nature energy

Article

https://doi.org/10.1038/s41560-025-01804-x

An actor–critic algorithm to maximize the power delivered from direct methanol fuel cells

Received: 10 August 2024

Accepted: 2 June 2025

Published online: 14 July 2025

Check for updates

Hongbin Xu[1,5], Yang Jeong Park[1,2,5], Zhichu Ren[1], Daniel J. Zheng[1], Davide Menga[1], Haojun Jia[3], Chenru Duan[3], Guanzhou Zhu[1], Yuriy Román-Leshkov[3], Yang Shao-Horn[1,4] ✉ & Ju Li[1,2] ✉

Optimizing nonlinear time-dependent control in complex energy systems such as direct methanol fuel cells (DMFCs) is a crucial engineering challenge. The long-term power delivery of DMFCs deteriorates as the electrocatalytic surfaces become fouled. Dynamic voltage adjustment can clean the surface and recover the activity of catalysts; however, manually identifying optimal control strategies considering multiple mechanisms is challenging. Here we demonstrated a nonlinear policy model (Alpha-Fuel-Cell) inspired by actor–critic reinforcement learning, which learns directly from real-world current–time trajectories to infer the state of catalysts during operation and generates a suitable action for the next timestep automatically. Moreover, the model can provide protocols to achieve the required power while significantly slowing the degradation of catalysts. Benefiting from this model, the time-averaged power delivered is 153% compared to constant potential operation for DMFCs over 12 hours. Our framework may be generalized to other energy device applications requiring long-time-horizon decision-making in the real world.

Human experience-based learning and decision-making prove increasingly inadequate in meeting the demands of society, especially when confronted with high-dimensional parameters in highly nonlinear and long-time-horizon dynamical systems[1–5]. A key priority is developing sustainable energy systems to address climate change and the energy crisis[6–9]. Direct methanol fuel cells (DMFCs) show great potential due to their high energy density and the ease of storage and transportation of the fuel form[10]. The practical performance of DMFCs is affected by multiple factors such as catalyst design (over-reliance on Pt group metals) and dynamic operating conditions[11,12]. Whereas commercial Pt-Ru systems are generally engineered to improve tolerance against poisoning when operated under ideal conditions, carefully managing operating parameters is essential to ensure long-term stability. But maintaining ideal external operating conditions for the optimized catalyst design (particularly for methanol oxidation (MOR)) is challenging in reality. The optimal conditions for one catalyst may not be applicable to others, leading to performance degradation[13]. One of the catalysts developed by us (Co-Pt-Ru/NC) can effectively reduce Pt usage while improving the activity of MOR; however, it still suffers from the poisoning problem and needs to be refreshed intermittently[14].

To solve the problem of gradual deactivation of the catalyst in a dynamic environment, voltage switching (that is, changing the potential as a function of time) is used to remove harmful adsorbents from the surface and extend catalysts' life[15–17]. Whereas refresh cycles (operation pauses and air bleed) have been reported to partially mitigate DMFC degradation, their arbitrariness emphasized the need for

[1]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]These authors contributed equally: Hongbin Xu, Yang Jeong Park. ✉e-mail: shaohorn@mit.edu; liju@mit.edu

optimization of the strategy for the anode[18–20]. Due to the complexity of DMFC recovery mechanisms which involve multiple factors, an electrochemistry-centred investigation offers a more tractable pathway. Obviously, the policy of changing the voltage $U(t)$ as a function of time t is critical[21]. However, manual parameter adjustments in experiments face many drawbacks[22,23]. Choosing the optimal policy for $U(t)$ considering complex time-dependent catalytic chemistry including memory effects and recovery mechanisms is challenging to achieve with traditional control theory[24].

Reinforcement learning (RL) is widely used to achieve human-level control[25] and complex decision-making tasks such as AlphaGo[26,27] and holds the potential to offer numerous advantages as a method for recovering catalyst activity and controlling the operation of DMFCs. First, the dynamic state of the fuel cell necessitates condition-specific interventions, with catalyst recovery required only under certain operating scenarios. RL enables real-time monitoring of the DMFC's state to make informed decisions accordingly. Moreover, compared to conventional methods such as PID (Proportional-Integral-Derivative) and MPC (Model Predictive Control), RL offers greater adaptability by learning optimal control policies through direct interaction with the system. Whereas PID and MPC rely on linear approximations or accurate system identification, RL continuously improves the performance of highly nonlinear and memory-dependent systems such as games and handles disturbances without predefined control laws or frequent retuning.

Nevertheless, applying RL for DMFCs in practice presents several challenges. Most advanced RL models are designed for virtual environments such as Atari games, where training can occur over millions of simulations at low cost[28–31]. However, DMFCs are real-world systems where the simulation-to-real (sim2real) gap necessitates training directly on experimental data to ensure reliability. Effective DMFC operation requires accurate state monitoring as different strategies should be employed according to various cell states. The RL faces limitations in goal adaptability[32] caused by the reliance on fixed reward function structures, which necessitate retraining for varied objectives, impeding adaptability in dynamic environments where goals may frequently change. Therefore, to maximize and control delivered power for DMFCs, a new RL-inspired system is necessary, which is data-efficient and goal-adaptive without relying on virtual simulations.

In this work, we developed a DMFC control system named Alpha-Fuel-Cell (αFC) inspired by RL, which can achieve goal adjustment by formulating the control strategy based on the states of cells while largely preserving the activity of catalysts. The electrochemical mass spectrometry (ECMS) was used to explore the hidden mechanism of the activity recovery. Data augmentation was employed to improve data efficiency. Our method harnesses stochastic gradient descent algorithms, leveraging the differentiability of neural networks to dynamically optimize actions towards the desired goals, offering an efficient pathway to achieving optimal behaviour. We also introduce a strategy for goal adaptability in RL. As such, the produced power controlled by αFC is 486% compared to the constant voltage strategy over 90 hours. And the output power of DMFC is maximized and stable over more than 250 hours by αFC. The model can seamlessly adapt to varying objectives without extensive retraining, which is particularly beneficial in controlling real-world systems.

Manually verifying the voltage-switching strategy

First, voltage switching was manually demonstrated to improve the activity of Co-Pt-Ru/NC catalyst. Resting potential and resting time were added to clean the catalyst surface and recover activity during chronoamperometry (CA) measurements. The resting potentials were chosen by comparing oxidation and reduction potentials (Supplementary Fig. 1). After applying MOR potential (0.7 V versus the reversible hydrogen electrode (V_{RHE})), the electrode was rested for 30 s and tested again at 0.7 V_{RHE}. The lower potential (0.1 V_{RHE}) can improve the activity when comparing the current density after resting at high potential (1.2 or 1.7 V_{RHE}) or integrating the charge over the same period (Supplementary Fig. 2). Additionally, high potentials could cause catalyst degradation due to carbon support corrosion and Ru leaching. Therefore, all the resting potentials during training and operation are below 0.6 V_{RHE}.

The slow linear sweep voltammetry (LSV) method was employed to exclude the polarization effect. Instead of changing the potential to 0.7 V_{RHE} directly after resting at 0.1 V_{RHE}, the potential was gradually increased to 0.7 V_{RHE} from 0.1 V_{RHE} at 10 mV s^{-1} after MOR measurements (0.7 V_{RHE}) and then kept at 0.7 V_{RHE} for the same duration. A similar CA current density of MOR indicated that the polarization effect can be ignored in this case (Supplementary Fig. 3). Taking 0.1 and 0.7 V_{RHE} as resting and working potentials, respectively, the constant strategy (0.1 V_{RHE} for 1 min and 0.7 V_{RHE} for 60 min) and switching strategy (0.1 V_{RHE} for 10 s and 0.7 V_{RHE} for 10 min, repeated six times) were compared (Supplementary Fig. 4). The switching strategy exhibited higher current density in the same operation time (Supplementary Fig. 5).

To verify the reproducibility, different catalyst batches were measured using the same measurement protocol. The difference in curves is small whether at the beginning of the test or after applying the switching strategy, suggesting the suitability for training the model (Supplementary Fig. 6). Although the catalytic activity decayed during long-term measurements, the switching strategy maintained a higher average current (Supplementary Fig. 7), proving its effectiveness. Therefore, it is promising to combine voltage switching and machine learning to control and maximize the delivered power by optimizing the parameters.

Workflow of Alpha-Fuel-Cell

The workflow of αFC is illustrated in Fig. 1. Four independent action variables were identified to control the DMFC: working time, working potential, resting time and resting potential. These variables shape a typical current–time trajectory (Fig. 1a), which is used to calculate the average power and state reflection. The proposed system consists of an actor module and a critic module, inspired by the actor–critic algorithm in RL (Fig. 1b).

