Regularity and Stability Properties of Selective SSMs with Discontinuous Gating

Anonymous authors
Paper under double-blind review

Abstract

Deep selective State-Space Models (SSMs), whose state-space parameters are modulated online by a *selection* signal, offer significant expressive power but pose challenges for stability analysis, especially under discontinuous gating. We study continuous-time selective SSMs through the lenses of passivity and Input-to-State Stability (ISS), explicitly distinguishing the *selection schedule* $x(\cdot)$ from the *driving (port) input* $u(\cdot)$. First, we show that state-strict dissipativity ($\beta > 0$) together with quadratic bounds on a storage functional implies exponential decay of homogeneous trajectories ($u \equiv 0$), yielding exponential forgetting. Second, by freezing the selection ($x(t) \equiv 0$) we obtain a passive LTV input–output subsystem and prove that its minimal available storage is necessarily quadratic, $V_{a,0}(t,h) = \frac{1}{2}h^H Q_0(t)h$, with $Q_0 \in \mathrm{AUC}_{\mathrm{loc}}$, accommodating discontinuities induced by gating. Third, under the strong hypothesis that a *single* quadratic storage certifies passivity uniformly over all admissible selection schedules, we derive a parametric LMI and universal kernel constraints on gating, formalizing an "irreversible forgetting" structure. Finally, we give sufficient conditions for global ISS with respect to the port input $u(\cdot)$, uniformly over admissible selection schedules, and we validate the main predictions in targeted simulation studies.

1 Introduction

The pursuit of stable, robust, and predictable behavior in dynamical systems is a cornerstone of scientific and engineering disciplines, with increasing relevance in the era of complex computational models like Neural ODEs (Chen et al., 2018) and Deep State-Space Models (SSMs) (Gu et al., 2022; Gu & Dao, 2024). Foundational paradigms of energy exchange, rooted in physics and later abstracted into powerful mathematical frameworks such as Willems' theory of dissipative systems (Willems, 1972) and Input-to-State Stability (ISS) (Sontag & Wang, 1995; Jiang et al., 1999), provide important tools for this pursuit. These energy-based and robustness-centric perspectives allow for rigorous analysis of a system's interaction with its environment and its inherent stability characteristics. Extending far beyond their physical origins, these principles are now crucial for understanding the conditions under which modern AI architectures operate reliably, maintain numerical stability, and process information effectively despite internal complexities and external disturbances. This paper leverages these well-established principles to investigate the regularity and stability properties of continuous-time deep selective SSMs. These models, whose parameters are dynamically modulated by their inputs, pose distinct analytical questions at the intersection of time-varying and nonlinear system dynamics, which this work aims to address.

Architectures such as Mamba (Gu & Dao, 2024), HGRN (Qin et al., 2024), and GLA (Yang et al., 2024) represent a new frontier in deep learning, termed selective SSMs (Cirone et al., 2024; Zubic et al., 2025; Soydan et al., 2024). Their defining characteristic, captured in our continuous-time model equation 1, is that their core state-space parameters (A, B, C) are not fixed but are dynamically shaped by both time-varying gating signals $\Delta(t)$ and a *selection (scheduling) signal* $x(t)$ (e.g., token embeddings) itself. For the passivity analysis, we distinguish the selection signal $x(t)$ from the driving (port) input $u(t)$. In many architectures, these coincide ($u \equiv x$), but keeping them distinct makes the quantifiers in the LMI and storage constructions explicit. This input-selectivity is key to their performance, enabling selective context processing. Analytically, however, this places these models in a complex domain: they are neither purely

Linear Time-Varying (LTV), due to the $x(t)$ dependence, nor are they classical nonlinear systems for which standard tools always apply without adaptation (Isidori, 1985; Khalil & Grizzle, 2002). While the theory for LTV system stability, including the behavior of quadratic "energy" functions V_Q (e.g., their regularity (Morandin & Hinsen, 2024)), and for ISS of input-driven nonlinear systems (Sontag, 1990) are individually well-understood, their integrated application to selective SSMs is not explored. Specifically, how does a unified "energy management" or passivity perspective constrain these input-dependent dynamics, especially when $\Delta(t)$ introduces abrupt changes? This paper addresses this crucial question by developing a rigorous framework for analyzing stability and structural properties in these advanced models. We bridge this gap by systematically applying and extending concepts from passivity theory and ISS to continuous-time selective SSMs. We investigate the conditions under which these models exhibit stable behavior and how a dissipativity assumption constrains their internal structure and gating mechanisms. Specifically, we demonstrate that even with input-dependent and potentially discontinuous gating, foundational properties of energy-like functions persist and impose significant structural constraints. Our contributions are as follows:

1. We show that strict dissipativity implies exponential forgetting of initial conditions (Theorem 4.1). Moreover, freezing the selection signal (e.g., $x(t) \equiv 0$) yields a passive LTV input–output subsystem whose minimal available storage is necessarily quadratic with an inducing matrix $Q_0 \in \mathrm{AUC}_{\mathrm{loc}}$, accommodating discontinuous gating (Theorem 4.2). Evaluating this storage along homogeneous trajectories ($u \equiv 0$) provides a canonical intrinsic energy for the frozen-selection dynamics.
2. Assuming a universal quadratic storage function certifies passivity uniformly over admissible selection schedules, we derive a parametric LMI and kernel constraints implying that “energy-less” state directions must be output-unobservable under any selection (Theorem 5.2).
3. We formalize an “irreversible forgetting” principle: once a state direction enters the kernel of the universal storage, it remains structurally forgotten over time (Theorem 5.4).
4. We provide sufficient conditions for global Input-to-State Stability (ISS) with respect to the driving (port) input $u(\cdot)$, linking uniform dissipativity to overall robustness (Theorem 6.1).

Our findings offer a rigorous framework for understanding and designing stable selective SSMs. The remainder of the paper is structured to build our argument progressively: we first establish the baseline stability of the unforced system (Section 4), then use the assumption of universal passivity to probe the system's structural constraints (Section 5), and finally provide conditions for robust stability under general inputs (Section 6). We conclude in Section 7 by summarizing our key insights and their practical implications. All formal proofs, the detailed empirical validations from our simulation studies, and an extended discussion of future directions are provided in the Appendix A.

2 Related Work

The analysis of stability and energy-based properties in dynamical systems has a rich history, providing crucial tools for understanding systems ranging from classical mechanics to modern AI. Our work on continuous-time selective State-Space Models (SSMs) builds upon several key research streams.

