

Winner-Take-All Computation in Spiking Neural Networks

Nancy Lynch
MIT
lynch@csail.mit.edu

Cameron Musco
Microsoft Research
camusco@microsoft.com

Merav Parter
Weizmann Institute of Science
merav.parter@weizmann.ac.il

April 30, 2019

Abstract

In this work we study biological neural networks from an algorithmic perspective, focusing on understanding tradeoffs between computation time and network complexity. Our goal is to abstract real neural networks in a way that, while not capturing all interesting features, preserves high-level behavior and allows us to make biologically relevant conclusions. Towards this goal, we consider the implementation of algorithmic primitives in a simple yet biologically plausible model of *stochastic spiking neural networks*. In particular, we show how the stochastic behavior of neurons in this model can be leveraged to solve a basic *symmetry-breaking task* in which we are given neurons with identical firing rates and want to select a distinguished one. In computational neuroscience, this is known as the winner-take-all (WTA) problem, and it is believed to serve as a basic building block in many tasks, e.g., learning, pattern recognition, and clustering. We provide efficient constructions of WTA circuits in our stochastic spiking neural network model, as well as lower bounds in terms of the number of auxiliary neurons required to drive convergence to WTA in a given number of steps. These lower bounds demonstrate that our constructions are near-optimal in some cases.

This work covers and gives more in-depth proofs of a subset of results originally published in [LMP17a]. It is adapted from the last chapter of C. Musco’s Ph.D. thesis [Mus18].

1 Background and Introduction to Results

Neural networks are studied in a number of academic communities from a wide range of perspectives. Significant work in computational neuroscience focuses on developing somewhat realistic mathematical models for these networks and generally studying their capacity to process information [Izh04, Tra09]. On the more theoretical side, a variety of artificial network models such as perceptron and sigmoidal networks, Hopfield networks, and Boltzmann machines have been developed. These models are tractable to theoretical analysis and studied in the context of their computational power, and applications to general function approximation, classification, and memory storage [HSW89, MSS91, SS95, Maa97]. In practical machine learning, biological fidelity and often theoretical tractability are put aside, and researchers study how neural-like networks and learning rules can be used to efficiently represent and learn complex concepts [Hay09, LBH15].

In contrast to the common approach in computational neuroscience and machine learning, in our work we focus not on general computation ability or broad learning tasks, but on specific algorithmic implementation and analysis. We define a model of neural computation along with

algorithmic problems that seem to be an important building blocks for higher level processing and learning tasks. We then design neural networks in our model that solve these problems, rigorously analyzing the complexity of our solutions in terms of asymptotic runtime and network size bounds. We hope that this new paradigm will provide new insights about computational tradeoffs, the power of randomness, and the role of noise in biological systems.

While focusing on somewhat different questions, our line of work is inspired by (1) work on the computational power of spiking neural networks, most notably by Maass et al. [Maa97, Maa99, Maa00] and (2) the work of Les Valiant [Val00a, Val00b, Val05], who defined the *neuroidal model* of computation and investigated implementations of basic learning modules within this model.

1.1 Spiking Neural Networks

We consider a model of *spiking neural networks* (SNNs) [Maa96, Maa97, GK02, Izh04, HJM13], defined formally in Section 2, in which neurons fire in discrete pulses, in response to a sufficiently high membrane potential. This potential is induced by spikes from neighboring neurons, which can have either an excitatory or inhibitory effect (increasing or decreasing the potential). Our model is *stochastic* – each neuron functions as a probabilistic threshold unit, spiking with probability given by applying a sigmoid function to the membrane potential. In this respect, our networks are similar to the popular Boltzmann machine [AHS85], with the important distinction that synaptic weights are not required to be symmetric and, as observed in nature, neurons are either strictly inhibitory (all outgoing edge weights are negative) or excitatory. Additionally, in this work, we focus on networks with fixed edge weights. The literature on Boltzmann machines tends to focus on learning, in which edge weights are adjusted iteratively until the network converges to some desired distribution on firing patterns. While a rich literature focuses on deterministic threshold circuits [MP69, HT⁺86] we employ a stochastic model as it is widely accepted that neural computation is inherently stochastic [AS94, SN94, FSW08], and that while this can lead to a number of challenges, it also affords significant computational advantages [Maa14].

