Winning a Won Game: Strict Reach–Avoid–Stay Control Barrier Functions for High-Dimensional Black-Box Systems

Donggeon David Oh[1], Duy P. Nguyen[1], Gongkai Yuan[2],
Qingchen Li[2], Jaime Fernández Fisac[1,†], and Haimin Hu[2,†]

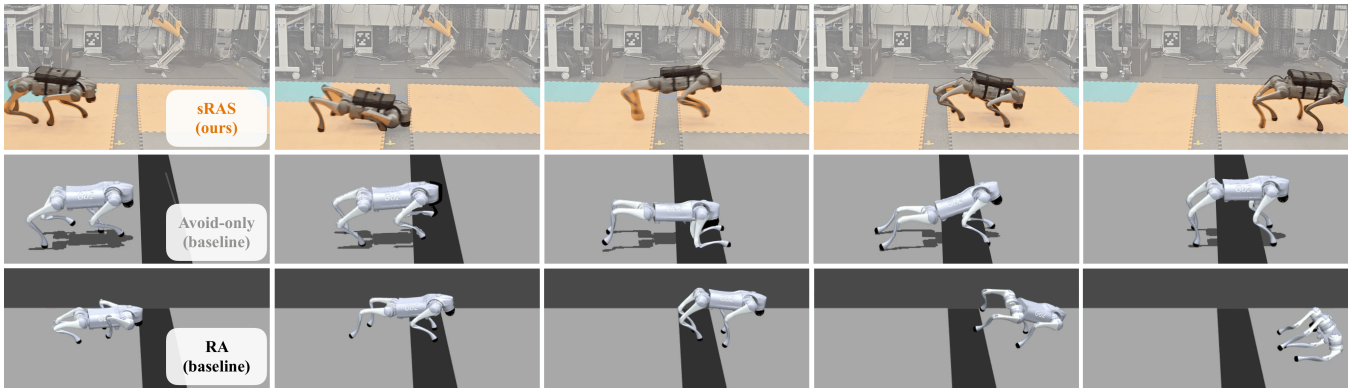

Fig. 1. A Unitree Go2 quadruped must jump across a 25 cm gap, land safely on the far side, and remain safe afterward. This is a strict reach–avoid–stay (**sRAS**) task, where a robot must safely reach a target and remain safely within it indefinitely after first target entry. All methods use the same task policy, which is a pure-pursuit policy that walks in the right direction. Sequences progress from left to right. **Top (hardware):** With our **sRAS** Q-control barrier function (CBF) safety filter, the robot builds momentum, crouches for takeoff, jumps across the gap, lands safely, and remains safe. The same successful behavior is observed in simulation. **Middle (simulation):** The **Avoid-only** safety filter does not encode reaching the far side; the robot fails to complete the jump, leaving its hind legs on the starting side. **Bottom (simulation):** The reach–avoid (**RA**) safety filter encodes reaching the target safely but not remaining safe afterward; the robot jumps across the gap but fails to land safely. The successful hardware execution illustrates the intended behavior of our **sRAS** Q-CBF filter: modifying task actions to keep the robot in **sRAS** winnable states and maintain safety within the target after first target entry.

***Abstract*— Robots must complete their tasks and maintain the achieved outcomes while avoiding safety failures at all times. Strict reach–avoid–stay (sRAS) formalizes this requirement: safely reaching a target and remaining there indefinitely after first entry. We propose an sRAS Q-control barrier function (CBF) safety filter for high-dimensional black-box systems under bounded uncertainty. Our construction combines a stay value encoding safe permanent residence in a target subset with a reach–avoid value encoding safe reachability of this subset while avoiding target states from which safe permanent residence cannot be guaranteed. We prove that these values jointly yield a valid robust discrete-time CBF and lift them to state–action Q-functions for runtime intervention. For exact values and under a measure-zero condition, our filter preserves sRAS feasibility from almost every winnable initial state and keeps the system safely within the target after first entry, against all admissible uncertainty realizations. We adopt reachability-based adversarial reinforcement learning for scalable value approximation using only black-box interactions. Notably, neither synthesis nor deployment of our filter requires known dynamics, affine structure, value derivatives, or hand-designed barriers. We validate our framework in quadruped gap jumping in simulation and hardware, where the robot crosses the gap, lands safely, and remains safe afterward. Simulated F1TENTH races further demonstrate safe overtaking and lead retention.**

[1]D. D. Oh, D. P. Nguyen, and J. F. Fisac are with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA. {do9948,duyn,jfisac}@princeton.edu

[2]G. Yuan, Q. Li, and H. Hu are with the Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA. {gyuan3,qli137}@jh.edu, haimin@cs.jhu.edu

[†]Equal advising.

The hardest game to win is a won game.

EMANUEL LASKER, WORLD CHESS CHAMPION

I. INTRODUCTION

As robots take on increasingly complex tasks in the real world, they must operate safely under uncertainty. *Safety filters* address this need by monitoring operation at runtime and intervening when necessary by modifying proposed controls to prevent catastrophic failures [1]. *Control barrier function (CBF)*-based filters are a popular choice because they can provide smooth safety interventions [1–3].

However, safety interventions should also *preserve the ability to complete the task and maintain the achieved outcome*. For example, a quadruped tasked with crossing a gap should build momentum, jump, land safely, and remain safe after touchdown, rather than stay on the starting side or jump only to fall after touchdown (Figure 1). In car racing, the car should overtake safely and retain the lead, rather than remain behind an opponent or get overtaken again (Figure 2). We formalize this requirement as *strict reach–avoid–stay (sRAS)*, which requires safely reaching a predefined target and remaining there indefinitely after the first entry.

We consider *high-dimensional systems* under bounded uncertainty with only *black-box access to their dynamics*. *Our goal is to devise a CBF-based safety filter for such systems that preserves sRAS feasibility, i.e., keeps the system in states from which the specification remains achievable.*

Existing work extends CBFs to encode similar temporal specifications through time-varying CBFs [4–6] and control Lyapunov–barrier methods [7, 8], but these approaches rely on known control-affine dynamics, preventing their direct application to black-box systems. On the other hand, reachability-based reinforcement learning (RL) enables scalable value learning for complex temporal specifications through black-box interaction [9–14]. In parallel, exact reachability values are valid CBFs in a generalized sense, with runtime filtering requiring known dynamics and value derivatives [15–17], while learned values have been used as discrete-time control barrier functions (DCBFs) for avoid-only tasks on black-box systems [2, 3]. We review these approaches and compare them with our work in Section II.

Building upon these works, we propose the *sRAS Q-CBF safety filter*. Our main contributions are as follows:

sRAS Q-CBF theory. We propose the *sRAS Q-CBF safety filter* for *high-dimensional black-box systems under bounded uncertainty*. We prove that the *stay value*, encoding safe permanent residence in a target subset, and the *reach–avoid value*, encoding reachability of this subset while avoiding failure states and target states from which safe permanent residence cannot be guaranteed, jointly yield a valid robust DCBF. For exact values and under a measure-zero condition, our filter 1) *is recursively feasible*, 2) *preserves sRAS feasibility from almost every winnable initial state*, and 3) *keeps the system safely within the target indefinitely after first entry* against all admissible uncertainty realizations.

Scalable black-box synthesis and deployment. We adopt reachability-based adversarial RL for scalable synthesis of stay and reach–avoid values and their state–action lifts using only black-box interactions. The lifted values, together with the learned worst-case response disturbance policies, enable direct evaluation of Q-CBF constraints for candidate controls. Thus, *neither synthesis nor deployment requires known dynamics, control- or disturbance-affine structure, value derivatives, or hand-designed barrier candidates*.

Robotic evaluation. We evaluate our sRAS Q-CBF safety filter on two robotic tasks: 1) quadruped gap jumping in simulation and hardware, where the robot crosses the gap, lands safely, and remains safe afterward, and 2) simulated F1TENTH endurance racing, where the ego vehicle overtakes opponents safely and retains the lead.

II. Related Work

CBFs for temporal specifications. Prior work extends CBFs beyond failure avoidance to encode temporal specifications. Time-varying CBFs are often adopted to capture the semantics of signal temporal logic (STL) tasks [4]. Das et al. [5, 6] synthesize spatiotemporal tubes for STL tasks, including prescribed-time reach–avoid–stay (RAS), and derive corresponding time-varying CBFs. Control Lyapunov–barrier methods that induce safe stabilization enable the synthesis of RAS controllers [7] and conservative approximations of local winning sets for decomposed temporal logic tasks [8]. These approaches require known control-affine dynamics for controller synthesis and deployment. By contrast, our sRAS Q-CBF framework enables end-to-end synthesis and deployment on high-dimensional black-box systems while preserving the ability to satisfy the sRAS specification from almost every state in its maximal robust winning set under a measure-zero assumption.

