VeRecycle: Reclaiming Guarantees from Probabilistic Certificates for Stochastic Dynamical Systems after Change

Lutz, S.; Lukina, A.; Spaan, M.T.J.

DOI
10.24963/ijcai.2025/52

Publication date
2025

Document Version
Final published version

Published in
Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence

VeRecycle: Reclaiming Guarantees from Probabilistic Certificates for Stochastic Dynamical Systems after Change

Sterre Lutz , **Matthijs T.J. Spaan** and **Anna Lukina**
Delft University of Technology, The Netherlands
{s.lutz, m.t.j.spaan, a.lukina}@tudelft.nl

Abstract

Autonomous systems operating in the real world encounter a range of uncertainties. Probabilistic neural Lyapunov certification is a powerful approach to proving safety of nonlinear stochastic dynamical systems. When faced with changes beyond the modeled uncertainties, e.g., unidentified obstacles, probabilistic certificates must be transferred to the new system dynamics. However, even when the changes are localized in a known part of the state space, state-of-the-art requires complete re-certification, which is particularly costly for neural certificates. We introduce VeRecycle, the first framework to formally reclaim guarantees for discrete-time stochastic dynamical systems. VeRecycle efficiently reuses probabilistic certificates when the system dynamics deviate only in a given subset of states. We present a general theoretical justification and algorithmic implementation. Our experimental evaluation shows scenarios where VeRecycle both saves significant computational effort and achieves competitive probabilistic guarantees in compositional neural control.

Code — https://github.com/SUMI-lab/VeRecycle

Extended version — https://doi.org/10.48550/arXiv.2505.14001

1 Introduction

Autonomous systems often exhibit nonlinear stochastic behavior. In safety-critical applications (e.g., robot navigation), it is crucial to provide provable guarantees for autonomous systems, both in terms of performance (e.g., reaching a target) and safety (e.g., avoiding material damage) [Kwiatkowska and Zhang, 2023]. The latter is particularly challenging to guarantee for the multitude of uncertainties of the real world. For instance, a spill of water may violate the friction threshold for a robot to move at the required safe speed.

Designing safe control for a stochastic nonlinear dynamical system, as a general model for many real-world autonomous systems, is enabled by the Lyapunov theory [Kalman and Bertram, 1959]. Initially developed for proving the stability of deterministic dynamical systems, Lyapunov theory prescribes the construction of a potential-energy function, which characterizes system dynamics and, when obeying a set of given conditions, becomes a formal certificate for the desired specification. Reaching the target while avoiding unsafe states can be naturally formulated as Lyapunov stability: the system's potential energy decreases towards the target region and increases around unsafe regions. Lyapunov certification extends to stochastic dynamical systems [Blumenthal and Getoor, 1968] and to proving reach-avoid properties thereof [Prajna *et al.*, 2004].

Overwhelming success of neural networks in visual recognition and especially feasibility of their verification [Katz *et al.*, 2017; Gehr *et al.*, 2018; Zhang *et al.*, 2018; Kouvaros and Lomuscio, 2021; Wu *et al.*, 2024] galvanized the research community into adoption of neural components in control design. Verification of dynamical systems under neural control called for an alternative to sum-of-square programming [Papachristodoulou and Prajna, 2002; Topcu *et al.*, 2008], limited to polynomial dynamics. Neural Lyapunov certification proved successful for deterministic systems [Richards *et al.*, 2018; Chang *et al.*, 2019; Abate *et al.*, 2021; Edwards *et al.*, 2023; Dawson *et al.*, 2023]. Recently, inspired by termination analysis of stochastic programs [Giacobbe *et al.*, 2022], probabilistic neural Lyapunov certification solidified its place as the state-of-the-art for reach-avoid properties of stochastic dynamical systems under neural control [Žikelić *et al.*, 2023a; Ansaripour *et al.*, 2023; Chatterjee *et al.*, 2023; Mathiesen *et al.*, 2023; Abate *et al.*, 2024; Badings *et al.*, 2024].

For a given model of dynamics and a control policy, state-of-the-art probabilistic reach-avoid certificates guarantee that with probability above a given threshold, the system will reach a target set of states while avoiding unsafe states. However, in scenarios where the system dynamics deviate from their stochastic model, even if only in a known subset of system states, the originally computed certificate requires revalidation to transfer its guarantee. Yet, there exists no principled way to infer if the changes in the dynamics affect the guarantees or to reuse the corresponding certificate, without repeating costly (neural) certification.

Example 1. *A robot navigates a warehouse with slippery floors, meaning the result of the robot's actions is probabilistic. To safely reach room 8 starting from room 0 (dotted squares in Figure 1 mark the centers of the rooms), the robot is controlled by a composition of neural policies, including*

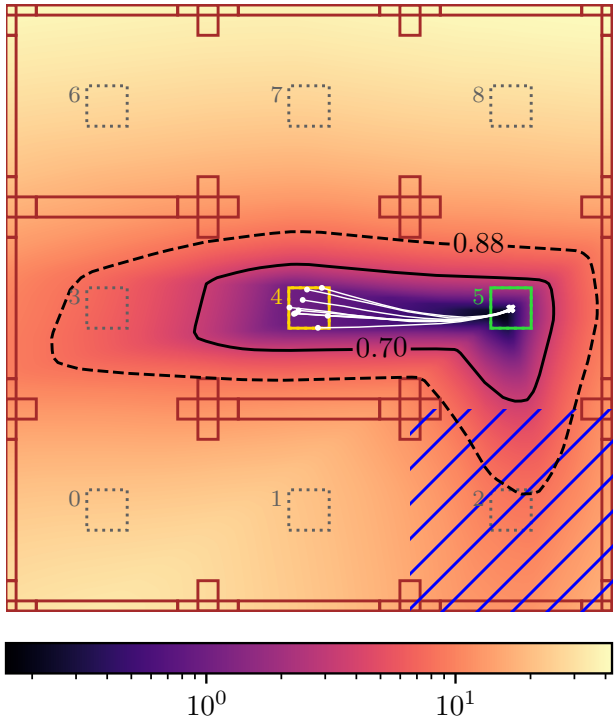

Figure 1: A probabilistic certificate (color gradient) for a navigation subtask 4 □ (yellow) → 5 □ (green), with darker color indicating higher probability of the system (with trajectories sampled in white) to complete the subtask while avoiding the walls (outlined in red). Dashed line shows original certified probability threshold, solid line shows threshold reclaimed by VeRecycle after a localized change in the system's dynamics (slanted blue pattern).

one guiding it from room 4 to 5, avoiding collision with the walls. Figure 1 visualizes a probabilistic neural certificate, which guarantees that starting anywhere close to the center of room 4, the robot will safely reach the center of room 5 with probability at least 0.88*. Suppose that an unidentified obstacle appeared in room 2 and the whole room is now deemed unsafe. Although the obstacle is not interfering with the given subtask, visual inspection shows the certificate is now invalid (dashed line crosses the area around the obstacle).*

In this work, we propose the first, to the best of our knowledge, theoretical justification for reclaiming probabilistic reach-avoid guarantees for neural control policies of discrete-time stochastic dynamical systems affected by a localized change, i.e., for which the dynamics changed in a given subset of the state space. In that, we assume no knowledge about the nature of the change.