Automated measurements (Supplementary Movie 1) were employed during the training and control process. The critic module serves as the action-value function in RL, which evaluates the value of each action in a given state (Fig. 1c). To design the highly predictive critic module, we conducted comparative studies of neural architectures (Supplementary Discussion 1 and Supplementary Figs. 8 and 9). A convolutional neural network (CNN) architecture was selected to process the current–time trajectory as the state, and a multi-layer perceptron (MLP) was chosen to handle the action variables. By using raw trajectories, the need for specialized measurements such as electrochemical impedance spectroscopy or complex feature engineering can be eliminated, thus improving computational efficiency and accuracy. The CNN's output is flattened and combined with MLP's output to predict the produced power.

The actor module determines control policies based on the desired power output (Fig. 1d). We propose leveraging the differentiability of the trained neural network to eliminate the need for retraining when the desired output changes. A batch of random actions is input into the critic module, and back propagation is applied to minimize the margin between the critic module's prediction and target outputs. This allows efficient optimization within a reasonable time of 0.3 s using only a central processing unit (CPU). Finally, the action with the smallest margin is selected.

Training and control process of αFC

To train the neural network of the critic module, we randomly sampled data in the action space. Many studies employ simulation-based surrogate models to learn policy[24,33]. However, developing surrogate models that perfectly replicate the behaviour of electrochemical cells is challenging. Considering the sim2real gap, real experimental data

https://doi.org/10.1038/s41560-025-01804-x

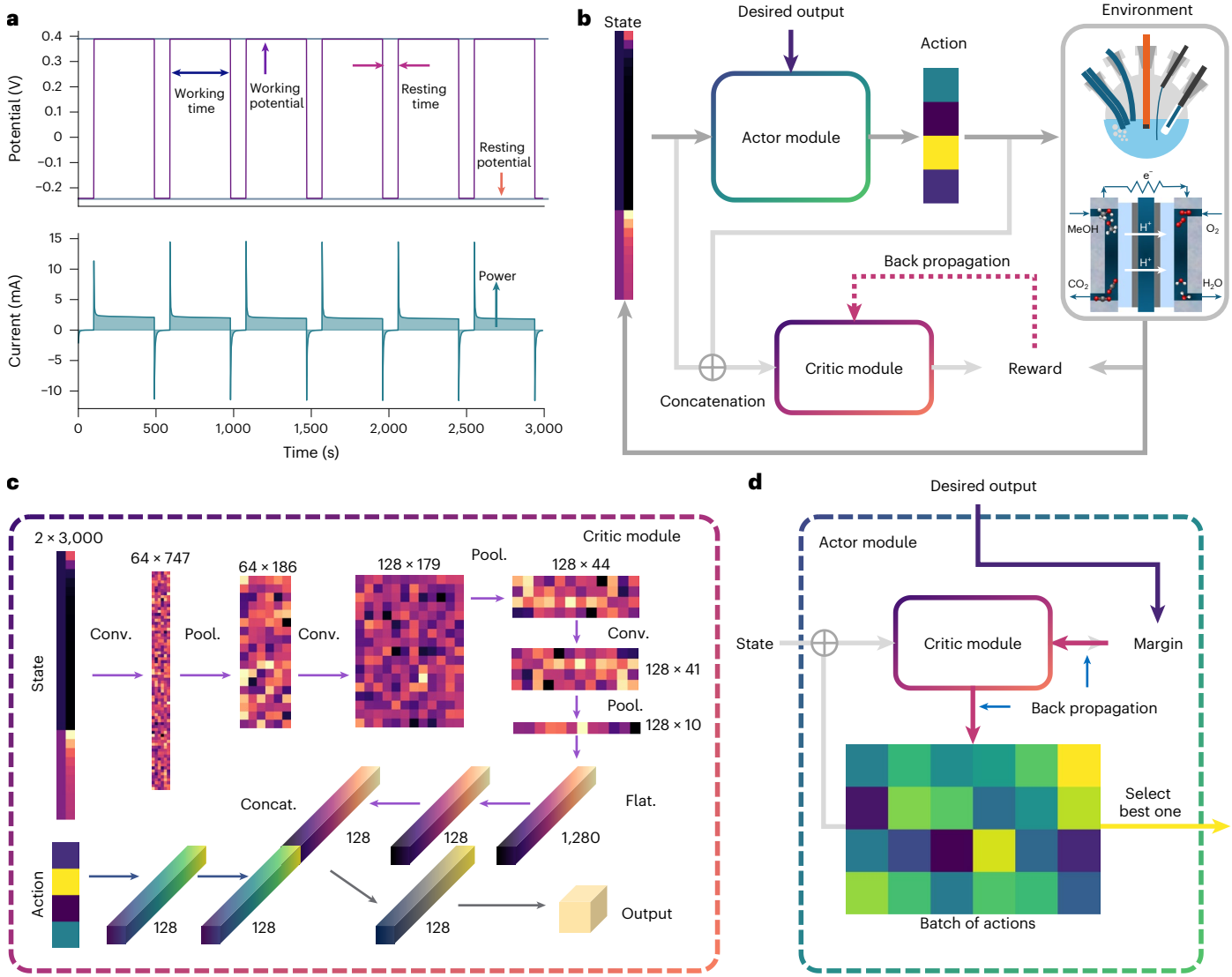

Fig. 1 | Schematic of our αFC system. a, The state of the DMFC system is defined as the trajectories of the potential and the current during a step lasting 5 minutes. The DMFC system is controlled by four action variables, working time, working potential, resting time and resting potential. **b**, αFC iteratively manipulates the DMFC by choosing the most appropriate actions, as determined by the actor module, to achieve the desired output for the given state. The critic module is trained to predict the reward, produced power of the DMFC, for the given state and the action. **c**, The critic module has two branches: 1D convolutional neural network for state input (images with 2 × 3,000 pixels) and action-head network for 4D action input. The numbers in the figure represent the size of the image or vector. The hidden states are used to predict the output power. Conv., convolution layer; Pool., pooling layer; Concat., concatenation. **d**, The actor module leverages the differentiability of the critic module to select the next action, minimizing the margin between the predicted and desired outputs.

were directly used for training. Because experimental data acquisition is time-consuming, data augmentation[34] was proposed to improve data efficiency, assuming that adjacent trajectories share similar states given a 5 min timestep size (Supplementary Fig. 10 and Supplementary Discussion 2). The learning curve of the critic module is illustrated in Fig. 2a, showing that the validation loss converges in 300 epochs. With data augmentation, the mean absolute error decreased to 0.011 mW and the Pearson correlation coefficient increased to 0.969 (Fig. 2b), demonstrating the accuracy of our model and the effectiveness of data augmentation.

For real experiments, an example of how αFC controls the system is depicted in Fig. 2c. After $step_{n-1}$, the state ($state_{n-1}$) of the cell was obtained from the trajectory curve. During this step, the current was 0.99 mA at 10 s (beginning current) and decreased to 0.81 mA at 300 s (ending current). Although the state change is difficult to see with naked eyes, it can be captured by αFC. The $state_{n-1}$ was passed to the actor module, which generates the action for the next step ($action_n$) to better achieve the desired output. It recommended performing low-potential cleaning once. At $step_n$, it showed the current dropping from 0.97 mA to 0.81 mA. Similarly, the $action_{n+1}$ was suggested by the αFC after receiving $state_n$ as input. The corresponding trajectory exhibited a current from 0.95 mA to 0.82 mA (Supplementary Fig. 11). This pattern suggests that the catalyst decays over time, but the final ending current can remain unchanged or even higher than in the previous two steps, indicating that the control of αFC can recover the active site of the catalyst and extend its life.