1.2 The Winner-Take-All Problem

In this work, we focus on the Winner-Take-All (WTA) problem, which is one of the most studied problems in computational neuroscience. A WTA network has n input neurons, n corresponding outputs, and a set of auxiliary neurons that facilitate computation. The goal is to pick a ‘winning’ input – that is, the network should produce a single firing output which corresponds to a firing input. Often the winning input is the one with the highest firing rate, in which case WTA serves as a neural max function. We focus on the case when all inputs have the same or similar firing rates, in which case WTA serves as a leader election unit. A formal definition of the WTA problem is given in Section 2.6.

WTA is widely applicable, including in circuits that implement visual attention via WTA competition between groups of neurons that process different input classes [KU87, LIKB99, IK01]. It is also the foundation of competitive learning [Now89, KK94, GL09], in which classifiers compete to respond to specific input types. More broadly, WTA is known to be a powerful computational primitive [Maa99, Maa00] – a network equipped with WTA units can perform some tasks significantly more efficiently than with just linear threshold neurons (McCulloch-Pitts neurons or perceptrons).

Due to its importance, there has been significant work on WTA, including in biologically plausible spiking networks [LRMM88, YG89, Tho90, CGL92, WS03, OL06, ODL09, ASNN⁺15]. This work is extremely diverse – while mathematical analysis is typically given, different papers show different guarantees and apply varying levels of rigor. To the best of our knowledge, prior to our

work, no asymptotic time bounds (e.g., as a function of the number of inputs n) for solving WTA in spiking neural networks have been established.¹ Additionally, previous analysis often requires a specific initial network state to show convergence and does not show that the network is self-stabilizing and converges from an arbitrary starting state, as is necessary in a biological system.

1.3 Our Contributions

We explore the tradeoff between the number of auxiliary neurons used in a WTA network (i.e., the complexity of the network) and the time required to select a winning output (to converge to a WTA state).

1.3.1 Network Constructions and Runtime Bounds

On the upper bound side, in Section 3 we describe, for any input size n and failure probability $\delta > 0$, a family of networks using just two auxiliary inhibitory neurons which solve the WTA problem in $O(\log n)$ steps in expectation, and with probability $\geq 1 - \delta$ in $O(\log n \cdot \log(1/\delta))$ steps (see Theorems 3.2 and 3.3).

Our two-inhibitor construction is based on a simple random competition idea. Outputs that fire in response to stimulation from their firing inputs excite two inhibitors, which, in turn, inhibit all the outputs. When more than one output fires, both inhibitors are excited. This leads to high levels of inhibition, causing firing outputs to stop firing and ‘drop out’ of the WTA competition. When exactly one output fires, just one of the inhibitors, known as the *stability inhibitor*, is excited. This inhibitor is responsible for maintaining a WTA steady-state: once a single output fires at a time step it becomes the winner of the network. It has a positive feedback self-loop that allows it to keep firing at subsequent times, while all other outputs do not fire due to inhibition from the stability inhibitor.

The basic network construction described above employs biologically plausible structures. In particular, convergence is driven using reciprocal excitatory-inhibitory connections, and stability is maintained via excitatory self-loops. Both these structures are used in many biological models of WTA computation [YG89, CGL92, RB15]. It is widely believed that inhibition is crucial for solving WTA – outputs compete for activation via *lateral inhibition* or *recurrent inhibition* [CGL92, RB15]. In our network, this inhibition is achieved through the two auxiliary inhibitors. Previous work has conjectured that widespread use of simple WTA implementations in the brain may explain how complex computation is possible even when inhibition is relatively limited and localized [Maa00]. Our work shows that WTA can be achieved and maintained efficiently using very few inhibitors and with a very simple connectivity structure.

We also demonstrate that, with a larger number of auxiliary neurons, it is possible to obtain faster convergence. In particular, in Section 5, we describe, for any input size n and failure probability $\delta > 0$, a family of networks using $O(\log n)$ auxiliary inhibitory neurons which solve the WTA problem in $O(1)$ steps in expectation, and with probability $\geq 1 - \delta$ in $O(\log(1/\delta))$ steps (see Theorems 5.2 and 5.3). At a high level, more auxiliary inhibitors allow for more fine-tuned levels of inhibition which drive faster convergence. In Section 5.7 we sketch two constructions that allow for more general runtime-inhibitor tradeoffs, interpolating between our two-inhibitor and $O(\log n)$ -inhibitor constructions. See [LMP17a] for more details.