Our approach, called VeRecycle, takes as input a neural control policy and a probabilistic neural reach-avoid certificate with a particular probability threshold. For an arbitrary, possibly unknown, change in the system dynamics affecting a known subset of the state space, VeRecycle automatically infers a new certified reach-avoid probability threshold.

In Example 1, VeRecycle quickly identifies the affected certificate and updates its threshold to 0.70 (solid line has no intersection with the unsafe area), without any additional neural training. This enables further use of existing neural policies and reclaiming the global probabilistic guarantee.

VeRecycle magnifies the advantages of tackling the problems with innate modular structure through the prism of localized analysis, for which compositional approaches to control and—with VeRecycle—certification demonstrate greater scalability [Neary *et al.*, 2023; Žikelić *et al.*, 2023b; Zhang *et al.*, 2023; Delgrange *et al.*, 2024]. In compositional control, our approach 1) automatically identifies which components of the compositional policy are affected by the change and 2) eliminates—when possible—the need for costly repairs of neural control policies, certificates, or system models by using unaffected components in a new compositional policy.

To summarize, we propose VeRecycle, a framework to formally reclaim guarantees from probabilistic neural certificates for discrete-time stochastic dynamical systems after localized changes. First, we formally prove that sound guarantees are inferred by VeRecycle from (approximated) infima of the original certificates. Second, our experimental evaluation demonstrates scenarios where VeRecycle's guarantees are competitive—particularly for compositional control—compared to complete re-certification, with computational cost reduced by three orders of magnitude. Thus, VeRecycle enables higher flexibility and scalability of the state-of-the-art probabilistic reach-avoid certification of discrete-time stochastic dynamical systems.

2 Preliminaries

We consider discrete-time stochastic dynamical systems $\Sigma = (\mathbb{X}, \mathbb{U}, \mathbb{W}, f, \mu)$, defined by the following equation $\forall t \in \mathbb{N}_{>0}$:

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t), \quad \mathbf{w}_t \sim \mu, \tag{1}$$

where $\mathbf{x}_t \in \mathbb{X} \subseteq \mathbb{R}^n$ is the n-dimensional state of Σ at time t with $\mathbf{x}_0$ as the initial state, $\mathbf{u}_t \in \mathbb{U} \subseteq \mathbb{R}^m$ is the m-dimensional control input Σ at time t, and $\mathbf{w}_t \in \mathbb{W} \subseteq \mathbb{R}^p$ is the p-dimensional realization of stochastic disturbance of Σ at time t. Transitions between states follow the dynamics function $f : \mathbb{X} \times \mathbb{U} \times \mathbb{W} \to \mathbb{X}$, and the probability distribution μ over $\mathbb{W}$, from which $\mathbf{w}_t$ is sampled independently at each time step. When the control inputs are determined by a policy $\pi : \mathbb{X} \to \mathbb{U}$, i.e., $\mathbf{u}_t := \pi(\mathbf{x}_t)$, Σ is said to be *controlled* by π.

An infinite sequence of state, action, and disturbance triples $(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t)_{t \in \mathbb{N}_0}$ is a *trajectory* if for all time steps $t \in \mathbb{N}_{>0}$, it holds that $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t)$, $\mathbf{u}_t = \pi(\mathbf{x}_t)$, and $\mathbf{w}_t \in \mathrm{support}(\mu)$. Given a fixed initial state $\mathbf{x}_0$, Markov decision process (MDP) semantics establish a probability space over all such trajectories [Puterman, 2014]. The probability measure and expectation within this probability space are denoted $\mathbb{P}^{\Sigma,\pi}_{\mathbf{x}_0}$ and $\mathbb{E}^{\Sigma,\pi}_{\mathbf{x}_0}$, respectively.

For the system semantics and probability measures to be well-defined, we assume $\mathbb{X}$, $\mathbb{U}$, and $\mathbb{W}$ to be Borel-measurable. Moreover, we follow the standard assumptions from control theory that f and π are Lipschitz continuous, and that $\mathbb{X}$ and $\mathbb{W}$ are closed and bounded.

2.1 Reach-avoid Verification

A common way to quantify the performance and safety of control policies for stochastic dynamical systems is via prob-

abilistic reach-avoid specifications, which enforce a threshold on the probability of reaching a target state without encountering unsafe states.

Definition 1. *A* reach-avoid specification *is a tuple* $(\mathbb{X}_\star, \mathbb{X}_\oslash, \mathbb{X}_0, \rho)$*, where* $\mathbb{X}_\star, \mathbb{X}_\oslash, \mathbb{X}_0 \subseteq \mathbb{X}$ *are target states, unsafe states, and initial states, respectively, and* $\rho \in [0,1]$ *is a probability threshold, i.e., minimally acceptable probability with which the specification holds. A system* Σ *controlled by a policy* π *satisfies a reach-avoid specification, denoted* $\Sigma, \pi \vDash (\mathbb{X}_\star, \mathbb{X}_\oslash, \mathbb{X}_0, \rho)$*, if and only if, for all* $\mathbf{x}_0 \in \mathbb{X}_0$*,*

$$\mathbb{P}^{\Sigma,\pi}_{\mathbf{x}_0}\{\exists \tau \in \mathbb{N}_0 : \mathbf{x}_\tau \in \mathbb{X}_\star \wedge (\forall t < \tau \in \mathbb{N}_0 : \mathbf{x}_t \notin \mathbb{X}_\oslash)\} \geq \rho.$$

Example 2. *Consider the original system described in Example 1, i.e., before water was spilled. This stochastic dynamical system* Σ *has a two-dimensional state space* $\mathbb{X} = [0,3]^2$*, the two features being x and y coordinates, respectively. The navigation task of safely moving from room 4 to 5 (yellow and green in Figure 1 respectively) with at least a probability of* 0.88*, is a reach-avoid specification* φ *where* $\mathbb{X}_\star = [2.4, 2.6] \times [1.4, 1.6]$*,* $\mathbb{X}_\oslash$ *is the union of the walls,* $\mathbb{X}_0 = [1.4, 1.6]^2$*, and* $\rho = 0.88$*.*

To verify that a reach-avoid specification holds for a system under a particular policy, one can construct probabilistic reach-avoid certificates, whose existence serves as proof.

