Combustion and Flame 282 (2025) 114406

Contents lists available at ScienceDirect

Combustion and Flame

journal homepage: www.sciencedirect.com/journal/combustion-and-flame

Deep reinforcement learning for adaptive control of thermoacoustic instabilities in a lean-premixed methane/hydrogen/air combustor

Bassem Akoush [a], Guillaume Vignat [a,*], Ryan Finley [a], Wai Tong Chung [a], Matthias Ihme [a,b,c,*]

[a] Department of Mechanical Engineering, Stanford University, Stanford CA 94305, USA
[b] Department of Photon Science, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[c] Department of Energy Science and Engineering, Stanford University, Stanford CA 94305, USA

ARTICLE INFO

Keywords:
Deep reinforcement learning
Machine learning
Combustion instability
Hydrogen combustion
Fuel-flexible combustion
Multi-sensor control

ABSTRACT

Thermoacoustic instabilities are a challenge in the design and operation of combustion systems. Addressing this challenge is becoming even more critical with the development of fuel-flexible combustors capable of operating with hydrogen and other sustainable fuel sources. While active control is a well-known method for damping combustion instabilities, identifying appropriate control parameters becomes increasingly complex in the presence of changing fuel composition and operating conditions. In this work, we present a model-free deep reinforcement learning (RL) technique to adaptively tune an active control system. We demonstrate that the RL-based active control system is able to adaptively suppress thermoacoustic instabilities over an extended range of operating conditions with minimal training. The demonstration is performed on a laboratory-scale bluff-body-stabilized premixed methane/hydrogen/air flame, at equivalence ratios ranging from 0.5 to stoichiometric, and with up to 80%$_{vol}$ hydrogen in the fuel. After training the RL system on a single operating condition, combustion instabilities can be mitigated over the entire operating range of the burner. Extending the training to three additional operating conditions allows the RL control system to fine-tune its policy and further reduce thermoacoustic instabilities, achieving a sixfold reduction in the acoustic source term over most of the operating range. We observe a reduction of up to 40 dB in acoustic pressure over 50% of the operating range. The proposed approach offers a promising path towards more efficient, adaptive control systems for thermoacoustic instabilities, demonstrating the potential of RL to address the operational challenges of fuel-flexible combustion systems.

Novelty and Significance Statement

We show the first experimental demonstration of a reinforcement learning-based control method for thermoacoustic instabilities. The experiments are performed on a laboratory-scale premixed methane/hydrogen/air bluff-body burner, which exhibits strong combustion instabilities over a wide range of operating conditions. Building upon a conventional control system, which utilizes a pressure sensor, an acoustic driver, and a gain- and phase-shift controller, the reinforcement learning-based controller is able to dampen instabilities over the entire operating range. This is achieved while training the controller on a single operating condition. Extending training to a total of four distinct operating conditions further fine-tunes the control policy and yields an additional reduction in the acoustic pressure amplitude. This research illustrates the potential of reinforcement learning for robust control in combustion systems - capable of addressing the challenges of complex combustion physics, adapting to unseen conditions, and merging information from heterogeneous sensors.

1. Introduction

Thermoacoustic instabilities are a major challenge in the design and operation of combustion systems, ranging from jet and rocket engines to industrial furnaces and stationary gas turbines [1]. Thermoacoustic instabilities typically occur due to the feedback loop between acoustic pressure fluctuations, injector response, and unsteady heat release rate, leading to undesirable effects, such as reduced efficiency, higher emissions, and structural damage to the combustor [2]. Increasingly stringent emission requirements have led to a shift towards lean-premixed combustors, which are more susceptible to thermoacoustic oscillations due to their sensitivity to acoustic perturbations [3] and higher power densities compared to conventional combustors [4]. With

* Correspondence to: Stanford University – Dept. of Mechanical Engineering, Bldg 530 Room 202, 440 Escondido Mall, Stanford CA 94305, USA.
E-mail addresses: gvignat@stanford.edu (G. Vignat), mihme@stanford.edu (M. Ihme).

https://doi.org/10.1016/j.combustflame.2025.114406
Received 8 April 2025; Received in revised form 5 August 2025; Accepted 7 August 2025
Available online 18 September 2025

Nomenclature

Latin

a	Agent action
$\mathcal{A}$	Action space
$\mathcal{B}$	Batch of transition tuples
$\mathcal{D}$	Replay buffer
D_c	Diameter of combustor tube (mm)
e'_s	Synthesized acoustic signal
f	Frequency (Hz)
f_0	Fundamental frequency of the thermoacoustic oscillation (Hz)
g	Gain of the controller
I	Chemiluminescence intensity
J	Objective function
L_c	Length of combustor tube (mm)
$\dot{m}$	Mass flow rate ($\mathrm{kg\,s^{-1}}$)
N_s	Maximum number of environment steps
$\mathcal{P}$	Transition dynamics
p'	Acoustic pressure (Pa)
p_0	Atmospheric pressure (Pa)
Q	Q-critic network
$\dot{Q}$	Volumetrically integrated heat release rate (W)
r	Reward
RI	Rayleigh integral (W)
s	State vector of the environment
$\mathcal{S}$	State space
S_{inj}	Injector cross-sectional area ($\mathrm{m^2}$)
$\mathfrak{T}_{\mathrm{rec}}$	Duration of data acquisition for state vector estimation (s)
$\mathfrak{T}_{\mathrm{wait}}$	Wait time before sampling signals (s)
u_b	Injector bulk velocity ($\mathrm{m\,s^{-1}}$)
V	Value network
$X_{\mathrm{H_2}}$	Hydrogen volume fraction (fuel)

Greek

β	Smoothing factor of the target value network
γ	Heat capacity ratio
λ	Learning rate of gradient descent
Φ	Trainable weights of Q-critic network
ϕ	Equivalence ratio
π	Policy network
ψ	Trainable weights of policy network
ρ_u	Density of unburnt reactants ($\mathrm{kg/m^3}$)
τ	Time delay of the controller (s)
θ	Trainable weights of value network
$\overline{\theta}$	trainable weights of target value network
ξ	Discount factor

Subscripts and Superscripts

$\bar{\cdot}$	Mean value
$\|\cdot\|$	Magnitude
$\tilde{\cdot}$	Evaluated with the current policy
ang $[\cdot]$	Phase angle
rms	Root mean square

Abbreviations

a.u.	Arbitrary unit
CDF	Cumulative distribution function
CPSD	Cross power spectral density
FLAME	Flame environment
FPGA	Field programmable gate arrays
HEX	Heat exchanger
MFC	Mass flow controller
NN	Neural network
PMT	Photo multiplier tube
PSD	Power spectral density
RL	Deep reinforcement learning
SAC	Soft actor-critic

the transition towards decarbonization in the energy and transportation sectors, there is also a growing interest in using carbon-free and sustainable fuels, such as hydrogen (H_2), ammonia (NH_3), and other synthetic fuels to power stationary gas turbines and jet engines. Hydrogen in particular exacerbates the challenges posed by thermoacoustic instabilities due to its high flame speed, higher power density, and the increased susceptibility of the flame to acoustic and convective perturbations [5,6].

Over the past decades, various methods have been proposed to suppress thermoacoustic instabilities, broadly categorized into passive and active methods [2]. Passive methods rely on altering either the acoustic properties of the combustor, or the flame response to acoustic perturbations. Acoustic methods passively mitigate thermoacoustic instabilities using baffles and other devices to modify the shape of acoustic eigenmodes [7], by controlling the acoustic coupling between cavities [8,9], or by increasing acoustic damping using perforated liners or resonators [10]. Tuning the flame response to dampen the thermoacoustic source term can be achieved by a variety of schemes: modifying the topology of the flame by changing the injector geometry [11], fuel composition [12], or fuel staging [13], which may affect combustion performance and pollutant emissions; or by creating a destructive interference in the flame and injector response to acoustic perturbations [14,15]. These passive control methods have been successfully integrated in practical systems [2,16], but limitations, such as weight, increased pollutant emissions, and degradation of other critical performance metrics can restrict their application. In addition, many of these passive methods must be tuned to target specific eigenfrequencies of the combustor, which reduces their range of applicability and adaptability for changing operating conditions and fuel composition.

In active control schemes, sensors are used to monitor the state of the flame in order to synthesize an input signal to actuators [17,18]. Active control systems fall into two broad categories depending on their response time [18]. "Slow" control systems have response times on the order of 100 ms or slower and essentially act as adjustable passive control schemes. In contrast, "fast" control systems respond to input perturbations within timescales much shorter than the period of the thermoacoustic oscillation. The present work focuses on "fast" control of thermoacoustic. Of particular relevance to the present study are early experimental demonstrations of active control of combustion instabilities using microphones, acoustic drivers, and phase-shift controllers [19,20]. In these control schemes, the signal e'_s used to drive the acoustic actuator is given by

$$e'_s(t) = g p'(t-\tau), \tag{1}$$

where p' is the acoustic pressure measured by the microphone, g is the controller's gain parameter, and τ is its delay parameter. This method has been extended to multiple actuators [21], and demonstrated in full-scale industrial applications [22].

Thermoacoustic oscillations are known to be particularly sensitive to even small changes in combustor geometry and operating conditions [23]. Given the growing emphasis on fuel flexibility in propulsion and power generation [5,6], the versatility of active control systems, whose response can be optimized by tuning control parameters, is an attractive prospect. This has sparked renewed interest towards the development of adaptive control schemes for thermoacoustic applications. Early attempts by Demayo et al. [24] focused on developing a robust control system that utilizes a stack of feedback sensor to monitor both thermoacoustics and pollutant emissions. This approach proved effective across different configurations, demonstrating the ability to optimize the system performance within 10 to 15 min. Liu et al. [25] used linear genetic programming to simultaneously suppress thermoacoustic instability and reduce emissions through modulation of the fuel stream. Dharmaputra et al. [26] used Bayesian optimization to optimize the parameters of a phase-shift controller in simulations and experiments, demonstrating the convergence to the global optimum for the controller's parameters. Additionally, their framework allowed to constrain the final acoustic pressure below a specified threshold and enabled knowledge transfer across different operating conditions. However, Bayesian optimization schemes do not adapt to unseen conditions and require re-running the optimization process, which can limit their deployment for real-time applications [27].

Over the past decade, deep reinforcement learning (RL) has demonstrated significant advances across various applications, including robotics control [28], drug discovery [29], natural language processing [30], flow control [31,32] and combustion [33–35]. RL integrates deep neural networks to handle high-dimensional state and action spaces, enabling the model to learn more complex representations from multi-modal inputs. Compared to Bayesian methods, RL is scalable and capable of handling larger state and action spaces. It has the advantage of dynamically adapting to unseen conditions and uncertain environments in contrast to conventional approaches. For example, Alhazmi and Sarathy [36] showed, on a reduced-order computational thermoacoustic model, that soft actor-critic (SAC), a model-free RL approach, was able to adjust the parameters of a phase-shift controller, and that SAC outperformed other control methods, such as extremum seeking control, H-infinity, and self-tuning regulator. The numerical results from their work illustrate the potential of using RL for suppressing combustion instabilities.

In the present work, we mitigate thermoacoustic instabilities in a laboratory-scale turbulent premixed burner, operated over a wide range of equivalence ratio and CH_4/H_2 mixing ratios, using a SAC-based adaptive control framework. Specifically, the objectives of this research are (1) to demonstrate experimentally the use of RL in a real-time combustion application; (2) to develop a highly adaptive control framework that is capable of dampening combustion instabilities over a large range of operating conditions; and (3) by developing this control framework, to address the technical challenges posed by thermoacoustic instabilities in fuel-flexible combustors that operate across a wide range of CH_4/H_2 fuel mixtures.

The structure of this article is outlined as follows: Section 2 presents a detailed description of the experimental configuration. Section 3 describes the thermoacoustic behavior of the burner in the absence of active control, providing a baseline for further discussion. In Section 4, we introduce the SAC model and present the active control framework. The results from training and testing the RL-SAC model are discussed in Section 5. Finally, Section 6 concludes the article with a summary of the key contributions of this work.

2. Experimental methods

This section discusses the key components of the experimental setup, which is used to evaluate the effectiveness of our adaptive control policy.

2.1. Burner setup

Experiments are conducted on a fully premixed turbulent bluff body burner, which is schematically shown in Fig. 1. This burner closely resembles experimental rigs investigated at Cambridge University [37] and the Norwegian University of Science and Technology (NTNU) [38].

