AIAA SciTech Forum
12-16 January 2026, Orlando, FL
AIAA SCITECH 2026 Forum

10.2514/6.2026-0292

Machine Learning-Based Active Control of Supersonic Twin-Rectangular Jet Flow

Brandon Yeung* and Oliver T. Schmidt†
University of California San Diego, La Jolla, California 92093

We investigate the feasibility of applying reinforcement learning (RL) to the active flow control of a supersonic twin-rectangular turbulent jet for noise reduction. Large-eddy simulations (LES) of the Mach 1.5 twin jet, with a Reynolds number of one million, constitute the learning environment. Control of the jet flow is realized through plasma actuators placed within the nozzles. The environment is observed via a three-dimensional grid of near-field pressure sensors. The reward signal is provided by an additional far-field pressure sensor, which measures the overall sound pressure level (OASPL) at a location relative to the nozzles that is representative of the final checker on an aircraft carrier. To learn a control policy that seeks to minimize the OASPL, we employ proximal policy optimization. For computational tractability, we train RL agents independently on low- and intermediate-resolution meshes. The agent trained on the low-resolution mesh is able to increase its reward over 1000 episodes of training, but its learned policy does not successfully generalize to a fully-resolved LES. The agent trained on the intermediate-resolution mesh is unable to outperform a random control policy after 300 episodes of training. In the context of numerical simulations, our results demonstrate that significant hurdles remain in the application of RL to turbulent jet noise reduction, and to the control of high-dimensional, highly stochastic systems more broadly.

Nomenclature

A_t	=	advantage
a	=	actuation signal
D_e	=	equivalent nozzle diameter, $D_\mathrm{e} = 1.6h$
E_t	=	expectation
f	=	frequency
h	=	nozzle height
L	=	objective function
St	=	Strouhal number, $fD_\mathrm{e}/u_\mathrm{j}$
s_t, a_t	=	state, action
t	=	time
t_on, t_off	=	time instants when actuator switches on or off
t_r	=	actuator rise time
x, y, z	=	Cartesian coordinates
ϵ	=	clipping parameter
ϕ	=	actuation phase shift
π	=	policy
τ	=	actuation period
θ	=	policy parameters

Subscript

j	=	fully-expanded nozzle exit property
max	=	maximum

*PhD Student, Department of Mechanical and Aerospace Engineering, Student Member AIAA.
†Associate Professor, Department of Mechanical and Aerospace Engineering, Senior Member AIAA.

I. Introduction

Among the most pressing challenges in fluid mechanics and its applications is the reduction of jet noise. While decades of jet noise research have significantly advanced the community's understanding of the sources of jet noise, including the critical role played by coherent structures [1] and the predictive ability of linearized models [2], new insights continue to be generated [3–5] that enrich this understanding but also highlight its incompleteness. In parallel with long-running efforts to understand jet noise are extensive studies into methods of jet noise reduction, which have met with only modest success [6, 7]. The need for robust noise control is particularly acute in high-Reynolds number, supersonic jets, found typically on tactical aircraft, that can lead to hearing loss as well as structural fatigue [8, 9]. For this class of jets, the difficulty of control is compounded by the desire to balance noise reduction with tactical performance, as well as the high forcing amplitude required to manipulate such flows [10].

For turbulent jet noise reduction, both open- and closed-loop control strategies have been attempted. The latter includes sensing of the output, generally yielding superior performance [11], and is the approach we take in this study. Closed-loop control has seen extensive use in subsonic jets; we refer the reader to Audiffred et al. [12] for a recent overview. By contrast, closed-loop control of supersonic free jets has been applied more sparingly. In particular, Doty et al. [13] conducted experiments with open- and closed-loop control of a supersonic rectangular jet using active chevrons, finding comparable and small broadband noise reduction in both cases. Zhou et al. [14] performed numerical simulations of resolvent-based closed-loop control of wavepackets. In the present study, we apply closed-loop control to high-fidelity numerical simulations of a supersonic jet. More specifically, we attempt to control the supersonic twin-rectangular jet developed by Samimy et al. [15] that was later simulated by Brès et al. [16, 17] and Yeung and Schmidt [18]. Given the high computational cost of numerical simulations, specifically large-eddy simulations (LES), a fundamental goal of this proof of concept is to address the feasibility of such an approach.

Since the work of Rabault et al. [19], there has been an explosion of interest in performing active flow control using deep reinforcement learning (RL), in which the control law, or RL agent, is parameterized by a neural network (see e.g. Schmidhuber [20] for a review, in addition to the standard text by Sutton and Barto [21]). This approach exploits the property of neural networks as universal function approximators [22] that can represent arbitrary control laws. Rather than analytically deriving a control law, it seeks to optimize one via trial-and-error learning. A further advantage of RL-based flow control is that it does not require the environment, in this case the LES solver, to be differentiable, instead treating the environment as a black box. In flow control, the most popular RL algorithm by far is proximal policy optimization, or PPO [23]. PPO has demonstrated state-of-the-art performance in a variety of control tasks, including stabilizing cylinder wakes [19, 24–26], falling films [27], and airfoil wakes [28], reducing the length of separation bubbles [29], and manipulating wall turbulence [30]. In this work, we employ PPO to control supersonic twin-rectangular jet noise. We measure the efficacy of PPO in terms of the amount of noise reduction it achieves relative to the natural jet. Because no benchmark for substantial jet noise reduction yet exists, the task is exceptionally challenging.