Definition 2 ([Žikelić *et al.*, 2023a])**.** *A continuous function* $C : \mathbb{X} \to \mathbb{R}_{\geq 0}$ *is a* reach-avoid certificate *for a system* Σ*, policy* π*, and reach-avoid specification* $\varphi = (\mathbb{X}_\star, \mathbb{X}_\oslash, \mathbb{X}_0, \rho)$*, if and only if the following conditions hold:*

1. *Initial condition: for all* $\mathbf{x} \in \mathbb{X}_0$*,* $C(\mathbf{x}) \leq 1$*;*
2. *Safety condition: for all* $\mathbf{x} \in \mathbb{X}_\oslash$*,* $C(\mathbf{x}) \geq \frac{1}{1-\rho}$*;*
3. *Decrease condition: there exists* $\varepsilon > 0$*, such that for all* $\mathbf{x} \in \mathbb{X} \backslash \mathbb{X}_\star$ *either* $\mathbb{E}_{\mathbf{w}\sim\mu}[C(\mathbf{x}) - C(f(\mathbf{x}, \pi(\mathbf{x}), \mathbf{w}))] \geq \varepsilon$ *or* $C(\mathbf{x}) \geq \frac{1}{1-\rho}$*.*

The existence of a probabilistic reach-avoid certificate proves (i.e., certifies*) that* $\Sigma, \pi \vDash \varphi$*. We use the shorthand* $C(S) := \{C(\mathbf{x}) \mid \mathbf{x} \in S\}$ *for the image of a subset* $S \subseteq \mathbb{X}$ *under* C*.*

The z axis (color gradient) in Figure 1 shows an example of a reach-avoid certificate for the specification φ and system Σ defined in Example 2. Intuitively, the third condition requires that, in any state, Σ is unlikely to transition to a state with a higher value in C, with larger differences being increasingly improbable. Consequently, the first and second conditions establish that from the initial states (with values ≤ 1), it is rare to reach the unsafe states (with values $\geq \frac{1}{1-\rho}$). Furthermore, Σ is likely to progress toward the lowest-valued part of the certificate. Since only the target area is permitted to violate the decrease condition and fall below the safety value $\frac{1}{1-\rho}$, the certificate takes its lowest value in the target area, to which the state of the system eventually converges.

Constructing probabilistic reach-avoid certificates analytically is challenging, especially for nonlinear system dynamics or neural policies. Thus, the state-of-the-art is a counterexample guided synthesis procedure that outputs a certificate in the form of a neural network—a *neural certificate* [Žikelić *et al.*, 2023a; Chatterjee *et al.*, 2023; Badings *et al.*, 2024]. The procedure loops between learning and verification until a correct certificate is found or the allotted time is exceeded; in the latter case, the result is inconclusive.

The learning procedure optimizes the candidate certificate on differentiable versions of conditions 1–3 from Definition 2 for a set of randomly sampled states and a set of counterexamples produced by the verification procedure. The verification procedure discretizes the continuous state space into "cells", verifying conditions 1–3 soundly for all states within the cells using Lipschitz constants and interval bound propagation (IBP) [Mirman *et al.*, 2018; Gowal *et al.*, 2019]. If the verification is inconclusive for a cell, i.e., when the bounds given by IBP and Lipschitz reasoning are not tight enough to prove or disprove the conditions for all states in the cell, the cell is split into smaller cells to attempt verification again. If a condition is certainly violated for a cell, the verification procedure feeds all states in the cell back into the learning procedure as counterexamples for further optimization.

3 Problem Statement

We formulate the general problem of reclaiming guarantees from probabilistic reach-avoid certificates after a change in system dynamics. Our problem formulation specifically addresses changes that affect the system dynamics only in a known subset of states $\mathbb{X}_?$. Importantly, this definition does not require additional knowledge of the new dynamics. Instead, the aim is to determine guarantees that are valid for any alternative dynamics function, as long as anywhere outside $\mathbb{X}_?$, the function is equivalent to the original dynamics.

This formulation captures a wide range of changes, such as the detection of unidentified obstacles, the appearance of materials with unknown friction coefficients within a specific region, or the onset of unexplained stochastic phenomena occurring exclusively within a certain velocity range.

Formally, we consider a stochastic dynamical system $\Sigma = (\mathbb{X}, \mathbb{U}, \mathbb{W}, f, \mu)$ controlled by a policy π, that satisfies a reach-avoid specification $\varphi = (\mathbb{X}_\star, \mathbb{X}_\oslash, \mathbb{X}_0, \rho)$, and a reach-avoid certificate C proving that relation, i.e., C certifies $\Sigma, \pi \vDash \varphi$.

Definition 3. *Given states* $\mathbb{X}_? \subset \mathbb{X}$*, a threshold* $\rho' \in [0,1]$ *is* reclaimable*, if and only if* C *certifies* $\widetilde{\Sigma}, \pi \vDash (\mathbb{X}_\star, \mathbb{X}_\oslash, \mathbb{X}_0, \rho')$ *for all* $\widetilde{\Sigma} = (\mathbb{X}, \mathbb{U}, \mathbb{W}, \widetilde{f}, \mu)$*, such that*

$$\{\mathbf{x} \in \mathbb{X} \mid \exists \mathbf{u}, \mathbf{w} : \widetilde{f}(\mathbf{x}, \mathbf{u}, \mathbf{w}) \neq f(\mathbf{x}, \mathbf{u}, \mathbf{w})\} \subseteq \mathbb{X}_?. \tag{2}$$

Problem 1. *Given a certificate* C*, original threshold* ρ*, and states* $\mathbb{X}_? \subset \mathbb{X}$*, determine the maximum reclaimable threshold* $\rho' \leq \rho$*, denoted* $\widetilde{\rho}$*.*

Example 3 (Example 2 continued)**.** *The spilled water in the blue hatched area in Figure 1 creates a new, alternative system with dynamics function* $\widetilde{f}$*,*

$$\widetilde{f}(\mathbf{x}, \mathbf{u}, \mathbf{w}) = \begin{cases} ? & \text{if } \mathbf{x} \in [2,3] \times [0,1] \\ f(\mathbf{x}, \mathbf{u}, \mathbf{w}) & \text{otherwise.} \end{cases} \tag{3}$$

The solution to Problem 1 for $\mathbb{X}_? = [2,3] \times [0,1]$ *gives us a threshold* $\widetilde{\rho}$ *on the probability of reaching green without colliding with walls. Whether the consequences of the spilled water are erratic movements, a complete shut-down, or barely noticeable, should not matter for this threshold:* C

should certify $\widetilde{\rho}$ for any possible realization of $\widetilde{f}$ and be the largest threshold $\leq \rho$ for which this holds.

We remark that Problem 1 defines a maximum threshold that is *always* valid, i.e., without any assumptions on the new dynamics function beyond Equation 2, and for a fixed C and π. While higher probability thresholds might be obtainable once the actual new dynamics are specified, such thresholds are not the generally valid thresholds of Problem 1: we are specifically interested in solving the problem for unknown dynamics in a known set of states. Similarly, updating the certificate or policy for the changed dynamics could also result in a higher threshold than $\widetilde{\rho}$. However, since the state-of-the-art certification of nonlinear, stochastic dynamics and/or neural policies is done with neural networks, changing them requires costly additional training and verification.

Thus, for applications where a new model of the dynamics can be obtained and alternative policies and certificates can be trained, the general solution to Problem 1 still serves a practical purpose. If the general threshold already matches the original specification or falls within some acceptable margin, there is no need to update the system model and repeat costly re-certification. Thus, a solution to Problem 1 serves as proof that such updates are truly necessary before spending significant computational resources.

4 VeRecycle

We propose VeRecycle, a framework to reclaim guarantees from reach-avoid certificates given a subset of states $\mathbb{X}_?$ in which dynamics may have changed, as formulated in Problem 1. VeRecycle exploits a relation between the maximum reclaimable probability threshold $\widetilde{\rho}$ and the infimum of a certificate on $\mathbb{X}_?$. First, we formalize this relation and prove that it provides an exact solution to Problem 1. Second, we discuss how VeRecycle reduces computational costs through approximations of infima for neural certificates while still providing soundness guarantees.