Methane (Linde, Danbury, CT, USA, > 99% purity), hydrogen (Linde, Danbury, CT, USA, > 99.99% purity) and laboratory clean dry air are premixed far upstream of the burner using an array of mass flow controllers (MFCs, Alicat Scientific, Tucson, AZ, USA, combined accuracy better than 0.8%). The reactants are not pre-heated and all experiments are conducted at ambient pressure. The mixture enters a settling plenum through four equally-spaced ports before passing into the injector through 72 ports, each having a diameter of 1 mm. The injector consists of a 19 mm-diameter cylindrical tube, with a 5 mm-diameter inner rod. A 45° conical bluff-body, made of 304L stainless steel, is used to anchor the flame with a diameter of 12.7 mm at the combustion chamber backplane. The bluff body assembly is centered using three grub screws located 51 mm upstream of the exit plane. The combustion chamber consists of a transparent quartz tube with an inner diameter $D_c = 70$ mm and length $L_c = 305$ mm.

2.2. Sensors

Combustion instabilities are commonly characterized by measuring the pressure fluctuations p' and volumetrically integrated heat release rate (HRR) $\dot{Q}$ in the flame region. In the present work, acoustic pressure fluctuations near the dump plane of the burner are measured using a high-dynamic range 1/4″ pressure field condenser microphone (model 378C10, PCB Piezotronics, Depew, NY, USA), connected to a model 482C15 signal conditioner (PCB Piezotronics, Depew, NY, USA). The microphone is placed on a non-reflecting semi-infinite water-cooled acoustic waveguide, which follows the design described by Rajendram Soundararajan et al. [39]. This waveguide introduces an acoustic delay of approximately 2 ms. To estimate the HRR, two photomultiplier tubes (PMTs, model H11902-110, Hamamatsu photonics, Hamamatsu, Japan) with a cut-off frequency of 20 kHz are used to record volume-integrated chemiluminescence from the flame. In the perfectly premixed flames investigated in the present work, time-resolved chemiluminescence measurements provide a qualitative indicator of the instantaneous heat release rate of the flame and, to some extent, allow to infer the equivalence ratio of the flame [40]. The first PMT is equipped with an optical bandpass filter with a center wavelength of 310 nm and a full width at half maximum (FWHM) of 10 nm (Asahi spectra, Tokyo, Japan) to record chemiluminescence from OH^* radicals in the flame. The second PMT is equipped with a 430 nm-centered optical bandpass filter (FWHM 10 nm, Asahi Spectra, Tokyo, Japan) to monitor CH^* chemiluminescence. Signals from the microphone and both PMTs are recorded on a National Instrument BNC-2110 data acquisition card at a sampling rate of 40 kHz. We also perform chemiluminescence imaging to examine the flame shape and its dynamics. The imaging setup and results are shown in Appendix A.

3. Thermoacoustic behavior of the burner

Even though self-excited thermoacoustic instabilities in confined premixed bluff-body-stabilized CH_4/H_2/air flames have been studied [38,41–43], the pronounced sensitivity of thermoacoustic instabilities to burner geometry and boundary conditions [23] warrants a thorough characterization of the thermoacoustic behavior of this specific burner across its operating range in order to establish a baseline for the active control results discussed in Section 5.

Fig. 2 illustrates a typical acoustic pressure signal recorded during a thermoacoustic instability. The acoustic pressure signal is harmonic, at a frequency close to $f_0 \approx 500$ Hz, marked with a red dot in the power spectral density in Fig. 2, corresponding to the first longitudinal mode

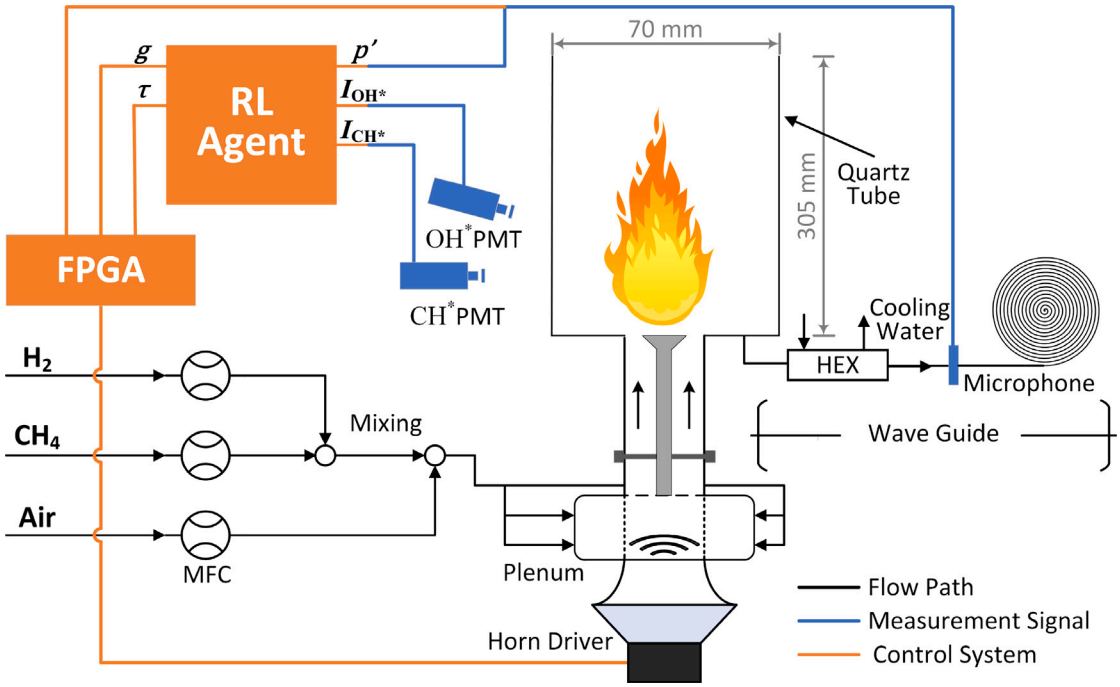

Fig. 1. Experimental setup of lab-scale burner. The RL agent operates on a standard desktop computer with a built-in data acquisition card. It communicates the control parameters (gain g and phase shift τ) to the real-time control loop, which operates on a field programmable gate array (FPGA) device. MFC: mass flow controller; PMT: photomultiplier tube; HEX: heat exchanger.

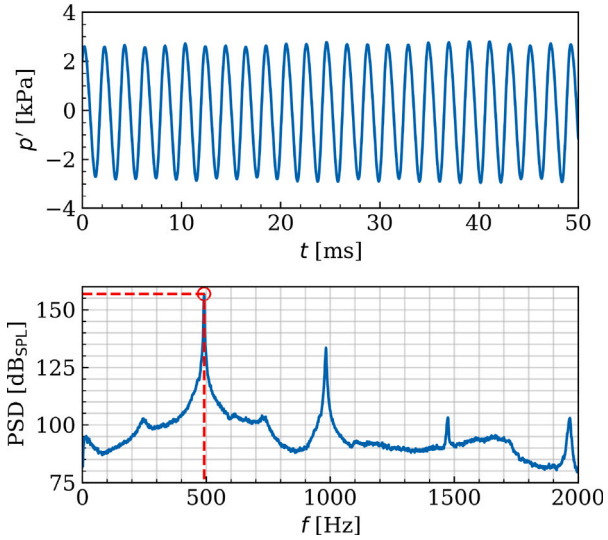

Fig. 2. Acoustic pressure signal measured by the microphone during a thermoacoustic oscillation. The burner was operated at a representative condition, $u_b = 22.5\,\mathrm{m\,s^{-1}}$, $\phi = 0.8$, and $X_{H_2} = 55\%$. Pressure signal in the time domain (top) and power spectral density of the pressure signal (bottom). The red circle and red dashed lines illustrate the fundamental frequency of the thermoacoustic oscillation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the combustion chamber. In what follows, we report the amplitude and frequency f_0 of the instability based on the maximum of the power spectral density of the acoustic pressure measured at the combustor backplane.

As a first step to characterize the burner, we examine its thermoacoustic behavior as a function of two key operating parameters: the equivalence ratio ϕ and the hydrogen volume fraction in the fuel X_{H_2} while keeping the bulk velocity of the reactants through the injection system constant at $u_b = \dot{m}/\left(S_{\mathrm{inj}}\rho_u\right) = 22.5\,\mathrm{m\,s^{-1}}$, with $\dot{m}$ the reactant mass flow rate, S_{inj} the cross-sectional area of the injection system at the dump plan, and ρ_u the density of the unburnt reactants. ϕ and X_{H_2} are known to affect the flame shape, length, and angle in this type of burner, therefore significantly affecting the flame transfer function and the thermoacoustic stability of the burner [38]. This is ideal to characterize an adaptive control scheme. The results are shown in stability maps in Fig. 3 in which we report the pressure amplitude at the fundamental frequency (Fig. 3(a)), the frequency of the pressure fluctuations (Fig. 3(b)), and the Rayleigh integral (Fig. 3(c)) of thermoacoustic oscillations as a function of operating conditions. The thermoacoustic source term, also known as Rayleigh integral RI, is calculated as the contribution of the heat release rate to the acoustic power radiated by the flame [44,45]:

$$RI = \frac{\gamma - 1}{\gamma}\frac{\overline{\dot{Q}}}{\mathcal{T}}\int_{t=0}^{\mathcal{T}}\frac{p'}{p_0}\frac{I'_{OH^*}}{\overline{I_{OH^*}}}dt, \tag{2}$$

where $'$, $\overline{\cdot}$, respectively, denote fluctuations and mean values of the signal, γ is the heat capacity ratio, $\overline{\dot{Q}}$ is the total time-averaged heat release rate of the flame, I_{OH^*} is the integrated OH^* chemiluminescence emission of the flame, assumed to be proportional to the instantaneous heat release rate for this premixed flame, and p_0 is the atmospheric pressure. $\mathcal{T}$ is the integration time used for the calculation. We compute $\overline{\dot{Q}}$ using the reactant mass flow rates while assuming complete combustion. High values of RI are observed during combustion instabilities characterized by high-amplitude pressure fluctuations that are in phase with the unsteady heat release. In contrast, under thermoacoustically stable conditions, the acoustic source term remains low and close to zero, indicating reduced fluctuation amplitudes and weak coupling between pressure oscillations and unsteady heat release.

The amplitude of the acoustic pressure fluctuations near the combustor dump plan, Fig. 3(a), shows a strong dependency on X_{H_2}, increasing from approximately 110 $\mathrm{dB_{SPL}}$ at the fundamental frequency with pure CH_4 to 150 $\mathrm{dB_{SPL}}$ for operation at higher H_2 mixture, $X_{H_2} \gtrsim 40\%$. This increase in acoustic pressure level is due to a supercritical Hopf bifurcation, a common occurrence in thermoacoustic

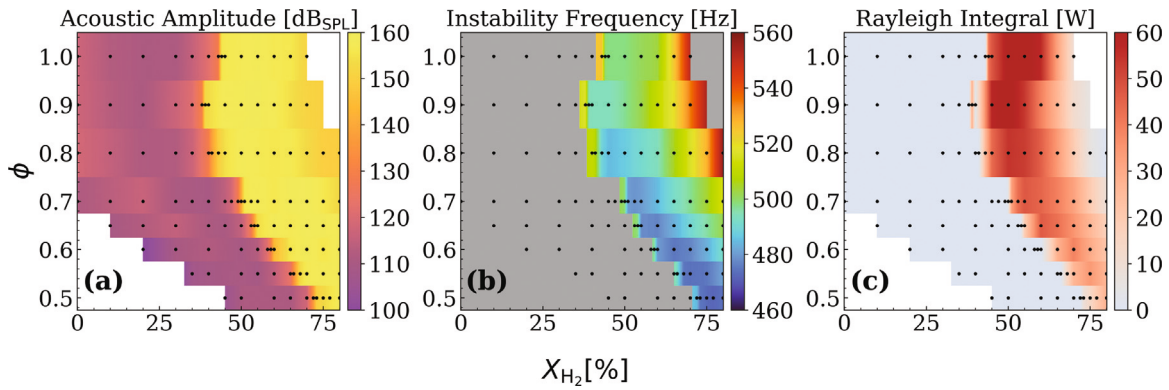

Fig. 3. Stability maps. (a) Acoustic pressure amplitude at the fundamental frequency, measured at the combustion chamber's dump plane [dB$_{SPL}$]; (b) fundamental frequency of the instability f_0 [Hz]; (c) Rayleigh integral RI [W]. In (b), gray color is used to indicate that no prominent peak is found in the PSD, due to the absence of thermoacoustic instabilities.