2.1 Dissipativity, Passivity, and LTV Systems

The foundational theory of dissipative systems, pioneered by Willems (Willems, 1972), provides a general framework for analyzing systems based on energy-like storage functions and supply rates. Passivity, a special case where the supply rate is the input-output inner product, is central to understanding robust stability and interconnection (Van der Schaft, 2000). The Kalman-Yakubovich-Popov (KYP) lemma established a vital link between frequency-domain passivity conditions and state-space properties for LTI systems (Kalman, 1963; Popov, 1961), which has been extended to LTV systems, often involving time-varying Riccati equations or differential/integral inequalities (Anderson & Vongpanitlerd, 2013). Crucially for our work, recent research by Morandin and Hinsen (Morandin & Hinsen, 2024) has rigorously investigated quadratic storage functions for passive LTV systems under minimal regularity assumptions. Our work leverages these findings

by demonstrating that freezing the selection signal (e.g., $x(t) \equiv 0$) yields a passive LTV *input–output* subsystem with port input $u(\cdot)$, which falls directly under the framework of (Morandin & Hinsen, 2024) and therefore inherits a minimal quadratic storage with $\mathrm{AUC}_{\mathrm{loc}}$ regularity, even with discontinuous gating.

2.2 Stability of Time-Varying and Switched Systems

The input-selectivity of modern SSMs creates dynamics that behave like switched systems, where the "switching signal" is the input data itself. This necessitates tools for analyzing systems with discontinuous parameter changes (Liberzon, 2003), such as methods involving common or multiple Lyapunov functions (Branicky, 1998). The formalisms of Filippov (Filippov, 1988) and Carathéodory (Coddington et al., 1956) provide the theoretical bedrock for guaranteeing that solutions to our system exist under the mild L^p_{loc} regularity we assume. Our work builds directly on this foundation. We move beyond proving solution existence to analyze how a single, coherent energy structure, an $\mathrm{AUC}_{\mathrm{loc}}$ quadratic storage function, can persist across these input-driven switches and what structural constraints this persistence imposes on the model's architecture.

2.3 Input-to-State Stability (ISS) for Nonlinear Systems

The ISS framework (Sontag et al., 1989; Sontag & Wang, 1995) provides a robust notion of stability for nonlinear systems subject to external inputs, characterized by ISS-Lyapunov functions. It quantifies how the system state is affected by both initial conditions and input magnitudes. While classical ISS theory often assumes smooth dynamics, its principles are highly relevant for our selective SSMs, where the selection signal $x(t)$ renders the dynamics schedule-dependent (and in the coupled case $u \equiv x$ recovers an input-driven nonlinear viewpoint). Our contribution (Theorem 6.1) adapts ISS concepts by seeking a common quadratic Lyapunov function that ensures stability uniformly across all input-induced dynamics, providing a specific condition under which these complex, input-modulated systems are globally robustly stable.

2.4 Stability and Dynamics of Modern SSMs in Deep Learning

Recent deep learning SSMs like S4 (Gu et al., 2022), S5 (Smith et al., 2023), and particularly Mamba (Gu & Dao, 2024) (and its variants), have shown remarkable performance. Theoretical analyses of these models are emerging. (Halloran et al., 2024) analyzed Mamba's stability via Lyapunov exponents, showing non-positive maximal exponents, implying robustness to small perturbations. Our work complements this by providing conditions for exponential decay and ISS from an energy/Lyapunov function perspective under specific passivity assumptions, offering a different angle on stability for the continuous-time selective formulation. Some other works connected selective SSMs to Controlled Differential Equations (CDEs) and Rough Path Theory to explain their expressivity (Kidger et al., 2020; Lyons, 2014; Cirone et al., 2024). While our focus is on stability and passivity using more classical control-theoretic tools, these works show the rich mathematical underpinnings of these models from another perspective.

3 Preliminaries: Selective SSMs and Passivity Framework

This section introduces the continuous-time selective State-Space Models (SSMs) that are the focus of our analysis, along with the fundamental concepts from control theory that support our approach.

3.1 System Definition and Notation

We consider dynamical systems evolving in continuous time $t \in \mathbb{T}$, where $\mathbb{T} \subseteq [0, +\infty)$ is a time interval. The internal state of the system at time t is denoted by $h(t) \in \mathbb{C}^N$ (or $\mathbb{R}^N$ in real-valued cases).

A key point for our passivity analysis is to distinguish between: (i) a *driving (port) input* $u(t) \in \mathbb{C}^{d_{\mathrm{in}}}$, which is the signal that appears in the supply rate $\Re\langle u, y\rangle$, and (ii) a *selection (scheduling) signal* $x(t) \in \mathbb{C}^{d_{\mathrm{in}}}$ (e.g., the token embedding) that parametrizes the time-varying matrices via input-dependent gating.[1]

[1]Throughout Sections 3–6 we assume $d_{\mathrm{out}} = d_{\mathrm{in}}$ so that $\langle u, y\rangle$ and the passivity LMI blocks are dimensionally well-defined (this is the common setting in sequence models where the recurrent channel has matched input/output width).

The system produces an output $y(t) \in \mathbb{C}^{d_{\text{in}}}$. For matrices and vectors, M^H denotes the Hermitian (conjugate) transpose. The standard Euclidean norm of a vector v is $\|v\|$, and the inner product of two vectors u, v is $\langle u, v\rangle = v^H u$. We denote the space of $n \times n$ Hermitian positive semidefinite matrices by $\mathbf{S}_+^n(\mathbb{C})$; for $Q \in \mathbf{S}_+^n(\mathbb{C})$, we write $Q \succeq 0$.

Our work focuses on Continuous-Time Selective SSMs. Inspired by architectures like Mamba (Gu & Dao, 2024), their defining characteristic is that the system's core parameters are dynamically modulated by the selection signal $x(t)$ and an auxiliary gating signal $\Delta(t)$. Formally, the dynamics are given by:

$$\begin{cases} \dot{h}(t) \;=\; A\Big(\Delta(t),\, x(t)\Big)\; h(t) \;+\; B\Big(\Delta(t),\, x(t)\Big)\; u(t), \\ y(t) \;=\; C\Big(\Delta(t),\, x(t)\Big)\; h(t). \end{cases} \tag{1}$$

Remark 3.1 (Coupled-input selective SSMs)**.** In many token-selective architectures, the same signal both *selects* the parameters and *drives* the dynamics; this corresponds to the special case $u(\cdot) \equiv x(\cdot)$. In this paper we keep $u(\cdot)$ and $x(\cdot)$ conceptually distinct to make the quantifiers in the passivity analysis explicit. Any condition proved for all admissible pairs $(x(\cdot), u(\cdot))$ remains valid (as a sufficient condition) under the coupling $u = x$.