Controlling ability and mechanism of αFC system

The control experiment to demonstrate the capability of αFC was conducted on a three-electrode system. Maximization can be achieved by setting a target much higher than the possible output at the current state of the cell. The control process was configured with different targets in the following sequence: maximum power, 0.2 mW, 0.1 mW, back to maximum power, −0.1 mW and back to maximum power (Fig. 3a).

https://doi.org/10.1038/s41560-025-01804-x

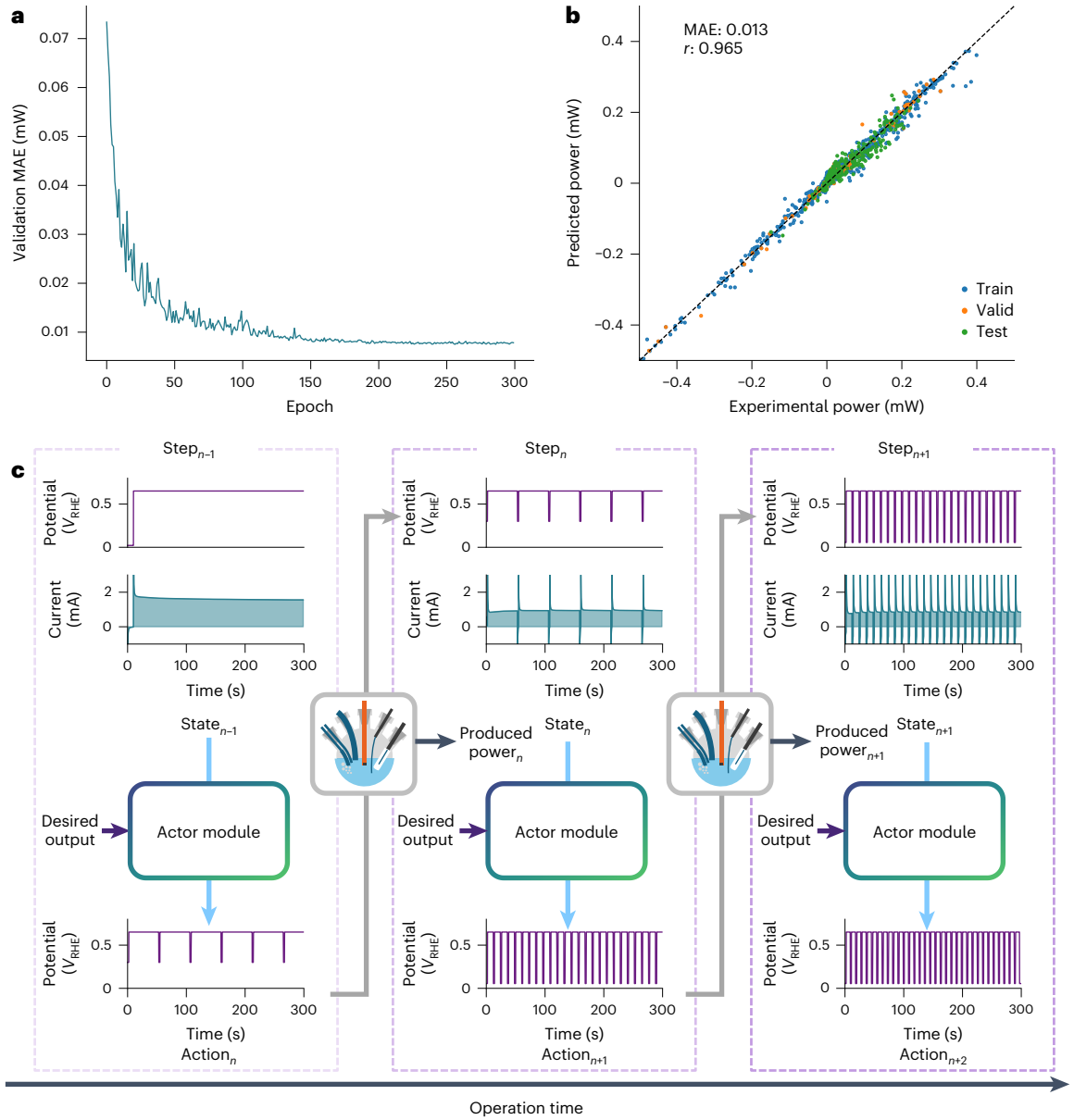

Fig. 2 | Result of training and the control process of αFC. a, The learning curve of the critic module from experimental data for 300 epochs. **b**, The parity plot between the experimental power and the critic model prediction. **c**, The example of controlling process of αFC. The state defined by trajectory is passed to the actor module and the next action expected to best fit the desired output is generated. The DMFC is then operated iteratively by this action, producing energy and the current trajectory of the next state.

A negative-power condition was used to evaluate if αFC could recover catalyst performance under extreme conditions. The DMFC system successfully reached the set targets, including the negative one, meaning that the fuel cell could be turned into an electrolyser. Notably, αFC always allowed the cell to return to its maximum output power when the target was switched to maximum power. Although initially the predicted power slightly exceeded the experimental power, it almost overlapped with maximized experimental power later. These results demonstrate the excellent goal adaptability of αFC. The specific parameters during operation with different targets were further analysed in Fig. 3b. As the target output changed, the control conditions were adjusted to match the produced output. For example, when the power is maximized, the working time is longer and the resting potential is lower, which allows for more energy production and better recovery of the catalytic activity. Energy production could be reduced by lowering the working potential, increasing the rest time and reducing the working time. The correlation coefficient between each variable and the final output is shown in Supplementary Fig. 12 (Supplementary Discussion 3).

A greedy algorithm (GA) was chosen as a baseline to demonstrate the superiority of our αFC using gradient-based optimization (GBO) (Supplementary Fig. 13). GBO minimizes the margin over 50 iterations using the critic module as a surrogate model. Three baseline GAs were set-up using trained critic modules: the same number of searches, the same number of iterations and the same running time. Gradient tracking takes roughly three times more computations, which justifies this comparison. The comparison results in Fig. 3c show that our GBO control strategy consistently achieves superior performance.

To investigate the control capability of the αFC and catalyst activity recovery mechanism, in situ ECMS was employed to monitor changes in key products during MOR. Four main products (H_2, CO, O_2 and CO_2) were examined. Background mass signals in 0.1 M $HClO_4$ electrolyte

