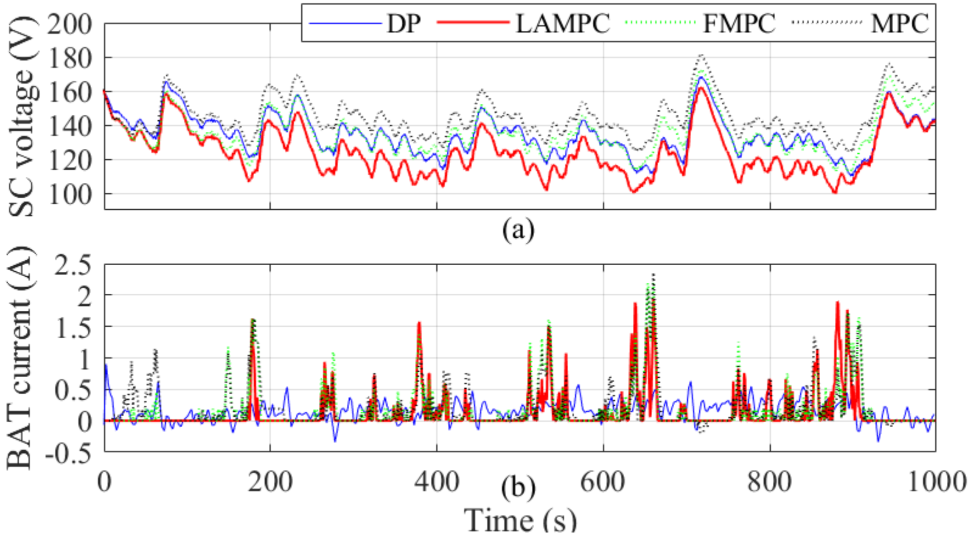
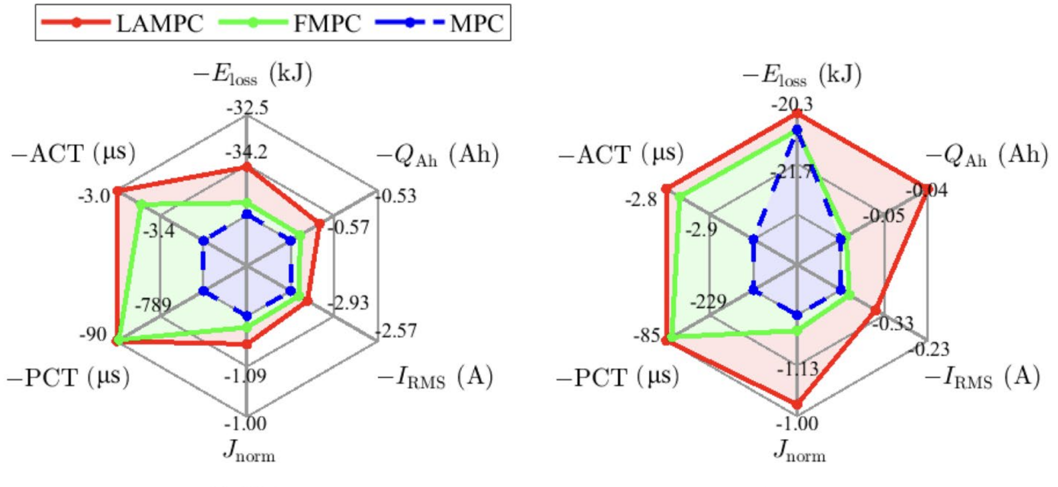
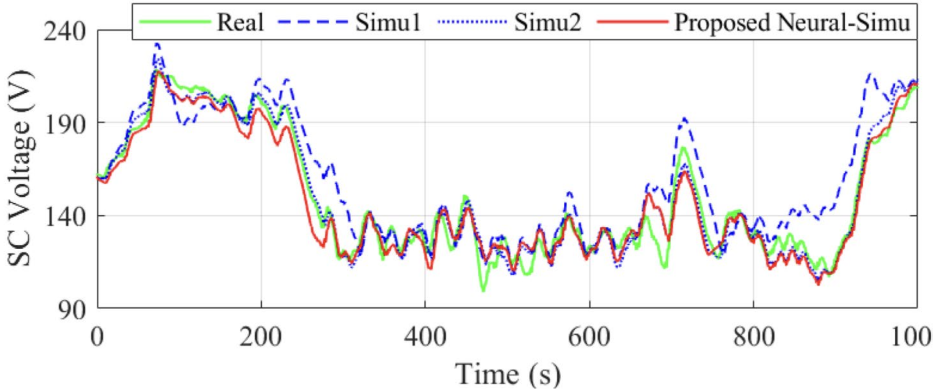


Figure: Control trajectories. DP: long-term optimal solved by DP. LAMPC: learning-augmented MPC. FMPC: a stripped-down version of LAMPC. MPC: normal MPC.



(a) SPP (b) FPP

Learning-augmented MPC outperforms other controllers in almost all objectives, close to long-term optimum (from DP).



The sim2real gap is effectively reduced.

Conclusion

- We propose a novel, MPC-based real-time EMS for HESS in WEC systems with **two innovations**: i) a fast real-time solution method using FDP, as the lower level, and ii) an online weighting factor tuning strategy using RL, as the upper level.
- The FDP control proves optimal (in short term) and computationally efficient.
- The upper-level RL control proves close to the long-term optimum.
- All methods have been **experimentally validated**.

For details of this work, please see:

[1] X. Zhu, Z. Lin, X. Huang, K. Chen, Y. Han and X. Xiao, "Real-Time Energy Management of Hybrid Energy Storage System With Application to Wave Energy Converters: A Learning-Augmented MPC Strategy," in *IEEE Transactions on Sustainable Energy*, vol. 17, no. 1, pp. 171-183, Jan. 2026

[2] X. Zhu, X. Huang and X. Xiao, "Fast Nonlinear Model Predictive Control for the Energy Management of Hybrid Energy Storage System in Wave Energy Converters," in *IEEE Transactions on Industrial Electronics*, vol. 72, no. 8, pp. 8154-8164, Aug. 2025

Thank you!