The remainder of this paper is organized as follows. Section II.A describes the set-up of the LES environment. Section II.B outlines the PPO algorithm and reports the hyperparameters used in training. Results are shown in section III, then summarized and discussed in section IV.

II. RL-based active flow control

A. Simulation environments

LES of the supersonic twin-rectangular jet, with which the RL agent interacts, are carried out using the flow solver 'Charles' [31]. The numerical set-up of the natural jet has been extensively validated by Brès et al. [16, 17] and documented by Yeung et al. [32] and Yeung and Schmidt [18]. For completeness, here we provide a brief overview. The nominally ideally-expanded and cold jet has a Mach number of 1.5 and a Reynolds number of approximately one million based on the equivalent nozzle diameter. The equivalent diameter is $D_\mathrm{e} = 1.6h$, where h is the nozzle height. In what follows, all lengths are non-dimensionalized by h. By convention, frequencies, f, are reported in terms of the Strouhal number, $St = fD_\mathrm{e}/u_\mathrm{j}$, where u_j is the nozzle exit velocity. For this study, three different computational grids are employed: a high-fidelity, validated grid containing 52 million control volumes, a medium-fidelity grid containing seven million control volumes, and a low-fidelity grid containing 241 thousand control volumes. The computational cost of training the RL agent on the most refined grid would be prohibitive. As such, the agent is first trained on the coarser grids. Once the agent is trained, we evaluate its performance on the refined grid in order to obtain an accurate estimate of the noise of the controlled jet.

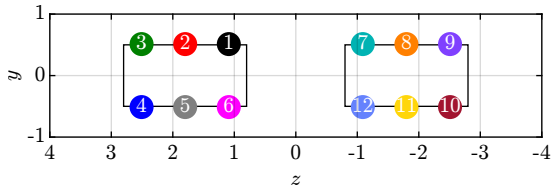

Fig. 1 The position of the centroid of each actuator in the y-z plane. The actuator numbering and colors match those in figures 4 and 5.

Active forcing of the twin jet is effected through localized arc filament plasma actuators [33] positioned immediately upstream of the nozzle lip, matching the set-up of companion experiments [15]. A total of 12 actuators are present, six in each nozzle. Their locations in the y-z plane are displayed in figure 1. The numerical implementation of the actuators is identical to that in Yeung and Schmidt [18, 34, 35]. Each of the 12 actuators can be controlled independently from the others. For the ith actuator, the actuation signal is a smoothed pulse wave expressed in the form

$$a_i(t) = \frac{1}{2}\left[\tanh\left(\frac{t - \lfloor t/\tau_i \rfloor \tau_i - t_{\mathrm{on},i}}{t_\mathrm{r}}\right) - \tanh\left(\frac{t - \lfloor t/\tau_i \rfloor \tau_i - (t_{\mathrm{on},i} + \delta t)}{t_\mathrm{r}}\right)\right], \tag{1}$$

where the rise time, t_r, and the duration of each pulse, δt, are empirical constants. The period, τ_i, and the time window, $t \in [t_{\mathrm{on},i}, t_{\mathrm{on},i} + \delta t]$, within each period that the actuator is switched on, are learned parameters. More concretely, the action space consists of the forcing frequency, $f_i = 1/\tau_i \in [0, f_\mathrm{max}]$, and phase shift, $\phi_i \in [0, 1]$, such that $t_{\mathrm{on},i} = \phi\tau_i$. The upper bound, f_max, on f_i corresponds to the cutoff frequency of the computational grid, $St \approx 0.5$ and 1 for the low- and medium-fidelity grids, respectively. Within these bounds, the action space is continuous. While more general control laws were considered (for example, the agent could learn to control a_i directly), we opt for a more restrictive functional form that mimics the behavior of experimental plasma actuators [15]. Moreover, constraining the control law to periodic forcing—albeit with time-varying frequencies and phase shifts—reduces the size of the search space.

The observation space consists of 470 scalar pressure probe measurements. Among these, 350 probes sample the volume given by $x \in [0, 30]$, $y \in [-5.3, 5.3]$, and $z \in [-6.6, 6.6]$. These dimensions approximately match the size of the Ffowcs Williams-Hawkings (FW-H) [36] surface in Brès et al. [16, 17] for the same twin-rectangular jet. The remaining 120 probes sample the wake of each of the 12 actuators. We apply a moving variance to the pressure to estimate the short-time-averaged overall sound pressure level (OASPL) at each probe location, expressed in units of decibels. Following standard practice, the observation is normalized before being fed to the agent.