Reachability analysis for temporal specifications. Decomposing complex temporal logic specifications into reachability subproblems has been adopted for devising least-restrictive feasible controller sets for STL [18] and constructing temporal logic trees for online control synthesis [19]. Recent advances in reachability-based RL enable scalable solutions to these subproblems. Time-discounted Bellman equations enable neural approximation of reachability values via RL [9, 20], and extensions to uncertain dynamics yield a predictive safety filter based on rollouts under disturbance [10]. These methods support learning optimal policies [12, 13] and constructing least-restrictive safety filters [14] for complex temporal logic tasks decomposed into graphs of reachability subproblems. RAS has also been decomposed into avoid and reach–avoid subproblems, with corresponding policies learned via RL [11]. Our work similarly decomposes the sRAS specification into two reachability subproblems and uses reachability-based RL to learn their values through black-box interactions. However, we differ from this line of work in that we devise a CBF safety filter that preserves sRAS feasibility under bounded uncertainty.

Reachability-derived CBFs. In continuous-time systems, the infinite-horizon avoid value is known to be a valid CBF when continuously differentiable [1]. Robust control barrier–value functions (CB-VFs), constructed as viscosity solutions of a modified Hamilton–Jacobi–Isaacs variational inequality, enable CBF-like filtering [15], while viscosity CBFs that satisfy the CBF condition in the viscosity sense are equivalent to time-invariant CB-VFs if nonnegative [16]. A finite-horizon reach–avoid value is also shown to be a valid time-varying viscosity-based CBF [17]. Although reachability values can recover maximal winning sets, runtime evaluation of the corresponding CBF constraints requires known dynamics and derivatives of the value or a viscosity test function. In discrete time, avoid values yield valid DCBFs, with the lifted value function enabling scalable black-box synthesis and deployment [2, 3]. However, these avoid-only values do not by themselves encode task-completion feasibility for temporal specifications. Our sRAS Q-CBF enables black-box synthesis and deployment without known dynamics, control- or disturbance-affine structure, or value derivatives while preserving the ability to satisfy the sRAS specification.

III. Preliminaries and Problem Formulation

A. Strict Reach–Avoid–Stay under Bounded Uncertainty

We consider a nonlinear discrete-time system

$$x_{t+1} = f(x_t, u_t, d_t), \tag{1}$$

where $x_t \in \mathcal{X} \subseteq \mathbb{R}^n$ is the state, $u_t \in \mathcal{U}$ is the control input, and $d_t \in \mathcal{D}$ is an unknown but bounded disturbance

input representing predictive uncertainty. We adopt a black-box viewpoint in which the transition mechanism in (1) can be queried for the next state given the current state, control, and disturbance, but need not be available in closed form or possess control- or disturbance-affine structure.

We consider sRAS tasks, with a *failure set* $\mathcal{F} \subset \mathcal{X}$ containing unacceptable states that must be avoided at all times, and a *target set* $\mathcal{T} \subset \mathcal{X}$ containing states that the system should reach and remain within after its first entry. Throughout, robust satisfaction requires a single control policy $\pi^u : \mathcal{X} \to \mathcal{U}$ to satisfy the specification against every disturbance policy $\pi^d : \mathcal{X} \to \mathcal{D}$.

Definition 1 (Maximal Robust sRAS Set)**.** *The maximal robust sRAS set is*

$$\Omega^*_{\mathrm{sRAS}} := \left\{ x \in \mathcal{X} \;\middle|\; \begin{array}{ll} \exists \pi^u,\ \forall \pi^d,\ \exists \tau \geq 0 \text{ s.t.} & \\ \mathbf{x}_x^{\pi^u,\pi^d}(t) \notin \mathcal{F}, & \forall t \geq 0, \\ \mathbf{x}_x^{\pi^u,\pi^d}(t) \notin \mathcal{T}, & \forall t < \tau, \\ \mathbf{x}_x^{\pi^u,\pi^d}(t) \in \mathcal{T}, & \forall t \geq \tau \end{array} \right\}. \quad (2)$$

Throughout, $\mathbf{x}_x^{\pi^u,\pi^d} : \mathbb{Z}_{\geq 0} \to \mathcal{X}$ *denotes the state trajectory initialized at* $\mathbf{x}_x^{\pi^u,\pi^d}(0) = x$ *and evolved according to* $\pi^u : \mathcal{X} \to \mathcal{U}$, $\pi^d : \mathcal{X} \to \mathcal{D}$, *and the dynamics* (1)*.*

Conventional RAS omits the condition $\mathbf{x}_x^{\pi^u,\pi^d}(t) \notin \mathcal{T}$, $\forall t < \tau$, in (2), allowing repeated target entry and exit before eventually remaining there permanently. This is undesirable for robotic tasks that require irreversible task completion—in car racing, the ego vehicle should not fall behind again after an overtake—making sRAS a more suitable specification. Any sRAS-feasible state $x \in \mathcal{X}$ is also RAS-feasible.

B. Robust Discrete-Time Control Barrier Function

A robust DCBF renders its 0-superlevel set robustly controlled invariant by requiring a control input whose worst-case next-step barrier value is lower-bounded by a class-$\mathcal{K}$ mapping of its current value. We adopt the definition in [3].

Definition 2 (Robust Discrete-Time CBF)**.** *A function* $h : \mathcal{X} \to \mathbb{R}$ *is a robust DCBF for system* (1) *if* $\Omega_h := \{x \in \mathcal{X} \mid h(x) \geq 0\}$ *satisfies* $\Omega_h \cap \mathcal{F} = \varnothing$*, and there exists a class-*$\mathcal{K}$ *function* β*, defined on a domain containing* $h(\Omega_h)$*, such that*

$$\sup_{u \in \mathcal{U}} \inf_{d \in \mathcal{D}} h\big(f(x,u,d)\big) \geq \beta\big(h(x)\big), \quad \forall x \in \Omega_h. \quad (3)$$

Ω_h may in general be a strict subset of the maximal robust safe set from which failure can be avoided indefinitely against every admissible disturbance policy [3].

IV. Decomposing the sRAS Game

We introduce a two-stage reachability-based value synthesis pipeline for robust sRAS. Stage I computes the maximal robust terminal safe set, the largest robust controlled invariant subset of $\mathcal{T} \setminus \mathcal{F}$. Stage II computes the maximal robust reach–avoid set from which this terminal safe set can be reached while avoiding $\mathcal{F}$ and interior target states outside the terminal safe set. For each stage, we lift the corresponding value into the state–control–disturbance space through Q-functions [21], which enables CBF constraint evaluation for black-box systems. Finally, we prove almost-everywhere recovery of the maximal robust sRAS set via our two-stage synthesis pipeline under an assumption. Throughout, we assume that all displayed maxima and minima are attained.

We choose continuous *safety and target margins* $g, \ell : \mathcal{X} \to \mathbb{R}$ satisfying

$$\mathcal{F} = \{x \in \mathcal{X} \mid g(x) < 0\}, \quad \mathcal{T} = \{x \in \mathcal{X} \mid \ell(x) \geq 0\}, \quad (4)$$

with $\mathrm{int}(\mathcal{T}) = \{\ell > 0\}$ and $\partial\mathcal{T} = \{\ell = 0\}$.

A. Stage I: Stay Value

We define the *stay-mode failure set* as

$$\mathcal{F}_{\mathrm{S}} := \mathcal{F} \cup \mathcal{T}^{\mathsf{c}}. \quad (5)$$

Definition 3 (Maximal Robust Terminal Safe Set)**.** *The maximal robust terminal safe set is*

$$\Omega^*_{\mathrm{S}} := \left\{ x \in \mathcal{X} \middle| \exists \pi^u,\ \forall \pi^d,\ \forall t \geq 0,\ \mathbf{x}_x^{\pi^u,\pi^d}(t) \notin \mathcal{F}_{\mathrm{S}} \right\}. \quad (6)$$

To transform this binary safety outcome into a "game of degree", we choose the *stay-mode safety margin*

$$g_{\mathrm{S}}(x) := \min\left\{g(x), \ell(x)\right\}, \quad (7)$$

which satisfies $\mathcal{F}_{\mathrm{S}} = \{g_{\mathrm{S}} < 0\}$.