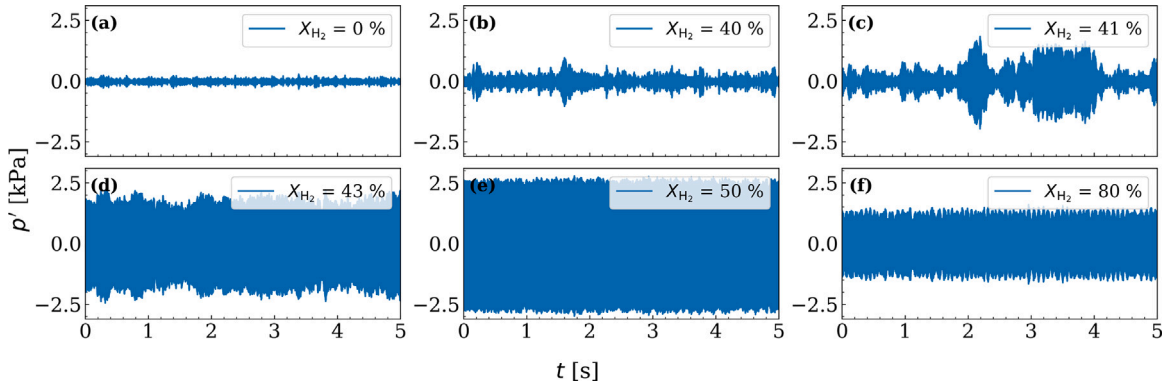

Fig. 4. Acoustic pressure fluctuations recorded near the backplane of the combustor for $u_b = 22.5\,\mathrm{m\,s^{-1}}$ and $\phi = 0.8$, with increasing hydrogen enrichment in the fuel: (a) $X_{H_2} = 0$; (b) $X_{H_2} = 40\%$; (c) $X_{H_2} = 41\%$; (d) $X_{H_2} = 43\%$; (e) $X_{H_2} = 50\%$; (f) $X_{H_2} = 80\%$.

systems [23], leading to a limit-cycle thermoacoustic instability as X_{H_2} is increased. The transition between stable and unstable operation follows a parabolic shape in $X_{H_2} - \phi$ space. This is consistent with the Rayleigh integral, Fig. 3(c), which features high values in the unstable region. The frequency of the instability spans the range between $480 \lesssim f_0 \lesssim 550\,\mathrm{Hz}$, Fig. 3(b), corresponding to the first longitudinal mode of the combustion chamber, $f_{1L} \approx c/[4(L_c + 0.4D_c)] \approx 525\,\mathrm{Hz}$. f_0 increases with increasing X_{H_2} and ϕ, most likely due to a higher temperature and sound speed within the combustion chamber and changes in the flame's describing function [38,46].

To examine in more detail the transition from stable to unstable operation, we keep both the injector bulk velocity and equivalence ratio constant at $u_b = 22.5\,\mathrm{m\,s^{-1}}$ and $\phi = 0.8$, and examine the effect of X_{H_2} as the control parameter. At this condition, $X_{H_2} \gtrsim 42\%$ are thermoacoustically unstable conditions. This trend is almost identical to the results by Aguilar et al. [47] in a similar configuration with a shorter and narrower combustion chamber.

As X_{H_2} is increased, the flame gradually shortens and transitions from a tulip shape at $X_{H_2} \lesssim 30\%$ to a V-shape at $30\% \lesssim X_{H_2} \lesssim 65\%$, and finally to a M-shape, $X_{H_2} \lesssim 65\%$, see Appendix A. In Fig. 4, characteristic acoustic pressure time traces are shown. Acoustic pressure fluctuations remain low at $X_{H_2} \lesssim 40\%$, until short, higher amplitude "bursts" are observed at $X_{H_2} = 41\%$ (Fig. 4(c)). At $X_{H_2} = 43\%$ (Fig. 4(d)), acoustic oscillations are sustained, but present cycle-to-cycle amplitude variations. These cycle-to-cycle variations, observed in Fig. 4(c–d), are likely one of the main reasons why the acoustic pressure fluctuations appear to gradually increase in the transition region in Fig. 3. At $X_{H_2} \gtrsim 50\%$, the thermoacoustic oscillations exhibit a limit cycle behavior.

The flame shape, its evolution as a function of X_{H_2}, and its impact on the thermoacoustic behavior of the burner are further discussed in Appendix A.

4. Active control using RL

Following the characterization of the thermoacoustic behavior of our test rig, we now discuss the implementation of active control using RL. RL is a subfield of ML in which an agent learns to make optimal decisions in an environment by receiving feedback in the form of rewards or penalties [33,48]. The goal is to learn a policy that maximizes the cumulative reward over time.

The RL problem can be formulated as a Markov decision process, defined by a tuple $(\mathcal{S}, \mathcal{A}, P, r)$, where the state space $\mathcal{S}$ and action space $\mathcal{A}$ are continuous. The reward function r quantifies the immediate feedback received by the agent after taking an action. The transition dynamics is governed by P, which defines the conditional probability distribution of the next state given the current state and action. The agent interacts with the environment over discrete time steps, generating trajectories $\tau_{\mathrm{traj}} = ([s, a, r]_0, [s, a, r]_1, \ldots, [s, a, r]_n)$, where each action a_t is selected according to the policy π as a function of the state s_t, $a_t \sim \pi(a_t \mid s_t)$. The objective of RL is to find an optimal policy π^* that maximizes the expected cumulative reward,

$$\pi^* = \arg\max_{\pi} \mathbb{E}(R \mid \pi), \tag{3}$$

where $R = \sum_{n=0}^{\infty} \xi^n r_n$ represents the long-term cumulative reward in which the future rewards are compounded over time with discount factor $\xi \in [0, 1]$. Generally, to quantify R, RL relies on the state value function (V) and state–action value function (Q) [49]. The state value

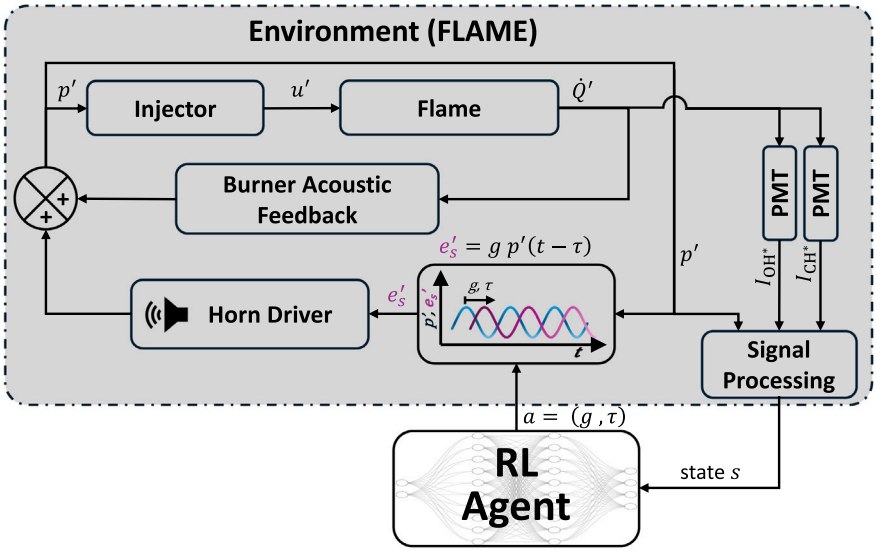

Fig. 5. Block diagram illustrating the control loop used in the present work.

function estimates the expected cumulative reward starting from state s and thereafter following policy π,

$$V^{\pi}(s) = \mathbb{E}[R \mid s, \pi]. \tag{4}$$

It measures the long-term benefit of being in state s under policy π, which enables comparisons between states in terms of long-term future rewards. Similarly, the state–action value function, also referred to as the Q-value, estimates the expected cumulative reward after taking action a in state s and subsequently following policy π,

$$Q^{\pi}(s, a) = \mathbb{E}[R \mid s, a, \pi]. \tag{5}$$

RL models can be broadly categorized into two main methods: model-free and model-based approaches [33]. In contrast to model-based RL, which utilizes an explicit model for the behavior of the system, $P(s_{t+1} \mid s_t, a_t)$, model-free methods directly learn a policy function through trial-and-error interactions with the environment [33]. Although this approach could potentially lead to high sample inefficiency compared to a model-based approach, it is more robust and favorable for challenging and difficult-to-model environments, such as thermoacoustically unstable combustion systems. Therefore, we use model-free RL in the present work.

In this section, we begin by describing the active control system, detailing its components and the environment setup. Then, we introduce the Soft Actor-Critic (SAC) agent, the RL algorithm employed for suppressing the combustion instabilities, explaining its architecture and training process.

4.1. Active control system

The present work uses the commonly employed microphone–loud speaker–phase-shift-controller architecture [19,20] for the active control of thermoacoustic instabilities. The block diagram of this system is shown in Fig. 5. The operating principle of this controller is to generate a synthetic acoustic wave using the actuator, which excites the flame almost in opposite phase to the pressure oscillation of the thermoacoustic instability, leading to destructive interference. The optimal values of the control parameters g and τ, see Eq. (1), are a function of the flame transfer function, of the burner's acoustic response function, and of the transfer functions of the measurement chain and actuator. Consequently, the optimal values of g and τ change as a function of the operating condition, ϕ and X_{H_2}. In the present work, we extend this controller architecture by dynamically tuning g and τ during burner operation using a RL approach.

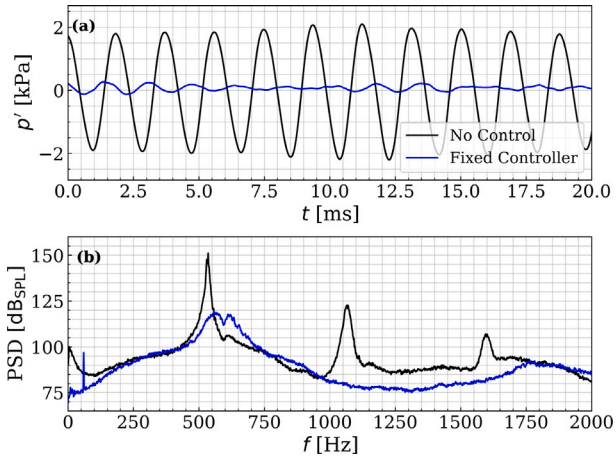

Fig. 6. Acoustic pressure traces (top) and power spectrum measured at $\phi = 0.9$ and $X_{H_2} = 70\%$ under different control conditions. In the "fixed controller" case, the control parameters are fixed and set to $g = 1.0$ and $\tau = 0.46$ ms.

The synthetic acoustic wave is generated using a compression horn driver (model D250-X, JBL, Los Angeles, CA, USA, with a frequency response from 0.4 to 9 kHz), directly connected to the injector tube using a catenoidal horn adaptor as illustrated in Fig. 1. The control signal to the loudspeaker is amplified using a KM750 audio amplifier (Behringer, Willich, Germany). The signal e'_s is synthesized by a real-time FPGA instrument (Moku:Go, Liquid Instruments, Lyneham, ACT, Australia, response time $\leq 14\,\mu$s) using Eq. (1).

Before training the RL agent, we evaluate the control setup using a "fixed controller", which uses fixed control parameters g and τ, Fig. 6. The "no control" case refers to the self excited instability. At $\phi = 0.9$ and $X_{H_2} = 70\%$, the acoustic pressure exhibits substantial harmonic oscillations, with a 2 kPa amplitude. To suppress these oscillations, we manually tune the controller, and operate it with fixed values of $g = 1.0$ and $\tau = 0.46$ ms. These parameters are found by trial and error to be optimal for suppressing oscillations at this operating condition. This "fixed controller" will be used as a baseline for comparison in subsequent analysis, Section 5.

4.2. Environment for combustion instability control

In RL, the environment represents the external system with which the RL agent interacts. In this work, the environment is shown in Fig. 5 and includes the burner with an active flame, sensors, actuators, and gain-delay function. The RL agent is implemented on a standard desktop computer. When the agent takes a new action, it updates the values of the control parameters g and τ on the FPGA board. This occurs at approximately 2 Hz.