For the reward, an additional pressure probe is placed at the location $x = 0$, $y = 0$, and $z = 12$, which is directly to the sideline of the nozzles. Assuming a dimensional equivalent nozzle diameter of 3.3 feet, the distance of the probe from the midpoint between the twin nozzles is approximately 25 feet. For an aircraft during takeoff from an aircraft carrier, this distance corresponds to the typical separation between the nozzle and the final checker, who is among the personnel exposed to the highest noise levels on a carrier deck [37]. At this distance, an acoustic analogy would improve our acoustic prediction. However, we have confirmed that the power spectrum obtained from the final checker probe accurately reproduces the FW-H prediction up to the grid cutoff frequency. To simplify the learning process, we thus opt to use direct probe measurements exclusively. Consistent with the observation-space probes, the pressure at the final checker probe is transformed into the short-time-averaged OASPL. The latter, subtracted from the OASPL of the natural jet at the same location, serves as the reward at each step. In other words, if the controlled jet is quieter than the natural jet, the reward will be positive.

B. Learning algorithm

To train the RL agent, we use the PPO implementation in Stable Baselines3 [38]. The PPO algorithm has been amply documented (see e.g. Schulman et al. [23] and Huang et al. [39]), so we will only outline the basics. PPO is an on-policy, policy gradient method that uses a clipped objective function to remove the incentive for overly aggressive model updates that might otherwise lead to worse performance. At its core, the algorithm seeks to maximize the objective

$$L(s_t, a_t, \theta, \theta_\mathrm{old}) = \mathrm{E}_t\left\{\min\left(\frac{\pi_\theta(s_t, a_t)}{\pi_{\theta_\mathrm{old}}(s_t, a_t)} A_t, \mathrm{clip}\left(\frac{\pi_\theta(s_t, a_t)}{\pi_{\theta_\mathrm{old}}(s_t, a_t)}, 1 - \epsilon, 1 + \epsilon\right) A_t\right)\right\}, \tag{2}$$

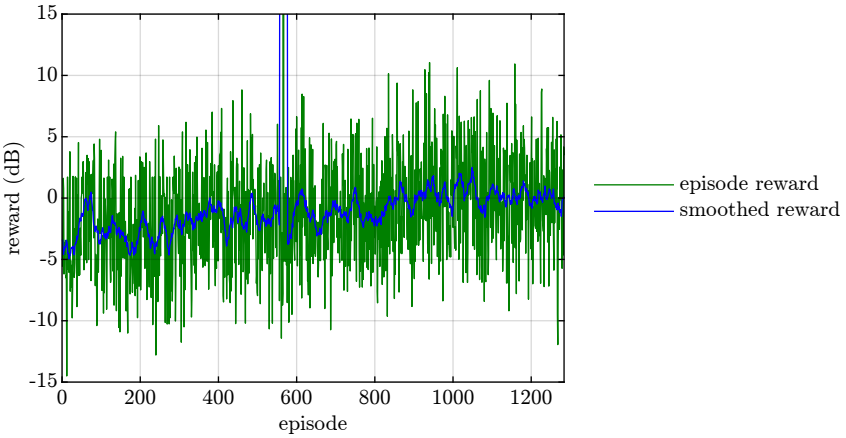

Fig. 2 Learning history over a single training run on the low-resolution grid.

where s_t and a_t are the state and action, respectively, at time step t, θ are the parameters of the policy π, $\mathrm{E}_t\{\cdot\}$ denotes the expectation, and A_t is the generalized advantage estimate [40]. The function $\mathrm{clip}(\cdot, 1-\epsilon, 1+\epsilon)$ clips its argument to the bounds given by $[1-\epsilon, 1+\epsilon]$, where ϵ is a hyperparameter. In essence, if $A_t > 0$, we wish to encourage the current state-action pair, i.e., we update the policy π_θ such that the ratio of probabilities, $\pi_\theta(s_t, a_t)/\pi_{\theta_{\mathrm{old}}}(s_t, a_t)$, is higher. However, due to the clipping function, a ratio that exceeds $1+\epsilon$ makes no additional contribution to the objective. Conversely, if $A_t < 0$, we wish to discourage the current state-action pair, i.e., we update the policy π_θ such that the ratio of action probabilities, $\pi_\theta(s_t, a_t)/\pi_{\theta_{\mathrm{old}}}(s_t, a_t)$, is lower. However, a ratio less than $1-\epsilon$ is clipped. The clipped objective thus disincentivizes large policy updates, with the size of the updates implicitly constrained by ϵ.

Interaction between the RL library and the LES solver is enabled by wrapping the solver in the standard Gymnasium interface [41]. The two components communicate with each other by reading and writing text files to disk. The agent interacts with the simulation environment once every 20 000 LES time steps. An episode terminates after 220 000 LES time steps, i.e., 11 interactions. Upon termination, the last snapshot of every episode is stored and set aside. Similar to Rabault et al. [19] and Li and Zhang [25], a new episode starts from a random member of the stored snapshots with 20% probability, or the last snapshot of the previous episode with 80% probability. The latter allows for control strategies that act over long time horizons. The rollout buffer has a length of 48 episodes, or 528 interactions. When the buffer is filled, the policy and value networks are updated using Adam [42], with minibatches of eight interactions. All other hyperparameters are set to the default values provided by Stable Baselines3.