4.1 Exact Solution

In this subsection, we prove that an exact solution $\widetilde{\rho}$ to Problem 1 depends only on the original bound ρ and the lowest value the certificate takes for states in $\mathbb{X}_?$: the infimum $\inf\{C(\mathbb{X}_?)\}$. In particular, we show that whenever $\inf\{C(\mathbb{X}_?)\} \geq 1$, $\widetilde{\rho}$ equals

$$\min\left(\rho, 1 - \frac{1}{\inf\{C(\mathbb{X}_?)\}}\right), \tag{4}$$

and that otherwise no reclaimable threshold exists.

We first show that whenever $\inf\{C(\mathbb{X}_?)\} \geq 1$, Equation 4 upper-bounds and lower-bounds $\widetilde{\rho}$ (Lemmas 1 and 2 respectively). Subsequently, we combine these results with a proof of the non-existence of $\widetilde{\rho}$ in case $\inf\{C(\mathbb{X}_?)\} < 1$, completing our proof of the relation.

Lemma 1 (Lower bound of $\widetilde{\rho}$)**.** *If $\inf\{C(\mathbb{X}_?)\} \geq 1$, then Equation 4 lower-bounds the maximum reclaimable threshold $\widetilde{\rho}$ defined in Problem 1.*

Proof sketch (proof in the extended version). We set ρ' to Equation 4 and prove that it is a reclaimable threshold for C. By definition of a maximum, it follows that the maximum of all such reclaimable thresholds, $\widetilde{\rho}$, is lower-bounded by Equation 4.

Threshold ρ' is reclaimable because it is chosen precisely such that for all states in $\mathbb{X}_?$, the certificate value is at least $\frac{1}{1-\rho'}$. Intuitively, this means that the states with unknown dynamics are visited rarely enough that we do not need to have an expected decrease of the certificate's value there: we reach those states with a probability $\leq 1 - \rho'$.

For states outside of the changed states $\mathbb{X}_?$, the expected value of the subsequent states remains unchanged since the dynamics function in $\mathbb{X} \backslash \mathbb{X}_?$ and the certificates' values remain unchanged. Note that even if one of the future subsequent states *is* within $\mathbb{X}_?$ and may therefore have new dynamics, this does not change the expectation in the current state, since the expectation is calculated over one time-step only. Conditions 1 and 2 follow similarly because they were satisfied for the original system and threshold, which concludes the proof of Lemma 1.

Lemma 2 (Upper bound of $\widetilde{\rho}$)**.** *If $\inf\{C(\mathbb{X}_?)\} \geq 1$, then Equation 4 upper-bounds the maximum reclaimable threshold $\widetilde{\rho}$ defined in Problem 1.*

Proof sketch (proof in the extended version). Any reclaimable threshold, including the maximum ($\widetilde{\rho}$), must be a valid threshold for all realizations of $\widetilde{f}$, including fully absorbing dynamics, i.e., where $\widetilde{f}(\mathbf{x}, \mathbf{u}, \mathbf{w}) = \mathbf{x}$, $\forall \mathbf{x} \in \mathbb{X}_?$. For such absorbing dynamics, it is impossible to prove the expected decrease of the certificate, since the expectation always equals the certificate's value in the current state. Thus, by condition 3 of Definition 2, the infimum of the certificate in $\mathbb{X}_?$ must be above $\frac{1}{1-\widetilde{\rho}}$, from which we derive the inequality in Lemma 2.

Lemma 3. *If $\inf\{C(\mathbb{X}_?)\} < 1$, then $\widetilde{\rho}$ does not exist.*

Proof sketch (proof in the extended version). We prove the implication of Lemma 3 by assuming that $\widetilde{\rho}$ exists, and concluding that $\inf\{C(\mathbb{X}_?)\} \geq 1$. If $\widetilde{\rho}$ exists, it must be valid for fully absorbing dynamics: $\widetilde{f}(\mathbf{x}, \mathbf{u}, \mathbf{w}) = \mathbf{x}$, $\forall \mathbf{x} \in \mathbb{X}_?$. Thus, by condition 3 of Definition 2, the infimum of the certificate in $\mathbb{X}_?$ must be above $\frac{1}{1-\widetilde{\rho}}$, from which we derive $\inf\{C(\mathbb{X}_?)\} \geq 1$.

Theorem 1 (Equality to $\widetilde{\rho}$)**.** *The maximum reclaimable threshold $\widetilde{\rho}$ defined in Problem 1 does not exist if $\inf\{C(\mathbb{X}_?)\} < 1$, and otherwise $\widetilde{\rho}$ equals Equation 4.*

Theorem 1 follows directly from Lemmas 1 to 3. A formal proof is given in the extended version.

4.2 Sound Approximation

Theorem 1 proves that the exact solution to Problem 1 only depends on the infimum of the certificate for $\mathbb{X}_?$ and the original certified threshold ρ. In practice, however, finding the infimum of a certificate can be expensive and difficult to scale, particularly for neural certificates. To address this, VeRecycle is equipped to work with approximations of the infimum, while still producing only sound thresholds.

Specifically, methods such as interval bound propagation (IBP) [Gowal *et al.*, 2019] can be used to determine values

I^- and I^+ such that $I^- \leq \inf\{C(\mathbb{X}_?)\} \leq I^+$. With these bounds on the infimum, we can determine guaranteed bounds on the exact solution in Equation 4:

$$\widetilde{\rho}^- = \min\left(\rho, 1 - \frac{1}{I^-}\right); \quad \widetilde{\rho}^+ = \min\left(\rho, 1 - \frac{1}{I^+}\right). \quad (5)$$

Theorem 2. *If $I^- \leq \inf\{C(\mathbb{X}_?)\} \leq I^+$ and $I^- \geq 1$, then $\widetilde{\rho}^- \leq \widetilde{\rho} \leq \widetilde{\rho}^+$.*

Proof sketch (proof in the extended version). We show that the function in Equation 4 is monotonically increasing for $\inf\{C(\mathbb{X}_?)\}$ on the interval $(0, \infty)$. Additionally, if $\inf\{C(\mathbb{X}_?)\} \geq 1$ the formula in Equation 4 is, by Theorem 1, equal to $\widetilde{\rho}$. Thus, if $1 \leq I^- \leq \inf\{C(\mathbb{X}_?)\} \leq I^+$, it follows that $\widetilde{\rho}^- \leq \widetilde{\rho} \leq \widetilde{\rho}^+$.

VeRecycle outputs $\widetilde{\rho}^-$ as a guaranteed threshold, certified by the original certificate for any realization of $\widetilde{f}$. The tightness of I^- to $\inf\{C(\mathbb{X}_?)\}$—and, consequently, $\widetilde{\rho}^-$ to $\widetilde{\rho}$—depends on the size of the interval used for IBP. Thus, VeRecycle uses a mesh to split up $\mathbb{X}_?$ into smaller intervals, increasing the tightness of $\widetilde{\rho}^-$ to $\widetilde{\rho}$.

If $\widetilde{\rho}^-$ is unsatisfactory, e.g., it does not match the original guarantee, or exceeds some permissible drop in probability, $\mathbb{X}_?$ can be split up using a finer mesh. Such a refinement is only done if the desired threshold is below $\widetilde{\rho}^+$: otherwise, the desired threshold cannot be obtained by refinement. In those cases, VeRecycle's output serves as proof that certification of the specification cannot be achieved without making additional assumptions on the changed dynamics, and/or performing additional training of the neural certificate.