Acoustic pressure p', and chemiluminescence I_{OH^*} and I_{CH^*} signals are recorded over a set period of time, $\mathfrak{T}_{rec}$. These signals are pre-processed to provide a comprehensive characterization of the environment to the RL agent: eight statistical quantities are computed to construct the environment's state space $s_t \in \mathcal{S}$ at time t, fusing information from multiple heterogeneous sensors and providing a multifaceted characterization of the instability's behavior. s_t is represented as a vector in $\mathbb{R}^8$:

- root mean square amplitude of the acoustic pressure, $s_{t,0} = p'_{rms}$;
- root mean square of chemiluminescence signals, $s_{t,1} = I'_{OH^*,rms}$ and $s_{t,2} = I'_{CH^*,rms}$;
- mean value of the chemiluminescence signals, $s_{t,3} = \overline{I_{OH^*}}$, and $s_{t,4} = \overline{I_{CH^*}}$, as well as their ratio, $s_{t,5} = \overline{I_{CH^*}}/\overline{I_{OH^*}}$, as an indicator of the burner's operating equivalence ratio;
- transfer function between pressure fluctuations p'_{rms} and flame chemiluminescence I'_{OH^*}, calculated using the cross power spectral density (CPSD) and power spectral density (PSD) based on the Welch periodogram method [50], and evaluated at the fundamental wave frequency f_0: $s_{t,6} = \| \text{CPSD}(p', I_{OH^*}, f_0)/\text{PSD}(p', f_0) \|$, the magnitude of the transfer function, and $s_{t,7} = \text{ang}\left[\text{CPSD}(p', I_{OH^*}, f_0)\right]$, the phase of the transfer function.

$\mathfrak{T}_{rec} = 100$ ms is chosen such that s_t can be estimated with sufficiently low statistical uncertainties. We note that the RL agent is isolated from the mass flow controllers and has no direct knowledge about the flow rates of individual reactants. The RL agent only operates based on acoustic pressure and chemiluminescence signals.

After evaluating s_t, the agent imposes its action $a_t \in \mathcal{A}$ on the environment. a_t contains two elements: the gain g and the time delay τ of the real-time gain-delay control loop. As illustrated in Fig. 5, the real-time gain-delay control loop is part of the environment and is operating at all times. The action $a_t = (g_t, \tau_t) \in \mathbb{R}^2$ of the RL agent therefore performs an update to the parameters running on the FPGA chip, on which the phase-shift controller is implemented. We constrain g within the range $[0, 3]$ (a.u.) and the time delay τ to $[0, 2.8]$ ms, with the upper bound corresponding to slightly more than the longest acoustic period observed in this system. These bounds ensure that the agent's action remains within expected limits during the control process. After the agent deploys an action, we wait $\mathfrak{T}_{wait} = 200$ ms before measuring a new state. $\mathfrak{T}_{wait}$ is chosen to be longer than the transient growth/decay of the pressure oscillation occurring when the controller is turned on or off. By using this delay $\mathfrak{T}_{wait}$ and by acquiring statistics over a duration much greater than both the acoustic period and the burner flow-through time, $\mathfrak{T}_{rec} = 100\,\text{ms} \gg f_0^{-1}$, we are able to achieve both statistical convergence when computing the state s_t, and a memory-less behavior for the environment, thereby ensuring that the environment behaves as a Markov process.

The reward function r is designed to simultaneously minimize the pressure fluctuations in the combustor and the amplitude of the signal sent to the loudspeaker, inspired by similar approaches found in the literature [36]:

$$r_t = C_r\left(p'_{rms,t}\right) - C_p\, p'_{rms,t} - C_a\left(\left(2\frac{g_t - g_{min}}{g_{max} - g_{min}} - 1\right)^2 + \left(2\frac{\tau_t - \tau_{min}}{\tau_{max} - \tau_{min}} - 1\right)^2\right), \tag{6}$$

where C_p and C_a are constants chosen, through hyperparameter tuning, to be $0.009\,\text{Pa}^{-1}$ and 0.05, respectively. The gain and time delay are scaled to range between $[-1, 1]$ using the action space bounds, $g_{min}, g_{max}, \tau_{min}$ and τ_{max}, before computing the reward. The coefficient C_r is chosen to take the following form:

$$C_r(p'_{rms}) = \begin{cases} 15 & \text{if} \quad p'_{rms} < 150\,\text{Pa}, \\ 10 & \text{if} \quad 150\,\text{Pa} \le p'_{rms} < 550\,\text{Pa}, \\ -5 & \text{if} \quad p'_{rms} \ge 550\,\text{Pa}. \end{cases} \tag{7}$$

The piecewise design for C_r allows the reward function to dynamically adjust based on p'_{rms}. This structure ensures that the reward provides significant and non-linear incentives or penalties depending on how far the agent deviates from complete suppression of the combustion instability.

4.3. SAC agent

SAC is a RL model introduced to address two major challenges in model-free RL: very high sampling complexity and weak convergence properties [51,52]. It is a hybrid approach that combines aspects of both policy gradient and value-based methods to find the optimal policy π^*. In SAC, neural networks (NN) are utilized to approximate the three key functions defined in Eq. (3)–(5), which are essential for learning and decision-making. The agent architecture, shown in Fig. 7, consists of five neural networks: a policy (*i.e.* actor) network π_ψ, which selects the action a_t as a function of the current state s_t, two Q-critic networks Q_{Φ_1} and Q_{Φ_2}, which approximate the state–action value function, Eq. (5), a value network V_θ, and a target value network $V_{\bar{\theta}}$, which both approximate the state value function, Eq. (4). Each NN consists of three layers and 256 hidden units per layer. The weights of these neural networks are denoted in their respective subscripts. For example, ψ represents the weights of the policy network π_ψ and Φ_1 is the weights of the first Q-critic network Q_{Φ_1}. The training process is divided into three main stages, discussed in detail in the following paragraphs and in Algorithm 1:

1. **Collect transitions**: at each step, the agent interacts with the environment to deploy an action a_t, observe the resulting state s_{t+1}, and generate a transition tuple $(s_t, a_t, r_t(s_t, a_t), s_{t+1})$. This transition is then stored in the replay buffer $\mathcal{D}$, a data structure with capacity of 10^6 transitions, which retains past experiences for future learning.
2. **Compute objective functions**: we randomly sample a subset $\mathcal{B}$ of the replay buffer containing 256 transition tuples to compute objective functions J_π, J_Q, and J_V (Lines 14–17 in Algorithm 1).
3. **Update neural networks:** using the objective functions J_π, J_{Q_j}, and J_V, we apply gradient descent to update the weights of the NNs.

These steps are iteratively repeated until convergence.

The policy network, colored in blue in Fig. 7, maps the state space $\mathcal{S}$ to probability distributions over the action space $\mathcal{A}$, which enables the agent to explore the action space efficiently. While collecting transitions (s_t, a_t, r_t, s_{t+1}), the action a_t is sampled from a normal distribution whose parameters, μ and σ, are computed using the policy network π_ψ. Unlike other RL approaches that maximize the cumulative reward in their policy objective function, SAC additionally maximizes the entropy of each policy output to explore more diverse actions [51]. This approach has been especially successful in learning more stochastic policies for complex environments [52]. By accounting for the log-probability distribution of the action conditioned by the state, $-\log \pi_\psi(\cdot \mid s)$, also known as entropy, in the calculation of the pol-

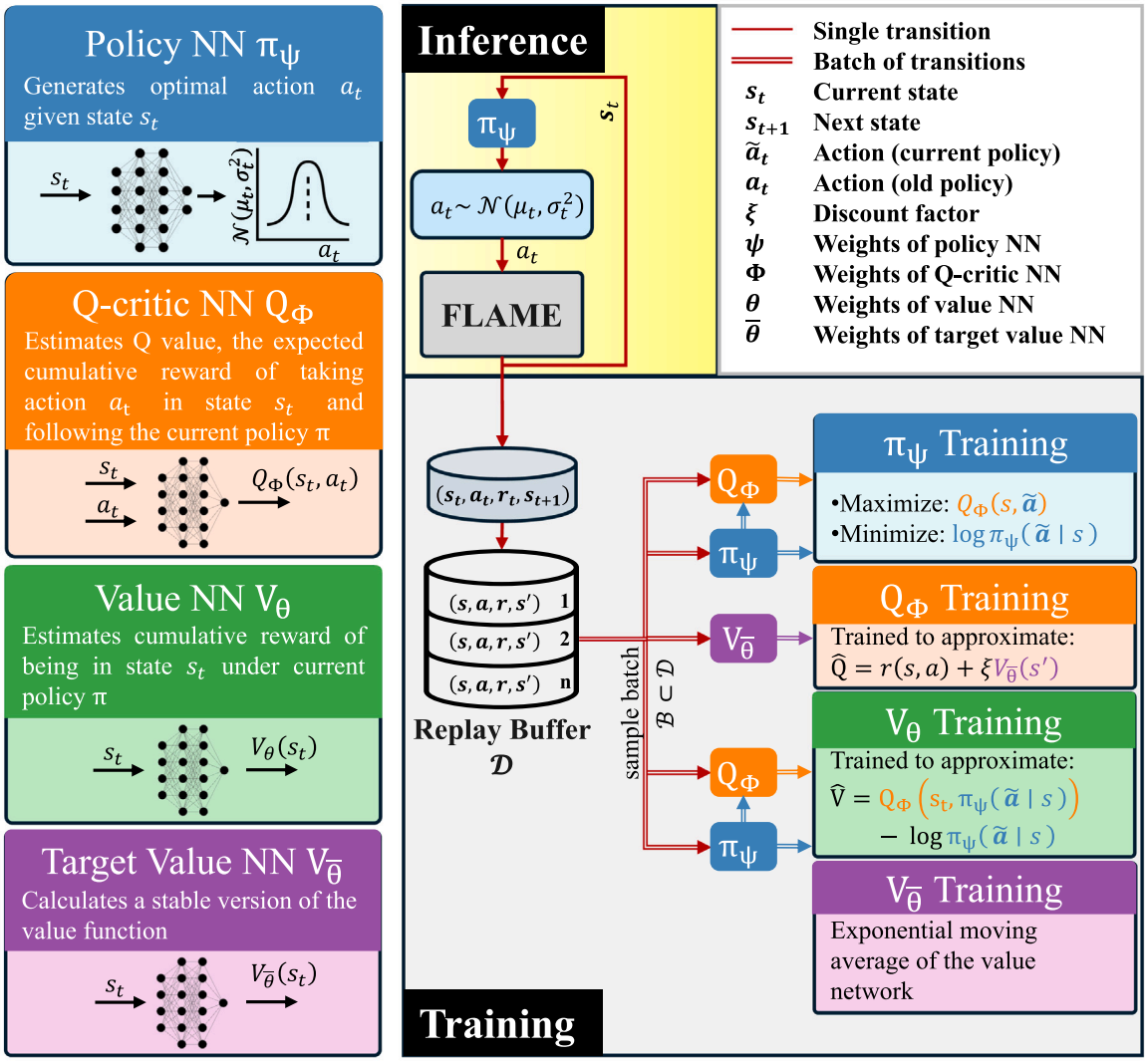

Fig. 7. SAC architecture. Each NN output is color-coded to represent its contribution to the agent's objective functions, where the color mapping facilitates understanding the interaction between different NNs. More details on the implementation are provided in Algorithm 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

icy objective function $J_\pi(\psi)$ (Line 14 in Algorithm 1), this approach balances the agent exploration and exploitation more effectively [51].

The Q-critic networks Q_{Φ_j}, colored in orange in Fig. 7, estimate the state–action value function Q, which quantifies the cumulative reward associated with taking an action a_t given a state s_t (Eq. (5)). We use two Q-critic networks, Q_{Φ_1} and Q_{Φ_2}, to reduce the overestimation bias when estimating the Q-value, which could otherwise lead to suboptimal policies or unstable learning [53]. Each Q-critic network is trained independently to generate separate outputs. The objective functions $J_Q(\Phi_j)$, Lines 15 and 16 in Algorithm 1, are computed to minimize the difference between the predicted Q and $\hat{Q}$, the cumulative future reward. $\hat{Q}$ is estimated using the target value network $V_{\bar{\theta}}$ [51,52], with a discount factor $\xi = 0.99$.

The value network V_θ estimates the state value function, *i.e.*, the expected long-term reward starting from state s_t and following the current policy π_ψ ((4)). The objective function of the value network $J_V(\theta)$, Line 17 in Algorithm 1, aims at minimizing the discrepancy between the value network and the expected value function. Starting from states $s \in \mathcal{B}$, actions $\tilde{a}$ are resampled using the current policy π_ψ, see Line 13 in Algorithm 1. By combining these resampled actions and the Q-critic networks, we build an estimator $\hat{V}$ for the expected value. At this step, we estimate the Q value as the minimum estimates from the two Q-critic networks [51,53].