III. Results

A. Training in low-resolution environment

In the initial training attempt, the policy and value networks are parameterized by multilayer perceptrons (MLPs), each with two hidden layers of 64 units each and activated by tanh nonlinearity. The agent is first trained on a single training run, with one random seed, on the 241 thousand-control volume grid. We carry out the training on a local workstation. The learning history, that is, the reward in each episode, is reported in figure 2. The episodic reward exhibits significant episode-to-episode variance, which is consistent with the highly stochastic nature of the turbulent twin-jet simulation. We have confirmed that the outlier at the 566th episode, while initially promising, is not reproducible. Also shown in figure 2 is the smoothed reward, obtained from a moving average of the episodic reward. The moving window—20 episodes—matches the buffer size of the PPO agent. The smoothed reward shows a clear upward trend, increasing approximately from -5 dB (i.e., 5 dB louder than the natural jet) to 0 dB (on a par with the natural jet) after 1000 episodes of training. Thereafter the smoothed reward reaches a plateau. Given that the episodic reward frequently exceeds 5 dB (i.e., 5 dB quieter than the natural jet), particularly in the latter half of the training, the plateau could indicate that the agent is trapped in a local minimum or saddle in the loss landscape.

To investigate the effectiveness of the learned policy in a high-fidelity environment that accurately captures noise emission, we evaluate the trained agent on the 52 million-control volume grid over six episodes. The large grid size requires substantial computing resources. We therefore perform the evaluation on a high-performance computing

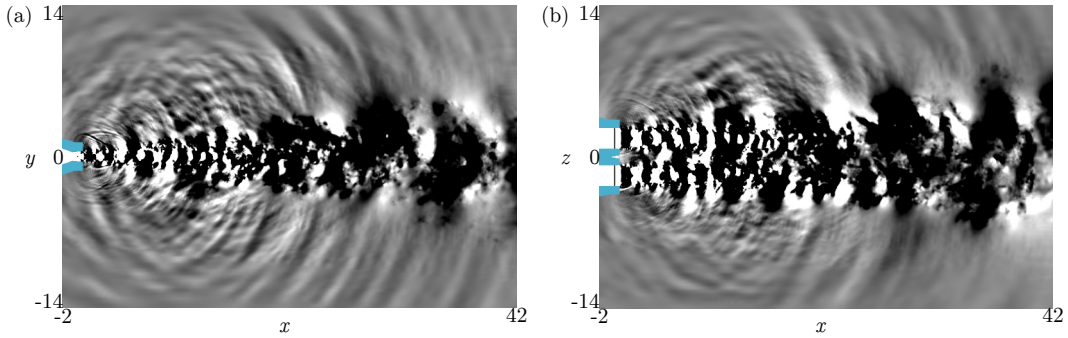

Fig. 3 Instantaneous pressure at the end of the fifth evaluation episode, corresponding to panel (e) in figures 4, 5, and 6, visualized on the $z = 1.8$ plane (a) and $y = 0$ plane (b).

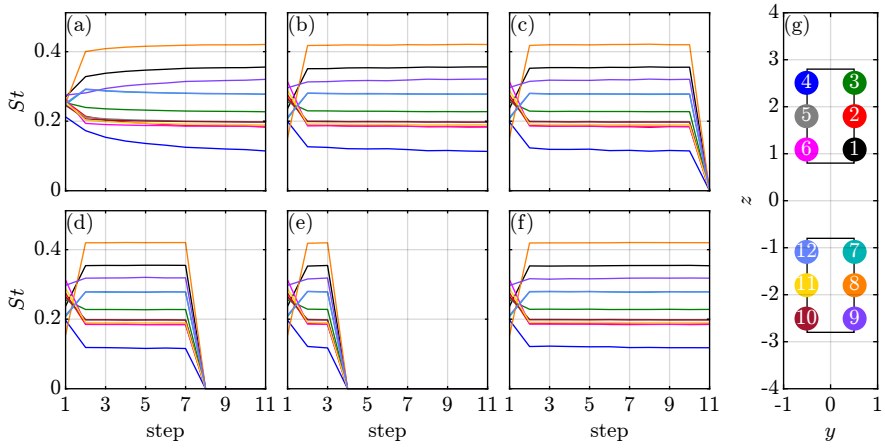

Fig. 4 Actuation frequencies when the trained model is evaluated on the fully-resolved grid over six episodes. Panels (a) through (f) correspond to episodes one through six, respectively. For reference, panel (g) reproduces figure 1. The frequencies of actuator two, five, and seven are nearly identical to those of actuators six, 10, and 12, respectively. The former are thus difficult to see.

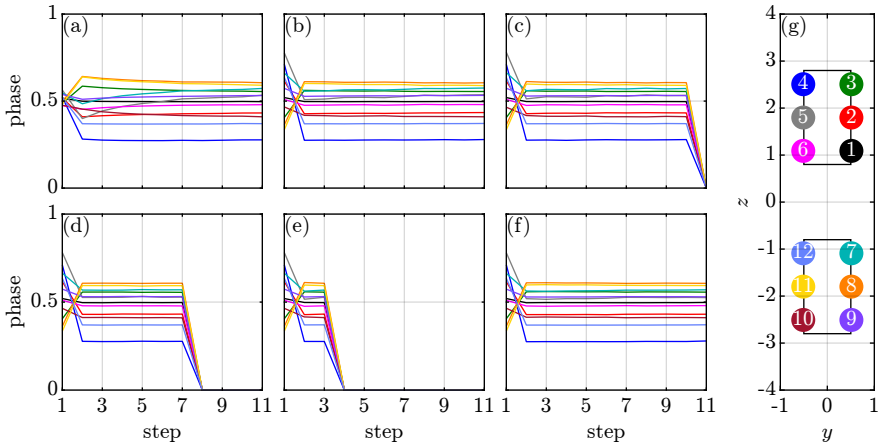

Fig. 5 Same as figure 4 but for actuation phase shifts.