The complexity of VeRecycle depends only on the complexity of determining (approximate) infima. For neural certificates, it is therefore determined by the complexity of IBP and the mesh size used.

5 VeRecycle for Compositional Control

VeRecycle is particularly valuable for systems under compositional control, in which a selection of policies is executed sequentially to complete a complex task. The modularity of such control allows VeRecycle to first reclaim guarantees for each component individually. Subsequently, the least affected components can be reused directly in a new composition, minimizing the need for costly repairs to neural control policies, certificates, and system models.

We consider compositional control based on a reach-avoid specification that is decomposed into several connected subtasks, either manually or automatically [Jothimurugan *et al.*, 2021]. This task decomposition is represented as a directed graph $G = (V, E, v_0, v_\star, \beta)$ where the mapping function $\beta : V \cup E \to \mathbb{X}$ relates the graph's elements to a stochastic dynamical system's state space: an edge e from vertex v_a to v_b represents the task of reaching $\beta(v_b)$ from $\beta(v_a)$ while avoiding $\beta(e)$.

A compositional policy for G then consists of a set of edge-associated policies $\{\pi_e \mid e \in E\}$ and a path $\pi^\star = (v_0, \ldots, v_k)$ in G, representing sequential execution of the policies π_e associated with each traversed edge e. Certification of the global reach-avoid specification φ with threshold ρ, requires that valid probability thresholds ρ_e and corresponding certificates C_e are given for all edge-associated policies π_e, and that multiplying them along the edges of the path $\pi^\star$ results in a value above the global probability threshold, i.e. $\prod_{i=1}^{k} \rho_{(v_{i-1}, v_i)} \leq \rho$.

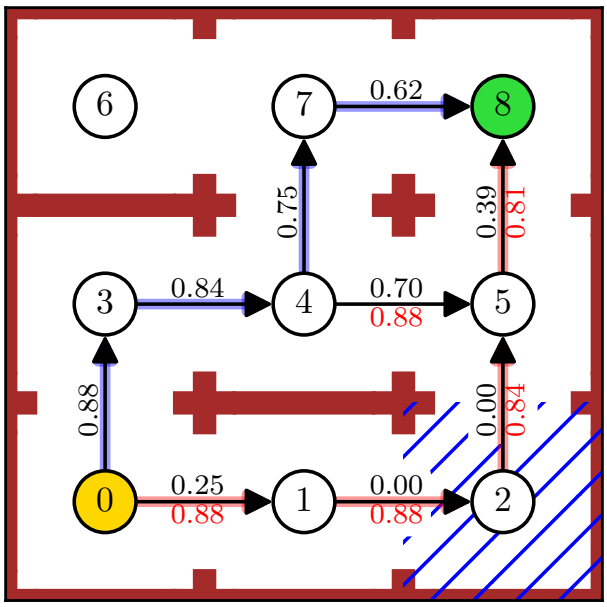

Figure 2: A compositional policy (blue) for a decomposed reach-avoid task: 0 (yellow) → 8 (green). Vertices are discrete abstractions of the rooms in the continuous environment. A directed edge signifies the existence of a certified policy between two rooms. Updated thresholds (black; original in red) reveal that the old path (red edges) no longer maximizes the global probability threshold.

Example 4. *We return to the running example of a robot navigating a warehouse with nine rooms. The goal—to navigate from the center of room 0 to the center of room 8—is broken down into subtasks of moving between two specific rooms. The specific task decomposition graph G is shown in Figure 2, with function β mapping vertices 0–8 to the dashed regions in Figure 1, and each edge to the walls ($\mathbb{X}_\oslash$). The compositional policy used to control the robot consists of the path $\pi^\star = (0, 1, 2, 5, 8)$ highlighted in red, and for each edge e, a policy π_e. This compositional policy is certified for the original dynamics of the system: for each edge e, some certificate C_e is given that proves a threshold ρ_e.*

Algorithm 1 shows how VeRecycle can be used to reclaim guarantees for such compositional policies, given the set $\mathbb{X}_?$ with changed dynamics, the graph G, and the sets of edge-associated certificates. For each edge in the task decomposition graph G used by the compositional controller, VeRecycle is invoked to reclaim a guaranteed threshold for the original certificate, ensuring validity for any change within $\mathbb{X}_?$ (Line 3). Each edge in G is assigned the weight $-\log(p)$, where p is the threshold obtained from VeRecycle (Line 4). The problem of maximizing the probability threshold of a path becomes a minimization problem that can be solved by

Algorithm 1: Recomposing Compositional Policies

Data: Set $\mathbb{X}_?$, graph $G = (V, E, v_0, v_\star, \beta)$, certificates $\{C_e \mid e \in E\}$ and thresholds $\{\rho_e \mid e \in E\}$.
Result: Path $\widetilde{\pi}^\star$ and threshold $\widetilde{\rho}$.

1 $W \leftarrow$ empty hash map
2 **foreach** $e \in E$ **do**
3 $\quad \widetilde{\rho}_e \leftarrow \textsc{VeRecycle}(C_e, \rho_e, \mathbb{X}_?)$
4 $\quad W[e] \leftarrow -\log \widetilde{\rho}_e$
5 $(v_0, \ldots, v_\star)$, weight $\leftarrow \textsc{ShortestPath}(G, W)$
6 $\widetilde{\rho} \leftarrow 2^{-\text{weight}}$
7 **return** $(v_0, \ldots, v_\star), \widetilde{\rho}$

an out-of-the-box shortest-path solver (Line 5). The sum of the weights along the shortest path is then converted into the probability threshold for that path (Line 6).

Example 5 (Example 4 continued)**.** *Once water has been spilled, the new dynamics $\widetilde{f}$—unknown in the states $\mathbb{X}_? = [2, 3] \times [0, 1]$—take effect. Given $\mathbb{X}_?, G, \{C_e \mid e \in E\}$, and $\{\rho_e \mid e \in E\}$, Algorithm 1—invoking VeRecycle for each edge—outputs an alternative blue path $\widetilde{\pi}^\star = (0, 3, 4, 7, 8)$ highlighted in Figure 2 which is certified to have at least $\widetilde{\rho} = 0.88 \cdot 0.84 \cdot 0.75 \cdot 0.62 \geq 0.33$, regardless of the effect of the spilled water. Although the certified threshold on the original path $\pi^\star$ has dropped to zero, an alternative with non-trivial guarantees could still be obtained. This illustrates how VeRecycle can exploit the modular nature of compositional policies while leaving policies and certificates (often neural networks) unaltered.*

The complexity of Algorithm 1 depends on the complexity of the calls to VeRecycle in Line 3, made once for each edge in the graph. While this complexity could likely be reduced further by minimizing the number of VeRecycle calls, e.g., a shortest-path solver with on-demand calls to VeRecycle only, this is outside the scope of this work. Algorithm 1 serves as a proof-of-concept implementation of VeRecycle applied to compositional control.