The target value network $V_{\bar{\theta}}$ provides a more stable version of the value function. This approach helps stabilize updates in the Q-critic networks and prevents Q-value overestimation, which could destabilize the learning process.

The weights of these neural networks are updated using mini batch-gradient descent based on objective functions $J_\pi(\psi), J_Q(\Phi_1), J_Q(\Phi_2)$, and $J_V(\theta)$, with one exception: the weights $\bar{\theta}$ of the target value network are updated by applying an exponential moving average of the value network weights θ with a smoothing factor $\beta = 0.005$. This approach allows for more stable gradient estimates than single-sample updates, which reduces the variance in the parameter updates. Lastly, our implementation uses Adam optimizer [54] with learning rate λ set to 10^{-4} for all NNs.

Unlike epoch terminology in ML training, RL models are trained based on the number of episodes, which represents a complete sequence of interactions between an agent and the environment. It begins from an initial state where $g = 0$ and $\tau = 0$ and proceeds through a series of

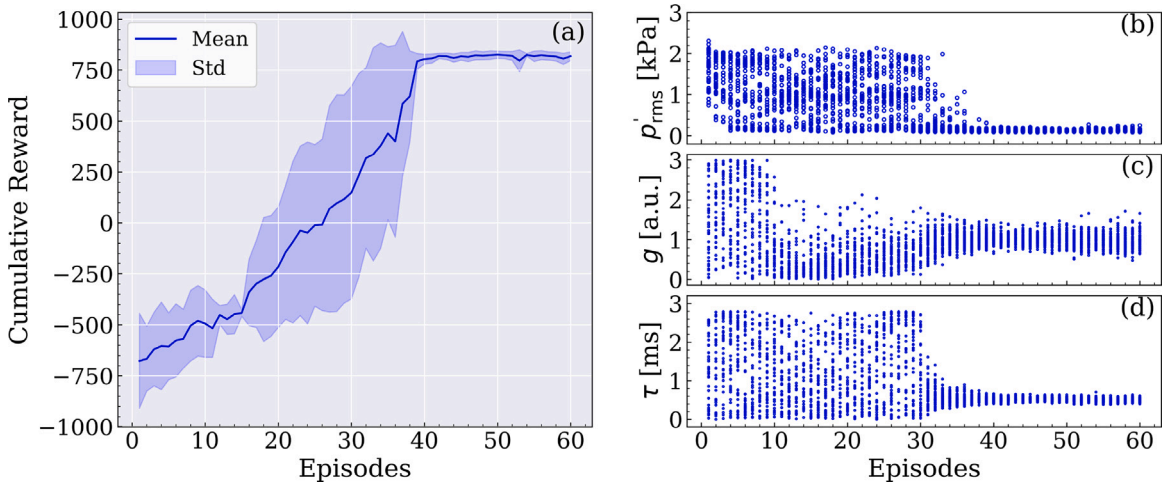

Fig. 8. (a) SAC's cumulative reward during training at $\phi = 0.9$ and $X_{\mathrm{H_2}} = 70\%$. The "Mean" and "Std" in this plot correspond to the mean and standard deviation across the five agents trained at this operating condition. (b–d) State and actions during the training process. Each dot corresponds to one environment step. (b) Acoustic pressure fluctuations p'_{rms}; (c) gain g; and (d) delay τ of the control loop. (b–d) The pressure fluctuation and action taken at each environment step are reported based on the episode to which the environment step belongs.

Algorithm 1 Soft Actor-Critic

Initialize NNs: π_ψ, $Q_{\Phi,1}$, $Q_{\Phi,2}$, V_θ, $V_{\overline{\theta}}$
Initialize replay buffer $\mathcal{D}$ and environment FLAME
for each episode **do**
  $s_t \leftarrow \mathrm{FLAME}(a_t = (0,0))$
  for each environment step $t = 1, N_s$ **do**
    Collect transitions:
    $a_t \sim \pi_\psi(\cdot \mid s_t)$
    $s_{t+1} \leftarrow \mathrm{FLAME}(a_t)$
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{(s_t, a_t, r(s_t, a_t), s_{t+1})\}$
    Compute objective functions:
    $\mathcal{B}$ = sample N transitions from $\mathcal{D}$
    sample actions $\tilde{a} \sim \pi_\psi(\cdot \mid s)$ for $s \in \mathcal{B}$
    compute log-probability $\log \pi_\psi(\tilde{a} \mid s)$ for $s \in \mathcal{B}$
    $J_\pi(\psi) = \frac{1}{|\mathcal{B}|} \sum_{s \in \mathcal{B}} \log \pi_\psi(\tilde{a} \mid s) - \min_{j=1,2} Q_{\Phi_j}(s, \tilde{a})$
    $J_Q(\Phi_j) = \frac{1}{|\mathcal{B}|} \sum_{(s,a,r,s') \in \mathcal{B}} \left(Q_{\Phi_j}(s,a) - \hat{Q}(s,a) \right)^2$, for $j \in \{1,2\}$
      with $\hat{Q}(s,a) = r(s,a) + \xi V_{\overline{\theta}}(s')$
    $J_V(\theta) = \frac{1}{|\mathcal{B}|} \sum_{s \in \mathcal{B}} \frac{1}{2} \left(V_\theta(s) - \hat{V}(s) \right)^2$
      with $\hat{V}(s) = \min_{j=1,2} Q_{\Phi_j}(s, \tilde{a}) - \log \pi_\psi(\tilde{a} \mid s)$
    Update neural networks:
    $\psi \leftarrow \psi - \lambda_\pi \nabla_\psi J_\pi(\psi)$
    $\Phi_j \leftarrow \Phi_j - \lambda_Q \nabla_{\Phi_j} J_Q(\Phi_j)$ for $j \in \{1,2\}$
    $\theta \leftarrow \theta - \lambda_V \nabla_\theta J_V(\theta)$
    $\overline{\theta} \leftarrow \beta\theta + (1-\beta)\overline{\theta}$
  end for
end for

actions and observations. Every interaction between the agent and the burner is referred to as an environment step. A complete episode spans $N_s = 60$ environment steps or transitions, with early-episode termination occurring after 35 steps if the agent fails to suppress the instability. Each environment step takes approximately 400 ms, contributing to the overall episode duration.

5. Results

In this section, we discuss our adaptive active control policy, which can effectively mitigate combustion instabilities across a wide range of fuel compositions and equivalence ratios. Section 5.1 discusses the training process of the RL model on a single operating condition. We then train the model across an extended operating range in Section 5.2. Finally, we discuss in Section 5.3 the optimal control parameters g and τ identified by the RL model.

5.1. Learning an optimal control policy for a single operating condition

This section presents results from training SAC to dampen combustion instabilities at a single operating condition of $\phi = 0.9$ and $X_{\mathrm{H_2}} = 70\%$. To assess the robustness of our results, we repeat the training five times, thereby creating five agents, each initialized with a different random seed, *i.e.*, different NN weights initialization. To evaluate the learning progress of the agent, we track the cumulative reward, defined as the sum of all rewards in one episode, Fig. 8(a). We also monitor, for each episode, the amplitude of the acoustic pressure fluctuations, p'_{rms}, Fig. 8(b), and the actions taken by the agent, Fig. 8(c,d). In the early stages of training, the agent has no knowledge of the behavior of the flame, the combustor, and the control system. It samples from a large range of both g and τ to explore the state–action space, thereby achieving highly variable levels of acoustic pressure fluctuations and a low cumulative reward. As training progresses, the cumulative reward increases noticeably, indicating that the agent is effectively learning and improving the policy. After approximately 10 episodes, we note that the range of gain parameter g explored by the agent reduces as it begins to focus on $g \lesssim 1.5$. After 34 episodes, we observe a pivotal shift: the agent begins to take relatively good actions that are able to consistently dampen the thermoacoustic instability and yield a higher reward. This transition aligns with the appreciable drop in the variance of the action space (g, τ) suggesting that the agent is narrowing its exploration range. In the later stages of training, episodes ≥ 40, the cumulative reward reaches a plateau as the agent converges to a bounded range of actions, indicating that the agent has learned to generate precise control parameters to effectively suppress the combustion instability. We refer to this model as RL-SAC I for later discussion.

Fig. 9 illustrates the damping of thermoacoustic oscillations by the RL-SAC I controller on two operating conditions. Initially, the controller is inactive and we activate it at $t = 0$ ms. The combustion system is initially at a limit cycle oscillation, with pressure fluctuation amplitudes on the order of 2 kPa. As the RL-SAC I controller is activated ($t = 0$ ms, red dashed line), the acoustic pressure initially shows a brief excursion after which the oscillation is dampened in $\mathcal{O}(100\,\mathrm{ms})$. After

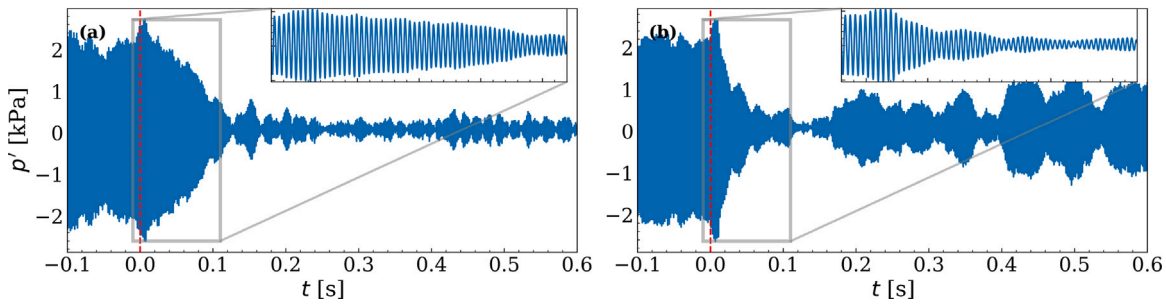

Fig. 9. Acoustic pressure traces recorded near the backplane of the combustor at (a) $\phi = 0.9$ and $X_{H_2} = 70\%$ (training condition) and (b) $\phi = 0.65$ and $X_{H_2} = 70\%$. The RL-SAC I controller is initially turned off and is activated at $t = 0$ (red dashed line). The inset is a zoom into the first 100 ms after the controller has been turned on. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

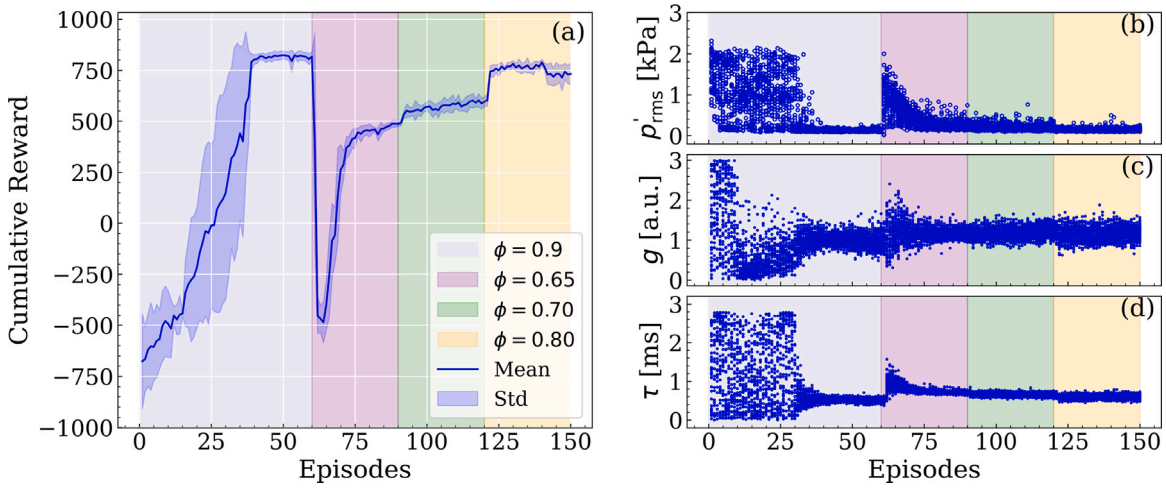

Fig. 10. (a) Cumulative reward of the RL-SAC IV agent during training. (b, c, d) State and actions observed during the training process. Each dot corresponds to one environment step. (b) Acoustic pressure fluctuations p'_{rms}; (c) gain g; and (d) delay τ of the control loop. (b–d) The pressure fluctuation and action taken at each environment step are reported based on the episode to which the environment step belongs.

this initial transient, $t \geq 100$ ms, the pressure trace shows that the RL-SAC I controller is able to fully suppress the instabilities at $\phi = 0.9$, Fig. 9(a). The small fluctuations in the pressure traces are primarily caused by combustion noise [55]. The Rayleigh integral, discussed in the following section, demonstrates that the pressure fluctuations and heat release are uncorrelated. In contrast, at $\phi = 0.65$, Fig. 9(b), we observe a noisy, bursting behavior, albeit at a significantly reduced amplitude, which the RL-SAC I active controller is not able to fully dampen.