The *stay value* V_{S} is then defined as the maximal g_{S} that can be maintained against the worst-case admissible disturbance policy over the infinite horizon,

$$V_{\mathrm{S}}(x) := \max_{\pi^u} \min_{\pi^d} \inf_{t \geq 0} g_{\mathrm{S}}\left(\mathbf{x}_x^{\pi^u,\pi^d}(t)\right), \quad (8)$$

and V_{S} satisfies the avoid-only Isaacs equation

$$V_{\mathrm{S}}(x) = \min\left\{g_{\mathrm{S}}(x), \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} V_{\mathrm{S}}\big(f(x,u,d)\big)\right\}. \quad (9)$$

Ω^*_{S} in (6) is recovered as the 0-superlevel set of V_{S}:

$$\Omega^*_{\mathrm{S}} = \{x \in \mathcal{X} \mid V_{\mathrm{S}}(x) \geq 0\}. \quad (10)$$

To directly evaluate candidate control inputs under disturbance realizations, we define the *lifted stay value*

$$\mathcal{Q}_{\mathrm{S}}(x,u,d) := \min\left\{g_{\mathrm{S}}(x), V_{\mathrm{S}}\big(f(x,u,d)\big)\right\}. \quad (11)$$

For fixed x, the map $z \mapsto \min\{g_{\mathrm{S}}(x), z\}$ is monotone nondecreasing, so it preserves the attained inner minimum and outer maximum, yielding

$$V_{\mathrm{S}}(x) = \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} \mathcal{Q}_{\mathrm{S}}(x,u,d). \quad (12)$$

The corresponding stay fallback policy is

$$\pi^{\mathbf{f}}_{\mathrm{S}}(x) \in \arg\max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} \mathcal{Q}_{\mathrm{S}}(x,u,d). \quad (13)$$

B. Stage II: Reach–Avoid Value

We define the *reach-mode target and failure sets* as

$$\mathcal{T}_{\mathrm{R}} := \Omega^*_{\mathrm{S}}, \qquad \mathcal{F}_{\mathrm{R}} := \mathcal{F} \cup \big(\mathrm{int}(\mathcal{T}) \setminus \mathcal{T}_{\mathrm{R}}\big), \tag{14}$$

so that the reach mode terminates only upon entry into Ω^*_{S}. $\mathcal{F}_{\mathrm{R}}$ does not include safe nonviable states on $\partial\mathcal{T}$, since $\mathcal{F} \cup \big(\mathcal{T} \setminus \mathcal{T}_{\mathrm{R}}\big)$ cannot in general be represented exactly as a strict sublevel set of a continuous safety margin. We quantify the resulting discrepancy in Subsection IV-C.

Definition 4 (Maximal Robust Reach–Avoid Set)**.** *The maximal robust reach–avoid set is*

$$\Omega^*_{\mathrm{RA}} := \left\{ x \in \mathcal{X} \;\middle|\; \begin{array}{l} \exists \pi^u,\ \forall \pi^d,\ \exists \tau \geq 0 \text{ s.t.} \\ \mathbf{x}_x^{\pi^u,\pi^d}(\tau) \in \mathcal{T}_{\mathrm{R}}, \\ \mathbf{x}_x^{\pi^u,\pi^d}(t) \notin \mathcal{F}_{\mathrm{R}},\ \ \forall 0 \leq t \leq \tau \end{array} \right\}. \tag{15}$$

Since $\mathcal{T}_{\mathrm{R}} = \Omega^*_{\mathrm{S}} \subseteq \mathcal{T} \setminus \mathcal{F}$ and $\Omega^*_{\mathrm{S}} \cap \mathcal{F}_{\mathrm{R}} = \varnothing$, every $x \in \Omega^*_{\mathrm{S}}$ satisfies Definition 4 with $\tau = 0$. Therefore,

$$\Omega^*_{\mathrm{S}} \subseteq \Omega^*_{\mathrm{RA}}. \tag{16}$$

We choose the *reach-mode target margin* ℓ_{R} and *reach-mode safety margin* g_{R} as

$$\ell_{\mathrm{R}}(x) := V_{\mathrm{S}}(x), \tag{17a}$$

$$g_{\mathrm{R}}(x) := \min\left\{ g(x), \max\left\{ -\ell(x), \ell_{\mathrm{R}}(x) \right\} \right\}. \tag{17b}$$

The target margin recovers the reach-mode target set,

$$\mathcal{T}_{\mathrm{R}} = \left\{ x \in \mathcal{X} \mid \ell_{\mathrm{R}}(x) \geq 0 \right\}, \tag{18}$$

while the safety margin recovers the reach-mode failure set,

$$\begin{aligned} \{g_{\mathrm{R}} < 0\} &= \{g < 0\} \cup \big(\{\ell > 0\} \cap \{\ell_{\mathrm{R}} < 0\}\big) \\ &= \mathcal{F} \cup \big(\mathrm{int}(\mathcal{T}) \setminus \mathcal{T}_{\mathrm{R}}\big) = \mathcal{F}_{\mathrm{R}}. \end{aligned} \tag{19}$$

The *reach–avoid value* V_{R} is defined by jointly considering the ℓ_{R} at arrival and g_{R} maintained up to arrival under the worst-case admissible disturbance policy:

$$\begin{aligned} V_{\mathrm{R}}(x) := \max_{\pi^u} \min_{\pi^d} \max_{\tau \geq 0} \min\Big\{ & \ell_{\mathrm{R}}\Big(\mathbf{x}_x^{\pi^u,\pi^d}(\tau)\Big), \\ & \min_{0 \leq t \leq \tau} g_{\mathrm{R}}\Big(\mathbf{x}_x^{\pi^u,\pi^d}(t)\Big) \Big\}, \end{aligned} \tag{20}$$

and V_{R} satisfies the reach–avoid Isaacs equation

$$V_{\mathrm{R}}(x) = \min\Big\{ g_{\mathrm{R}}(x), \max\Big\{ \ell_{\mathrm{R}}(x), \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} V_{\mathrm{R}}\big(f(x,u,d)\big) \Big\} \Big\}. \tag{21}$$

Ω^*_{RA} in (15) is recovered as the 0-superlevel set of V_{R}:

$$\Omega^*_{\mathrm{RA}} = \left\{ x \in \mathcal{X} \mid V_{\mathrm{R}}(x) \geq 0 \right\}. \tag{22}$$

To directly evaluate candidate control inputs under disturbance realizations, we define the *lifted reach–avoid value*

$$\mathcal{Q}_{\mathrm{R}}(x,u,d) := \min\big\{ g_{\mathrm{R}}(x), \max\big\{ \ell_{\mathrm{R}}(x), V_{\mathrm{R}}\big(f(x,u,d)\big) \big\} \big\}. \tag{23}$$

For fixed x, the map $z \mapsto \min\{g_{\mathrm{R}}(x), \max\{\ell_{\mathrm{R}}(x), z\}\}$ is monotone nondecreasing, and therefore

$$V_{\mathrm{R}}(x) = \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} \mathcal{Q}_{\mathrm{R}}(x,u,d). \tag{24}$$

The corresponding reach–avoid fallback policy is

$$\pi^{\mathbf{u}}_{\mathrm{R}}(x) \in \arg\max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} \mathcal{Q}_{\mathrm{R}}(x,u,d). \tag{25}$$

C. Maximal Robust sRAS Set Characterization

To characterize the difference between Ω^*_{RA} and Ω^*_{sRAS} caused by the target-boundary discrepancy, we isolate the relevant portion of the 0-level set of V_{R} as

$$\mathcal{Z}_{\mathrm{R}} := \{ x \in \Omega^*_{\mathrm{RA}} \setminus \Omega^*_{\mathrm{S}} \mid V_{\mathrm{R}}(x) = 0 \}. \tag{26}$$

Proposition 1 (sRAS Set Characterization)**.**

$$\Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}} \subseteq \Omega^*_{\mathrm{sRAS}} \subseteq \Omega^*_{\mathrm{RA}}. \tag{27}$$

Proof. We first show $\Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}} \subseteq \Omega^*_{\mathrm{sRAS}}$. Fix $x \in \Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}}$. If $x \in \Omega^*_{\mathrm{S}}$, then (10)–(13) imply that $\pi^{\mathbf{u}}_{\mathrm{S}}$ robustly renders $\Omega^*_{\mathrm{S}} \subseteq \mathcal{T} \setminus \mathcal{F}$ invariant, so Definition 1 holds with $\tau = 0$.