6 Evaluation

We demonstrate a VeRecycle implementation with interval bound propagation [Gowal *et al.*, 2019] to approximate infima of neural certificates. We analyze the efficiency and quality of reclaimed guarantees by VeRecycle on both standalone and compositional neural control in the "Stochastic Nine Rooms" benchmark (Figure 2) introduced by Žikelić *et al.* [2023b], with new, unknown dynamics in one of the rooms (e.g., the blue hatched area).

Out-of-the-box baselines are unavailable since, to the best of our knowledge, VeRecycle is the first framework to reclaim sound guarantees without requiring updated models of the dynamics. As a baseline, we therefore use state-of-the-art neural certification [Badings *et al.*, 2024] on worst-case dynamics in $\mathbb{X}_?$ (absorbing, i.e., $\mathbf{x}_{t+1} = \mathbf{x}_t$), treating $\mathbb{X}_?$ as part of the unsafe set $\mathbb{X}_\oslash$. We consider three settings for this baseline. First (B-I), verifying that a given new threshold—output by VeRecycle—is valid for the original certificate, i.e., given C and ρ', decide correctness. Second (B-II), finding a correct threshold given the original certificate as a candidate, i.e., given C, find a valid ρ' and C'. Third (B-III), finding a correct threshold and certificate from scratch, i.e., determine ρ' and C', without C. All experiments are run on Delft High Performance Computing Centre (DHPC) [2024] with NVIDIA A100 GPUs.

		Threshold $\widetilde{\rho}^-$		
Method	Time (s)	Loose C	Tight C	All
VeRecycle	**0.33** ± 0.66	0.58 ± 0.35	**0.65** ± 0.35	0.59 ± 0.35
B-I	9.22 ± 2.79	0.58 ± 0.35	**0.65** ± 0.35	0.59 ± 0.35
B-II	191.73 ± 161.64	**0.61** ± 0.34	0.64 ± 0.34	**0.62** ± 0.34
B-III	918.96 ± 190.64	0.51 ± 0.31	0.53 ± 0.29	0.51 ± 0.31

Table 1: Runtime and reclaimed probability thresholds, averaged (with standard deviation) over 5 random seeds and 63 tasks (9 specifications $\times$ 7 changes). Thresholds are grouped by the quality of original certificate C: tight certificates (available for 2 out of 9 specifications) are below $\frac{1}{1-\rho}$ only inside the rooms containing $\mathbb{X}_0$ and $\mathbb{X}_\star$. Best result per column is marked in bold.

6.1 Experiment 1: Individual Certificates

We analyze VeRecycle's reclaimed thresholds for 63 unique tasks, combining 9 reach-avoid specifications (graph edges in Figure 2) with 7 different subsets $\mathbb{X}_?$ (rooms). As input, we use neural policies obtained through proximal policy optimization (PPO) [Schulman *et al.*, 2017] with neural reach-avoid certificates [Badings *et al.*, 2024]. Technical details are provided in the extended version.

In runtime (Table 1, column 2), VeRecycle outperforms the baseline on all settings. Searching for a new threshold and certificate from scratch (B-III) requires the highest computation time. Using the original certificate as a candidate (B-II) reduces computation time by an order of magnitude, emphasizing the benefit of reusing original verification results. When informed of a correct threshold by VeRecycle (B-I) runtime decreases further, demonstrating the benefit of inferring thresholds from certificates rather than trial-and-error search. VeRecycle outperforms B-I by two orders of magnitude, showing the efficiency of ignoring unaffected states.

In addition to requiring less computational effort, VeRecycle produces tighter probability thresholds (Table 1, columns 3–5) than re-verification from scratch (B-III). While re-verification with the original certificate (B-II) produces higher thresholds on average, VeRecycle performs comparably on high-quality input certificates—where $C(\mathbf{x}) \geq \frac{1}{1-\rho}$ for all rooms outside the initial and target rooms. As illustrated in Figure 3, B-II can refine unnecessary low-valued states in "loose" certificates to achieve a higher threshold than VeRecycle. However, this refinement process is computationally expensive, making VeRecycle a lightweight alternative to identify when certificate refinement is necessary. Future work on improving certification techniques for generating higher-quality certificates would further enhance VeRecycle's performance and applicability.

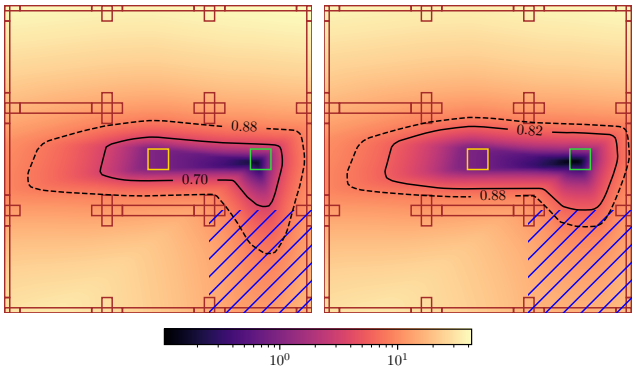

Figure 3: Example of a "loose" original certificate (gradient left) refined by baseline algorithm (gradient right) to obtain an updated threshold $\widetilde{\rho}$ (solid line) closer to the original ρ (dashed line). Yellow ($\mathbb{X}_0$), green ($\mathbb{X}_\star$), red ($\mathbb{X}_\varnothing$), and blue ($\mathbb{X}_?$) denote task sets.

	Avg. $\widetilde{\rho}^-$ per $\mathbb{X}_?$ location		
Method	Target adj.	Other	All
VeRecycle	$0.24_{\pm 0.09}$	$\mathbf{0.44}_{\pm 0.06}$	$0.38_{\pm 0.12}$
B-I	$0.24_{\pm 0.09}$	$\mathbf{0.44}_{\pm 0.06}$	$0.38_{\pm 0.12}$
B-II	$\mathbf{0.40}_{\pm 0.07}$	$0.42_{\pm 0.06}$	$\mathbf{0.42}_{\pm 0.06}$
B-III	$0.19_{\pm 0.09}$	$0.21_{\pm 0.06}$	$0.20_{\pm 0.07}$

Table 2: Reclaimed probability bounds for compositional policy (Experiment 2), averaged over 5 random seeds for 7 different sets $\mathbb{X}_?$ with new dynamics. Grouped by whether $\mathbb{X}_?$ is adjacent to the target room, i.e., in room 5 or 7. Best results per column are in bold.

6.2 Experiment 2: Compositional Control

Next, we compare the global thresholds obtained with Algorithm 1 on the graph from Figure 2, using VeRecycle and baselines to reclaim guarantees for each graph edge, and NetworkX's Dijkstra [Hagberg *et al.*, 2008] as shortest-path solver. The results (Table 2) show that VeRecycle performs competitively on the majority of tasks.

In two particular tasks—rooms with vertex 5 and 7 having unknown dynamics—the effect of low-quality certificates on VeRecycle's performance can still be observed. In those cases, the edges (7,8) and (5,8)—both with "loose" certificates—are simultaneously impacted, thus impacting the reclaimed guarantee on all feasible paths to the target states.

Overall, we observe that VeRecycle is particularly powerful in compositional control settings, because the impact of low-quality input certificates is mitigated by the availability of alternative policies.