3.2 Well-Posedness and Function Space Assumptions

To ensure the system equation 1 is well-defined even with discontinuous parameter changes, we adopt mild regularity conditions. For any admissible input trajectory $x(\cdot)$, let $A_{\text{eff}}(t) := A(\Delta(t), x(t))$ (and similarly for $B_{\text{eff}}, C_{\text{eff}}$). We assume the effective matrices belong to standard Lebesgue spaces: $A_{\text{eff}}(\cdot) \in L^1_{\text{loc}}$, $B_{\text{eff}}(\cdot) \in L^2_{\text{loc}}$, and $C_{\text{eff}}(\cdot) \in L^2_{\text{loc}}$, with the selection signal $x(\cdot) \in L^2_{\text{loc}}$ and the driving input $u(\cdot) \in L^2_{\text{loc}}$. These conditions ensure that the ODE for $h(t)$ satisfies Carathéodory conditions, guaranteeing the local existence and uniqueness of an absolutely continuous solution $h(\cdot) \in W^{1,1}_{\text{loc}}$ (Filippov, 1988; Coddington et al., 1956).

Remark 3.2 (Feasibility of Regularity Assumptions)**.** These assumptions are mathematically mild and readily satisfied by modern selective SSMs. In token-based architectures like Mamba, the selection mechanism results in system matrices that are **piecewise-constant**. Such functions are well-behaved members of the L^p_{loc} spaces, confirming that our framework is grounded in practical implementations while being general enough for future architectures.

3.3 Passivity and Dissipativity

We analyze the system equation 1 using the powerful energy-based frameworks of passivity and dissipativity (Willems, 1972). Intuitively, a passive system cannot generate its own energy. It can only store or dissipate energy supplied from its input. A strictly dissipative system is even stronger, as it inherently loses energy over time. These concepts are formalized in the following definitions.

Definition 3.3 (Storage Function)**.** A function $V : \mathbb{T} \times \mathbb{C}^N \to \mathbb{R}$ is a *storage function* if it is non-negative, i.e., $V(t, h) \geq 0$ for all (t, h), and typically $V(t, 0) = 0$.

Definition 3.4 (Passivity)**.** The system equation 1 is *passive* if there exists a storage function V such that for all admissible trajectories on any interval $[t_0, T]$:

$$V\big(T, h(T)\big) - V\big(t_0, h(t_0)\big) \leq \int_{t_0}^{T} \operatorname{Re}\langle u(\tau), y(\tau)\rangle \; d\tau. \tag{2}$$

The term $\operatorname{Re}\langle u(\tau), y(\tau)\rangle$ is the instantaneous power, or *supply rate*, provided to the system.

Definition 3.5 (Strict Dissipativity)**.** The system is *strictly dissipative* with rate $\beta > 0$ if it satisfies the stronger inequality:

$$V\big(T, h(T)\big) - V\big(t_0, h(t_0)\big) \leq \int_{t_0}^{T} \operatorname{Re}\langle u(\tau), y(\tau)\rangle \; d\tau - \beta \int_{t_0}^{T} \|h(\tau)\|^2 \, d\tau. \tag{3}$$

A positive β implies an intrinsic rate of energy dissipation, which is crucial for proving strong stability properties like exponential decay. Given the absolute continuity of $h(\cdot)$ and local Lipschitzness of V in h,

these integral inequalities have corresponding differential forms: $\frac{d}{dt}V \leq \operatorname{Re}\langle u, y\rangle$ for passivity, and $\frac{d}{dt}V \leq \operatorname{Re}\langle u, y\rangle - \beta\|h\|^2$ for strict dissipativity.

3.4 Quadratic Storage Functions and AUC Regularity

A significant portion of our analysis focuses on *quadratic storage functions* of the form $V_Q(t, h) = \frac{1}{2}h^H Q(t)h$, where $Q : \mathbb{T} \to \mathbf{S}_+^N(\mathbb{C})$ is a time-varying matrix. The regularity of $Q(t)$ is critical in our setting, as the gating mechanism can induce discontinuities in the system dynamics.

To handle this robustly, we consider $Q(t)$ to belong to the class of Locally Absolutely Upper Semicontinuous matrix functions, denoted $Q \in \mathrm{AUC}_{\mathrm{loc}}(\mathbb{T}, \mathbf{S}_+^N(\mathbb{C}))$, following the framework of (Morandin & Hinsen, 2024). For our purposes, $\mathrm{AUC}_{\mathrm{loc}}$ allows controlled discontinuities and ensures $\dot{Q}(t)$ exists a.e. Under the additional hypothesis that Q induces a universal quadratic storage for a passive LTV subsystem implies further structural constraints, including rank monotonicity (Theorem 5.1). This property is the foundation for the "irreversible forgetting" we analyze later.

Formally, a function $Q \in \mathrm{AUC}_{\mathrm{loc}}$ is characterized by being of locally bounded variation ($\mathrm{BV}_{\mathrm{loc}}$), having a derivative $\dot{Q}(t)$ that exists almost everywhere, and satisfying certain integral and jump conditions. The key property that enforces rank monotonicity is that the singular part of its decomposition must be weakly monotonically decreasing in the Loewner order (e.g., at any jump, $\lim_{\tau\to t^-} Q(\tau) \succeq Q(t)$). This mathematical structure is precisely what makes the framework suitable for analyzing switched systems while retaining essential properties about their energy evolution.

4 Fundamental Stability and Structural Properties from Passivity

This section establishes the foundational stability guarantees and explores the core structural properties inherent in passive selective SSMs. We begin by analyzing the *homogeneous frozen-selection* case, where the selection signal is fixed to a reference value $x(t) \equiv 0$ and the port input is turned off $u(t) \equiv 0$. The resulting dynamics $\dot{h}(t) = A(\Delta(t), 0)h(t)$ are determined entirely by the model's initialization, allowing us to isolate and understand its intrinsic stability and memory properties. As we demonstrate in our simulation study in Appendix A.2.1, this baseline behavior is a powerful predictor of a model's practical performance. We first show how strict dissipativity leads to exponential forgetting, and then prove that the unforced dynamics possess a fundamental, well-behaved quadratic energy structure.

4.1 Exponential Decay from Strict Dissipativity

The most fundamental stability guarantee arises when the system intrinsically dissipates energy faster than it stores it, even with zero input. This leads to exponential convergence of the state to the origin.