https://doi.org/10.1038/s41560-025-01804-x

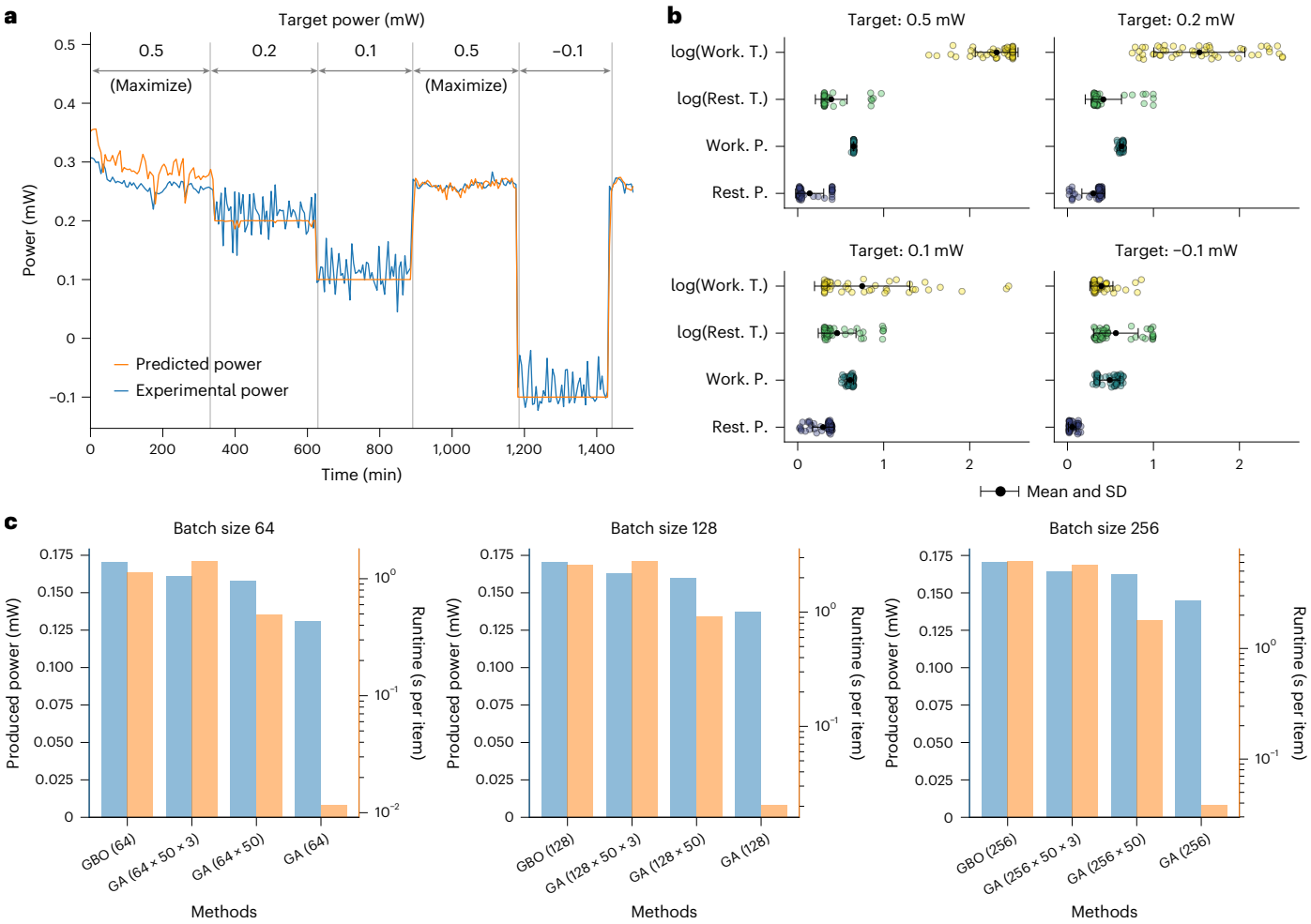

Fig. 3 | Controlling ability and analysis of the αFC system. a, Produced power of the DMFC according to the time. The target power is controlled. The green line represents the experimental power of DMFC, and the orange line represents the prediction of the critic module. **b**, Distribution of the action variables according to the target with a sample size of 50. Work. T., working time; Rest. T., resting time; Work. P., working potential; Rest. P., resting potential. **c**, Comparison of optimizing strategies for the actor module. Predictions of produced power by GBO and GA are plotted. GBO and GA were compared at the same batch size, same iteration and same run-time level.

with CV scans are illustrated in Fig. 4a. Weak signals of CO_2 and CO were observed as the potential reached higher levels, probably due to the gradual oxidation of carbon into CO_2 at elevated potential, with CO serving as a fragment of CO_2. The faint O_2 signal is attributed to a small amount of oxygen evolution occurring at high potentials. The clear H_2 signal can be attributed to hydrogen evolution during the CV scan at low potentials in acidic conditions. Upon adding 0.1 M methanol into the electrolyte, the CO_2 and CO signals became stronger, reflecting methanol oxidation (Fig. 4b). Unlike CA measurements, CV scans showed slower current drop, suggesting that the catalyst surface is refreshed during the hydrogen evolution. Small pumps of CO signal appeared during hydrogen production (circled in red dashes), probably due to the bound CO being replaced by H intermediates and released to the bulk solution. Switching to CA at a potential of 0.65 V_{RHE}, the signals of CO_2 and CO initially peaked and then declined, similar to current behaviour (Fig. 4c), due to the accumulation of harmful intermediates on the catalyst surface, which block active sites. Furthermore, no H_2 signal was observed as the constant potential did not allow catalyst cleaning like in the CV scan. This also explains why in typical MOR tests, the performance in CV scans is more stable compared to CA measurements.

Figure 4d shows the signal of these four products when αFC controls the reaction. As expected, H_2 signals reappeared when applying resting potential to MOR, during which the surface of the catalyst was covered by H intermediates (clean the surface to recover the activity). Moreover, a small peak around 100 s (red circle) could be attributed to the accumulated CO being released to the electrolyte. Additionally, the signal decay ratio of CO_2 and CO during experiments was calculated (Supplementary Fig. 14). During CV scans, CO appears as a fragment of CO_2, so the ratio of their decline should remain constant. However, during the resting potential, the changes of CO signal are influenced not only by the decrease in CO_2 but also by the release of CO. By comparing the decline ratio of CO_2 to CO signals, it is smaller under resting potential conditions, indicating CO release from the catalyst surface in this state. This further confirms that the αFC effectively cleans harmful species from the catalyst surface, thereby recovering its performance.

Maximizing the delivered power by αFC system

To further demonstrate the enhancement of delivered power by αFC, it was compared with different operational strategies. Initially, in a three-electrode system (Supplementary Fig. 15), a constant potential (0.65 V_{RHE}) was applied to the commercial PtRu/C catalyst for four hours, serving as a benchmark, because the model indicated the power is maximum for DMFCs at this potential for MOR. Additionally, three different strategies were employed for our Co-Pt-Ru/NC catalyst: a constant potential strategy, a manual switching strategy and αFC to maximize mean power (Supplementary Fig. 16). To avoid unfair

https://doi.org/10.1038/s41560-025-01804-x

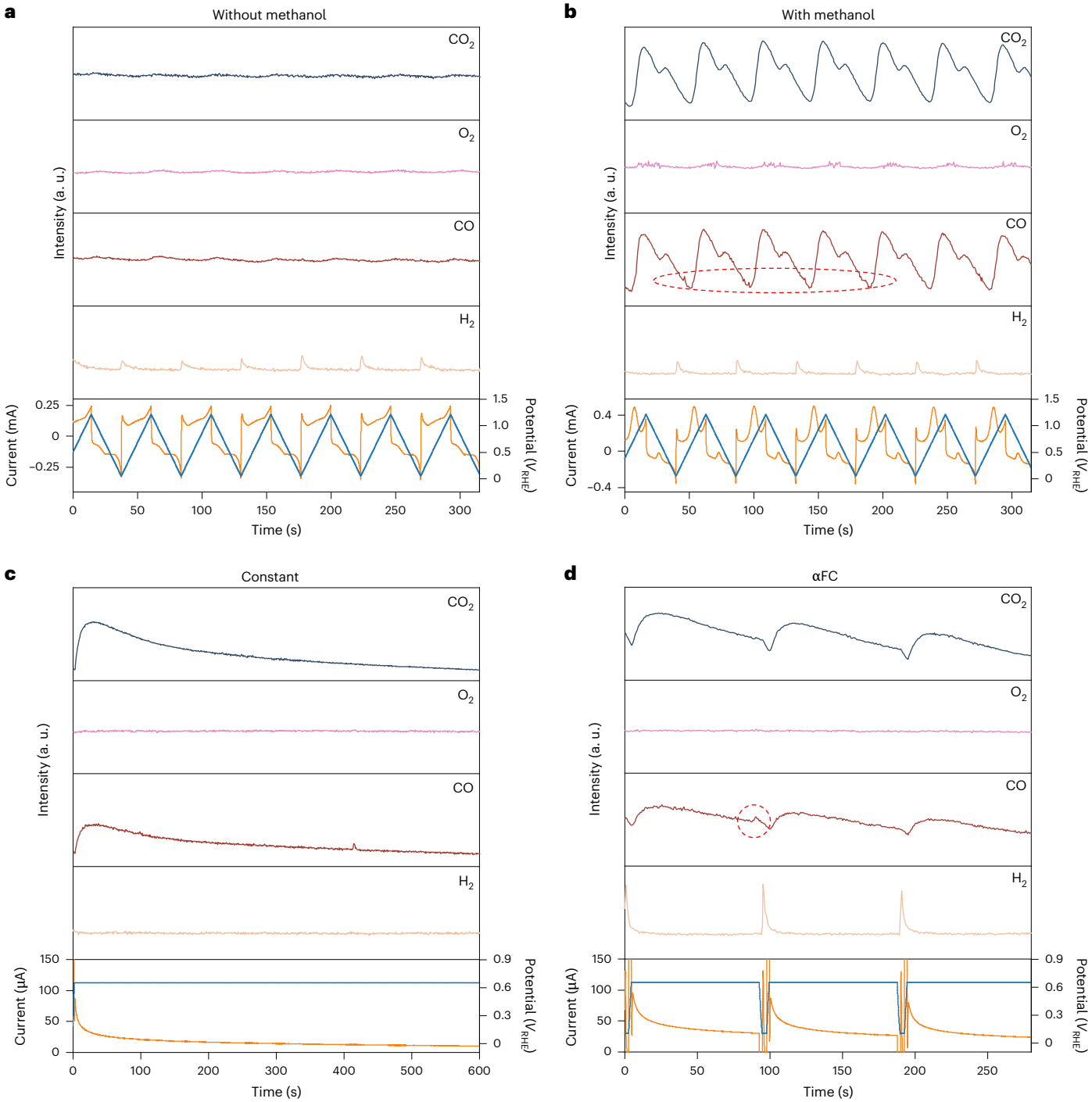

Fig. 4 | Mechanism exploration of αFC system by ECMS. ECMS results of H_2 (light orange), CO (brown), O_2 (pink) and CO_2 (dark blue) mass signal with different conditions. **a**, Background signal collection with CV scans at 0.1 M $HClO_4$ solution. **b**, The mass signal with the same CV scans with methanol (0.1 M methanol in 0.1 M $HClO_4$). **c**, The mass signal using constant potential strategy (holding at 0.65 V_{RHE}). **d**, Mass signal using αFC to control the reaction. Small pumps of CO signal while hydrogen is produced (circled by red dash). The orange line and blue line represent current and potential during measurements, respectively.

comparison, the parameters for the constant and switching strategies were determined by the model (Fig. 3b).