Otherwise if $x \in \Omega^*_{\mathrm{RA}} \setminus (\Omega^*_{\mathrm{S}} \cup \mathcal{Z}_{\mathrm{R}})$, $V_{\mathrm{R}}(x) > 0$ by (22) and (26), so (20) gives a policy π^u_{R} such that, for every π^d, the trajectory $x_t := \mathbf{x}_x^{\pi^u_{\mathrm{R}},\pi^d}(t)$ admits $\bar{\tau} \geq 0$ with $\ell_{\mathrm{R}}(x_{\bar{\tau}}) > 0$ and $g_{\mathrm{R}}(x_t) > 0,\ \forall t \leq \bar{\tau}$. Let $\tau \leq \bar{\tau}$ be the first entry time into Ω^*_{S}, which exists since $x_{\bar{\tau}} \in \Omega^*_{\mathrm{S}}$. For all $t < \tau$, $\ell_{\mathrm{R}}(x_t) < 0$, so (17b) gives $g(x_t) > 0$ and $\ell(x_t) < 0$, hence $x_t \notin \mathcal{F} \cup \mathcal{T}$. The policy applying π^u_{R} outside Ω^*_{S} and $\pi^{\mathbf{u}}_{\mathrm{S}}$ inside Ω^*_{S} follows x_t through τ and remains in Ω^*_{S} thereafter, satisfying Definition 1 with τ, thus proving $x \in \Omega^*_{\mathrm{sRAS}}$.

To show $\Omega^*_{\mathrm{sRAS}} \subseteq \Omega^*_{\mathrm{RA}}$, fix $x \in \Omega^*_{\mathrm{sRAS}}$ and a policy π^u satisfying Definition 1. For any π^d, let $x_t := \mathbf{x}_x^{\pi^u,\pi^d}(t)$ and let τ be its first entry time into $\mathcal{T}$. To prove $x_\tau \in \Omega^*_{\mathrm{S}}$, take any $\tilde{\pi}^d$ and define $\hat{\pi}^d$ to equal π^d on $\mathcal{T}^{\mathrm{c}}$ and $\tilde{\pi}^d$ on $\mathcal{T}$. Since $x_t \notin \mathcal{T},\ \forall t < \tau$, $\hat{x}_t := \mathbf{x}_x^{\pi^u,\hat{\pi}^d}(t) = x_t,\ \forall t \leq \tau$ and $\hat{x}_t \in \mathcal{T} \setminus \mathcal{F},\ \forall t \geq \tau$ by Definition 1. Now let $\tilde{x}_{\tilde{t}} := \mathbf{x}_{x_\tau}^{\pi^u,\tilde{\pi}^d}(\tilde{t})$. Since $\hat{x}_t \in \mathcal{T},\ \forall t \geq \tau$ and $\hat{\pi}^d = \tilde{\pi}^d$ on $\mathcal{T}$, we have $\tilde{x}_{\tilde{t}} = \hat{x}_{\tau+\tilde{t}} \in \mathcal{T} \setminus \mathcal{F},\ \forall \tilde{t} \geq 0$. Since $\tilde{\pi}^d$ was arbitrary, Definition 3 gives $x_\tau \in \Omega^*_{\mathrm{S}} = \mathcal{T}_{\mathrm{R}}$. For all $t < \tau$, $x_t \notin \mathcal{F} \cup \mathcal{T}$ and $\mathrm{int}(\mathcal{T}) \setminus \mathcal{T}_{\mathrm{R}} \subseteq \mathcal{T}$, so $x_t \notin \mathcal{F}_{\mathrm{R}}$ by (14). At $t = \tau$, $x_\tau \in \mathcal{T}_{\mathrm{R}}$ and $x_\tau \notin \mathcal{F}_{\mathrm{R}}$ by (14). Thus the same π^u and τ satisfy Definition 4, proving $x \in \Omega^*_{\mathrm{RA}}$. □

In practice, 0-level sets of reachability values often appear as boundary-like lower-dimensional sets, both in numerical solutions and learned approximations [3, 9, 20]. This motivates our assumption that $\mathcal{Z}_{\mathrm{R}}$ has Lebesgue measure zero.

Corollary 1 (Almost-Everywhere sRAS Characterization)**.** *If $\mathcal{Z}_{\mathrm{R}}$ has Lebesgue measure zero, then Ω^*_{RA} and Ω^*_{sRAS} coincide almost everywhere.*

V. Robust Switched sRAS Q-CBF

In this section, we show that V_{R} and V_{S} jointly form a valid robust DCBF for the state-dependent switched system that represents the transition from the reach mode to the stay mode. The resulting sRAS Q-CBF safety filter, under every admissible disturbance realization, is recursively feasible and preserves membership in Ω^*_{sRAS} for almost every initial state in Ω^*_{sRAS} under a measure-zero condition. Moreover, if the filtered trajectory enters Ω^*_{S} in finite time, it remains there thereafter and satisfies the sRAS specification in Definition 1.

A. State-Dependent Switched-System Formulation

We formalize the reach-to-stay transition using a state-dependent switched discrete-time control system [22].

Definition 5 (State-Dependent Switched Discrete-Time Control System)**.** *A state-dependent switched discrete-time control system consists of a finite mode set $\mathcal{M}$, subsystem dynamics $\{f_m\}_{m\in\mathcal{M}}$ with $f_m : \mathcal{X}\times\mathcal{U}\times\mathcal{D}\to\mathcal{X}$, and a switching law $\sigma : \mathcal{X}\to\mathcal{M}$. The active mode at time t is $\sigma_t := \sigma(x_t)$, and the state evolves as*

$$x_{t+1} = f_{\sigma_t}(x_t, u_t, d_t). \tag{28}$$

For each $m\in\mathcal{M}$, the operating region is $\mathcal{X}_m := \{x\in\mathcal{X} \mid \sigma(x) = m\}$. A switch occurs when $\sigma_{t+1}\neq\sigma_t$.

Definition 6 (Robust Switched DCBF)**.** *For the switched system in Definition 5, let $\mathcal{F}_m\subseteq\mathcal{X}$ and $h_m : \mathcal{X}\to\mathbb{R}$ denote the failure set and barrier function of each $m\in\mathcal{M}$, with $\Omega_{h_m} := \{x\in\mathcal{X}_m \mid h_m(x)\geq 0\}$. The family $\{h_m\}_{m\in\mathcal{M}}$ is a valid robust switched DCBF if, for every $m\in\mathcal{M}$, there exists a class-$\mathcal{K}$ function β_m, defined on a domain containing $h_m(\Omega_{h_m})$, such that the following conditions hold.*
(i) Mode-wise safety: *$\Omega_{h_m}\cap\mathcal{F}_m = \varnothing$.*
(ii) Switched barrier condition: *For every $x\in\Omega_{h_m}$, there exists $u\in\mathcal{U}$ such that, for every $d\in\mathcal{D}$, letting $x' := f_m(x,u,d)$ and $m' := \sigma(x')$,*

$$h_{m'}(x') \geq \begin{cases} \beta_m\big(h_m(x)\big), & m' = m,\\ 0, & m'\neq m. \end{cases} \tag{29}$$

Condition (ii) requires one control to satisfy the conventional robust DCBF condition in Definition 2 for same-mode successors and place switched successors in the destination certificate's 0-superlevel set, for every disturbance. The latter is analogous to the jump condition of hybrid CBFs [23].

B. Robust Switched sRAS Q-CBF

We represent the sRAS problem with a transition from reach to stay mode as a state-dependent switched system.

Definition 7 (sRAS Switched System)**.** *Let $\mathcal{M} := \{\mathrm{R},\mathrm{S}\}$ and $f_{\mathrm{R}} = f_{\mathrm{S}} := f$, and associate the reach and stay modes with failure sets $\mathcal{F}_{\mathrm{R}}$ and $\mathcal{F}_{\mathrm{S}}$, respectively. The state-dependent switching law is*

$$\sigma(x) := \begin{cases} \mathrm{R}\ \textit{(reach mode)}, & x\notin\Omega^*_{\mathrm{S}},\\ \mathrm{S}\ \textit{(stay mode)}, & x\in\Omega^*_{\mathrm{S}}. \end{cases} \tag{30}$$

We establish feasibility of the mode-dependent Q-CBF constraints and their equivalence to robust DCBF constraints.