Theorem 4.1 (Exponential Decay from Strict Dissipativity)**.** Consider the continuous-time selective state-space model defined in Eq. equation 1. Suppose there exists a storage functional $V : [0, \infty) \times \mathbb{C}^N \to \mathbb{R}_{\geq 0}$ satisfying:

(i) **Strict Dissipativity Inequality:** For some constant $\beta > 0$, every admissible pair $(x(\cdot), u(\cdot))$ and the corresponding state–input–output trajectory $\{h(\tau), u(\tau), y(\tau)\}$ on any interval $[t_0, T]$ satisfies:

$$V\big(T,\, h(T)\big) - V\big(t_0,\, h(t_0)\big) \leq \int_{t_0}^{T} \operatorname{Re}\langle u(\tau), y(\tau)\rangle \, d\tau - \beta \int_{t_0}^{T} \|h(\tau)\|^2 \, d\tau. \tag{4}$$

(ii) **Regularity and Quadratic Bounds:** $V(t, h)$ is locally Lipschitz in h, absolutely continuous in t along system trajectories, and there exist constants $k_2 \geq k_1 > 0$ such that for all $t \geq 0$ and $h \in \mathbb{C}^N$:

$$k_1\|h\|^2 \leq V(t, h) \leq k_2\|h\|^2. \tag{5}$$

Then, for any admissible selection schedule $x(\cdot)$, the corresponding homogeneous (zero port-input) trajectory (i.e., when $u(t) \equiv 0$ for $t \geq 0$) exhibits exponential decay: there exist constants $C \geq 1$ and $\gamma > 0$ such that

for any initial state $h(0)$, the solution satisfies:

$$\|h(t)\| \leq C\, e^{-\gamma t}\, \|h(0)\| \quad \text{for all } t \geq 0. \tag{6}$$

Proof Sketch. The proof considers the system without port input, i.e., $u = 0$ (so the supply term $\operatorname{Re}\langle u, y\rangle$ vanishes). In this case, the strict dissipativity inequality simplifies to $\frac{dV}{dt} \leq -\beta\|h\|^2$. Since the energy function V is assumed to be quadratically bounded, satisfying $k_1\|h\|^2 \leq V \leq k_2\|h\|^2$, this implies $\frac{dV}{dt} \leq -\gamma V$ for some constant $\gamma > 0$. This is a classic differential inequality whose solution is an exponential decay. The bounds on V then directly translate this exponential decay of energy into an exponential decay of the state norm $\|h(t)\|$. Check Appendix A.3 for the full proof. □

4.2 Inherent Quadratic Structure and Regularity in Frozen-Selection LTV Dynamics

While Theorem 4.1 guarantees stability under strict dissipativity, it does not specify the form of the storage function V. We now connect selective SSMs to the passive LTV framework of Morandin & Hinsen (2024). The key point is that the results of Morandin & Hinsen (2024) apply to *input–output* LTV systems of the form

$$\dot{h}(t) = A(t)h(t) + B(t)u(t), \qquad y(t) = C(t)h(t) + D(t)u(t), \tag{7}$$

with mild regularity on (A, B, C, D) and passivity defined via the supply rate $\Re\langle u, y\rangle$ (Morandin & Hinsen, 2024, Eq. (1), Def. 1.1). Accordingly, instead of removing the input channel, we *freeze the selection signal* (e.g., set $x(t) \equiv 0$) and obtain a genuine LTV input–output subsystem with port input $u(\cdot)$. For passive LTV systems, the available storage is finite and coincides with a minimal quadratic storage $V_a(t, h) = \frac{1}{2}h^H Q(t)h$ with $Q \in \mathrm{AUC}_{\mathrm{loc}}$ (Morandin & Hinsen, 2024, Def. 4.1, Cor. 4.3). Evaluating this quadratic storage along homogeneous (zero port-input) trajectories $u \equiv 0$ yields a canonical quadratic energy functional for the frozen-selection baseline dynamics (i.e., $x \equiv 0$ and $u \equiv 0$).

To prove this, we leverage the concept of the *available storage* function, which represents the maximum energy that can be extracted from the system from a given state. It is a standard result in dissipativity theory that if any storage function exists, then the available storage function is well-defined and is itself a valid storage function (see Appendix A.4). The following theorem connects this concept to the structure of selective SSMs.

Theorem 4.2 (Existence and $\mathrm{AUC}_{\mathrm{loc}}$ regularity of a minimal quadratic storage for the frozen-selection LTV subsystem)**.** Consider the selective SSM equation 1 with selection signal $x(\cdot)$ and driving input $u(\cdot)$. Assume the coefficient maps satisfy the standard well-posedness regularity: for admissible $x(\cdot)$ we have

$$A(\Delta(\cdot), x(\cdot)) \in L^1_{\mathrm{loc}}(\mathbb{T}), \quad B(\Delta(\cdot), x(\cdot)) \in L^2_{\mathrm{loc}}(\mathbb{T}), \quad C(\Delta(\cdot), x(\cdot)) \in L^2_{\mathrm{loc}}(\mathbb{T}),$$

and $u(\cdot) \in L^2_{\mathrm{loc}}(\mathbb{T})$. (Here $D \equiv 0$, hence $D \in L^\infty_{\mathrm{loc}}$ holds trivially.)

Suppose there exists a storage functional $V : \mathbb{T} \times \mathbb{C}^N \to \mathbb{R}_{\geq 0}$ with $V(t, 0) = 0$ such that for some $\beta \geq 0$ the dissipativity inequality

$$V\big(T, h(T)\big) - V\big(t_0, h(t_0)\big) \leq \int_{t_0}^{T} \operatorname{Re}\langle u(\tau), y(\tau)\rangle\, d\tau - \beta \int_{t_0}^{T} \|h(\tau)\|^2\, d\tau \tag{8}$$

holds for all admissible pairs $(x(\cdot), u(\cdot))$ and all corresponding trajectories of equation 1. (In particular, dropping the nonpositive term $-\beta \int \|h\|^2$ yields the standard passivity inequality with supply rate $\Re\langle u, y\rangle$.)

Fix the *frozen-selection* signal $x(t) \equiv 0$ and define the associated LTV input–output system

$$\begin{cases} \dot{h}(t) = A_0(t)h(t) + B_0(t)u(t), \\ y_0(t) = C_0(t)h(t), \end{cases} \qquad \text{where} \quad A_0(t) = A(\Delta(t), 0),\ B_0(t) = B(\Delta(t), 0),\ C_0(t) = C(\Delta(t), 0). \tag{9}$$

Then the following hold:

(a) **Passivity of the frozen-selection LTV system.** The LTV system equation 9 is passive in the sense of Morandin & Hinsen (2024, Def. 1.1). Indeed, restricting equation 8 to $x(\cdot) \equiv 0$ yields a dissipation inequality for every state–input–output solution (h, u, y_0) of equation 9.