Results over four hours showed that the power output increased in the following order: PtRu/C_constant, Co-Pt-Ru/NC_constant, Co-Pt-Ru/NC_switch and Co-Pt-Ru/NC_αFC (Fig. 5a). The αFC-controlled strategy achieved a power of 0.284 ± 0.013 mW, which is 2.15 times and 4.64 times more than Co-Pt-Ru/NC_constant and PtRu/C_constant, respectively. On the basis of this, the operation time was extended to 12 hours (Supplementary Fig. 17). In Fig. 5b, under the constant potential strategy, although Co-Pt-Ru/NC showed better catalytic activity than PtRu/C, power decreased rapidly due to CO or other species poisoning in both cases. Switching strategy reduced catalyst deactivation, showing better sustained performance than the constant strategy. More importantly, αFC achieved a 30.4% increase in power compared to the switching strategy. Further comparison with constant strategies for our catalyst and commercial catalyst showed

https://doi.org/10.1038/s41560-025-01804-x

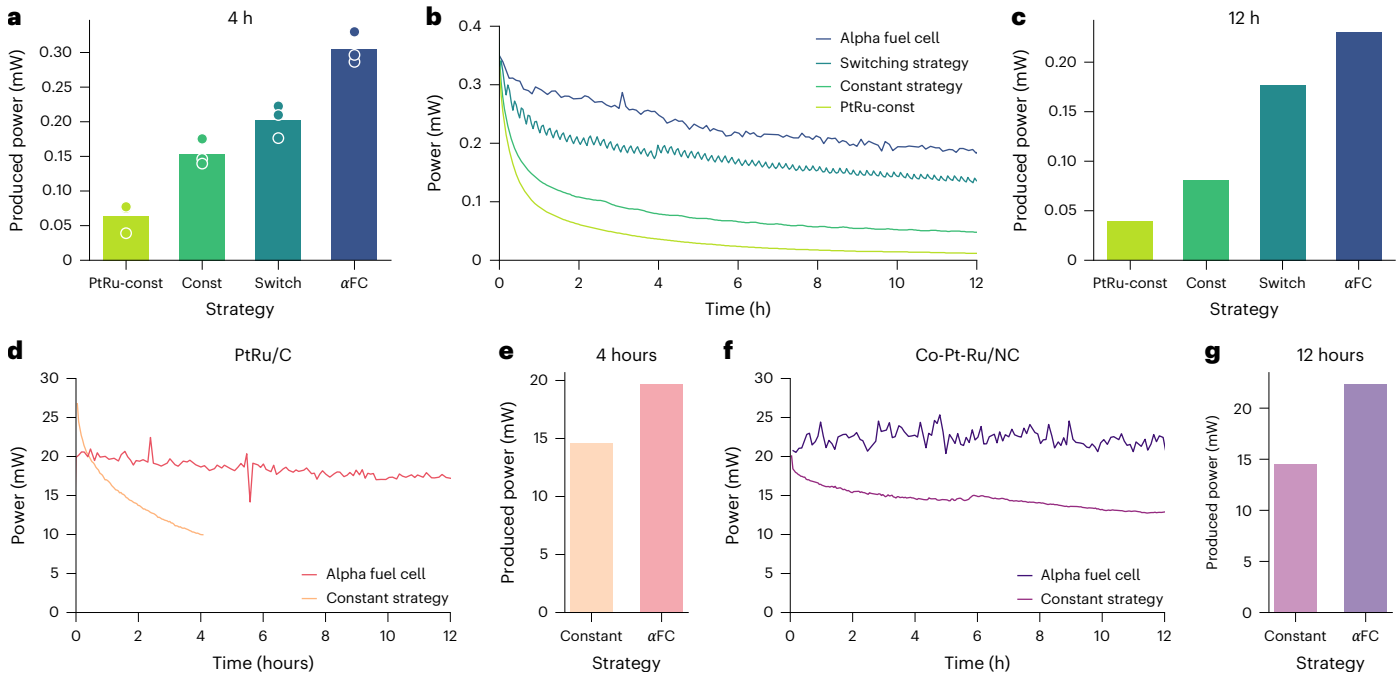

Fig. 5 | Controlling and maximizing the delivered power by αFC system. **a**, The comparison of the produced power of each strategy in the three-electrode system. PtRu-Const means a system with PtRu/C catalyst controlled by constant potential. Const and Switch mean systems with Co-Pt-Ru/NC catalyst controlled by constant and switching potential, respectively. αFC means the system is freely controlled by αFC with Co-Pt-Ru/NC. Each experiment lasted 4 hours. **b**, The comparison of the produced power according to the time in 12 h in three-electrode system. The PtRu-Const means PtRu/C catalyst controlled by constant potential, while all other strategies were using Co-Pt-Ru/NC catalyst. **c**, The produced power of different strategies for 12 hours. **d**, The produced power of DMFC using PtRu/C as the catalyst according to the time controlled by constant potential and αFC. **e**, The total produced power for 4 hours. **f**, The produced power of DMFC using Co-Pt-Ru/NC as catalyst according to the time and controlled by constant potential and αFC. **g**, The total produced power for 12 hours. Test conditions for DMFC: 1 × 1 cm^2 MEA, methanol flow rate: 5 ml min^{-1}, O_2 flow rate: 40 ml min^{-1}, 1 M methanol, Nafion117 as membrane, Pt/C as the ORR catalyst, the cell is operated at 60 °C. Details can be found in Methods.

αFC's ability to increase power by 185.2% and 486.1%, respectively (Fig. 5c).

To explore strategy adaptability, principal component analysis was employed to visualize changes in the optimal driving strategy as cell states changed (Supplementary Fig. 18 and Supplementary Discussion 4). The ageing trajectories obtained from the constant strategy deteriorate rapidly, while staying longer in a fresher cell state of the switching strategy. This supports that dynamic adjustment using αFC can effectively increase energy production while restoring activity.

Additional comparative experiments were conducted to validate the effectiveness of αFC by fixing the resting potential at 0.04 V_{RHE} and 0.3 V_{RHE}, while allowing αFC to determine the remaining variables (Supplementary Fig. 19). Over the 12-hour test, the output power under both fixed resting potentials was higher than that of the constant potential strategy but still underperformed compared to full αFC control. This is because in a complex and dynamically changing environment, the resting potential should be determined based on the real-time state of the catalyst rather than being fixed at a specific value, which makes it difficult to adapt to different catalyst states. For example, under a fixed resting potential of 0.04 V_{RHE}, the catalyst may already be in a favourable state, making resting or cleaning unnecessary. Enforcing such an action could reduce output power (because the current is not high during resting) and potentially disrupt a stable catalyst state. Conversely, at a fixed resting potential of 0.3 V_{RHE}, surface cleaning may be insufficient when needed, leading to the accumulation of harmful species and subsequent activity decline. Therefore, allowing αFC to autonomously assess and control the catalyst's state is essential for maintaining optimal performance.

Furthermore, as fuel cells are typically required to have a long lifetime, long-term comparative experiments (αFC vs constant potential) were conducted. After 90 hours of testing, the average power output under αFC control was 4.86 times that of the constant potential (Supplementary Fig. 20). Additionally, CO stripping experiments under the two control conditions showed that the catalyst under αFC control exhibited an earlier onset potential and higher current (Supplementary Fig. 21), confirming that over long-term testing, αFC more effectively preserves catalyst activity, thereby extending its lifespan and maximizing power output.

Other control methods, such as PID and MPC, were also employed to demonstrate the superiority of αFC. PID, as a conventional linear control method, struggled with complex systems such as fuel cells due to its reliance on precise parameter tuning. As a result, PID produced only about 50% of the power achieved by αFC (Supplementary Fig. 22). MPC method relies on a difficult-to-achieve prerequisite: highly accurate simulation. In the context of fuel cells, efforts are made to simulate their electrochemical behaviour over time, such as through the Butler–Volmer equation. However, the simulation results often fail to align closely with experimental observations (Supplementary Fig. 23) due to the difficulty of accurately modelling variables such as catalyst degradation (which varies between catalysts), changes in the electrolyte, membrane effects and variations on the oxygen reduction (ORR) side and so on. Consequently, when the simulation is inaccurate, the control results produced by MPC can deviate substantially from the expected outcomes.

The universality of αFC was further examined by applying it to a DMFC without retraining (Supplementary Fig. 24), although the model was trained by a three-electrode system. An important premise was the model's conclusion that the potential for achieving maximum power is 0.65 V_{RHE}, based on our assumption that the potential for the ORR is at 0.9 V_{RHE}, resulting in a cell operating voltage of 0.25 V.