Lemma 1 (Feasibility of Q-CBF Constraints)**.** *Let β_{R} and β_{S} be class-$\mathcal{K}$ functions satisfying $\beta_{\mathrm{R}}(\rho)\leq\rho$ and $\beta_{\mathrm{S}}(\rho)\leq\rho$ for all $\rho\geq 0$. For every $x\in\Omega^*_{\mathrm{RA}}\setminus\Omega^*_{\mathrm{S}}$, there exists $u\in\mathcal{U}$ such that*

$$\min_{d\in\mathcal{D}} \mathcal{Q}_{\mathrm{R}}(x,u,d) \geq \beta_{\mathrm{R}}\big(V_{\mathrm{R}}(x)\big). \tag{31}$$

For every $x\in\Omega^*_{\mathrm{S}}$, there exists $u\in\mathcal{U}$ such that

$$\min_{d\in\mathcal{D}} \mathcal{Q}_{\mathrm{S}}(x,u,d) \geq \beta_{\mathrm{S}}\big(V_{\mathrm{S}}(x)\big). \tag{32}$$

Proof. For $x\in\Omega^*_{\mathrm{RA}}\setminus\Omega^*_{\mathrm{S}}$, choose $u = \pi^{\blacktriangledown}_{\mathrm{R}}(x)$. By (24) and (25), $\min_d \mathcal{Q}_{\mathrm{R}}(x,\pi^{\blacktriangledown}_{\mathrm{R}}(x),d) = V_{\mathrm{R}}(x)\geq\beta_{\mathrm{R}}(V_{\mathrm{R}}(x))$, where the inequality follows from $V_{\mathrm{R}}(x)\geq 0$ and $\beta_{\mathrm{R}}(\rho)\leq\rho$. The stay constraint follows identically from (12)–(13). □

Lemma 2 (Equivalence of Q-CBF Constraints)**.** *Let β_{R} and β_{S} be class-$\mathcal{K}$ functions satisfying $\beta_{\mathrm{R}}(\rho)\leq\rho$ and $\beta_{\mathrm{S}}(\rho)\leq\rho$ for all $\rho\geq 0$. For every $x\in\Omega^*_{\mathrm{RA}}\setminus\Omega^*_{\mathrm{S}}$ and $u\in\mathcal{U}$,* (31) *is equivalent to*

$$\min_{d\in\mathcal{D}} V_{\mathrm{R}}\big(f(x,u,d)\big) \geq \beta_{\mathrm{R}}\big(V_{\mathrm{R}}(x)\big). \tag{33}$$

For every $x\in\Omega^*_{\mathrm{S}}$ and $u\in\mathcal{U}$, (32) *is equivalent to*

$$\min_{d\in\mathcal{D}} V_{\mathrm{S}}\big(f(x,u,d)\big) \geq \beta_{\mathrm{S}}\big(V_{\mathrm{S}}(x)\big). \tag{34}$$

Proof. Fix $x\in\Omega^*_{\mathrm{RA}}\setminus\Omega^*_{\mathrm{S}}$, $u\in\mathcal{U}$, and let $c_{\mathrm{R}} := \beta_{\mathrm{R}}(V_{\mathrm{R}}(x))$. Equations (22), (18), (21), and $\beta_{\mathrm{R}}(\rho)\leq\rho$ give $g_{\mathrm{R}}(x)\geq V_{\mathrm{R}}(x)\geq c_{\mathrm{R}}\geq 0 > \ell_{\mathrm{R}}(x)$. Thus, for every $d\in\mathcal{D}$, (23) gives $\mathcal{Q}_{\mathrm{R}}(x,u,d)\geq c_{\mathrm{R}}$ if and only if $V_{\mathrm{R}}(f(x,u,d))\geq c_{\mathrm{R}}$.

For $x\in\Omega^*_{\mathrm{S}}$, the stay-mode equivalence follows analogously from (10), (9), $\beta_{\mathrm{S}}(\rho)\leq\rho$, and (11). □

These results allow us to show that V_{R} and V_{S} jointly form a robust switched DCBF for the sRAS switched system.

Theorem 1 (sRAS Q-CBF)**.** *Let β_{R} and β_{S} be class-$\mathcal{K}$ functions satisfying $\beta_{\mathrm{R}}(\rho)\leq\rho$ and $\beta_{\mathrm{S}}(\rho)\leq\rho$ for all $\rho\geq 0$. For the sRAS switched system in Definition 7, let $h_{\mathrm{R}} := V_{\mathrm{R}}$ and $h_{\mathrm{S}} := V_{\mathrm{S}}$. Then the family $\{h_{\mathrm{R}}, h_{\mathrm{S}}\}$ is a valid robust switched DCBF in the sense of Definition 6.*

Proof. We verify the two conditions in Definition 6.
(i) Mode-wise safety. By (30), (22), and (10), $\Omega_{h_{\mathrm{R}}} = \Omega^*_{\mathrm{RA}}\setminus\Omega^*_{\mathrm{S}}$ and $\Omega_{h_{\mathrm{S}}} = \Omega^*_{\mathrm{S}}$. By Definition 4 and Definition 3, $\Omega^*_{\mathrm{RA}}\cap\mathcal{F}_{\mathrm{R}} = \varnothing$ and $\Omega^*_{\mathrm{S}}\cap\mathcal{F}_{\mathrm{S}} = \varnothing$, establishing mode-wise safety.
(ii) Switched barrier condition. For $x\in\Omega^*_{\mathrm{RA}}\setminus\Omega^*_{\mathrm{S}}$, Lemma 1 provides a control satisfying (31), and Lemma 2 ensures, for every $d\in\mathcal{D}$, $V_{\mathrm{R}}(x')\geq\beta_{\mathrm{R}}(V_{\mathrm{R}}(x))\geq 0$, where $x' := f(x,u,d)$. If $x'\notin\Omega^*_{\mathrm{S}}$, this is the same-mode condition of (29); if $x'\in\Omega^*_{\mathrm{S}}$, (10) gives $V_{\mathrm{S}}(x')\geq 0$, satisfying the switched condition of (29). For $x\in\Omega^*_{\mathrm{S}}$, the two lemmas provide a control satisfying (32) and ensure $V_{\mathrm{S}}(x')\geq\beta_{\mathrm{S}}(V_{\mathrm{S}}(x))\geq 0$ for every $d\in\mathcal{D}$. Thus $x'\in\Omega^*_{\mathrm{S}}$ by (10) and the same-mode condition of (29) is satisfied. □

C. Robust Switched sRAS Q-CBF Safety Filter

Building on Theorem 1, we now propose the sRAS Q-CBF safety filter for black-box systems under uncertainty.

Definition 8 (sRAS Q-CBF Safety Filter)**.** *Let β_{R} and β_{S} be class-$\mathcal{K}$ functions satisfying $\beta_{\mathrm{R}}(\rho)\leq\rho$ and $\beta_{\mathrm{S}}(\rho)\leq\rho$ for all $\rho\geq 0$, and let $\pi^{\textit{task}} : \mathcal{X}\to\mathcal{U}$ be any task policy. For each $x\in\Omega^*_{\mathrm{RA}}$, the sRAS Q-CBF safety filter solves*

$$\phi(x,\pi^{\textit{task}}(x)) \in \arg\min_{u\in\mathcal{U}} \big\|u-\pi^{\textit{task}}(x)\big\|_2^2 \tag{35a}$$

$$\text{s.t. } \min_{d\in\mathcal{D}} \mathcal{Q}_{\mathrm{R}}(x,u,d)\geq\beta_{\mathrm{R}}(V_{\mathrm{R}}(x)),\ x\in\Omega^*_{\mathrm{RA}}\setminus\Omega^*_{\mathrm{S}} \tag{35b}$$

$$\min_{d\in\mathcal{D}} \mathcal{Q}_{\mathrm{S}}(x,u,d)\geq\beta_{\mathrm{S}}(V_{\mathrm{S}}(x)),\ x\in\Omega^*_{\mathrm{S}}. \tag{35c}$$

Crucially, given V_{R}, V_{S} and their lifted forms $\mathcal{Q}_{\mathrm{R}}$, $\mathcal{Q}_{\mathrm{S}}$, (35b)–(35c) can be evaluated for black-box systems without known dynamics, control-affinity, or value gradients.

Finally, we show that the filter is recursively feasible and preserves the ability to satisfy the sRAS specification.