(b) **Existence of minimal quadratic available storage.** Let $V_{a,0}$ denote the available storage of equation 9 in the sense of Morandin & Hinsen (2024, Def. 4.1). Since equation 9 is passive and satisfies the regularity assumptions of Morandin & Hinsen (2024, Eq. (1)), its available storage is finite and coincides with a quadratic storage,

$$V_{a,0}(t, h) = \tfrac{1}{2} h^H Q_0(t) h,$$

for some matrix function $Q_0 : \mathbb{T} \to \mathbf{S}_+^N(\mathbb{C})$, and $V_{a,0}$ is minimal among all storage functions (Morandin & Hinsen, 2024, Cor. 4.3).

(c) $\mathrm{AUC}_{\mathrm{loc}}$ **regularity of** $Q_0(t)$**.** The inducing matrix function can be chosen with $\mathrm{AUC}_{\mathrm{loc}}$ regularity:

$$Q_0 \in \mathrm{AUC}_{\mathrm{loc}}(\mathbb{T}, \mathbf{S}_+^N(\mathbb{C})),$$

again by Morandin & Hinsen (2024, Cor. 4.3).

Finally, the homogeneous (unforced) dynamics of the frozen-selection system correspond to setting $u \equiv 0$ in equation 9, i.e., $\dot{h} = A_0(t)h$. If, in addition, V satisfies the hypotheses of Theorem 4.1 (strict dissipativity $\beta > 0$ and quadratic bounds), then these homogeneous trajectories decay exponentially as stated in Theorem 4.1.

Proof Sketch. The frozen-selection system equation 9 is an LTV input–output system of the form

$$\dot{h}(t) = A_0(t)h(t) + B_0(t)u(t), \qquad y_0(t) = C_0(t)h(t) + D_0(t)u(t),$$

with $D_0 \equiv 0$. Under our assumptions we have $A_0 \in L^1_{\mathrm{loc}}$, $B_0, C_0 \in L^2_{\mathrm{loc}}$, and $D_0 \in L^\infty_{\mathrm{loc}}$, hence equation 9 fits the setting of Morandin & Hinsen (2024, Eq. (1)).

Since equation 8 holds for all admissible $(x(\cdot), u(\cdot))$, it holds in particular when restricting to $x(\cdot) \equiv 0$. Dropping the nonpositive term $-\beta \int \|h\|^2$ yields the passivity dissipation inequality of Morandin & Hinsen (2024, Def. 1.1) for equation 9, so the frozen-selection LTV system is passive.

For passive LTV systems, Morandin & Hinsen (2024, Cor. 4.3) shows that the available storage (Morandin & Hinsen, 2024, Def. 4.1) is finite and coincides with a minimal quadratic storage $V_{a,0}(t, h) = \frac{1}{2} h^H Q_0(t) h$ with $Q_0 \in \mathrm{AUC}_{\mathrm{loc}}$. Finally, the homogeneous dynamics correspond to fixing $u \equiv 0$ (cf. Morandin & Hinsen (2024, Sec. 2.1)), and exponential decay under strict dissipativity follows from Theorem 4.1. □

5 Constraints Imposed by Universal Quadratic Passivity on Gating Mechanisms

Building on the foundational properties of the unforced system, this section studies the inherent design constraints of selective SSMs. We investigate the consequences of a strong but informative assumption: that a *single*, universal quadratic storage function $V_Q(t, h)$ guarantees passivity across all possible selection-driven (schedule-dependent) dynamics. The power of this assumption lies not in its generality, but in the strict, necessary consequences it imposes on the gating mechanism. This approach allows us to derive concrete constraints in the form of LMIs and to formalize a notion of "irreversible forgetting".

5.1 Regularity and Rank Monotonicity of Universal Quadratic Storage

If a single quadratic function V_Q is capable of certifying passivity regardless of the input-driven variations in system parameters, the matrix $Q(t)$ defining this function must possess inherent structural regularity and obey specific monotonicity properties.

Theorem 5.1 (Regularity and Rank of Universal Quadratic Storage)**.** Consider the continuous-time selective state-space model equation 1 under the standard regularity assumptions. Suppose there exists a time-varying

quadratic storage function $V_Q(t,h) = \frac{1}{2}h^H Q(t)h$ (with $Q : \mathbb{T} \to \mathbf{S}_+^N(\mathbb{C})$) that satisfies the dissipativity inequality (which implies passivity by dropping the nonpositive term):

$$V_Q\big(T,\, h(T)\big) - V_Q\big(t_0,\, h(t_0)\big) \leq \int_{t_0}^{T} \operatorname{Re}\langle u(\tau), y(\tau)\rangle \, d\tau - \beta \int_{t_0}^{T} \|h(\tau)\|^2 \, d\tau \tag{10}$$

for some $\beta \geq 0$, valid for *all* admissible pairs $(x(\cdot), u(\cdot))$ and all corresponding state–input–output trajectories $\{h(\tau), u(\tau), y(\tau)\}$ of equation 1. Then, the matrix function $Q(t)$ must satisfy the following properties:

(a) **AUC Regularity:** $Q(t)$ must belong to the class of locally absolutely upper semicontinuous matrix functions, i.e., $Q \in \mathrm{AUC}_{\mathrm{loc}}(\mathbb{T}, \mathbf{S}_+^N(\mathbb{C}))$.

(b) **Rank Monotonicity:** The rank of $Q(t)$, denoted $r(t) = \operatorname{rank}(Q(t))$, must be weakly monotonically non-increasing over time $t \in \mathbb{T}$.

Proof Sketch. The logic here is similar to the previous theorem but applied to the storage matrix $Q(t)$ itself rather than the minimal one. If a single quadratic function V_Q works for all admissible pairs $(x(\cdot), u(\cdot))$, then in particular it works for the frozen-selection LTV subsystem obtained by fixing $x(\cdot) \equiv 0$. Hence V_Q is a valid quadratic storage function for that passive LTV system, and the structural results of Morandin & Hinsen (2024) imply $Q \in \mathrm{AUC}_{\mathrm{loc}}$ and that $t \mapsto \operatorname{rank}(Q(t))$ is weakly non-increasing. The same theoretical results from Morandin & Hinsen (2024) that impose structure on the minimal storage function also apply to any valid quadratic storage function. This forces $Q(t)$ to have $\mathrm{AUC}_{\mathrm{loc}}$ regularity and, crucially, a non-increasing rank over time. Check Appendix A.6 for the full proof. □

5.2 Parametric LMI and Kernel Constraints on Gating

The existence of a universal quadratic storage function V_Q translates into concrete algebraic constraints that the selection-dependent matrices $A(t,x), B(t,x), C(t,x)$ must satisfy *parametrically* for all admissible *selection values* x. Equivalently, for each fixed selection value x and almost every time t, the resulting quadratic dissipation inequality must hold for all states h and all port inputs u. These constraints take the form of a parametric Linear Matrix Inequality (LMI) and lead to important conditions on how the output matrix $C(t,x)$ interacts with the kernel of the storage matrix $Q(t)$.