5.2. Learning an optimal adaptive control policy

In this section, we extend the training process of our RL agent on four operating conditions, in an effort to develop an adaptive control system that can operate over a wide range of ϕ and X_{H_2} with minimal training. During training, we keep $X_{H_2} = 70\%$ constant to test the RL's capability in extending to other fuel compositions outside of the training range. We consider four distinct equivalence ratios. We use the following notations:

- Agent RL-SAC I is the one discussed in Section 5.1. It is trained until convergence on a single condition $\phi = 0.9$.
- Agent RL-SAC II is initially trained at $\phi = 0.9$, and then at $\phi = 0.65$.
- Agent RL-SAC III's training spans three conditions $\phi = 0.9, 0.65$, and 0.7.
- Agent RL-SAC IV's training expands to a fourth conditions $\phi = 0.9, 0.65, 0.7$, and 0.8.

Fig. 10(a) shows the cumulative reward at each episode while extending the SAC training across these four conditions. During the first 60 episodes, the agent is trained on a single operating condition, $\phi = 0.9$ and $X_{H_2} = 70\%$, see discussion in Section 5.1. At the 60th episode, the equivalence ratio is changed to $\phi = 0.65$ and the cumulative reward suddenly drops, indicating that the RL agent has identified that the control policy it learned previously at $\phi = 0.9$ is no longer suitable at this new operating condition. This is consistent with the increased variance in the gain and the delay (see Fig. 10(c–d)) along with higher pressure fluctuations in Fig. 10(b). Therefore, the policy needs additional tuning to generalize to a new behavior of the flame. Around the 85th episode, as the cumulative reward reaches a new plateau and the action taken by the controller converges toward the optimum value, we again change the operating conditions of the burner to $\phi = 0.7$. We observe that the agent is capable of adapting to a new operating condition within approximately 15 episodes. In contrast to the change of operating condition $\phi = 0.9 \rightarrow 0.65$, during the transition $\phi = 0.65 \rightarrow 0.7$, the cumulative reward increases slightly. This behavior is even more striking during the transition $\phi = 0.7 \rightarrow 0.8$. This indicates that the agent is exploiting more than exploring after these changes of operating conditions. The policy developed at the 90th episode ($\phi = 0.65 \rightarrow 0.7$) already shows a good degree of generalizability and only requires minor additional tuning unlike the policy at the 60th episode. The small adjustments in operating conditions, $\phi = 0.65 \rightarrow 0.7$, and $\phi = 0.7 \rightarrow 0.8$, refine the policy incrementally. Consequently, the agent changes its focus from learning and exploration to refining and exploiting the learned control strategies. This is clearly marked by the absence of any sudden drops in the cumulative reward, combined with low variance

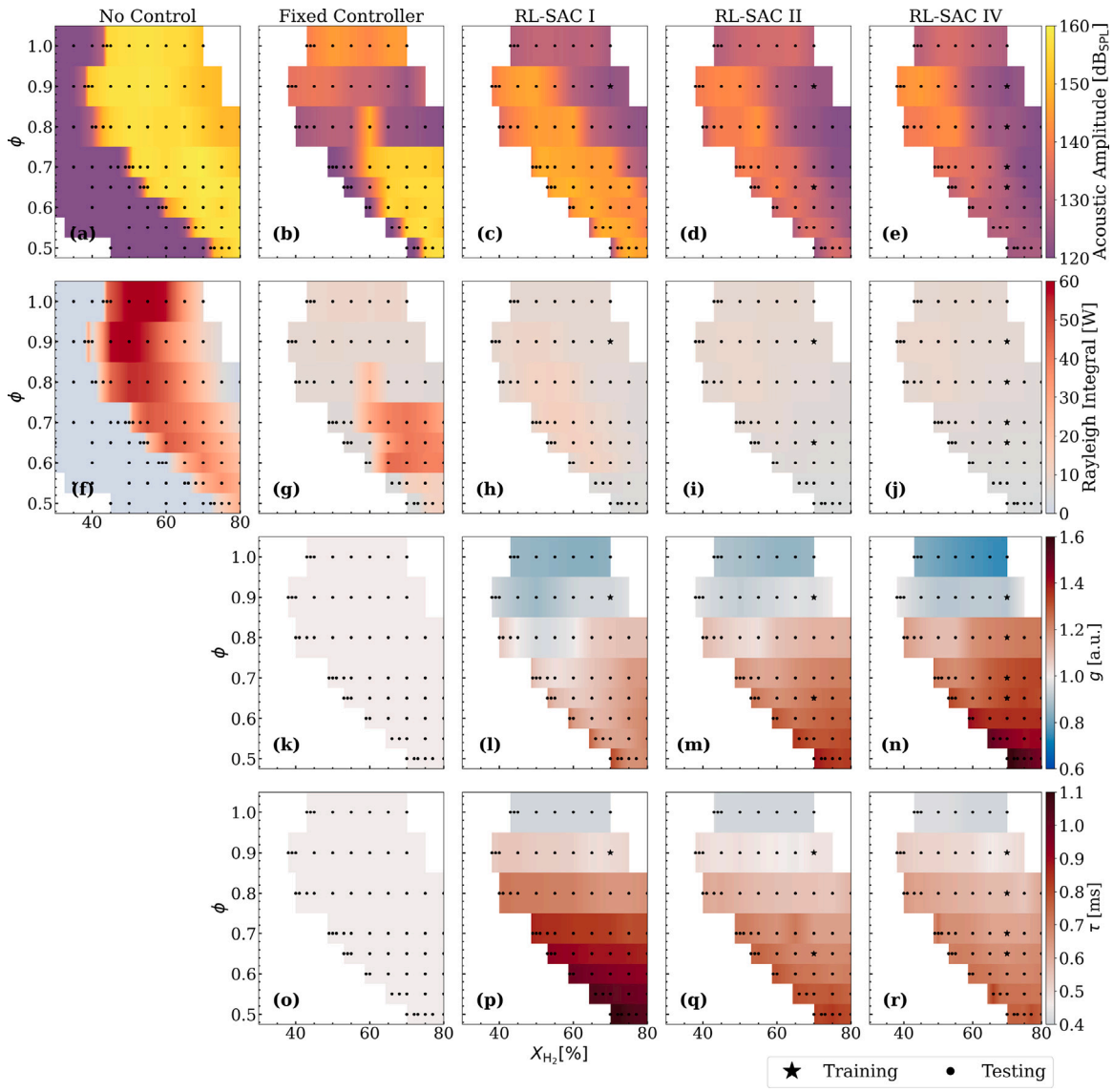

Fig. 11. Testing of the active control schemes at different operating conditions and learning stages. (a–e) Amplitude of the harmonic acoustic pressure fluctuation near the combustor backplane. (f–j) Rayleigh integral (Eq. (2)). (k–n) Median value of the controller gain g; (o–r) delay τ. (k–r) Gray corresponds to $g = 1.0$ and $\tau = 0.46$ ms, the control parameters used in the "fixed controller". (Column 1) No active control. (Column 2) Fixed control parameters tuned for optimal suppression at $\phi = 0.9$ and $X_{H_2} = 70\%$. (Columns 3–5) SAC based controllers trained on an increasing numbers of operating conditions. The operating conditions used for training of the SAC controllers are shown with stars.

in the actions taken by the agent. The consistent performance of the agent across five different random seeds and four operating conditions highlights the stability and reliability of the learned policy.

The strength of our RL approach, however, mostly resides in its adaptability. In Fig. 11, we test our RL agents across a wide range of ϕ and X_{H_2} conditions, corresponding to all conditions at which thermoacoustic instabilities are observed in the absence of active control (see Fig. 3). During the measurements leading to Fig. 11, the SAC controllers are operated in testing mode, *i.e.*, all the trainable parameters in the NNs are kept constant. While the fixed parameter controller is able to dampen the thermoacoustic oscillations at certain conditions (dark regions in Fig. 11(b)), its performance is sub-optimal and uneven across the operating range of the burner. At lean conditions, the acoustic pressure amplitude remains high and a strong combustion instability is present, as illustrated by the elevated Rayleigh integral in this region, Fig. 11(g). This lack of generalizability of fixed-parameter gain-delay controllers is consistent with the literature [21]. In contrast, the simplest RL-based controller, RL-SAC I, is able to substantially reduce the amplitude of thermoacoustic fluctuations across almost the entire operating range, Fig. 11(c). In regions of the operating map closer to the training condition, RL-SAC I achieves a comparable suppression in the acoustic amplitude in the range of 120−130 dB$_{SPL}$ similar to the training condition. Even in regions in which acoustic pressure fluctuations remain elevated, the oscillation amplitude and thermoacoustic source term RI are both significantly reduced compared to the "no control" and "fixed controller" cases. This happens as RL-SAC I continuously refines its policy to maximize the reward by reducing the acoustic amplitude. It effectively explores diverse actions that contribute to the goal of suppressing the combustion instabilities even when trained on a single operating condition compared to the fixed action in the "fixed controller" case. For example, in certain regimes of operating conditions where the effective control demands that the gain exceeds unity, the fixed controller fails to achieve suppression due to its fixed setting. In contrast, RL-SAC I adapts to these conditions, selecting the necessary actions dynamically to maintain effective control, discussed in Section 5.3. In addition, the ability of RL-SAC I to perform better

than a fixed-parameter controller likely indicates that RL-SAC I is able to leverage the expanded state space s_t to infer a reasonable black-box model for the behavior of the burner and its response to a policy, thereby extrapolating better control parameters even at conditions differing significantly from its training.

As we fine-tune the SAC controller by training at additional operating conditions, the overall performance of the model improves rapidly, as shown in Fig. 11 (columns 4–5). RL-SAC IV, shown in the rightmost column of Fig. 11 and trained 2.5 times longer than RL-SAC I, on a total of four conditions, is the most robust controller, consistently suppressing combustion instabilities across the majority of the operating range, as shown by the Rayleigh integral in Fig. 11(j).

Catastrophic forgetting [56] is generally a major concern when an agent is trained on multiple conditions. It occurs when learning a new condition interferes or replaces the knowledge learned from previous conditions, resulting in forgetting the policies that worked well under previous conditions. However, our agent has maintained a balance between integrating new information and retaining useful strategies from previous conditions due to the replay buffer [57]. This is consistent with the enhanced performance from RL-SAC I to RL-SAC IV as shown in Fig. 11. This suggests that the policy has been fine-tuned, such that it mitigates the risk of catastrophic forgetting, thereby preserving the robustness and efficiency of the learned control scheme.

In Fig. 11(e), we nevertheless observe that the acoustic pressure amplitude remains slightly more elevated at certain conditions, especially at lower X_{H_2} near the boundary of the tested domain. These regions of lower efficacy of the RL-SAC IV controller are found near the location where we observe a bifurcation between stable and unstable operation in the absence of active control, Fig. 11(a). Although the acoustic pressure fluctuations are reduced by approximately 20 dB_{SPL} when using active control, this lower performance is unexpected and should be investigated in future work.

In the supplementary information, Fig. S1, we present a more quantitative assessment of the reduction in oscillation amplitude achieved by each controller over the entire test set. For RL-SAC IV, 50% of the test conditions exhibit reductions of at least 26 dB in the pressure fluctuations, with some exceeding 40 dB. This represents a notable improvement compared to RL-SAC I, which achieves at least 26 and 14 dB reduction on only 25% and 50% of the test conditions, respectively. These results show the convergence of the controller with additional training and demonstrate the suitability of our approach in mitigating combustion instabilities across a large range of operating conditions and flame types.