¹Aside from immediate bounds for deterministic circuits using many ($\Omega(n)$) auxiliary neurons [LRMM88, Maa00].

1.3.2 Lower Bounds

Aside from the above network constructions and runtime analysis, we also prove lower bounds, showing that these constructions are optimal or near optimal. In Section 4 we prove that no network can solve WTA (in a reasonable parameter regime) using just a single auxiliary neuron (see Theorem 4.3). Roughly, it is not possible for a single neuron to both drive fast convergence and maintain stability of a valid WTA configuration once one has been reached. The dual role that inhibition plays in two-inhibitor construction of driving convergence and maintaining stability requires at least two inhibitors. We also show that, considering a slightly restricted class of networks, our two-inhibitor construction is near-optimal. No network with just two-auxiliary neurons can solve WTA with constant probability in $o\left(\frac{\log n}{\log \log n}\right)$ steps (see Theorem 4.14). This matches the runtime of our network up to a $O(\log \log n)$ factor. See [LMP17a] for additional lower bounds on the convergence time for networks using $\alpha > 2$ inhibitors, in a similar model to the one presented here.

1.4 Road Map

In Section 2 we describe our spiking neural network model and specify the WTA problem. In Section 3 we describe and analyze the convergence of our simple family of two-inhibitor WTA networks. In Section 4 we prove lower bounds that show the near optimality of our two-inhibitor construction. In Section 5 we show how to obtain faster convergence using a network construction with $O(\log n)$ auxiliary inhibitors. This construction also requires generalizing our model to allow for a *history period*, over which the firing of a neuron’s neighbors can affect its membrane potential. Finally, in Section 6 we conclude by discussing open questions arising from our work and possible directions for future work.

2 Spiking Neural Network Model

In this section we describe our basic neural network model, which consists of a set of neurons connected by weighted synaptic connections. Each neuron fires (spikes) stochastically at each time step, with probability dependent on the firing of its neighbors in the previous time step. These neighbors may have either an excitatory (inducing more firing) or inhibitory (suppressing firing) effect. In Section 5 we describe a variation on this model in which the probability that a neuron spikes depends not just on the spiking of its neighbors in the previous time step, but on the spikes during some history period preceding the current time.

2.1 Network Structure

We first describe the basic network structure and parameters. A *Spiking Neural Network* (SNN) $\mathcal{N} = \langle N, w, b, f \rangle$ consists of:

- N , a set of neurons, partitioned into a set of input neurons X , a set of output neurons Y , and a set of auxiliary neurons A . N is also partitioned into a set of *excitatory* and *inhibitory* neurons E and I . All input and output neurons are excitatory. That is, $X \cup Y \subseteq E$.
- $w : N \times N \rightarrow \mathbb{R}$, a weight function describing the weighted synaptic connections between the neurons in the network. w is restricted in a few notable ways:

$$- w(u, x) = 0 \text{ for all } u \in N, x \in X.$$

- Each excitatory neuron $v \in E$ has $w(v, u) \geq 0$ for every u . Each inhibitory neuron $v \in I$ has $w(v, u) \leq 0$ for every u .
- $b : N \rightarrow \mathbb{R}$, a bias function, assigning an activation bias to each neuron.
- $f : \mathbb{R} \rightarrow [0, 1]$, a spike probability function, satisfying a few restrictions:
 - f is continuous and monotonically increasing.
 - $\lim_{x \rightarrow \infty} f(x) = 1$ and $\lim_{x \rightarrow -\infty} f(x) = 0$.

2.1.1 Remarks on Network Structure

Before describing the dynamics of our neural network, we give a few remarks on, and explanations of, the above parameters determining the network structure.

Weight Function (w): The weight function w describes the strength of the synaptic connections between neurons in N . The restriction that $w(u, x) = 0$ for every input neuron $x \in X$ is motivated by the desire for networks to be composable. The input neurons in X may be output neurons of another network, and so incoming connections are avoided to simplify definitions and analysis when networks are composed in higher level modular designs.