Theorem 2 (sRAS Feasibility Preservation)**.** *For any task policy $\pi^{task} : \mathcal{X} \to \mathcal{U}$, let $\pi^u_{\mathrm{F}}(x) := \phi(x, \pi^{task}(x))$ denote the filtered policy from Definition 8. For every initial state $x_0 \in \Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}} \subseteq \Omega^*_{\mathrm{sRAS}}$ and every disturbance policy $\pi^d : \mathcal{X} \to \mathcal{D}$, the sRAS Q-CBF safety filter is recursively feasible and satisfies $\mathbf{x}_{x_0}^{\pi^u_{\mathrm{F}},\pi^d}(t) \in \Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}} \subseteq \Omega^*_{\mathrm{sRAS}}$, $\forall t \geq 0$. Moreover, if $\mathbf{x}_{x_0}^{\pi^u_{\mathrm{F}},\pi^d}$ enters Ω^*_{S} at $\tau \geq 0$, then $\mathbf{x}_{x_0}^{\pi^u_{\mathrm{F}},\pi^d}(t) \in \Omega^*_{\mathrm{S}}$, $\forall t \geq \tau$, and $\mathbf{x}_{x_0}^{\pi^u_{\mathrm{F}},\pi^d}$ satisfies the sRAS conditions in Definition 1.*

Proof. By Proposition 1, $\Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}} \subseteq \Omega^*_{\mathrm{sRAS}}$. Fix arbitrary $x_0 \in \Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}}$, π^{task}, and π^d, and let $x_t := \mathbf{x}_{x_0}^{\pi^u_{\mathrm{F}},\pi^d}(t)$.

If $x_t \in \Omega^*_{\mathrm{S}} \subseteq \Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}}$, Lemma 1 gives feasibility of (35c), and Lemma 2 gives $V_{\mathrm{S}}(x_{t+1}) \geq \beta_{\mathrm{S}}(V_{\mathrm{S}}(x_t)) \geq 0$. Thus $x_{t+1} \in \Omega^*_{\mathrm{S}}$ by (10), and by induction $x_{\bar{t}} \in \Omega^*_{\mathrm{S}}$, $\forall \bar{t} \geq t$.

If $x_t \in \Omega^*_{\mathrm{RA}} \setminus (\Omega^*_{\mathrm{S}} \cup \mathcal{Z}_{\mathrm{R}})$, Lemma 1 gives feasibility of (35b), (22) and (26) give $V_{\mathrm{R}}(x_t) > 0$, and Lemma 2 gives $V_{\mathrm{R}}(x_{t+1}) \geq \beta_{\mathrm{R}}(V_{\mathrm{R}}(x_t)) > 0$. Hence $x_{t+1} \in \Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}}$ by (22) and (26). Together with $x_t \in \Omega^*_{\mathrm{S}}$ case, induction yields recursive feasibility and $x_t \in \Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}} \subseteq \Omega^*_{\mathrm{sRAS}}$, $\forall t \geq 0$.

Finally, whenever $x_t \in \Omega^*_{\mathrm{RA}} \setminus (\Omega^*_{\mathrm{S}} \cup \mathcal{Z}_{\mathrm{R}})$, (22) and (26) give $V_{\mathrm{R}}(x_t) > 0$, so (21) gives $g_{\mathrm{R}}(x_t) > 0$. Since (18) gives $\ell_{\mathrm{R}}(x_t) < 0$, (17b) implies $g(x_t) > 0$ and $\ell(x_t) < 0$. Therefore, if x_t reaches Ω^*_{S}, it satisfies Definition 1. □

Corollary 2 (Almost-Everywhere Maximality)**.** *If $\mathcal{Z}_{\mathrm{R}}$ has Lebesgue measure zero, then, by Proposition 1 and Corollary 1, the guarantees of Theorem 2 apply to almost every initial state in Ω^*_{sRAS} and any task policy against every admissible disturbance policy.*

Corollary 3 (Conservative sRAS Feasibility Preservation)**.** *Without the measure-zero condition, the filter still preserves sRAS feasibility on $\Omega^*_{\mathrm{RA}} \setminus \mathcal{Z}_{\mathrm{R}} \subseteq \Omega^*_{\mathrm{sRAS}}$, although this certified domain may be conservative relative to Ω^*_{sRAS}.*

VI. Synthesis and Deployment of sRAS Q-CBF

The sRAS Q-CBF filter in Definition 8 requires 1) access to V_{R}, V_{S} and their lifted forms $\mathcal{Q}_{\mathrm{R}}$, $\mathcal{Q}_{\mathrm{S}}$, and 2) solving an inner minimization over disturbances. We address both requirements via game-theoretic RL [3, 10, 24]: we sequentially learn the values and the corresponding fallbacks, and then train the best-response disturbances for runtime filtering.

A. Sequential Value Synthesis

For each mode $m \in \{\mathrm{S}, \mathrm{R}\}$, we jointly train a neural critic $\mathcal{Q}_{\omega_m}$, a best-effort fallback $\pi^u_{\theta_m}(\cdot \mid x)$, and an adversarial disturbance $\pi^d_{\psi_m}(\cdot \mid x, u)$. The disturbance retains an instantaneous informational advantage by observing and reacting to the control, following the $\max_u \min_d$ game structure. Writing the actors' deployment outputs as $\pi^u_{\theta_m}(x)$ and $\pi^d_{\psi_m}(x, u)$, these networks provide the value estimate

$$\widehat{V}_m(x) := \mathcal{Q}_{\omega_m}\Big(x, \pi^u_{\theta_m}(x), \pi^d_{\psi_m}\big(x, \pi^u_{\theta_m}(x)\big)\Big). \tag{36}$$

The learned fallback approximates (13) or (25), while the disturbance approximates a locally worst-case response to it.

Using transitions (x, u, d, x') collected through *black-box queries* of (1) and sampled from a replay buffer $\mathcal{B}_m$, we train the critic against the *discounted* Isaacs target [3, 10, 24]:

$$\mathcal{L}^Q_m(\omega_m, \theta_m, \psi_m) := \mathbb{E}\Big[\big(\mathcal{Q}_{\omega_m}(x, u, d) - y_m\big)^2\Big], \tag{37a}$$

$$y_{\mathrm{S}} := (1-\gamma) g_{\mathrm{S}}(x) + \gamma \min\{g_{\mathrm{S}}(x), q'_{\mathrm{S}}\}, \tag{37b}$$

$$y_{\mathrm{R}} := (1-\gamma) \min\{\hat{\ell}_{\mathrm{R}}(x), \hat{g}_{\mathrm{R}}(x)\} + \gamma \min\{\hat{g}_{\mathrm{R}}(x), \max\{\hat{\ell}_{\mathrm{R}}(x), q'_{\mathrm{R}}\}\}, \tag{37c}$$

where $\gamma \in (0,1)$ is the discount factor and $q'_m := \mathcal{Q}_{\omega'_m}(x', u', d')$ is the output of a slowly updated target critic, with $u' \sim \pi^u_{\theta_m}(\cdot \mid x')$ and $d' \sim \pi^d_{\psi_m}(\cdot \mid x', u')$. We first train for the stay mode, then freeze $\widehat{V}_{\mathrm{S}}$ and train for the reach mode using $\hat{\ell}_{\mathrm{R}} := \widehat{V}_{\mathrm{S}}$ and $\hat{g}_{\mathrm{R}} := \min\{g, \max\{-\ell, \hat{\ell}_{\mathrm{R}}\}\}$.

For each mode, we train the fallback to maximize the critic and the disturbance to minimize it, with entropy regularization encouraging exploration for both actors. All updates use sampled transitions and neural network gradients, without differentiating through the dynamics. In doing so, we adopt gradient descent-ascent (GDA) with *finite timescale separation*: ψ_m is updated on a faster timescale than θ_m, allowing $\pi^d_{\psi_m}$ to track a locally best-effort worst-case response to the slowly evolving fallback $\pi^u_{\theta_m}$. The underlying game-theoretic RL framework guarantees local convergence to a minimax equilibrium in the policy space [24].

B. Universal Worst-Case Response Disturbance Policy

While $\pi^d_{\psi_m}$ approximates a locally worst-case response to $\pi^u_{\theta_m}$, it need not do so for the *filtered policy* from Definition 8. For each mode, we therefore freeze ω_m and further train a copy of $\pi^d_{\psi_m}$ to minimize the critic $\mathcal{Q}_{\omega_m}$ over state–control pairs sampled using a family of diverse control policies. The resulting *universal best-effort worst-case response disturbance policy*, denoted by $\pi^d_{\tilde{\psi}_m}(\cdot \mid x, u)$, is trained to approximate the pointwise minimizer of $\mathcal{Q}_{\omega_m}(x, u, \cdot)$ across states and candidate controls, providing a learned surrogate for the inner disturbance minimization. Using the deployment output of $\pi^d_{\tilde{\psi}_m}$, we replace the constraint in Definition 8 with

$$\mathcal{Q}_{\omega_m}\big(x, u, \pi^d_{\tilde{\psi}_m}(x, u)\big) \geq \beta_m\big(\widehat{V}_m(x)\big). \tag{38}$$

Evaluating (38) requires only neural network evaluations, enabling runtime filtering on black-box systems without nested disturbance optimization, known dynamics, control- or disturbance-affine structure, or value derivatives.