(HPC) cluster. Unlike training, during evaluation the agent no longer needs to explore the action space. As such, for evaluation only, the agent acts deterministically, always returning the action with the highest probability for a given observation. Figure 3 shows a snapshot of the instantaneous pressure field. The actions of the agent, consisting of actuation frequencies and phase shifts, are reported in figures 4 and 5, respectively. All six episodes show similar behaviors. In figure 4, at the start of each episode, the 12 actuators fire at comparable frequencies, $0.2 \lesssim St \lesssim 0.3$. Starting in the second step, the actuation frequencies diverge in the same pattern. In particular, from highest to lowest frequency, the actuators are always ordered consistently: actuators eight, one, nine, 12 and seven (tied), three, five and 10 (tied), 11, six and two (tied), and four. Out of the first six on this list, five actuators are on the upper nozzle lip, $y = 0.5$. In other words, actuators on the upper lip fire at higher frequency than those on the lower lip, $y = -0.5$. There is no obvious difference between frequencies on the left and right nozzles. Actuators that fire at nearly identical frequencies do not show a clear pattern: seven and 12 share the same z-coordinate, but neither two and six nor five and 10 do.

As shown in figure 5, the actuation phase shifts are also consistent between episodes. Starting in the second step, each actuator maintains approximately constant phase. The learned policy appears to include pairs of actuators firing at close to the same phase. Specifically, actuators eight and 11, three and seven, five and nine, one and six, and two and 10 are roughly paired. Actuators four and 12 are unpaired and distinct from the rest. This stands in contrast to the actuation frequencies, which are mostly distinct. The actuators with paired phases appear to follow a pattern. Actuators eight and 11, and one and six, share the same z-coordinate. Three and seven share the same y-coordinate, and are in the same location on their respective nozzles. The pairs five and nine, and two and 10, are mirrored about the major axis, $y = 0$. From largest to smallest phase, the actuators (or actuator pairs) are ordered as follows: eight and 11 (tied), three and seven (tied), five and nine (tied), one and six (tied), two and 10 (tied), 12, and four.

For each episode, the reward received from the environment at each step is shown in figure 6, along with the average reward over each episode. While some episodes, particularly the first, produce instantaneous positive rewards, none of the episodes yield a positive average reward. In other words, the controlled jet is louder on average than the natural jet in all evaluation episodes. As shown in figures 4(d,e), 5(d,e) and 6(d,e), in the fourth and fifth episodes, the agent switches off all actuators before the end of each episode. Once the actuators are off, the instantaneous reward is zero by definition. A hypothesis for the premature stopping of active control is that, because the reward is almost always negative during evaluation, to maximize the return the trained agent has an incentive to simply stop actuating, whereupon the reward increases to zero.

The low- and high-resolution grids used to train and evaluate the agent, respectively, differ in size by a factor of over 200. It is clear from these results that the gap is far too large for an agent trained on the smaller grid to generalize to the larger grid successfully. This motivates the use of an intermediate-fidelity, and much costlier, simulation for training.

B. Training in intermediate-resolution environment

As described in section II.A, the intermediate-resolution training environment has a grid size of approximately seven million, 30 times the size of the previous, low-resolution environment. To ensure there is no influence from the low-resolution environment on the new RL agent, the policy and value networks are re-initialized and re-trained from

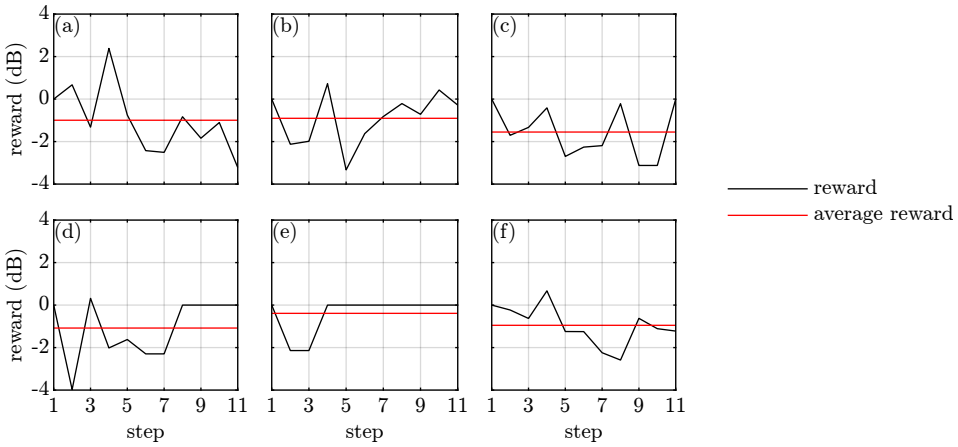

Fig. 6 Same as figure 4 but for rewards. The black curves indicate the instantaneous rewards. The red lines mark the reward averaged over each episode.