The restriction that each neuron v is either inhibitory or excitatory is motivated by the observation, known as *Dale's principle*, that neurons typically employ the same neurotransmitter at each outgoing synapse, regardless of its target [Osb13]. Thus, all outgoing connections are either inhibitory or excitatory, depending on the transmitter used. For example, inhibitory connections predominantly stem from inhibitory *GABAergic neurons*, which employ the neurotransmitter *gamma-Aminobutyric acid* (GABA) [WMK⁺02, RFLHL11].

We often view the weight function as defining the edge weights of a directed graph, whose edges are the synaptic connections. Formally we can define:

Definition 2.1 (Synaptic Connection Graph). *Given spiking neural network $\mathcal{N} = \langle N, w, b, f \rangle$, let $G(\mathcal{N})$ be the weighted directed graph with vertex set N and a directed edge (u, v) with weight $w(u, v)$ for all u, v with $w(u, v) \neq 0$.*

Note that the weight function $w(u, v)$ need not be symmetric, and typically will not be. Additionally, we allow $u \in N$ with $w(u, u) \neq 0$. That is, $G(\mathcal{N})$ may have self-loops.

Bias Function (b): The bias function, along with the spike probability function, determines how large a neuron's membrane potential must be for the neuron to spike with good probability. The larger the bias, the more excited the neuron must be before it fires. We will see in Section 2.2 exactly how the bias affects the spiking probability.

Spike Probability Function (f): Common choices for the spike probability function f are symmetric functions with $f(0) = \frac{1}{2}$. For example, we will typically set f to the sigmoid function $f(x) = \frac{1}{1+e^{-x/\lambda}}$ for some *temperature parameter* $\lambda > 0$.

2.2 Network Dynamics

We now describe in detail the dynamics of our neural network model.

A *configuration* $C : N \rightarrow \{0, 1\}$ of an SNN $\mathcal{N} = \langle N, w, b, f \rangle$ is a mapping from each neuron in the network to a firing state. $C(u) = 1$ indicates that u fires (i.e., generates a spike). $C(u) = 0$ indicates that it does not fire. We similarly define an *input configuration* $C_X : X \rightarrow \{0, 1\}$ to be a mapping from each input neuron to a firing state and an *output configuration* $C_Y : Y \rightarrow \{0, 1\}$

to be a mapping from each output neuron to a firing state. For any configuration C and set of neurons $M \subseteq N$, we let $C(M)$ be the restriction of C to the domain M .

An SNN evolves in a sequence of discrete, synchronous times, which we label with integers $t = 0, 1, \dots$. We denote the configuration at time t by N^t . Similarly, we denote the input and output configurations at time t by $X^t \stackrel{\text{def}}{=} N^t(X)$ and $Y^t \stackrel{\text{def}}{=} N^t(Y)$ respectively.

Formally, an *execution* is a finite or infinite sequence of configurations. The *length* of a finite execution $N^0 N^1 \dots N^t$ is defined to be $t + 1$. The length of an infinite execution $N^0 N^1 \dots$ is defined to be ∞ . We analogously define an *input execution* and an *output execution* as a sequence of input configurations $X^0 X^1 \dots$ and output configurations $Y^0 Y^1 \dots$ respectively.

For each neuron $u \in N$, we use the notation $u^t \stackrel{\text{def}}{=} N^t(u)$ to denote the firing state of neuron u in the configuration N^t . More generally, for any ordered set of neurons $M = \{m_1, \dots, m_n\}$ we let $M^t \in \{0, 1\}^n$ denote the binary vector with m_j^t as its j^{th} entry. For any configuration C , we let $\|C\|_1 = |\{u \in N : C(u) = 1\}|$ denote the number of firing neurons in C .

We will typically use α to denote an execution, and α_X, α_Y to denote an input or output execution respectively. We will use a superscript to denote the length of a finite execution. For any execution α let $\text{output}(\alpha)$ be the output execution of the same length obtained by restricting each configuration in α to the output neurons Y .