Remark 1 (Post-hoc Verification)**.** *Guarantees in Section V do not automatically extend to learned values and policies. Offline conformal prediction offers probabilistic guarantees for learned reachable sets [25] and safety filters [26].*

VII. Experimental Results

In this section, we evaluate the proposed sRAS filter on two robotic tasks: quadruped gap jumping in simulation and hardware, and simulated F1TENTH endurance racing.

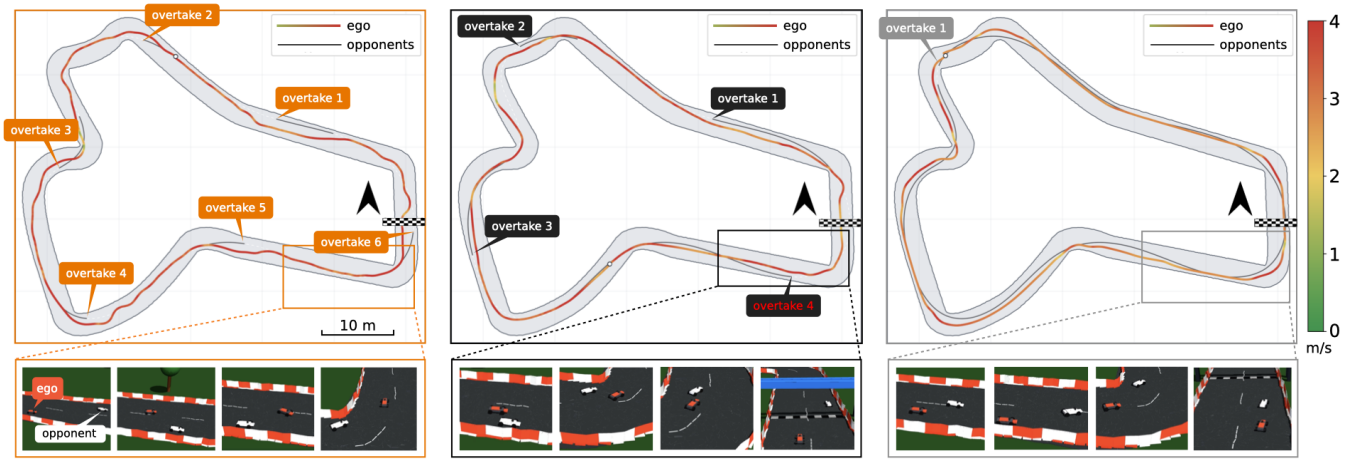

Fig. 2. Representative F1TENTH races, with ego trajectories colored by speed. Our **sRAS** filter (left) maintains safety while enabling successful overtakes and lead retention, whereas **RA** (middle) can lose the lead after initially getting ahead and **Avoid-only** (right) remains blocked behind a slower opponent.

Method	Safe Rate [%] ↑	Overtakes/Race ↑	2-Lap Time [s] ↓
Avoid-only	**98.2** [**97.2**, **98.9**]	1.00 ± 0.03	111.40 ± 3.32
RA	56.8 [53.7, 59.8]	3.59 ± 0.90	116.43 ± 8.74
sRAS	**98.0** [**96.9**, **98.7**]‡	$\mathbf{9.63 \pm 0.99}^{\dagger\ddagger}$	$\mathbf{97.37 \pm 2.87}^{\dagger\ddagger}$

TABLE I: F1TENTH results from 1000 simulated races per filter. Our **sRAS** filter achieves a 98.0% safe rate with significantly more overtakes per race and shorter 2-lap times than both baselines. Safe rates include 95% Wilson confidence intervals; other metrics show mean $\pm$ standard deviation. For **sRAS**, $\dagger$ and $\ddagger$ indicate significant differences ($p < 0.01$) from **Avoid-only** and **RA**, respectively.

Method	Safe Rate [%] ↑	Cross Rate [%] ↑	Success Rate [%] ↑
Avoid-only	85.6 [83.3, 87.6]	0.8 [0.4, 1.6]	0.7 [0.3, 1.4]
RA	63.4 [60.4, 66.3]	87.4 [85.2, 89.3]	63.3 [60.3, 66.2]
sRAS	**86.9** [84.7, 88.9]‡	**89.7** [87.7, 91.4]†	**86.9** [84.7, 88.9]†‡

TABLE II: Go2 gap-jumping results from 1000 simulated episodes per filter. Our **sRAS** filter achieves an 86.9% success rate (percentage of episodes with safe crossing, safe landing, and continued safety), significantly higher than both baselines. All metrics show 95% Wilson confidence intervals. For **sRAS**, $\dagger$ and $\ddagger$ denote significant differences ($p < 0.01$) from **Avoid-only** and **RA**, respectively.

Baselines. We compare against Avoid-only and reach–avoid (RA) baselines. The Avoid-only filter uses an avoid-only value trained with safety margin g, whereas the RA filter uses a reach–avoid value trained with safety margin g and target margin ℓ. Both baseline filters apply robust Q-CBF interventions [3]. These comparisons assess the benefits of encoding target reaching and safe retention after first entry.

A. F1TENTH Endurance Race

We first evaluate our sRAS filter in a high-fidelity F1TENTH simulator without disturbances (singleton $\mathcal{D}$), built in MuJoCo [27] based on `F1TENTH Gym` [28]. The ego must race for two laps, safely overtake as many opponents as possible, and minimize completion time. An overtake counts as successful only after the ego gets ahead and retains the lead for 5 s, after which the opponent is replaced by a new one spawned at a random location ahead of the ego.

Setup. The safety margin $g(x) := \min\{g_{\text{wall}}(x), g_{\text{oppo}}(x)\}$ is the minimum of the ego's signed clearance margins to the track boundary and opponent. The target margin $\ell(x) := CP_{\text{ego}} - CP_{\text{opp}} - 0.58\,\text{m}$ is the difference in unwrapped centerline progress in meters, with a one-vehicle-length lead buffer. We use a fixed model predictive path integral (MPPI) policy [29] as the task policy for both ego and opponent, with an additional filter applied to the ego. The common 90-dimensional base observation includes normalized LiDAR measurements, ego velocity and yaw rate, previous controls, and ego-centric opponent features. The sRAS filter additionally observes the normalized target margin in both modes and the frozen stay value in reach mode. The control inputs are commanded steering angle and speed.

Evaluation and Metrics. We evaluate all filters on the same 1000 random seeds. After a collision, the ego is held at the impact position for 3 s before respawning at the nearest track centerline point. We report *safe rate* (percentage of races completing two laps without collision), *mean successful overtakes per race*, and *mean two-lap completion time*.

Results. Table I reports the results, with significance assessed using exact McNemar tests for safe rate and paired t-tests for other metrics (Holm-adjusted $p < 0.01$). The sRAS filter achieves a high safe rate similar to Avoid-only and significantly higher than RA. It also achieves significantly more successful overtakes and significantly shorter two-lap completion times than both baselines. As Figure 2 illustrates, sRAS repeatedly passes opponents and retains the lead, whereas Avoid-only does not encode passing and remains blocked behind a slower opponent. RA does not require safe lead retention, so it can collide after briefly getting ahead or lose the lead and enter battles that slow down both vehicles.

B. Quadruped Gap Jumping

We next evaluate the sRAS filter on a 36-D Unitree Go2 quadruped in hardware and in MuJoCo simulation [27]. The robot must jump across a 25 cm gap, land safely on the far side, and remain safe afterward. To test robustness against adversarial uncertainty realizations in simulation, we apply external forces up to 30 N in arbitrary directions at arbitrary torso locations, selected by the disturbance policy trained following Subsection VI-B. Accounting for these disturbances also aims to mitigate the sim-to-real gap.

Setup. The safety margin $g(x)$ is the minimum of normalized margins for trunk clearance above local terrain, body tilt, and non-foot contact force. The target margin $\ell(x)$ combines progress beyond the gap, foot-to-terrain proximity, uprightness, and a planar-speed margin modulated by foot proximity to encode an upright arrival near the ground on the far side. In both simulation and hardware, we use a pure-pursuit walking task policy toward the far side of the gap. The observation consists of a 5-frame history of body-frame angular velocity, projected gravity vector, velocity command, gait phase, joint positions and velocities, and a 17×11 local height scan. The control inputs are 12 joint-position targets relative to a nominal standing pose, tracked by a PD controller at $50\,\mathrm{Hz}$.