Theorem 5.2 (Passivity Constraints on Selection via a Parametric LMI)**.** Consider the selective SSM equation 1 with selection signal $x(\cdot)$ and driving input $u(\cdot)$ under the standard regularity assumptions. Assume the system's passivity is guaranteed by a *single*, time-varying quadratic storage function

$$V_Q(t,h) = \frac{1}{2}h^H Q(t)h, \qquad Q : \mathbb{T} \to \mathbf{S}_+^N(\mathbb{C}),$$

such that for some $\beta \geq 0$ the dissipativity inequality

$$V_Q\big(T,\, h(T)\big) - V_Q\big(t_0,\, h(t_0)\big) \leq \int_{t_0}^{T} \operatorname{Re}\langle u(\tau), y(\tau)\rangle \, d\tau - \beta \int_{t_0}^{T} \|h(\tau)\|^2 \, d\tau \tag{11}$$

holds for *all* initial states and for *all* admissible pairs $(x(\cdot), u(\cdot))$.

Then, for almost every $t \in \mathbb{T}$ and for every admissible selection value $x \in \mathbb{C}^{d_{\mathrm{in}}}$, the following parametric LMI holds:

$$\mathcal{L}(t,x) \coloneqq \begin{bmatrix} \dot{Q}(t) + Q(t)A(t,x) + A(t,x)^H Q(t) + 2\beta I & Q(t)B(t,x) - C(t,x)^H \\ B(t,x)^H Q(t) - C(t,x) & 0 \end{bmatrix} \preceq 0, \tag{12}$$

where $A(t,x) = A(\Delta(t),x)$, $B(t,x) = B(\Delta(t),x)$, $C(t,x) = C(\Delta(t),x)$, and $\dot{Q}(t)$ exists a.e.

In particular, the following universal kernel condition must hold:

$$\forall v \in \operatorname{Ker}\big(Q(t)\big), \quad C\big(\Delta(t), x\big)v = 0 \quad \text{for all admissible } x \text{ at time } t, \text{ a.e. } t \in \mathbb{T}. \tag{13}$$

Remark 5.3 (Coupled case $u = x$)**.** If the architecture uses a single signal (i.e., $u(\cdot) \equiv x(\cdot)$), then equation 12 remains a strong *sufficient* condition for passivity and motivates the regularizer used in Appendix A.2.3. Necessity holds under the stronger schedule-independent hypothesis stated in the theorem (passivity for all admissible pairs $(x(\cdot), u(\cdot))$).

Proof Sketch. This proof translates the differential form of the passivity inequality, $\frac{dV_Q}{dt} \leq \ldots$, into a matrix inequality. By substituting the system dynamics and the quadratic form of V_Q, we arrive at an expression that must be nonpositive for all states h and port inputs u, for each fixed selection value x (and for a.e. time t). Such an expression can be elegantly represented as a Linear Matrix Inequality (LMI). For the LMI to hold for all h and u (uniformly over admissible selection values x), specific conditions must be met. By choosing a state v from the kernel of Q (where $Qv = 0$), the LMI simplifies dramatically, revealing that the term involving the output matrix C must be zero to prevent the inequality from being violated. This leads to the necessary condition that $C(t, x)v = 0$. Check Appendix A.7 for the full proof. □

5.3 Irreversible Forgetting and Energy Consistency in Kernel Subspace

The constraints derived above, particularly the non-increasing rank of $Q(t)$ and the universal kernel condition for $C(t, x)$, suggest a form of structural irreversibility associated with the modes that fall into the kernel of the storage matrix $Q(t)$. We now explore this concept of "irreversible forgetting" from the perspective of the universal storage function V_Q and examine the energy balance required within this forgotten subspace.

Theorem 5.4 (Inertness of Forgotten Modes and Gating Constraints from Passivity)**.** Consider the continuous-time selective state-space model equation 1 under standard regularity assumptions. Assume its passivity is guaranteed by a *single*, time-varying quadratic storage function $V_Q(t, h) = \frac{1}{2}h^H Q(t)h$ with $Q : \mathbb{T} \to \mathbf{S}_+^N(\mathbb{C})$, satisfying the dissipativity inequality equation 11 for some $\beta \geq 0$ and for *all* admissible pairs $(x(\cdot), u(\cdot))$. Then, the following properties regarding "forgotten modes" (states in $\mathrm{Ker}(Q(t))$) hold:

1. **Irreversibility in storage dimension (rank monotonicity):** By Theorem 5.1, the matrix function Q can be chosen in $\mathrm{AUC}_{\mathrm{loc}}(\mathbb{T}, \mathbf{S}_+^N(\mathbb{C}))$ and the rank $r(t) = \mathrm{rank}(Q(t))$ is weakly monotonically non-increasing over time. Equivalently, since $Q(t) \in \mathbb{C}^{N\times N}$, $\dim \mathrm{Ker}(Q(t)) = N - r(t)$ is weakly non-decreasing.

2. **Energy Consistency within the Kernel:** For almost every $t \in \mathbb{T}$ where $\dot{Q}(t)$ exists, every direction $v \in \mathrm{Ker}(Q(t))$ satisfies the kernel energy inequality

$$v^H \dot{Q}(t)v + 2\beta\|v\|^2 \leq 0. \tag{14}$$

This ensures the evolution $\dot{Q}(t)$ within the kernel is consistent with the required dissipation β. If $\beta = 0$, $h^H \dot{Q}h \leq 0$; if $\beta > 0$, $h^H \dot{Q}h \leq -2\beta\|h\|^2 < 0$ (for $h \neq 0$).

3. **Constraint on selection schedules:** Under the universal passivity hypothesis, the parametric LMI equation 12 must hold for almost every t and for every admissible selection value x. Hence any selection schedule $x(\cdot)$ that attains values for which equation 12 fails on a set of times of positive measure would contradict the hypothesis. The dynamics must respect the energy structure defined by $Q(t)$ and $\dot{Q}(t)$ within the kernel.