5.3. Optimum actions chosen by RL controller

Fig. 11(k–r) shows the parameters g and τ determined by all four controllers as a function of operating conditions. For the reader's convenience, the color map in Fig. 11 is centered around the values identified in Section 4.1 for the condition $\phi = 0.9$ and $X_{H_2} = 70\%$ and used for the "fixed controller". The range of action used by the RL-SAC IV controller is quite large: depending on operating condition, the gain parameter spans values ranging from 0.75 to 1.6, Fig. 11(n), while the delay parameter ranges from under 0.4 ms to 0.8 ms, Fig. 11(r), illustrating the benefit of an adaptive controller to address thermoacoustic instabilities in highly fuel-flexible combustion applications. Fig. 11(n, r) reveals distinct regimes identified by the RL-SAC IV controller: at $\phi \geq 0.9$ the optimal gain value to suppress the instability is on the order of $g \lesssim 1$. In contrast, at very lean conditions $\phi \lesssim 0.6$, a much higher gain is required. A similar pattern can be observed for the delay parameter τ: the controller uses longer τ at very lean conditions, which is expected as the flame and its associated convective timescale increases at leaner operating conditions. The dependency of the delay parameter on X_{H_2} is less pronounced, with a general trend of slightly lower τ being used at high X_{H_2}. This behavior might be due to the training being conducted at constant hydrogen enrichment $X_{H_2} = 70\%$. RL-SAC II's optimal actions differ only slightly from RL-SAC IV's, indicating that additional training only leads to minor modifications to the control policy to further optimize and improve the generalizability of the control policy. Notably, for $\phi \leq 0.55$, an increase in the gain values is observed. This increase correlates with a significant reduction in the acoustic amplitude during testing in this region of the operating map, leading to further improvements in the agent's performance. It is remarkable to observe that RL-SAC I and RL-SAC IV follow very similar trends in the actions they take as ϕ and X_{H_2} are varied, Fig. 11(i, p). Despite being trained on a single operating condition, RL-SAC I has correctly identified that g should be decreased (resp. increased) at richer (resp. leaner) equivalence ratios, and that τ should be shorter (resp. longer) at richer (resp. leaner) equivalence ratios, further illustrating the capability of our SAC controller to reasonably extrapolate a suitable control policy from minimal training.

6. Conclusions

In this work, we experimentally demonstrate that a deep RL based active control system can suppress combustion instabilities over a wide range of operating conditions in a premixed turbulent CH_4/H_2/air bluff-body-stabilized flame. The deep RL controller uses a SAC architecture. The state space used as input by the controller leverages both acoustic and chemiluminescence sensors. By combining these heterogeneous sensors and using domain-specific knowledge to preprocess their signals, at its training condition, the SAC controller is able to match and outperform the performance of a fixed parameter controller. More importantly, our extensive testing shows that it is able to tune itself to adapt to unseen operating conditions. A controller trained on a single operating condition is able to significantly dampen thermoacoustic oscillations over the extensive operating range of the burner. With additional training on four distinct operating conditions, the SAC controller's performance is further improved, and it is able to achieve at least 26 dB reduction in acoustic pressure fluctuations on more than 50% of the investigated operating conditions.

This SAC controller represents a promising technique to reduce the amplitude of acoustic pressure fluctuations in fuel flexible systems. Future research should focus on fully suppressing weak thermoacoustic bursts, which occur in certain regions of the operating map despite active control. These bursts, which are highly stochastic in nature, have proven a challenge for the current controller. This could be addressed by considering a control system with additional degrees of freedom in their real-time control system.

CRediT authorship contribution statement

Bassem Akoush: Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Guillaume Vignat:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation, Conceptualization. **Ryan Finley:** Software. **Wai Tong Chung:** Conceptualization. **Matthias Ihme:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the U.S Department of Energy's office of Energy Efficiency and Renewable Energy (EERE) under the Industrial Efficiency and Decarbonization office's award number DE-EE0011202.

Appendix A. Flame shape and flame dynamics

To examine the flame shape, a high speed camera (Phantom TE2010, Vision Research, Wayne, NJ, USA), equipped with an optical filter (centered at 425 nm with a FWHM of 50 nm, Edmund Optics, Barrington, NJ, USA), is used to capture line-of-sight-integrated CH* chemiluminescence images at a repetition rate of 6 kHz. The exposure time is 166.4 µs. In addition, a Tucsen Dhyana 400BSI V3 sCMOS camera, equipped with a 58 mm lens (Voigtlander, Germany), with an aperture set at $f/8$, is used to record the flame images under stable conditions. A bandpass filter centered at 430 nm (10 nm FWHM, Edmund Optics) is used to record CH*. The exposure time is 5 s.

In Fig. A.1, we show CH* chemiluminescence images of the flame at different levels of hydrogen in the fuel stream. An inverse Abel transform is used to deconvolve the shape of the flame under an assumption of axisymmetry. For Fig. A.1, we use the active control system with fixed parameters to suppress all thermoacoustic oscillations. We show the state of the flame at rest, in the absence of combustion instabilities. At $X_{H_2} = 0$, the pure methane flame has a characteristic "tulip" shape. As X_{H_2} is increased, the flame shortens and transitions to a "V" shape. A very weak chemiluminescence signal is visible in the outer shear layer of the jet, which is far weaker than the flame anchored along the inner shear layer between the main jet and the inner recirculation zone located above the bluff-body. Between $40\% \leq X_{H_2} \leq 50\%$, in the region in which the bifurcation to thermoacoustic instability occurs, we do not observe any sudden change in the topology of the flame. We only observe a gradual shortening of the flame, accompanied by a minor increase in the strength of the chemiluminescence from the outer shear layer. Between $50\% \leq X_{H_2} \leq 80\%$, the flame gradually transitions to an "M" shape. For all operating conditions investigated, the flame is located away from the wall, in contrast to previous work [47].

Regarding flame dynamics during combustion instabilities, we show high speed chemiluminescence imaging of the flame during limit cycle oscillations at two conditions, $X_{H_2} \in \{50\%; 80\%\}$, in Fig. A.2. At each condition, oscillations are recorded for 1.16 s at 6 kHz. Images are processed using dynamic mode decomposition [58] to reconstruct phase-averaged images at the main frequency of oscillation, to which an inverse Abel transform is subsequently applied. The reader is referred to Ref. [59] for the detailed procedure.

At both conditions, we observe that the flame's shape response to acoustic pressure follows a well-known behavior for M- and V-shaped flames identified by Schuller et al. [60]. In these types of flames, the flame behavior is dictated by vortex shedding from the injection system. At $X_{H_2} = 50\%$, most of the unsteady heat release rate fluctuations occur near the tip of the flame, while the flame angle does not show significant fluctuations. At $X_{H_2} = 80\%$, the "M"-shaped flame shows a cyclic lengthening and shortening. Similar behaviors have been observed in the literature, for laminar, turbulent, and swirled premixed flames [61–64].

Observations of the flame shape are helpful to interpret the thermoacoustic behavior in Fig. 4. Let us first consider the transition occurring between $40\% \leq X_{H_2} \leq 43\%$. In this interval, the flame is "V"-shaped. The only significant change is a gradual shortening caused by the increased flame speed of the CH_4/H_2/air mixture [38]. Well-established scalings indicate that the characteristic time delay associated with the flame transfer functions of such premixed "V" flames gradually decreases [38,60,65], leading the flame's transfer function into a region of higher gain and shorter phase with regards

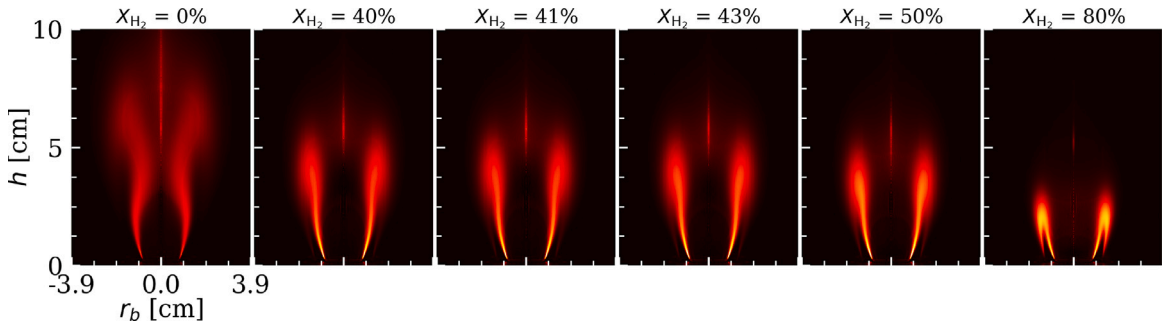

Fig. A.1. CH* chemiluminescence images of the flames at $u_b = 22.5\,\mathrm{m\,s^{-1}}$ and $\phi = 0.8$. From left to right, the volume fraction of hydrogen in the fuel, X_{H_2}, is gradually increased. These images are time-averaged and an inverse Abel transform is used to simplify interpretation. For $X_{H_2} \geq 41\%$, the active control system is used to suppress thermoacoustic oscillations. Note that, in all of these images, the walls of the confinement tube are located outside of the field of view, far from the flame itself. The vertical line near the centerline is caused by spurious reflections in the quartz tube used for the combustion chamber.

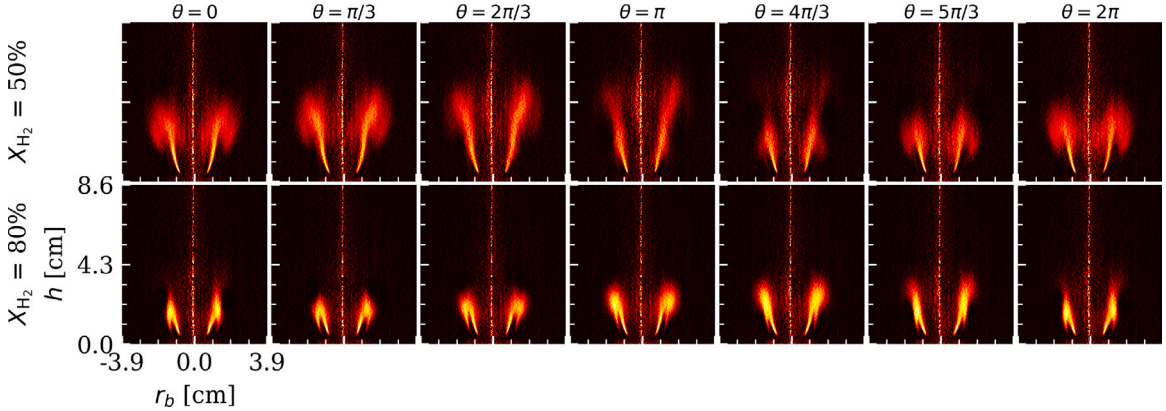

Fig. A.2. Phase-averaged chemiluminescence images of the flame at $u_b = 22.5\,\mathrm{m\,s^{-1}}$, $\phi = 0.8$, $X_{H_2} = 50\%$ (top), and $X_{H_2} = 80\%$ (bottom). These images are obtained during a limit cycle oscillation. Phase-averaging is performed using a dynamic mode decomposition procedure [58], followed by an inverse Abel transform (see [59] for a detailed description of the methodology). Note that the phase reference is arbitrary.

to perturbations, which can be more favorable for thermoacoustic oscillations [66]. For higher H_2 content, $50\% \leq X_{H_2} \leq 80\%$, in a region where thermoacoustic instabilities exhibit a limit-cycle behavior, whose amplitude slightly decreases with X_{H_2}, the flame adopts a distinct "M" shape. The reduced gain of the flame-describing function of "M" flames [64] could explain the decrease in the amplitude of the limit cycle observed at high X_{H_2} in Fig. 3.

Appendix B. Supplementary data

The supplementary information presents additional quantitative data to assess the performance improvements achieved with the RL controllers. The unprocessed experimental data used for characterizing the thermoacoustic instabilities in Section 3 is publicly available through the Stanford Digital Repository [67].

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.combustflame.2025.114406.