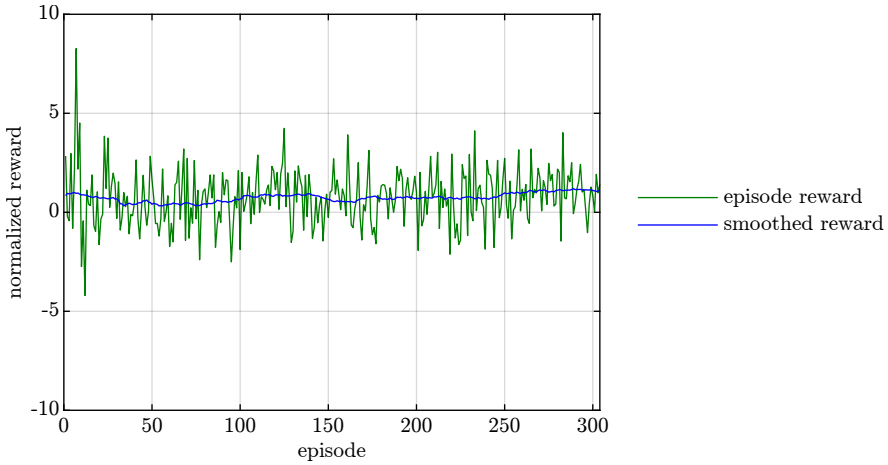

Fig. 7 Learning history over a single training run on the intermediate-resolution grid.

scratch. Increasing the mesh resolution improves the modeling of the plasma actuators [16], in addition to that of the acoustics. The larger training mesh size now requires both the training and the evaluation to take place on the HPC. To accelerate learning, the agent interacts with and collects experiences from multiple independent environments running in parallel. In addition, we modify the architecture of the neural networks to allow them to represent more complex control laws as well as better align with established practice [19]. Specifically, we represent the PPO agent by MLPs with two hidden layers of 512 units each, compared to 64 units in the initial training. Nonlinearity is provided by ReLU activation.

The learning history is reported in figure 7. During training, the reward is normalized by the running standard deviation. Accordingly, the normalized reward is shown. To aid interpretation, we also show the smoothed reward by applying a moving average filter to the episode reward. As before, the moving window is equal in length to the rollout buffer size, that is, 48 episodes. Over the course of about 300 episodes of training, the learning history shows no sign of a sustained increase in reward. The PPO agent is thus unable to make progress towards noise reduction in the intermediate-resolution environment. Given the lack of training progress, we choose not to evaluate the agent further.

IV. Summary and discussion

We conduct a feasibility study of RL-based closed-loop control of a supersonic twin-rectangular turbulent jet, with a Reynolds number of approximately one million, in an LES environment. The RL agent is parameterized by MLPs and is trained using the PPO algorithm with the objective of minimizing jet noise. To keep computational costs tractable, the

training is carried out on a low- and an intermediate-resolution grid, each containing 241 thousand and seven million grid points, respectively. When the agent is trained, we evaluate its performance on a refined grid containing 52 million grid points, where the hydrodynamic and acoustic fluctuations of the jet are fully resolved. The control input consists of 12 plasma actuators in the nozzles. Each actuator may be independently controlled by the RL agent. Sensing of the observation space is provided by 470 pressure probes distributed throughout the computational domain. The OASPL at an additional pressure probe, whose location relative to the nozzle is representative of the final checker on an aircraft carrier deck, supplies the reward signal.

Given the constraints on computational resources, we perform a total of two independent training runs, one on the low-resolution grid and one on the intermediate-resolution grid. On the low-resolution grid, the episode reward displays high variance and frequently exceeds 5 dB of noise increase as well as reduction. When averaged over the length of the rollout buffer, the smoothed reward grows from around -5 dB (i.e., 5 dB louder than the natural jet) at the start of training to around 0 dB after approximately 1000 episodes, beyond which it reaches a plateau. It remains unclear if the agent has converged to the optimal policy or is merely trapped in a local maximum. When the trained agent is evaluated on the fully-resolved grid over six episodes, it consistently yields negative average rewards, indicating increases in noise compared to the natural jet. The inability of the trained agent to achieve any noise reduction on the refined grid suggests the learned policy exhibits poor generalizability. To seek an improved control law, we train an independent PPO agent on the intermediate-resolution grid. However, after 300 episodes of training, the reward has shown no significant increase. In other words, the agent has not discovered a control strategy that performs better than a random policy.

It is well-known that the performance of an RL set-up should ideally be measured over a large number of trials [43]. The learning histories demonstrated in this proof of concept therefore may not be statistically robust. In particular, it is entirely possible that the agent could achieve greater noise reduction if, for example, we employ hyperparameter tuning, experiment with another RL algorithm, restart the LES from different flow states, or even just vary the random seed. In practice, it is far too computationally costly to conduct repeated trials involving high-fidelity simulations of high-Reynolds number flows, though this may be more attainable in an experimental setting [44]. Despite these caveats, it is clear from our feasibility study that RL-based active control is severely challenged by high-dimensional stochastic systems such as turbulent flows. Developing low-cost surrogate models of the flow and techniques to accelerate convergence of flow statistics, including the OASPL, could facilitate future RL control efforts. An open question is whether it is practical to implement output sensing for closed-loop control on a real aircraft, which could be more amenable to open-loop strategies.