Moreover, during LSV testing of the DMFC, the peak power density was also observed around 0.25 V (Supplementary Fig. 25), demonstrating the model's universality and its accuracy. As operation conditions in three-electrode systems and DMFCs are quite different, such as counter reaction, pH of electrolyte, temperature and so on, αFC successfully adapted across platforms.

For the commercial PtRu/C catalyst, traditional constant voltage testing resulted in a rapid decay in device current within four hours, leading to a sharp decrease in output power. However, when the same membrane electrode assembly (MEA) after being measured by constant strategy was further controlled by αFC, output power could be maximized and maintained for over 12 hours (Fig. 5d). Within the initial four-hour run time, αFC could increase output by 34.91% (Fig. 5e). When Co-Pt-Ru/NC was employed as a lower-precious-metal alternative, it demonstrated higher stability in constant voltage testing, yet still exhibited noticeable output power decay, highlighting the utility of αFC (Fig. 5f). When αFC controlled the DMFC with Co-Pt-Ru/NC as the anode catalyst, the output power could be sustained at maximum levels (around 20 mW) with minimal decay. Over a 12-hour period, αFC further increased output power by 53.79%, providing compelling evidence of its effective control over DMFC (Fig. 5g).

Typically, the performance of a DMFC is influenced by various factors, not only the degradation and deactivation of the internal catalyst but also external operating conditions such as temperature, methanol concentration and flow rate. RL, particularly the αFC-controlled system, proves advantageous by adapting to environmental changes and maintaining optimal power output. To validate this capability, experiments were conducted using PtRu/C as the catalyst, where external conditions were varied while the system was controlled by αFC. In the previous tests, the methanol flow rate was set at 5 ml min^{-1} and the O_2 flow rate was 40 ml min^{-1}. First, keeping all other conditions unchanged, the methanol flow rate was increased to 20 ml min^{-1}, the LSV results indicated the peak power density was around 12.8 mW cm^{-2} (Supplementary Fig. 26a). When αFC (without any modification) was used for long-term control, the power output remained stable at this level (Supplementary Fig. 26b), in contrast to the rapid decline observed with the constant strategy used before (Fig. 5b). Similarly, when the O_2 flow rate was increased from 40 ml min^{-1} to 100 ml min^{-1} while maintaining all other parameters under αFC control, the peak power density also decreased (13.9 mW cm^{-2}). However, αFC still ensured stable operation at the corresponding maximum power output over an extended period (Supplementary Fig. 27).

These results demonstrate that while external factors influence DMFC performance, they do not affect the robustness of αFC. The effectiveness of αFC is independent of changes in external conditions: when external conditions are optimal, αFC maintains the maximum power output stably under those conditions; when conditions are suboptimal, αFC can still achieve the maximum possible power output for that specific scenario. Notably, whether in a three-electrode system, an actual DMFC device, using the self-developed Co-Pt-Ru/NC catalyst, or the commercial PtRu/C catalyst and under various operating conditions, αFC consistently exhibits stable performance, highlighting its strong adaptability and broad applicability. Although we did not extensively discuss external factors in this study, combining our approach with other machine-learning-guided designs of DMFC that account for these external factors could lead to the development of more efficient, stable and controllable DMFCs capable of maximizing performance[35]. Finally, long-term measurements controlled by αFC were conducted (Supplementary Fig. 28). The output power is maximized and stable over more than 250 hours. The power dropped after around 240 h of operation due to the depletion of oxygen. However, the power could be recovered after restoring the oxygen supply, which further demonstrates the robustness of αFC.

Conclusions

In summary, we successfully developed Alpha-Fuel-Cell controller inspired by reinforcement learning to maximize the time-averaged power delivered by DMFCs. Our work is a proof-of-concept study demonstrating how to combine energy storage/conversion devices with edge artificial intelligence to tune an individual device's performance, highlighting the neural network's power in experimental science. In addition to industrial use, we note that research devices are frequently modified with unknown characteristics, so using a self-learning controller and non-constant operation conditions can facilitate more rapid identification of any research device's true optimized performance. This idea can be generalized to multiple fields, such as battery formation protocol and charging policy, electrodeposition, temperature/fluid flow controls in reactors and so on.

Methods

Synthesis of Co-Pt-Ru/NC

All chemicals were used as received without further purification. Milli-Q deionized water was used in all experiments. First, ZIF-67 was prepared by the precipitation reaction between $Co(NO_3)_2{\cdot}6H_2O$ (1.23 g) and 2-methylimidazole (1.46 g) in 50 ml methanol at ambient conditions. After 20 hours of reaction, ZIF-67 was obtained by centrifugation and then washed with methanol three times, followed by drying in the oven overnight. To prepare Co NPs/NC, an alumina boat crucible loaded with ZIF-67 was placed into a tube furnace, heated to 1,000 °C at the rate of 5 °C min^{-1} and then kept at that temperature for 2 h under N_2 flow. After that, 100 mg Co NPs/NC was dispersed in 50 ml H_2O and sonicated for 30 min. Five ml H_2PtCl_6 hydrate (4 mg ml^{-1}) and 5 ml $RuCl_3$ hydrate (4 mg ml^{-1}) were added into the above suspension and followed by stirring for 2 h. The suspension was centrifuged with water three times and dried overnight to obtain Co+Pt+Ru/NC. The final product Co-Pt-Ru/NC was prepared by pyrolysis of Co+Pt+Ru/NC at 900 °C with a rate of 5 °C min^{-1} for 2 h under N_2 flow.

Electrochemical measurements

Electrodes were prepared by drop-casting an ink containing catalyst powder with Nafion on a glassy carbon (GC) disk electrode (Pine Research, 5 mm diameter). Typically, 5 mg of catalyst was dispersed in a mixture with 980 µl of ethanol and 20 µl of Nafion (5 wt%) via ultrasonication for 60 min to form the catalyst ink. Ten µl of catalyst ink was dropped on the GC with a nominal loading of 250 $\mu g_{catalyst}\ cm^{-2}{}_{geo}$ and 46 $\mu g_{Nafion}\ cm^{-2}{}_{geo}$. Electrochemical measurements were carried out in a three-electrode set-up in a glass electrochemical cell with Ag/AgCl reference electrode and carbon rod counter electrode. The Ag/AgCl reference electrode was calibrated in the same electrolyte by measuring the hydrogen oxidation/evolution using a platinum working electrode and defining the point of zero current as 0 V vs RHE. The potential was controlled using a Biologic VSP-300 potentiostat. Cyclic voltammetry (CV) could be recorded in the solution of Ar-saturated 1 M CH_3OH in 0.1 M $HClO_4$ with a potential range between ~0.05-1.2 V vs. RHE at a scan rate of 50 mV/s. The background current was collected from the CV of catalysts in Ar-saturated 0.1 M $HClO_4$ before CH_3OH addition. The chronoamperometry (CA) measurements were conducted for data collection, model training and verification. Two pumps were used to refresh the electrolyte during measurements.

DMFCs measurements

The as-synthesized Co-Pt-Ru/NC and commercial Pt/C were used as catalysts at the anode and cathode to fabricate the membrane electrode assembly (MEA), respectively. First, 40 mg catalyst powder was dispersed in 2 ml isopropanol with 200 µl Nafion (5 wt%) and sonicated for 1 h. The catalyst was drop-cast on carbon cloth to reach a precious metal loading of ~1 mg cm^{-2} for Co-Pt-Ru/NC and a precious metal loading of 4 $mg_{Pt}\ cm^{-2}$ for Pt/C as anode and cathode, respectively. The MEAs were fabricated by sandwiching the Nafion117 membrane between the anode and cathode and followed by hot pressing at 120 °C under a pressure of 0.2 MPa for 3 min. The assembled DMFC was tested at 60 °C by feeding 1 M CH_3OH at a flow rate of 5 ml min^{-1} and purging

humidified oxygen at a flow rate of 40 ml min^{-1}. The commercial PtRu/C was used for comparison with a precious metal loading of 2 mg_{PtRu} cm^{-2}. Sigracet 22 BB from Fuel Cell Store was used as the gas diffusion layer. The active area of MEA is 1 × 1 cm^2. When changing the external conditions, the methanol flow rate was changed to 20 ml min^{-1} or O_2 flow rate was changed to 100 ml min^{-1}.