7 Related Work

Probabilistic neural reach-avoid certification. Probabilistic neural certification is the state-of-the-art for verifying reach-avoid specifications for stochastic dynamical systems under neural control [Žikelić *et al.*, 2023a; Ansaripour *et al.*, 2023; Chatterjee *et al.*, 2023; Mathiesen *et al.*, 2023; Abate *et al.*, 2024; Badings *et al.*, 2024]. VeRecycle complements these methods by efficiently deciding whether such guarantees are maintained when system dynamics change, minimizing the high computational cost of re-certifying the complete updated system model.

Robust control synthesis. Robust control synthesis addresses changes in dynamics by designing control policies that perform reliably under a modeled range of disturbances. Techniques include adaptations of Markov Decision Process (MDP) solvers for richer models expressing multiple variations of the system: e.g., parametric MDPs [Badings *et al.*, 2022], interval MDPs [Jiang *et al.*, 2022; Coppola *et al.*, 2024], uncertain MDPs [Wolff *et al.*, 2012], and multi-environment MDPs [Raskin and Sankur, 2014]. While the type of changes addressed by VeRecycle could theoretically be addressed by robust synthesis techniques as well, this would require either knowing all possible changes in advance or updating and solving the adapted MDP again. Thus, we see VeRecycle as a lightweight alternative that does not require prior knowledge of possible changes.

Reusing policies in compositional control. Compositional control as described in Section 5 stems from the concept of options [Sutton *et al.*, 1999], adapted for safety-critical applications through correct-by-design synthesis methods for a high-level plan [Jothimurugan *et al.*, 2021; Neary *et al.*, 2023] and formal verification of used components [Ivanov *et al.*, 2021; Delgrange *et al.*, 2024]. Notably, [Žikelić *et al.*, 2023b] verify each component with probabilistic neural certification, inspiring our application of VeRecycle to compositional control. In addition to solving more difficult, long-horizon tasks, a motivation for compositional control is reusing the components in alternative high-level plans. This is not straightforward in case of changes in system dynamics, however, since guarantees on components must still be reclaimed before composing a (new) high-level plan, making VeRecycle complementary to compositional control.

8 Conclusion

VeRecycle is the first framework to formally reclaim guarantees from probabilistic neural certificates for discrete-time stochastic dynamical systems after localized changes. We formally prove that VeRecycle infers sound guarantees from (approximated) infima of the original certificates. Our experimental evaluation demonstrates scenarios in which VeRecycle's guarantees are competitive—particularly for compositional control—compared to complete re-certification, with computational cost reduced by three orders of magnitude.

Future work. We plan to extend the presented theoretical foundations to probabilistic neural certificates for continuous-time systems [Neustroev *et al.*, 2025]. Additionally, we aim to explore how the certificate learning process can be optimized to maximize the guarantees reclaimed with VeRecycle. Another avenue of future research is synergies between VeRecycle and learning-enabled certification [Kolter and Manek, 2019; Wu *et al.*, 2023; Takeishi and Kawahara, 2021; Schlaginhaufen *et al.*, 2021], using VeRecycle to pinpoint crucial states for which new dynamics must be learned.

Acknowledgments

This research was funded by the NWO Veni grant Explainable Monitoring (222.119). This work was done in part while Anna Lukina and Sterre Lutz were visiting the Simons Institute for the Theory of Computing.

References

[Abate *et al.*, 2021] Alessandro Abate, Daniele Ahmed, Mirco Giacobbe, and Andrea Peruffo. Formal synthesis of Lyapunov neural networks. *IEEE Control. Syst. Lett.*, 5(3):773–778, 2021.

[Abate *et al.*, 2024] Alessandro Abate, Mirco Giacobbe, and Diptarko Roy. Stochastic omega-regular verification and control with supermartingales. In *CAV (3)*, pages 395–419, 2024.

[Ansaripour *et al.*, 2023] Matin Ansaripour, Krishnendu Chatterjee, Thomas A. Henzinger, Mathias Lechner, and Dorde Žikelić. Learning provably stabilizing neural controllers for discrete-time stochastic systems. In *ATVA (1)*, pages 357–379, 2023.

[Badings *et al.*, 2022] Thom Badings, Murat Cubuktepe, Nils Jansen, Sebastian Junges, Joost-Pieter Katoen, and Ufuk Topcu. Scenario-based verification of uncertain parametric MDPs. *International Journal on Software Tools for Technology Transfer*, 24(5):803–819, 2022.

[Badings *et al.*, 2024] Thom Badings, Wietze Koops, Sebastian Junges, and Nils Jansen. Learning-Based Verification of Stochastic Dynamical Systems with Neural Network Policies. *arXiv preprint arXiv:2406.00826*, 2024.

[Blumenthal and Getoor, 1968] R. M. C. Blumenthal and R. K. Getoor. *Markov Processes and Potential Theory*. Pure and applied mathematics: A series of monographs and textbooks. Academic Press, 1968.

[Chang *et al.*, 2019] Ya-Chien Chang, Nima Roohi, and Sicun Gao. Neural Lyapunov control. In *NeurIPS*, pages 3240–3249, 2019.

[Chatterjee *et al.*, 2023] Krishnendu Chatterjee, Thomas A. Henzinger, Mathias Lechner, and Dorde Žikelić. A learner-verifier framework for neural network controllers and certificates of stochastic systems. In *TACAS (1)*, pages 3–25, 2023.

[Coppola *et al.*, 2024] Rudi Coppola, Andrea Peruffo, Licio Romao, Alessandro Abate, and Manuel Mazo Jr. Data-driven Interval MDP for Robust Control Synthesis. *arXiv preprint arXiv:2404.08344*, 2024.

[Dawson *et al.*, 2023] Charles Dawson, Sicun Gao, and Chuchu Fan. Safe control with learned certificates: A survey of neural Lyapunov, barrier, and contraction methods for robotics and control. *IEEE Trans. Robotics*, 39(3):1749–1767, 2023.

[Delft High Performance Computing Centre (DHPC), 2024] Delft High Performance Computing Centre (DHPC). DelftBlue Supercomputer (Phase 2). https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2, 2024.

[Delgrange *et al.*, 2024] Florent Delgrange, Guy Avni, Anna Lukina, Christian Schilling, Ann Nowe, and Guillermo Pérez. Controller synthesis from deep reinforcement learning policies. In *EWRL*, pages 1–34, 2024.

[Edwards *et al.*, 2023] Alec Edwards, Andrea Peruffo, and Alessandro Abate. A general verification framework for dynamical and control models via certificate synthesis. *CoRR*, abs/2309.06090, 2023.

[Gehr *et al.*, 2018] Timon Gehr, Matthew Mirman, Dana Drachsler-Cohen, Petar Tsankov, Swarat Chaudhuri, and Martin T. Vechev. AI2: safety and robustness certification of neural networks with abstract interpretation. In *IEEE Symposium on Security and Privacy*, pages 3–18, 2018.

[Giacobbe *et al.*, 2022] Mirco Giacobbe, Daniel Kroening, and Julian Parsert. Neural termination analysis. In *ESEC/SIGSOFT FSE*, pages 633–645, 2022.

[Gowal *et al.*, 2019] Sven Gowal, Krishnamurthy Dvijotham, Robert Stanforth, Rudy Bunel, Chongli Qin, Jonathan Uesato, Relja Arandjelovic, Timothy Mann, and Pushmeet Kohli. On the Effectiveness of Interval Bound Propagation for Training Verifiably Robust Models. *arXiv preprint arXiv:1810.12715*, 2019.