Acknowledgments

We gratefully acknowledge support from Office of Naval Research award N00014-23-1-2457, under the supervision of Dr. Steve Martens. LES calculations were carried out on the "Carpenter" Cray EX4000 systems in ERDC DSRC, using allocations provided by DoD HPCMP.

References

[1] Jordan, P., and Colonius, T., "Wave packets and turbulent jet noise," *Annu. Rev. Fluid Mech.*, Vol. 45, 2013, pp. 173–195.

[2] Tam, C. K. W., and Hu, F. Q., "On the three families of instability waves of high-speed jets," *J. Fluid Mech.*, Vol. 201, 1989, pp. 447–483.

[3] Towne, A., Cavalieri, A. V. G., Jordan, P., Colonius, T., Schmidt, O., Jaunet, V., and Brès, G. A., "Acoustic resonance in the potential core of subsonic jets," *J. Fluid Mech.*, Vol. 825, 2017, pp. 1113–1152.

[4] Schmidt, O. T., Towne, A., Colonius, T., Cavalieri, A. V. G., Jordan, P., and Brès, G. A., "Wavepackets and trapped acoustic modes in a turbulent jet: coherent structure eduction and global stability," *J. Fluid Mech.*, Vol. 825, 2017, pp. 1153–1181.

[5] Tam, C. K. W., "New Noise Component of Supersonic Jets: Laboratory Jets to High-Performance Aircraft," *AIAA J.*, Vol. 0, No. 0, 2025, pp. 1–12.

[6] Brown, C. A., "Acoustics of excited jets—a historical perspective," *J. Acoust. Soc. Am.*, Vol. 118, No. 3, 2005, p. 1863.

[7] Henderson, B., "Fifty years of fluidic injection for jet noise reduction," *Int. J. Aeroacoust.*, Vol. 9, No. 1–2, 2010, pp. 91–122.

[8] Tam, C. K. W., "Supersonic jet noise," *Annu. Rev. Fluid Mech.*, Vol. 27, 1995, pp. 17–43.

[9] Wall, A. T., Gee, K. L., Morris, P. J., Colonius, T., and Lowe, K. T., "Introduction to the special issue on supersonic jet noise," *J. Acoust. Soc. Am.*, Vol. 151, No. 2, 2022, pp. 806–816.

[10] Samimy, M., Adamovich, I., Webb, B., Kastner, J., Hileman, J., Keshav, S., and Palm, P., "Development and characterization of plasma actuators for high-speed jet control," *Exp. Fluids*, Vol. 37, 2004, pp. 577–588.

[11] Brunton, S. L., and Noack, B. R., "Closed-Loop Turbulence Control: Progress and Challenges," *Appl. Mech. Rev.*, Vol. 67, No. 5, 2015, p. 050801.

[12] Audiffred, D. B. S., Cavalieri, A. V. G., Maia, I. A., Martini, E., and Jordan, P., "Reactive experimental control of turbulent jets," *J. Fluid Mech.*, Vol. 994, 2024, p. A15.

[13] Doty, M. J., Fuller, C. R., Schiller, N. H., and Turner, T. L., "Active Noise Control of Radiated Noise From Jets," Tech. Rep. TM-2013-218041, NASA, September 2013.

[14] Zhou, Y., Jung, J., Bhagwat, R., and Towne, A., "Estimation and closed-loop control of supersonic jet noise," *APS Division of Fluid Dynamics Meeting Abstracts*, APS, 2025.

[15] Samimy, M., Webb, N., Esfahani, A., and Leahy, R., "Perturbation-based active flow control in overexpanded to underexpanded supersonic rectangular twin jets," *J. Fluid Mech.*, Vol. 959, No. A13, 2023, pp. 1–34.

[16] Brès, G. A., Yeung, B. C. Y., Schmidt, O. T., Esfahani, A., Webb, N., Samimy, M., and Colonius, T., "Towards large-eddy simulations of supersonic jets from twin rectangular nozzle with plasma actuation," *AIAA paper 2021-2154*, 2021.

[17] Brès, G. A., Bose, S. T., Ivey, C. B., Emory, M., and Ham, F., "GPU-accelerated large-eddy simulations of supersonic jets from twin rectangular nozzle," *AIAA paper 2022-3001*, 2022.

[18] Yeung, B., and Schmidt, O. T., "Spectral dynamics of natural and forced supersonic twin-rectangular jet flow," *J. Fluid Mech.*, Vol. 1018, 2025, p. A34.

[19] Rabault, J., Kuchta, M., Jensen, A., Réglade, U., and Cerardi, N., "Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control," *J. Fluid Mech.*, Vol. 865, 2019, pp. 281–302.

[20] Schmidhuber, J., "Deep learning in neural networks: An overview," *Neural Netw.*, Vol. 61, 2015, pp. 85–117.

[21] Sutton, R. S., and Barto, A. G., *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, 2018.

[22] Hornik, K., Stinchcombe, M., and White, H., "Multilayer feedforward networks are universal approximators," *Neural Netw.*, Vol. 2, No. 5, 1989, pp. 359–366.