Data acquisition

To collect training data more efficiently and accurately, we employed a three-electrode system to do CA measurements for MOR. To mimic or calculate 'pseudo power' for the fuel cell, we assumed the potential for the other half reaction (oxygen reduction reaction) is 0.9 V_{RHE}. The power can be calculated by $P = (0.9 - E_{RHE}) \times I$, where E_{RHE} represents the potential for CA measurements of MOR and I is the current. During switching CA measurements, there are two steps including the working time for working potential and the resting time for resting potential. To minimize the effect of methanol evaporation during long-time measurements, two pumps were used to refresh the electrolyte; one is used to feed the new electrolyte and the other is used to pump the tested electrolyte to the waste liquid tank, thus maintaining the methanol concentration in the electrolyte at a constant value (1 M CH_3OH).

Data collection is conducted using our proposed automated controller system. This automated system receives 4D action variables as input every 5 minutes and performs the given control for 5 minutes. This means that the control conditions change every 5 minutes and we call this 'step'. We assume that our controller follows a Markov decision process (MDP) where an action only depends on its previous state. The action space at ith step is set as a 4D vector consisting of resting potential, resting time, working potential and working time. The state at ith step is defined as the concatenation of the trajectory of $(i-1)$th action and the current trajectory as the cell response to that. The reward (power) at ith step R_i is calculated as the produced power during ith step, which is the area under the curve of the cell over time.

$$R_i = \frac{1}{T}\int_0^T I_i(t)\, V_i(t)\, \mathrm{d}t$$

where T is the time interval at ith timestep, I_i is the current trajectory at ith timestep, V_i is the trajectory of potential controlled at ith timestep.

Like typical RL, our model for controlling the DMFC system is trained from accumulated data according to the cell operation. We acquired 1,000 random action data points as warm-up.

Data augmentation

In developing RL framework for the control of direct methanol fuel cells, a data augmentation strategy was employed to enhance data efficiency. Central to our approach was the assumption that adjacent states in the operational space share similar cell conditions. Our model was trained using an augmented dataset, wherein each known state's data was used to generate additional, synthetic data points for its adjacent states. This data-efficient approach improved the model's performance and reduced the need for extensive data collection, accelerating the training process and increasing the feasibility of implementing RL for fuel cell control in practical settings.

Alpha-Fuel-Cell algorithm

Our αFC algorithm is inspired by the actor–critic method[35], which is a temporal difference variant of policy gradient, having two neural networks: an actor and a critic. While the actor is trained to optimize the goal using a policy gradient approach, the critic is trained to calculate the action-value function so that the actor returns better actions. Our critic module similarly predicts rewards from state and action but sets the discount factor to 0 so that it only considers immediate reward. Instead of directly predicting action from the state using a neural network, our actor calculates the desired action, which is optimized based on the gradient to achieve the target reward by utilizing the differentiability of the neural network.

Critic module

In our model for direct methanol fuel cell control, a Convolutional Neural Network (CNN) was employed to process the state inputs. The primary input to the CNN comprised the cell's current response to voltage manipulation over a period of five minutes. This time-series data, capturing the intricate dynamics of the fuel cell, provided a comprehensive view of its current state. Following the CNN processing, the model focused on a 4D action space, defined by resting time, resting potential, working time and working potential. These action parameters were fed into an action-head network, where they underwent embedding to capture the nuanced relationships and operational constraints inherent in the fuel cell's functioning. The embedded action representations were then concatenated with the CNN's output, forming a rich, integrated feature set. This combined feature set was further processed using a Multi-Layer Perceptron (MLP). The MLP, leveraging its deep learning capability, was tasked with predicting the cell's power output for the given operational interval.

The state input to the CNN at a certain step i, denoted as S_i, is composed of the cell's current (I) and voltage (V) responses, collected over a step of five minutes. This is represented as a multivariate time-series data:

$$S_i \equiv \{(I_1, V_1), (I_2, V_2), \ldots, (I_T, V_T)\}$$

where (I_t, V_t) represents the current and voltage at time t, and T is the total number of time steps. The CNN processes this input to extract relevant features:

$$\mathbf{z}_{i,\mathrm{state}} = f_\theta(S_i)$$

where f_θ is a CNN parameterized by a parameter set θ and $\mathbf{z}_{i,\mathrm{state}}$ is the hidden state as CNN output at the ith step. The action space is a 4D vector $\mathbf{A}_i = (a_{\mathrm{rest.time}}, a_{\mathrm{rest.pot}}, a_{\mathrm{work.time}}, a_{\mathrm{work.pot}})$, representing resting time, resting potential, working time and working potential, respectively. This action vector at step i is embedded using an action-head network to capture the complex dynamics of the fuel cell:

$$\mathbf{z}_{i,\mathrm{action}} = g_\varphi(\mathbf{A}_i)$$

where g_φ is an action-head network parameterized by a parameter set φ and $\mathbf{z}_{i,\mathrm{action}}$ is the hidden state as the output of the action-head network. The embedded action representation $\mathbf{z}_{i,\mathrm{action}}$ is concatenated with the CNN output $\mathbf{z}_{i,\mathrm{state}}$ and fed into an MLP for predicting the power output:

$$\hat{R}_i = h_\xi([\mathbf{z}_{i,\mathrm{state}}, \mathbf{z}_{i,\mathrm{action}}])$$

where h_ξ is an MLP parameterized by a parameter set ξ and $\hat{R}_i$ is the reward prediction of the model.

This model is trained to predict the cell's power output during the specified interval. The training process involves optimizing the parameters θ, φ, ξ of the CNN, action-head network and MLP, respectively, to maximize the operational efficiency of the fuel cell.

$$\mathcal{L}(\theta, \varphi, \xi) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{R}_i - R_i\right)^2}$$

where n is a number of data points.

The optimization is typically performed using a gradient-based method, aiming to reduce the prediction error of the power output,

which directly correlates with improving the fuel cell's performance under various operating conditions.

$$\theta, \varphi, \xi = \underset{\theta,\varphi,\xi}{\operatorname{argmin}} \mathcal{L}(\theta, \varphi, \xi)$$

This targeted optimization helps in fine tuning the control strategies for the fuel cell, enhancing both its efficiency and longevity.

Action module operation

Our control system controls the DMFC system by changing action variables every 5 minutes. By utilizing the differentiability of the neural network, we adopted gradient-based optimization for selecting the next action variable. The initial action variable $\mathbf{A}_{\text{seed}}$ is defined by a uniform distribution function from its minimum $\mathbf{A}_{\text{min}}$ and maximum $\mathbf{A}_{\text{max}}$.

$$\mathbf{A}_{\text{seed}} \sim \mathcal{U}(\mathbf{A}_{\text{min}}, \mathbf{A}_{\text{max}})$$

$$\mathbf{A}_i = \underset{\mathbf{A}_{\text{seed}}}{\operatorname{argmin}} \left(h_\xi\left(\left[f_\theta(S_i), g_\varphi(\mathbf{A}_{\text{seed}})\right]\right) - R_{\text{target}}\right)^2$$

$$S_{i+1} \sim P(S_{i+1}, |, S_i, A_i)$$

where R_{target} is the desired power to be controlled and $\mathbf{A}_i$ is the action variable for the step. Next state $(i+1)$th step is sampled according to the probability distribution P defined by DMFC cell reaction kinetics. Our automated control framework iteratively performs this optimization to continuously control the DMFC system.

Automated control and measurement

To automatically control and measure the DMFC systems, EC-Lab software is combined with PyAutoGUI[36] using Python programming language. Once the programme starts, αFC conducts the first action defined as applying a constant 0.65 V for 5 min. The produced power during 5 min is calculated from the raw trajectories of potential and current. Raw trajectories and produced power are saved as a JSON file. The actor module receives the state as input and selects the most appropriate action for the desired output, considering the state of the cell at that time. Then, the critic module receives both action and state as input and predicts the DMFC output for the next 5 minutes. The actions selected by the actor module are used to control the connected DMFC, which acts as an environment that responds to actions and returns rewards, through EC-Lab. The trajectory over 5 minutes is recorded to serve as the next state, and the power produced is used as labels to train the output of the critic module. This process iterates every 5 minutes until the experiment ends.

Real applications of αFC

After the training, several experiments were conducted to demonstrate the ability of αFC. First, the controlling experiment was carried out in a three-electrode system by setting different targets. As the highest reward of that condition is around 0.3 mW, therefore, the target was set as 0.5 mW to maximize power. The controlling process was in the following order: maximum power, 0.2 mW, 0.1 mW, back to maximum power, −0.1 mW and back to maximum power. To compare the effectiveness of αFC, different strategies were employed in three-electrode systems, including PtRu/C with constant potential, Co-Pt-Ru/NC with constant potential, switching strategy and αFC. The final demonstration was on DMFCs by using PtRu/C and Co-Pt-Ru/NC with αFC.