Remark 5.5 (On strict dissipativity and rank-deficient storage)**.** If $\beta > 0$ and $Q(t) \succeq 0$, then the kernel inequality $v^H \dot{Q}(t)v + 2\beta\|v\|^2 \leq 0$ implies that any nonzero $v \in \mathrm{Ker}(Q(t))$ would satisfy $v^H \dot{Q}(t)v < 0$ at times where $\dot{Q}(t)$ exists, which would contradict the requirement $Q(t + \varepsilon) \succeq 0$ for small $\varepsilon > 0$. Thus, under state-strict dissipativity ($\beta > 0$), $Q(t)$ must be positive definite for almost every t, so the kernel-based "irreversible forgetting" discussion is primarily meaningful in the passive case $\beta = 0$.

Proof Sketch. This theorem's proof is a direct consequence of the preceding results.

1. **Kernel Non-Shrinking:** This follows immediately from the rank monotonicity established in Theorem 5.1; if the rank cannot increase, the dimension of the kernel ($N -$ rank) cannot decrease.

2. **Energy Consistency:** This is derived by taking the (1,1) block of the LMI from Theorem 5.2 and evaluating it for a state h in the kernel of Q. This isolates the term $h^H \frac{dQ}{dt} h$ and shows it must be negative enough to account for the required energy dissipation.

3. **Constraint on Dynamics:** This is a logical conclusion: any selection signal x that would induce dynamics violating the LMI contradicts the fundamental assumption of universal passivity, and is therefore inadmissible.

Check Appendix A.8 for the full proof. □

6 Robust Stability under Input-Driven Dynamics

Having established the baseline properties of the unforced system and the structural constraints imposed by passivity, we now arrive at the ultimate goal: analyzing the selective SSM in the fully general and realistic case where it is driven by persistent, non-zero *port inputs* $u(\cdot)$ under an exogenous selection schedule $x(\cdot)$. This section provides sufficient conditions for Input-to-State Stability (ISS), the gold standard for robust stability in nonlinear systems. To achieve this strong guarantee, we must impose conditions that ensure the system's internal dynamics are powerful enough to overcome any disturbance from the input. These conditions, while strong, provide a clear, verifiable pathway to designing certifiably robust models.

6.1 Global ISS from Uniform Dissipativity

To guarantee robust stability in the presence of arbitrary bounded inputs, basic passivity is not enough. We need to ensure the system's internal dynamics are uniformly contractive, that is, they dissipate energy at a guaranteed rate, regardless of which operating mode the *selection signal* $x(\cdot)$ selects. This ensures the system can actively counteract the energy being injected by the *port input* $u(\cdot)$.

With this subsection, we present such a condition based on the existence of a common quadratic Lyapunov function demonstrating uniform dissipativity.

Theorem 6.1 (Global Stability from Uniform Local Dissipativity)**.** Consider the continuous-time selective state-space model equation 1 under the standard regularity assumptions. Assume further that:

(i) There exists a time-varying quadratic Lyapunov function candidate $V_Q(t,h) = \frac{1}{2}h^H Q(t)h$, where $Q : \mathbb{T} \to \mathbf{S}_+^N(\mathbb{C})$ satisfies:

- $Q \in W^{1,1}_{\text{loc}}(\mathbb{T}, \mathbf{S}_+^N(\mathbb{C}))$ (absolutely continuous locally).
- $Q(t)$ is uniformly positive definite and bounded: there exist constants $k_2 \geq k_1 > 0$ such that $k_1 I \preceq Q(t) \preceq k_2 I$ for all $t \in \mathbb{T}$.

(ii) The dynamics satisfy a uniform dissipativity condition with respect to V_Q: There exists a constant $\delta > 0$ such that for almost every $t \in \mathbb{T}$ and for **all** admissible *selection* values $x \in \mathbb{C}^{d_{\text{in}}}$ that can occur at time t, the following matrix inequality holds:

$$\dot{Q}(t) + Q(t)A\big(\Delta(t), x\big) + A\big(\Delta(t), x\big)^H Q(t) \preceq -2\delta Q(t) \tag{15}$$

(This implies that for any fixed selection value x, the homogeneous LTV system $\dot{h} = A(\Delta(t), x)h$ satisfies $\dot{V}_Q \leq -2\delta V_Q$ and hence is uniformly exponentially stable; equivalently, $\|h(t)\| \leq \sqrt{k_2/k_1}\, e^{-\delta(t-t_0)}\|h(t_0)\|$.)

(iii) The input matrix $B(\Delta(t), x)$ is uniformly bounded: There exists a constant $M_B > 0$ such that $\|B(\Delta(t), x)\|_2 \leq M_B$ for all t and admissible x.

Then, the selective state-space system equation 1 is globally ISS with respect to the *port input* $u(t)$, uniformly over all admissible selection schedules $x(\cdot)$. Specifically, there exist constants $\tilde{C} \geq 1$, $\tilde{\gamma} > 0$, and a class $\mathcal{K}$ gain function σ such that for any initial state $h(t_0)$, any admissible selection schedule $x(\cdot)$, and any admissible port input $u(\cdot)$, the solution satisfies:

$$\|h(t)\| \leq \tilde{C}e^{-\tilde{\gamma}(t-t_0)}\|h(t_0)\| + \sigma\left(\|u\|_{[t_0,t],\infty}\right), \qquad \|u\|_{[t_0,t],\infty} := \operatorname*{ess\,sup}_{t_0 \leq \tau \leq t} \|u(\tau)\|. \quad \text{for all } t \geq t_0. \tag{16}$$

Proof Sketch. We analyze the time derivative of the Lyapunov function V_Q along the system trajectories. The derivative splits into two parts: one from the internal dynamics (A matrix) and one from the external *port input* (B matrix). The uniform dissipativity assumption provides a strong negative bound on the internal part ($\leq -2\delta V_Q$). The input part is bounded using the Cauchy-Schwarz inequality. Combining these results in a differential inequality of the form $\frac{dV}{dt} \leq -aV + b\sqrt{V}$. By a change of variables, $\Psi = \sqrt{V}$, this transforms into a standard linear differential inequality, $\frac{d\Psi}{dt} \leq -c\Psi + d$, whose solution is bounded. Translating this bound back to the state norm $\|h(t)\|$ yields the classic Input-to-State Stability (ISS) estimate. Check Appendix A.9 for the full proof. □

Remark 6.2 (Relationship to the Passivity LMI)**.** The ISS condition in Eq. equation 15 is conceptually stronger than the passivity LMI from Theorem 5.2. The passivity LMI describes an **energy balance**, ensuring the system does not generate energy. In contrast, the ISS condition demands a forced **energy decay**, ensuring the system's internal dynamics actively dissipate energy at a uniform exponential rate.