The behavior of an SNN is determined as follows:

- **Input Neurons:** For each problem we consider, we will specify how the infinite input execution $X^0 X^1 \dots$ is determined. In this work, we will typically fix the input so that for each $u \in X$, u^t is constant for all $t \geq 0$. However, we may also specify a distribution from which $X^0 X^1 \dots$ is drawn. For example, this sequence may be generated by setting $u^t = 1$ with some probability p_u and $u^t = 0$ with probability $1 - p_u$, independently at random for each $u \in X$ and each time t .
- **Initial Firing States:** For each non-input $u \in N \setminus X$, the firing state u^0 is arbitrary. In this work, we typically show convergence results that hold for all possible settings of these initial states, giving our networks a self-stabilizing property, since they will converge from any arbitrary perturbation of the state (see e.g., Theorem 2.8).
- **Firing Dynamics:** For each non-input neuron $u \in N \setminus X$ and every time $t \geq 1$, let $\text{pot}(u, t)$ denote the membrane potential at time t and $p(u, t)$ denote the corresponding firing probability. These values are calculated as:

$$\text{pot}(u, t) = \left(\sum_{v \in N} w(v, u) \cdot v^{t-1} \right) - b(u) \text{ and } p(u, t) = f(\text{pot}(u, t)) \quad (1)$$

where f is the spike probability function. At time t , each non-input neuron u fires independently with probability $p(u, t)$.

Any SSN $\mathcal{N} = \langle N, w, b, f \rangle$, initial configuration N^0 , and infinite input execution α_X define a probability distribution over infinite executions, $\mathcal{D}(\mathcal{N}, N^0, \alpha_X)$. This distribution is the natural distribution that follows from applying the stochastic firing dynamics of (1). Formally, for any finite execution α , we define the *cone* of executions extending α , $A(\alpha)$, to be the set of all infinite executions that start with α . $\mathcal{D}(\mathcal{N}, N^0, \alpha_X) : \mathcal{F} \rightarrow [0, 1]$ is a probability measure where the σ -algebra $\mathcal{F}$ consists of all such cones, closed under complement, countable unions, and countable intersections.

Given $\mathcal{D}(\mathcal{N}, N^0, \alpha_X)$ we can also define a distribution $\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X)$ on infinite output executions. Given any finite output execution α_Y , we define the cone $A(\alpha_Y)$ to be the set of all infinite

output executions extending α_Y . We define the σ -algebra $\mathcal{F}_Y$ to be the set of all such cones, closed under complement, countable union, and countable intersection. Finally, for $F_Y \in \mathcal{F}_Y$, we define $\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X) : \mathcal{F}_Y \rightarrow [0, 1]$ by:

$$\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X)[F_Y] = \mathcal{D}(\mathcal{N}, N^0, \alpha_X)[\{\alpha : \text{output}(\alpha) \in F_Y\}].$$

2.3 Problems and Solving Problems

A problem P is a mapping from an infinite input execution α_X to a set of output distributions. A network $\mathcal{N}$ is said to *solve problem P on input α_X* if, for any initial configuration N^0 , the output distribution $\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X)$ is an element of $P(\alpha_X)$. A network $\mathcal{N}$ is said to *solve problem P* if it solves P on every infinite input execution α_X . For an example of such a problem definition see Section 2.6, where we formally define the winner-take-all problem.

2.4 Basic Results and Properties of the Model

In this section we prove some basic properties of the spiking neural network model described in the preceding sections. The first property is a simple Markov independence claim: conditioned on the configuration at time $t - 1$, a network's execution from time t on is independent of all times before $t - 1$. Formally:

Lemma 2.2 (Markov Property). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be an SNN. For any time $t \geq 1$, and finite execution $C^0 \dots C^{t-1}$ of $\mathcal{N}$, and any configuration C of $\mathcal{N}$:*

$$\mathbb{P}[N^t = C | N^{t-1} N^{t-2} \dots N^0 = C^{t-1} C^{t-2} \dots C^0] = \mathbb{P}[N^t = C | N^{t-1} = C^{t-1}].$$

Proof. The potential of every $u \in N$ at time t as computed in (1) is determined by N^{t-1} . Thus, the spike probability $p(u, t) = f(\text{pot}(u, t))$ is fully determined by N^{t-1} .