CO stripping experiments

The CO stripping activities of Co-Pt-Ru/NC after long-term measurements controlled by constant potential and αFC were compared. Typically, after the long-term test, the catalysts were subjected to CO adsorption at 0.05 V vs RHE for 30 minutes and the solution was then purged with Ar to remove unbound CO. Stripping data were collected by CV at a scan rate of 10 mV s^{-1}. The current was corrected with background current.

In situ electrochemical mass spectrometer

All electrochemical measurements with mass spectra were carried out with a commercially available microchip-based electrochemistry mass spectrometry (ECMS) set-up (SpectroInlets ApS, Denmark). Five-mm diameter glassy carbon was used as the working electrode. The counter (carbon rod) and reference (Ag/AgCl) electrodes were inserted in a glass tube with a ceramic frit on the tip, respectively. The 0.1 M $HClO_4$ was used as an electrolyte first to collect the background mass signal. The mass spectra of different species (H_2, H_2O, O_2, CO_2) were recorded while the CV of MOR was measured from ~0.05 V_{RHE} to ~1.2 V_{RHE}. The mass spectra of these species during CA test and controlled by αFC were also recorded.

Proportional-Integral-Derivative control of DMFCs

PID control (Proportional-Integral-Derivative Control) is a control technique that generates a control signal so that the output of the system reaches the desired value. The PID control signal input $u(t)$ is defined as follows:

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau)\, d\tau + K_d \frac{de(t)}{dt}$$

where $e(t)$ is the error between the output of the system and the desired value at time t, K_p is the proportional gain, K_i is the integral gain and K_d is the derivative gain.

In our DMFC system, each gain is a 4D vector. In general, there are various tuning methods such as the Ziegler–Nichols method, which determines the gain from the vibration of the system to obtain each gain. In our study, the gains were determined empirically through simulation.

Data availability

The data that support the findings of this study are available within the Article and its Supplementary Information. Source data are provided with this paper.

Code availability

The αFC code associated with this manuscript is available via GitHub at https://github.com/parkyjmit/alphaFC.

References

1. Yao, Z. et al. Machine learning for a sustainable energy future. *Nat. Rev. Mater.* **8**, 202–215 (2022).
2. Attia, P. M. et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **578**, 397–402 (2020).
3. Jordan, M. I. & Mitchell, T. M. Machine learning: trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).
4. Rosen, A. S. et al. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter* **4**, 1578–1597 (2021).
5. Yao, Z. et al. Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nat. Mach. Intell.* **3**, 76–86 (2021).
6. Jiao, K. et al. Designing the next generation of proton-exchange membrane fuel cells. *Nature* **595**, 361–369 (2021).
7. Zhu, G. et al. Rechargeable Na/Cl_2 and Li/Cl_2 batteries. *Nature* **596**, 525–530 (2021).
8. Wu, Y., Jiang, Z., Lu, X., Liang, Y. & Wang, H. Domino electroreduction of CO_2 to methanol on a molecular catalyst. *Nature* **575**, 639–642 (2019).

9. Davidson, D. J. Exnovating for a renewable energy transition. *Nat. Energy* **4**, 254–256 (2019).
10. Feng, Y., Liu, H. & Yang, J. A selective electrocatalyst–based direct methanol fuel cell operated at high concentrations of methanol. *Sci. Adv.* **3**, e1700580 (2017).
11. Martín, A. J., Mitchell, S., Mondelli, C., Jaydev, S. & Pérez-Ramírez, J. Unifying views on catalyst deactivation. *Nat. Catal.* **5**, 854–866 (2022).
12. Poerwoprajitno, A. R. et al. A single-Pt-atom-on-Ru-nanoparticle electrocatalyst for CO-resilient methanol oxidation. *Nat. Catal.* **5**, 231–237 (2022).
13. Wang, J. et al. Toward electrocatalytic methanol oxidation reaction: longstanding debates and emerging catalysts. *Adv. Mater.* **35**, 2211099 (2023).
14. Xu, H. et al. A cobalt-platinum-ruthenium system for acidic methanol oxidation. *Chem. Mater.* **36**, 6938–6949 (2024).
15. Timoshenko, J. et al. Steering the structure and selectivity of CO_2 electroreduction catalysts by potential pulses. *Nat. Catal.* **5**, 259–267 (2022).
16. Xu, L. et al. In situ periodic regeneration of catalyst during CO_2 electroreduction to C_{2+} products. *Angew. Chem. Int. Ed.* **61**, e202210375 (2022).
17. Zhang, X.-D. et al. Asymmetric low-frequency pulsed strategy enables ultralong CO_2 reduction stability and controllable product selectivity. *J. Am. Chem. Soc.* **145**, 2195–2206 (2023).
18. Rabissi, C., Brightman, E., Hinds, G. & Casalegno, A. In operando investigation of anode overpotential dynamics in direct methanol fuel cells. *Int. J. Hydrogen Energy* **41**, 18221–18225 (2016).
19. Rabissi, C., Brightman, E., Hinds, G. & Casalegno, A. In operando measurement of localised cathode potential to mitigate DMFC temporary degradation. *Int. J. Hydrogen* **43**, 9797–9802 (2018).
20. Bresciani, F., Casalegno, A., Bonde, J. L., Odgaard, M. & Marchesi, R. A comparison of operating strategies to reduce DMFC degradation: a comparison of operating strategies to reduce DMFC degradation. *Int. J. Energy Res.* **38**, 117–124 (2014).
21. Kusne, A. G. et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, 5966 (2020).
22. Szymanski, N. J. et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 86–91 (2023).
23. Severson, K. A. et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy* **4**, 383–391 (2019).
24. Seo, J. et al. Avoiding fusion plasma tearing instability with deep reinforcement learning. *Nature* **626**, 746–751 (2024).
25. Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
26. Silver, D. et al. Mastering the game of go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
27. Silver, D. et al. Mastering the game of go without human knowledge. *Nature* **550**, 354–359 (2017).
28. Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. Proximal policy optimization algorithms. Preprint at https://arxiv.org/abs/1707.06347 (2017).
29. Haarnoja, T., Zhou, A., Abbeel, P. & Levine, S. Soft actor–critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proc. 35th International Conference on Machine Learning* 1861–1870 (PMLR, 2018).
30. Hessel, M. et al. Rainbow: combining improvements in deep reinforcement learning. In *Thirty-second AAAI Conference on Artificial Intelligence (AAAI-18)* 3215–3222 (AAAI, 2018).
31. Mnih, V. et al. Playing Atari with deep reinforcement learning. Preprint at https://arxiv.org/abs/1312.5602 (2013).
32. Panzer, M. & Bender, B. Deep reinforcement learning in production systems: a systematic literature review. *Int. J. Prod. Res.* **60**, 4316–4341 (2022).
33. Degrave, J. et al. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602**, 414–419 (2022).
34. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019).
35. Kort-Kamp, W. J. M. et al. Machine learning-guided design of direct methanol fuel cells with a platinum group metal-free cathode. *J. Power Sources* **626**, 235758 (2025).
36. Sweigart, A. PyAutoGUI documentation. *Read the Docs* https://pyautogui.readthedocs.org (2020).

Acknowledgements

We acknowledge support by the Defense Advanced Research Projects Agency (DARPA) under agreement number HR00112490369. This work made use of the Materials Research Laboratory Shared Experimental Facilities at Massachusetts Institute of Technology. This work was carried out in part through the use of MIT.nano's facilities. This work was performed in part at the Center for Nanoscale Systems of Harvard University. H.X. acknowledges Y. Rao for discussion and support. Y.J.P. also acknowledges support by a grant from the National Research Foundation of Korea (NRF) funded by the Korean government, Ministry of Science and ICT (MSIT) (number RS-2024-00356670).

Author contributions

J.L., H.X. and Y.J.P. conceived the original idea. H.X. performed the synthesis, the electrochemical measurements, data collection and model construction. Y.J.P. developed and trained the αFC model. Z.R. developed automated measurements for experiments. H.X., D.J.Z. and D.M. performed ECMS measurements. D.Z., H.J., C.D. and G.Z. participated in the discussion. H.X. and Y.J.P. verified and analysed data. H.X., Y.J.P., J.L., Y.S.-H. and Y.R.-L. drafted the manuscript. All authors contributed to the revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41560-025-01804-x.

Correspondence and requests for materials should be addressed to Yang Shao-Horn or Ju Li.

Peer review information *Nature Energy* thanks Zhongbao Wei, Lei Xing and the other, anonymous, reviewer for their contribution to the peer review of this work.