Formally, if the ISS condition holds (with $\delta > 0$ and a positive definite Q), it is sufficient to satisfy the internal stability portion (the (1,1) block) of the passivity LMI. However, a system can be passive (energy-balanced) without being uniformly contractive in this stricter sense. Thus, the ISS condition is a specialized and powerful tool for proving robust stability, while the passivity LMI is a more general tool for analyzing a system's energy flow.

7 Conclusion

In this paper, we established a rigorous bridge between the control-theoretic frameworks of passivity and Input-to-State Stability (ISS) and the complex, input-dependent dynamics of modern selective SSMs. Our work provides a new language for analyzing these models and demonstrates that even with discontinuous gating, their stability is governed by well-defined principles of energy management.

Our theoretical findings translate directly into three key practical insights. First, we provide a formal rationale for the critical role of principled initialization (e.g., HiPPO), explaining it as the design of a desirable baseline energy landscape for the system's intrinsic dynamics (see Appendix A.1.1). Second, we have shown a demonstrated path to robust model design by deriving a parametric passivity LMI and implementing it as a practical regularizer. While the LMI is necessary under the schedule-independent universal-passivity hypothesis, it remains a strong sufficient constraint in the common coupled setting $u \equiv x$. Our experiments confirm this regularizer dramatically improves training stability with negligible impact on task performance, offering a concrete tool for building more reliable models (see Appendix A.1.2). Finally, our analysis of structural constraints and "irreversible forgetting" (under the strong universal quadratic storage hypothesis, and primarily in the passive case $\beta = 0$) offers a new lens through which to understand the fundamental trade-offs between stability and expressivity in these powerful architectures (see Appendix A.1.3).

While significant challenges remain, particularly in translating these guarantees to the discrete-time domain and scaling our methods, this work lays a critical foundation. By connecting abstract stability concepts to concrete, verifiable properties of the model, our framework paves the way for a new class of certifiably robust and interpretable models for sequential data processing.

References

Brian DO Anderson and Sumeth Vongpanitlerd. *Network analysis and synthesis: a modern systems theory approach*. Courier Corporation, 2013.

Michael S Branicky. Multiple lyapunov functions and other analysis tools for switched and hybrid systems. *IEEE Transactions on automatic control*, 43(4):475–482, 1998.

Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David Kristjanson Duvenaud. Neural ordinary differential equations. In *Neural Information Processing Systems*, 2018. URL `https://api.semanticscholar.org/CorpusID:49310446`.

Nicola Muca Cirone, Antonio Orvieto, Benjamin Walker, Cristopher Salvi, and Terry Lyons. Theoretical foundations of deep selective state-space models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL `https://openreview.net/forum?id=3SzrqwupUx`.

Earl A Coddington, Norman Levinson, and T Teichmann. Theory of ordinary differential equations, 1956.

AF Filippov. Existence and general properties of solutions of discontinuous systems. In *Differential Equations with Discontinuous Righthand Sides*, pp. 48–122. Springer, 1988.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. URL `https://openreview.net/forum?id=tEYskw1VY2`.

Albert Gu, Tri Dao, Stefano Ermon, Atri Rudra, and Christopher Ré. Hippo: Recurrent memory with optimal polynomial projections. *Advances in Neural Information Processing Systems*, 33, 2020.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *The International Conference on Learning Representations (ICLR)*, 2022.

John T Halloran, Manbir Gulati, and Paul F Roysdon. Mamba state-space models are lyapunov-stable learners. *arXiv preprint arXiv:2406.00209*, 2024.

Alberto Isidori. *Nonlinear control systems: an introduction*. Springer, 1985.

Zhong-Ping Jiang, Eduardo D. Sontag, and Yuan Wang. Input-to-state stability for discrete-time nonlinear systems. *IFAC Proceedings Volumes*, 32:2403–2408, 1999. URL `https://api.semanticscholar.org/CorpusID:12463317`.

Rudolf E Kalman. Lyapunov functions for the problem of lur'e in automatic control. *Proceedings of the national academy of sciences*, 49(2):201–205, 1963.

Hassan K Khalil and Jessy W Grizzle. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ, 2002.

Patrick Kidger, James Morrill, James Foster, and Terry Lyons. Neural controlled differential equations for irregular time series. *Advances in neural information processing systems*, 33:6696–6707, 2020.

Daniel Liberzon. *Switching in systems and control*, volume 190. Springer, 2003.

Terry Lyons. Rough paths, signatures and the modelling of functions on streams. *arXiv preprint arXiv:1405.4537*, 2014.

Riccardo Morandin and Dorothea Hinsen. Dissipative energy functionals of passive linear time-varying systems. In *ArXiv*, 2024. URL `https://api.semanticscholar.org/CorpusID:274981788`.

Vasile-Mihai Popov. Absolute stability of nonlinear systems of automatic control. *Avtomatika i telemekhanika*, 22(8):961–979, 1961.

Zhen Qin, Songlin Yang, Weixuan Sun, Xuyang Shen, Dong Li, Weigao Sun, and Yiran Zhong. Hgrn2: Gated linear rnns with state expansion. In *Proceedings of COLM*, 2024.

Jimmy T.H. Smith, Andrew Warrington, and Scott Linderman. Simplified state space layers for sequence modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Ai8Hw3AXqks.

Eduardo Sontag. Mathematical control theory: Deterministic finite dimensional systems. In *Springer*, 1990. URL https://api.semanticscholar.org/CorpusID:18201935.

Eduardo Sontag and Yuan Wang. On characterizations of the input-to-state stability property. *Systems & Control Letters*, 24:351–359, 1995. URL https://api.semanticscholar.org/CorpusID:13893551.

Eduardo D Sontag et al. Smooth stabilization implies coprime factorization. *IEEE transactions on automatic control*, 34(4):435–443, 1989.

Taylan Soydan, Nikola Zubic, Nico Messikommer, Siddhartha Mishra, and Davide Scaramuzza. S7: Selective and simplified state space layers for sequence modeling. *ArXiv*, abs/2410.03464, 2024. URL https://api.semanticscholar.org/CorpusID:273162247.

Arjan Van der Schaft. *L2-gain and passivity techniques in nonlinear control*. Springer, 2000.

Jan C. Willems. Dissipative dynamical systems part i: General theory. *Archive for Rational Mechanics and Analysis*, 45:321–351, 1972. URL https://api.semanticscholar.org/CorpusID:123076101.

Songlin Yang, Bailin Wang, Yikang Shen, Rameswar Panda, and Yoon Kim. Gated linear attention transformers with hardware-efficient training. In *Proceedings of ICML*, 2024.

Nikola Zubic, Federico Soldà, Aurelio Sulser, and Davide Scaramuzza. Limits of deep learning: Sequence modeling through the lens of complexity theory. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=DhdqML3FdM.