Evaluation and Metrics. We evaluate each filter in 1000 simulation episodes. We report *safe rate* (percentage avoiding failure), *cross rate* (percentage reaching the far side, regardless of landing safety), and *success rate* (percentage crossing and landing safely, then remaining safe until episode end).

Simulation Results. Table II reports simulation results with significance assessed using exact McNemar tests (Holm-adjusted $p < 0.01$). The Avoid-only filter does not encode crossing and rarely completes the jump. Although RA often reaches the far side, it does not require continued safety after target entry, leaving the robot vulnerable to failure during landing or subsequent stabilization. Figure 1 illustrates such failure modes. Our sRAS filter modifies actions from the walking policy to preserve the feasibility of safe crossing, landing, and continued safety afterward, and achieves a significantly higher success rate than both baselines.

Hardware Results. Figure 1 (top) shows the sRAS-filtered Go2 reproducing the successful behavior observed in simulation: crossing, landing safely, and remaining safe afterward.

VIII. Conclusion

We presented an sRAS Q-CBF safety filter for high-dimensional black-box systems under bounded uncertainty. For exact values and under a measure-zero condition, the filter is recursively feasible, preserves sRAS feasibility from almost every winnable initial state, and keeps the system safely within the target after first entry against all admissible uncertainty realizations. Reachability-based adversarial RL and Q-function lifting enable scalable synthesis and deployment without known dynamics, affine structure, value derivatives, or hand-designed barriers. We validated our method on two robotic tasks: quadruped gap jumping in simulation and hardware, and simulated F1TENTH racing. The filter enabled safe crossing, landing, and continued safety for the quadruped, and safe overtaking and lead retention for the F1TENTH race, demonstrating the intended sRAS behavior.

Acknowledgment

Claude (Anthropic) and Codex (OpenAI) were used to generate code for the experiments in Section VII.

References

[1] K.-C. Hsu, H. Hu, and J. F. Fisac. "The safety filter: A unified view of safety-critical control in autonomous systems". *Annual Review of Control, Robotics, and Autonomous Systems* 7 (2024), pp. 47–72.

[2] D. D. Oh, J. Lidard, H. Hu, et al. "Safety with agency: Human-centered safety filter with application to AI-assisted motorsports". *Robotics: Science and Systems (RSS)*. 2025.

[3] D. D. Oh, D. P. Nguyen, H. Hu, and J. F. Fisac. "Synthesis and deployment of maximal robust control barrier functions through adversarial reinforcement learning". *IEEE Conference on Decision and Control (CDC)*. IEEE. 2026.

[4] L. Lindemann and D. V. Dimarogonas. "Control barrier functions for signal temporal logic tasks". *IEEE Control Systems Letters* 3.1 (2018), pp. 96–101.

[5] R. Das, P. Bakshi, and P. Jagtap. "Control barrier functions for prescribed-time reach-avoid-stay tasks using spatiotemporal tubes". *European Control Conference (ECC)*. IEEE. 2025, pp. 398–403.

[6] R. Das, S. Choudhury, and P. Jagtap. "Control barrier functions for the full class of signal temporal logic tasks using spatiotemporal tubes". 2025. arXiv: 2510.19595.

[7] Y. Meng, Y. Li, and J. Liu. "Control of nonlinear systems with reach-avoid-stay specifications: A Lyapunov-barrier approach with an application to the Moore-Greizer model". *American Control Conference (ACC)*. IEEE. 2021, pp. 2284–2291.

[8] R. Zhou, Y. Yuan, H. Chang, et al. "Patching control Lyapunov barrier functions for temporal logic specifications with bounded controls". 2026. arXiv: 2606.12768.

[9] K.-C. Hsu, V. Rubies-Royo, C. J. Tomlin, and J. F. Fisac. "Safety and liveness guarantees through reach-avoid reinforcement learning". *Robotics: Science and Systems (RSS)*. 2021.

[10] D. P. Nguyen, K.-C. Hsu, W. Yu, et al. "Gameplay filters: Robust zero-shot safety through adversarial imagination". *Conference on Robot Learning*. PMLR. 2025, pp. 387–407.

[11] G. Chenevert, J. Li, A. Kannan, et al. "Deep reinforcement learning for reach-avoid-stay problems". 2026. arXiv: 2410.02898.

[12] W. Sharpless, D. Hirsch, S. Tonkens, et al. "Dual-objective reinforcement learning with novel Hamilton-Jacobi-Bellman formulations". *International Conference on Learning Representations*. Vol. 2026. 2026, pp. 138080–138121.

[13] W. Sharpless, O. So, D. Hirsch, et al. "Bellman value decomposition for task logic in safe optimal control". 2026. arXiv: 2602.19532.

[14] O. So, W. Sharpless, S. Herbert, and C. Fan. "Value functions for temporal logic: Optimal policies and safety filters". 2026. arXiv: 2605.01051.

[15] J. J. Choi, D. Lee, K. Sreenath, et al. "Robust control barrier–value functions for safety-critical control". *IEEE Conference on Decision and Control (CDC)*. IEEE. 2021, pp. 6814–6821.

[16] D. Hirsch, J. F. Fisac, and S. Herbert. "Viscosity CBFs: Bridging the control barrier function and Hamilton-Jacobi reachability frameworks in safe control theory". *American Control Conference (ACC)*. IEEE. 2026, pp. 593–600.

[17] A. Begzadic, N. Shinde, S. Tonkens, et al. "Back to base: Towards hands-off learning via safe resets with reach-avoid safety filters". *Learning for Dynamics and Control Conference*. Vol. 283. Proceedings of Machine Learning Research. PMLR, 2025, pp. 1154–1166.

[18] M. Chen, Q. Tam, S. C. Livingston, and M. Pavone. "Signal temporal logic meets reachability: Connections and applications". *International Workshop on the Algorithmic Foundations of Robotics*. Springer. 2018, pp. 581–601.

[19] Y. Gao, A. Abate, F. J. Jiang, et al. "Temporal logic trees for model checking and control synthesis of uncertain discrete-time systems". *IEEE Transactions on Automatic Control* 67.10 (2021), pp. 5071–5086.

[20] J. F. Fisac, N. F. Lugovoy, V. Rubies-Royo, et al. "Bridging Hamilton-Jacobi safety analysis and reinforcement learning". *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2019, pp. 8550–8556.

[21] C. J. C. H. Watkins and P. Dayan. "Q-learning". *Machine Learning* 8 (1992), pp. 279–292.

[22] D. Liberzon and A. S. Morse. "Basic problems in stability and design of switched systems". *IEEE Control Systems Magazine* 19.5 (1999), pp. 59–70.

[23] L. Lindemann, H. Hu, A. Robey, et al. "Learning hybrid control barrier functions from data". *Conference on Robot Learning*. PMLR. 2021, pp. 1351–1370.

[24] J. S. Wang, H. Hu, D. P. Nguyen, and J. F. Fisac. "MAGICS: Adversarial RL with minimax actors guided by implicit critic Stackelberg for convergent neural synthesis of robot safety". *International*

Workshop on the Algorithmic Foundations of Robotics. Springer. 2024, pp. 459–480.

[25] A. Lin and S. Bansal. "Verification of neural reachable tubes via scenario optimization and conformal prediction". *Annual Learning for Dynamics & Control Conference*. PMLR. 2024, pp. 719–731.

[26] H. Hu. "Permissive safety through trusted inference: Verifiable belief-space neural safety filters for assured interactive robotics". *World Symposium on the Algorithmic Foundations of Robotics*. 2026.

[27] E. Todorov, T. Erez, and Y. Tassa. "MuJoCo: A physics engine for model-based control". *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2012, pp. 5026–5033.

[28] M. O'Kelly, H. Zheng, D. Karthik, and R. Mangharam. "F1TENTH: An open-source evaluation environment for continuous control and reinforcement learning". *NeurIPS 2019 Competition and Demonstration Track*. Vol. 123. PMLR, 2020, pp. 77–89.

[29] G. Williams, A. Aldrich, and E. A. Theodorou. "Model predictive path integral control: From theory to parallel computation". *Journal of Guidance, Control, and Dynamics* 40.2 (2017), pp. 344–357.