References

[1] S. Candel, Combustion dynamics and control: Progress and challenges, Proc. Combust. Inst. 29 (2002) 1–28.
[2] T.C. Lieuwen, V. Yang, Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms, and Modeling, American Institute of Aeronautics and Astronautics, 2005.
[3] S. Candel, D. Durox, T. Schuller, J.-F. Bourgouin, J.P. Moeck, Dynamics of swirling flames, Annu. Rev. Fluid Mech. 46 (2014) 147–173.
[4] Y. Méry, Dynamical response of a perfectly premixed flame and limit behavior for high power density systems, Combust. Flame 192 (2018) 410–425.
[5] S. Taamallah, K. Vogiatzaki, F.M. Alzahrani, E.M. Mokheimer, M. Habib, A.F. Ghoniem, Fuel flexibility, stability and emissions in premixed hydrogen-rich gas turbine combustion: Technology, fundamentals, and numerical simulations, Appl. Energy 154 (2015) 1020–1047.
[6] J. Beita, M. Talibi, S. Sadasivuni, R. Balachandran, Thermoacoustic instability considerations for high hydrogen combustion in lean premixed gas turbine combustors: A review, Hydrogen 2 (2021) 33–57.
[7] J.C. Oefelein, V. Yang, Comprehensive review of liquid-propellant combustion instabilities in F-1 engines, J. Propul. Power 9 (1993) 657–677.
[8] T. Schuller, D. Durox, P. Palies, S. Candel, Acoustic decoupling of longitudinal modes in generic combustion systems, Combust. Flame 159 (2012) 1921–1931.
[9] G. Vignat, D. Durox, K. Prieur, S. Candel, An experimental study into the effect of injector pressure loss on self-sustained combustion instabilities in a swirled spray burner, Proc. Combust. Inst. 37 (2019) 5205–5213.
[10] D. Zhao, X.Y. Li, A review of acoustic dampers applied to combustion chambers in aerospace industry, Prog. Aerosp. Sci. 74 (2015) 114–130.
[11] D. Durox, J.P. Moeck, J.-F. Bourgouin, P. Morenton, M. Viallon, T. Schuller, S. Candel, Flame dynamics of a variable swirl number system and instability control, Combust. Flame 160 (2013) 1729–1742.
[12] K.T. Kim, J.G. Lee, H.J. Lee, B.D. Quay, D.A. Santavicca, Characterization of forced flame response of swirl-stabilized turbulent lean-premixed flames in a gas turbine combustor, J. Eng. Gas Turbines Power 132 (2010) 041502.
[13] X. Wang, X. Han, H. Song, C. Zhang, J. Wang, X. Hui, Y. Lin, D. Yang, C.J. Sung, Combustion instabilities with different degrees of premixedness in a separated dual-swirl burner, J. Eng. Gas Turbines Power 142 (2020) 061012.
[14] E. Æsøy, G.K. Jankee, S. Yadala, N.A. Worth, J.R. Dawson, Suppression of self-excited thermoacoustic instabilities by convective-acoustic interference, Proc. Combust. Inst. 39 (4) (2023) 4611–4620.
[15] N. Noiray, D. Durox, T. Schuller, S. Candel, A novel strategy for passive control of combustion instabilities through modification of flame dynamics, in: Proc. ASME Turbo Expo, Paper GT2008-51520, Berlin, Germany, Jun. 9–13, 2008, pp. 1133–1144.
[16] T. Poinsot, Prediction and control of combustion instabilities in real engines, Proc. Combust. Inst. 36 (2017) 1–28.
[17] D. Zhao, Z. Lu, H. Zhao, X.Y. Li, B. Wang, P. Liu, A review of active control approaches in stabilizing combustion systems in aerospace industry, Prog. Aerosp. Sci. 97 (2018) 35–60.
[18] K. McManus, T. Poinsot, S.M. Candel, A review of active control of combustion instabilities, Prog. Energy Combust. Sci. 19 (1) (1993) 1–29.
[19] W. Lang, T. Poinsot, S. Candel, Active control of combustion instability, Combust. Flame 70 (1987) 281–289.
[20] T. Poinsot, F. Bourienne, S. Candel, E. Esposito, W. Lang, Suppression of combustion instabilities by active control, J. Propul. Power 5 (1989) 14–20.
[21] J.P. Moeck, M.R. Bothien, D. Guyot, C.O. Paschereit, Phase-shift control of combustion instability using (combined) secondary fuel injection and acoustic forcing, in: Active Flow Control, Berlin, Germany, Sept. 27-29, 2006, pp. 408–421.
[22] J. Seume, N. Vortmeyer, W. Krause, J. Hermann, C.-C. Hantschk, P. Zangl, S. Gleis, D. Vortmeyer, A. Orthmann, Application of active combustion instability control to a heavy duty gas turbine, J. Eng. Gas Turbines Power 120 (4) (1998) 721–726.
[23] M.P. Juniper, R.I. Sujith, Sensitivity and nonlinearity of thermoacoustic oscillations, Annu. Rev. Fluid Mech. 50 (2018) 661–689.
[24] T.N. Demayo, V.G. McDonell, G.S. Samuelsen, Robust active control of combustion stability and emissions performance in a fuel-staged natural-gas-fired industrial burner, Proc. Combust. Inst. 29 (2002) 131–138.
[25] Y. Liu, J. Tan, H. Li, Y. Hou, D. Zhang, B.R. Noack, Simultaneous control of combustion instabilities and NO_x emissions in a lean premixed flame using linear genetic programming, Combust. Flame 251 (2023) 112716.
[26] B. Dharmaputra, P. Reckinger, B. Schuermans, N. Noiray, BOATS: Bayesian optimization for active control of thermoacoustic, J. Sound Vib. 582 (2024) 118415.
[27] M. Malu, G. Dasarathy, A. Spanias, Bayesian optimization in high-dimensional spaces: A brief survey, in: 12th International Conference on Information, Intelligence, Systems & Applications, IISA, IEEE, 2021, pp. 1–8.
[28] X.B. Peng, A. Kanazawa, J. Malik, P. Abbeel, S. Levine, SFV: Reinforcement learning of physical skills from videos, ACM Trans. Graph. 37 (6) (2018) 1–14.
[29] M. Korshunova, N. Huang, S. Capuzzi, D.S. Radchenko, O. Savych, Y.S. Moroz, C.I. Wells, T.M. Willson, A. Tropsha, O. Isayev, Generative and reinforcement learning approaches for the automated de novo design of bioactive compounds, Commun. Chem. 5 (1) (2022) 129.
[30] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, R. Lowe, Training language models to follow instructions with human feedback, Adv. Neural Inf. Process. Syst. 35 (2022) 27730–27744.
[31] D. Fan, L. Yang, Z. Wang, M.S. Triantafyllou, G.E. Karniadakis, Reinforcement learning for bluff body active flow control in experiments and simulations, Proc. Natl. Acad. Sci. USA 117 (42) (2020) 26091–26098.
[32] C. Vignon, J. Rabault, R. Vinuesa, Recent advances in applying deep reinforcement learning for flow control: Perspectives and future directions, Phys. Fluids 35 (3) (2023) 031301.
[33] M. Ihme, W.T. Chung, A.A. Mishra, Combustion machine learning: Principles, progress and prospects, Prog. Energy Combust. Sci. 91 (2022) 101010.
[34] X. Zhan, H. Xu, Y. Zhang, X. Zhu, H. Yin, Y. Zheng, Deepthermal: Combustion optimization for thermal power generating units using offline reinforcement learning, in: Proc. AAAI Conf. Artif. Intell., vol. 36, 2022, pp. 4680–4688.
[35] M. Ihme, W.T. Chung, Artificial intelligence as a catalyst for combustion science and engineering, Proc. Combust. Inst. 40 (2024) 105730.
[36] K. Alhazmi, S.M. Sarathy, Adaptive phase shift control of thermoacoustic combustion instabilities using model-free reinforcement learning, Combust. Flame 257 (2023) 113040.
[37] S. Ayache, J.R. Dawson, A. Triantafyllidis, R. Balachandran, E. Mastorakos, Experiments and large-eddy simulations of acoustically forced bluff-body flows, Int. J. Heat Fluid Flow 31 (2010) 754–766.
[38] E. Æsøy, J.G. Aguilar, S. Wiseman, M.R. Bothien, N.A. Worth, J.R. Dawson, Scaling and prediction of transfer functions in lean premixed H_2/CH_4-flames, Combust. Flame 215 (2020) 269–282.
[39] P. Rajendram Soundararajan, D. Durox, A. Renaud, G. Vignat, S. Candel, Swirler effects on combustion instabilities analyzed with measured FDFs, injector impedances and damping rates, Combust. Flame 238 (2022) 111947.
[40] J. Ballester, T. García-Armingol, Diagnostic techniques for the monitoring and control of practical flames, Prog. Energy Combust. Sci. 36 (2010) 375–411.
[41] C.Y. Lee, L.K.B. Li, M.P. Juniper, R.S. Cant, Nonlinear hydrodynamic and thermoacoustic oscillations of a bluff-body stabilised turbulent premixed flame, Combust. Theor. Model. 20 (1) (2016) 131–153.
[42] S. Nakaya, K. Omi, T. Okamoto, Y. Ikeda, C. Zhao, M. Tsue, H. Taguchi, Instability and mode transition analysis of a hydrogen-rich combustion in a model afterburner, Proc. Combust. Inst. 38 (4) (2021) 5933–5942.
[43] G. Singh, S. Mariappan, Experimental investigation on the route to vortex-acoustic lock-in phenomenon in bluff body stabilized combustors, Combust. Sci. Technol. 193 (9) (2021) 1538–1566.
[44] T. Poinsot, D. Veynante, Theoretical and Numerical Combustion, third ed., self-published, 2016.
[45] B. Schuermans, J. Moeck, A. Blondé, B. Dharmaputra, N. Noiray, The Rayleigh integral is always positive in steadily operated combustors, Proc. Combust. Inst. 39 (4) (2023) 4661–4669.
[46] N. Noiray, D. Durox, T. Schuller, S. Candel, A unified framework for nonlinear combustion instability analysis based on the flame describing function, J. Fluid Mech. 615 (2008) 139–167.
[47] J.G. Aguilar, E. Æsøy, J.R. Dawson, The influence of hydrogen on the stability of a perfectly premixed combustor, Combust. Flame 245 (2022) 112323.
[48] C.J. Watkins, P. Dayan, Q-learning, Mach. Learn. 8 (1992) 279–292.
[49] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, Human-level control through deep reinforcement learning, Nature 518 (7540) (2015) 529–533.

[50] P. Welch, The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoust. 15 (2) (1967) 70–73.
[51] T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, in: Proc. Int. Conf. Mach. Learn., PMLR, 2018, pp. 1861–1870.
[52] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, S. Levine, Soft actor-critic algorithms and applications, 2018, arXiv preprint arXiv:1812.05905.
[53] H. Hasselt, Double Q-learning, Adv. Neural Inf. Process. Syst. 23 (2010).
[54] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
[55] M. Ihme, Combustion and engine-core noise, Annu. Rev. Fluid Mech. 49 (1) (2017) 277–310.
[56] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A.A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, R. Hadsell, Overcoming catastrophic forgetting in neural networks, Proc. Natl. Acad. Sci. USA 114 (13) (2017) 3521–3526.
[57] D. Rolnick, A. Ahuja, J. Schwarz, T. Lillicrap, G. Wayne, Experience replay for continual learning, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, R. Garnett (Eds.), in: Advances in Neural Information Processing Systems, vol. 32, 2019.
[58] P.J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28.
[59] G. Vignat, N. Minesi, P. Rajendram Soundararajan, D. Durox, A. Renaud, V. Blanchard, C.O. Laux, S. Candel, Improvement of lean blow out performance of spray and premixed swirled flames using nanosecond repetitively pulsed discharges, Proc. Combust. Inst. 38 (2021) 6559–6566.
[60] T. Schuller, D. Durox, S. Candel, A unified model for the prediction of laminar flame transfer functions: Comparisons between conical and V-flame dynamics, Combust. Flame 134 (2003) 21–34.
[61] M. Stöhr, K. Oberleithner, M. Sieber, Z. Yin, W. Meier, Experimental study of transient mechanisms of bistable flame shape transitions in a swirl combustor, J. Eng. Gas Turbines Power 140 (1) (2018) 011503.
[62] G. Bonciolini, D. Ebi, U. Doll, M. Weilenmann, N. Noiray, Effect of wall thermal inertia upon transient thermoacoustic dynamics of a swirl-stabilized flame, Proc. Combust. Inst. 37 (4) (2019) 5351–5358.
[63] Q. An, W.Y. Kwong, B.D. Geraedts, A.M. Steinberg, Coupled dynamics of lift-off and precessing vortex core formation in swirl flames, Combust. Flame 168 (2016) 228–239.
[64] D. Durox, T. Schuller, N. Noiray, S. Candel, Experimental analysis of nonlinear flame transfer functions for different flame geometries, Proc. Combust. Inst. 32 I (2009) 1391–1398.
[65] W. Polifke, Modeling and analysis of premixed flame dynamics by means of distributed time delays, Prog. Energy Combust. Sci. 79 (2020) 100845.
[66] T. Schuller, T. Poinsot, S. Candel, Dynamics and control of premixed combustion systems based on flame transfer and describing functions, J. Fluid Mech. 894 (2020) P1.
[67] B. Akoush, G. Vignat, R. Finley, W.T. Chung, M. Ihme, Thermoacoustic instability experimental dataset for fuel-flexible combustion systems, Stanf. Digit. Repos. (2025).