[23] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O., "Proximal Policy Optimization Algorithms," arXiv, 2017.

[24] Tang, H., Rabault, J., Kuhnle, A., Wang, Y., and Wang, T., "Robust active flow control over a range of Reynolds numbers using an artificial neural network trained through deep reinforcement learning," *Phys. Fluids*, Vol. 32, No. 5, 2020, p. 053605.

[25] Li, J., and Zhang, M., "Reinforcement-learning-based control of confined cylinder wakes with stability analyses," *J. Fluid Mech.*, Vol. 932, 2022, p. A44.

[26] Wang, Q., Yan, L., Hu, G., Li, C., Xiao, Y., Xiong, H., Rabault, J., and Noack, B. R., "DRLinFluids: An open-source Python platform of coupling deep reinforcement learning and OpenFOAM," *Phys. Fluids*, Vol. 34, No. 8, 2022, p. 081801.

[27] Belus, V., Rabault, J., Viquerat, J., Che, Z., Hachem, E., and Reglade, U., "Exploiting locality and translational invariance to design effective deep reinforcement learning control of the 1-dimensional unstable falling liquid film," *AIP Adv.*, Vol. 9, No. 12, 2019, p. 125014.

[28] Paris, R., Beneddine, S., and Dandois, J., "Reinforcement-learning-based actuator selection method for active flow control," *J. of Fluid Mech.*, Vol. 955, 2023, p. A8.

[29] Font, B., Alcántara-Ávila, F., Rabault, J., Vinuesa, R., and Lehmkuhl, O., "Active flow control of a turbulent separation bubble through deep reinforcement learning," *J. Phys.: Conf. Ser.*, Vol. 2753, No. 1, 2024, p. 012022.

[30] Cavallazzi, G. M., Guastoni, L., Vinuesa, R., and A., P., "Deep Reinforcement Learning for the Management of the Wall Regeneration Cycle in Wall-Bounded Turbulent Flows," *Flow Turbulence Combust.*, 2024.

[31] Brès, G. A., Ham, F. E., Nichols, J. W., and Lele, S. K., "Unstructured Large-Eddy Simulations of Supersonic Jets," *AIAA J.*, Vol. 55, No. 4, 2017, pp. 1164–1184.

[32] Yeung, B. C. Y., Schmidt, O. T., and Brès, G. A., "Three-Dimensional Spectral POD of Supersonic Twin-Rectangular Jet Flow," *AIAA paper 2022-3345*, 2022.

[33] Samimy, M., Kim, J.-H., Kastner, J., Adamovich, I., and Utkin, Y., "Active control of high-speed and high-Reynolds-number jets using plasma actuators," *J. Fluid Mech.*, Vol. 578, 2007, pp. 305–330.

[34] Yeung, B. C. Y., and Schmidt, O. T., "Plasma Actuation and Bispectral Mode Decomposition of Supersonic Twin-Rectangular Jet Flow," *AIAA paper 2023-4177*, 2023.

[35] Yeung, B. C. Y., and Schmidt, O. T., "Space-Time Proper Orthogonal Decomposition of Actuation Transients of a Plasma-Controlled Twin-Rectangular Jet," *AIAA paper 2024-4192*, 2024.

[36] Ffowcs Williams, J. E., Hawkings, D. L., and Lighthill, M. J., "Sound generation by turbulence and surfaces in arbitrary motion," *Phil. Trans. R. Soc. A*, Vol. 264, No. 1151, 1969, pp. 321–342.

[37] McKinley, R. L., Wall, A. T., Aubert, A. C., Smith, M. J., and Nantz, R. E., "CVN Carrier Flight Deck Noise Levels During F-35C Operations," Tech. Rep. JSF20-201, F-35 Joint Program Office, December 2019.

[38] Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., and Dormann, N., "Stable-Baselines3: Reliable Reinforcement Learning Implementations," *J. Mach. Learn. Res.*, Vol. 22, No. 268, 2021, pp. 1–8.

[39] Huang, S., Dossa, R. F. J., Raffin, A., Kanervisto, A., and Wang, W., "The 37 Implementation Details of Proximal Policy Optimization," *ICLR Blog Track*, 2022.

[40] Schulman, J., Moritz, P., Levine, S., Jordan, M., and Abbeel, P., "High-Dimensional Continuous Control Using Generalized Advantage Estimation," arXiv, 2018.

[41] Towers, M., Kwiatkowski, A., Terry, J., Balis, J. U., De Cola, G., Deleu, T., ao, M. G., Kallinteris, A., Krimmel, M., KG, A., Perez-Vicente, R., Pierré, A., Schulhoff, S., Tai, J. J., Tan, H., and Younis, O. G., "Gymnasium: A Standard Interface for Reinforcement Learning Environments," arXiv, 2024.

[42] Kingma, D. P., and Ba, J., "Adam: A Method for Stochastic Optimization," arXiv, 2017.

[43] Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., and Meger, D., "Deep Reinforcement Learning That Matters," *AAAI*, Vol. 32, No. 1, 2018.

[44] Zhang, J., Xia, C., Jiang, X., Fumarola, I., and Rigas, G., "Autonomous real-time control of turbulent dynamics," arXiv, 2025.