[Hagberg *et al.*, 2008] Aric Hagberg, Pieter J. Swart, and Daniel A. Schult. Exploring network structure, dynamics, and function using NetworkX. Technical report, Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2008.

[Ivanov *et al.*, 2021] Radoslav Ivanov, Kishor Jothimurugan, Steve Hsu, Shaan Vaidya, Rajeev Alur, and Osbert Bastani. Compositional Learning and Verification of Neural Network Controllers. *ACM Transactions on Embedded Computing Systems*, 20(5s):92:1–92:26, 2021.

[Jiang *et al.*, 2022] Jesse Jiang, Ye Zhao, and Samuel Coogan. Safe Learning for Uncertainty-Aware Planning via Interval MDP Abstraction. *IEEE Control Systems Letters*, pages 2641–2646, 2022.

[Jothimurugan *et al.*, 2021] Kishor Jothimurugan, Suguman Bansal, Osbert Bastani, and Rajeev Alur. Compositional Reinforcement Learning from Logical Specifications. In *NeurIPS*, pages 10026–10039, 2021.

[Kalman and Bertram, 1959] Rudolf E Kalman and John E Bertram. Control system analysis and design via the second method of Lyapunov: (i) continuous-time systems (ii) discrete time systems. *IRE Transactions on Automatic Control*, 4(3):112, 1959.

[Katz *et al.*, 2017] Guy Katz, Clark W. Barrett, David L. Dill, Kyle Julian, and Mykel J. Kochenderfer. Reluplex: An efficient SMT solver for verifying deep neural networks. In *CAV (1)*, pages 97–117, 2017.

[Kolter and Manek, 2019] J. Zico Kolter and Gaurav Manek. Learning stable deep dynamics models. In *NeurIPS*, pages 11126–11134, 2019.

[Kouvaros and Lomuscio, 2021] Panagiotis Kouvaros and Alessio Lomuscio. Towards scalable complete verification

of ReLU neural networks via dependency-based branching. In *IJCAI*, pages 2643–2650, 2021.

[Kwiatkowska and Zhang, 2023] Marta Kwiatkowska and Xiyue Zhang. When to Trust AI: Advances and Challenges for Certification of Neural Networks. In *FedCSIS*, pages 25–37, 2023.

[Mathiesen *et al.*, 2023] Frederik Baymler Mathiesen, Simeon C. Calvert, and Luca Laurenti. Safety certification for stochastic systems via neural barrier functions. *IEEE Control. Syst. Lett.*, 7:973–978, 2023.

[Mirman *et al.*, 2018] Matthew Mirman, Timon Gehr, and Martin Vechev. Differentiable Abstract Interpretation for Provably Robust Neural Networks. In *ICML*, pages 3578–3586, 2018.

[Neary *et al.*, 2023] Cyrus Neary, Aryaman Singh Samyal, Christos Verginis, Murat Cubuktepe, and Ufuk Topcu. Verifiable Reinforcement Learning Systems via Compositionality. *arXiv preprint arXiv:2309.06420*, 2023.

[Neustroev *et al.*, 2025] Grigory Neustroev, Mirco Giacobbe, and Anna Lukina. Neural continuous-time supermartingale certificates. In *AAAI*, pages 27538–27546, 2025.

[Papachristodoulou and Prajna, 2002] Antonis Papachristodoulou and Stephen Prajna. On the construction of Lyapunov functions using the sum of squares decomposition. In *CDC*, pages 3482–3487, 2002.

[Prajna *et al.*, 2004] Stephen Prajna, Ali Jadbabaie, and George J. Pappas. Stochastic safety verification using barrier certificates. In *CDC*, pages 929–934, 2004.

[Puterman, 2014] Martin L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[Raskin and Sankur, 2014] Jean-François Raskin and Ocan Sankur. Multiple-Environment Markov Decision Processes. *arXiv preprint arXiv:1405.4733*, 2014.

[Richards *et al.*, 2018] Spencer M. Richards, Felix Berkenkamp, and Andreas Krause. The Lyapunov neural network: Adaptive stability certification for safe learning of dynamical systems. In *CoRL*, pages 466–476, 2018.

[Schlaginhaufen *et al.*, 2021] Andreas Schlaginhaufen, Philippe Wenk, Andreas Krause, and Florian Dorfler. Learning stable deep dynamics models for partially observed or delayed dynamical systems. In *NeurIPS*, pages 11870–11882, 2021.

[Schulman *et al.*, 2017] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[Sutton *et al.*, 1999] Richard S. Sutton, Doina Precup, and Satinder Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1-2):181–211, 1999.

[Takeishi and Kawahara, 2021] Naoya Takeishi and Yoshinobu Kawahara. Learning dynamics models with stable invariant sets. In *AAAI*, pages 9782–9790, 2021.

[Topcu *et al.*, 2008] Ufuk Topcu, Andrew K. Packard, and Peter J. Seiler. Local stability analysis using simulations and sum-of-squares programming. *Automatica*, 44(10):2669–2675, 2008.

[Wolff *et al.*, 2012] Eric M Wolff, Ufuk Topcu, and Richard M Murray. Robust control of uncertain Markov decision processes with temporal logic specifications. In *CDC*, pages 3372–3379, 2012.

[Wu *et al.*, 2023] Junlin Wu, Andrew Clark, Yiannis Kantaros, and Yevgeniy Vorobeychik. Neural Lyapunov control for discrete-time systems. In *NeurIPS*, pages 2939–2955, 2023.

[Wu *et al.*, 2024] Haoze Wu, Omri Isac, Aleksandar Zeljic, Teruhiro Tagomori, Matthew L. Daggitt, Wen Kokke, Idan Refaeli, Guy Amir, Kyle Julian, Shahaf Bassan, Pei Huang, Ori Lahav, Min Wu, Min Zhang, Ekaterina Komendantskaya, Guy Katz, and Clark W. Barrett. Marabou 2.0: A versatile formal analyzer of neural networks. In *CAV (2)*, pages 249–264, 2024.

[Zhang *et al.*, 2018] Huan Zhang, Tsui-Wei Weng, Pin-Yu Chen, Cho-Jui Hsieh, and Luca Daniel. Efficient neural network robustness certification with general activation functions. In *NeurIPS*, pages 4944–4953, 2018.

[Zhang *et al.*, 2023] Songyuan Zhang, Yumeng Xiu, Guannan Qu, and Chuchu Fan. Compositional neural certificates for networked dynamical systems. In *L4DC*, pages 272–285, 2023.

[Žikelić *et al.*, 2023a] Đorđe Žikelić, Mathias Lechner, Thomas A. Henzinger, and Krishnendu Chatterjee. Learning Control Policies for Stochastic Systems with Reach-Avoid Guarantees. In *AAAI*, pages 11926–11935, 2023.

[Žikelić *et al.*, 2023b] Đorđe Žikelić, Mathias Lechner, Abhinav Verma, Krishnendu Chatterjee, and Thomas Henzinger. Compositional Policy Learning in Stochastic Control Systems with Formal Guarantees. In *NeurIPS*, pages 47849–47873, 2023.