So, conditioned on $N^{t-1} = C^{t-1}$, $p(u, t)$ is a deterministic function of C^{t-1} . We can compute:

$$\mathbb{P}[N^t = C | N^{t-1} = C^{t-1}] = \prod_{u \in N} [C(u) \cdot p(u, t) + (1 - C(u)) \cdot (1 - p(u, t))].$$

So $\mathbb{P}[N^t = C | N^{t-1} = C^{t-1}]$ is a deterministic function of C and all the $p(u, t)$ collectively and thus of C and C^{t-1} . So for *any* $C^0 \dots C^{t-2}$,

$$\mathbb{P}[N^t = C | N^{t-1} = C^{t-1}] = \mathbb{P}[N^t = C | N^{t-1} N^{t-2} \dots N^0 = C^{t-1} C^{t-2} \dots C^0],$$

giving the lemma. □

In our proofs, we will often bound the probability of some event $\mathcal{E}_t$ occurring at time t , giving a bound independent of the preceding network configuration N^{t-1} . However, $\mathcal{E}_t$ itself will depend on N^{t-1} , and there may be correlations between $\mathcal{E}_t$ and $\mathcal{E}_{t'}$ for $t \neq t'$. Below, we give a useful lemma which allows us to bound the number of times that $\mathcal{E}_t$ occurs over a given time period by comparing to the number of times that a coin tossed independently at each time would come up heads in the same time period.

Lemma 2.3. *For every $t \in \mathbb{Z}^{>0}$ let $A_t \in \mathcal{A}$ be a random variable in some domain $\mathcal{A}$, $f : \mathcal{A} \rightarrow \{0, 1\}$ be any function, and $B_t = f(A_t)$. Let $Z_t \in \{0, 1\}$ be a set of independent random variables. Suppose:*

1. $\mathbb{P}[B_1 = 1] \geq \mathbb{P}[Z_1 = 1]$.

2. For every $t \geq 2$, $\mathbb{P}[B_t = 1 | A_{t-1}, \dots, A_1] \geq \mathbb{P}[Z_t = 1]$.

Then for every t and $d \in \mathbb{Z}^{\geq 0}$,

$$\mathbb{P} \left[\sum_{i=1}^t B_i \geq d \right] \geq \mathbb{P} \left[\sum_{i=1}^t Z_i \geq d \right]. \quad (2)$$

Lemma 2.3 and its proof are similar to Lemma 2.2 of [KKKL11]. However, we include a full proof for completeness and since we are in a slightly different setting, where we condition on the full past state, rather than just the preceding values of $B_t = f(X_t)$.

Proof. We prove the result via induction on t . The base case with $t = 1$ is given by assumption (1). For any $t > 1$, assuming that (2) holds for all $t' < t$, we have:

$$\begin{aligned} \mathbb{P} \left[\sum_{i=1}^t B_i \geq d \right] &= \mathbb{P} \left[B_t = 1 \mid \sum_{i=1}^{t-1} B_i = (d-1) \right] \cdot \mathbb{P} \left[\sum_{i=1}^{t-1} B_i = (d-1) \right] + \mathbb{P} \left[\sum_{i=1}^{t-1} B_i \geq d \right] \\ &\geq \mathbb{P}[Z_t = 1] \cdot \mathbb{P} \left[\sum_{i=1}^{t-1} B_i = d-1 \right] + \mathbb{P} \left[\sum_{i=1}^{t-1} B_i \geq d \right] \\ &= \mathbb{P}[Z_t = 1] \cdot \left(\mathbb{P} \left[\sum_{i=1}^{t-1} B_i \geq d-1 \right] - \mathbb{P} \left[\sum_{i=1}^{t-1} B_i \geq d \right] \right) + \mathbb{P} \left[\sum_{i=1}^{t-1} B_i \geq d \right] \\ &= \mathbb{P}[Z_t = 1] \cdot \mathbb{P} \left[\sum_{i=1}^{t-1} B_i \geq d-1 \right] + \mathbb{P}[Z_t = 0] \cdot \mathbb{P} \left[\sum_{i=1}^{t-1} B_i \geq d \right]. \end{aligned}$$

By the inductive assumption we can then bound:

$$\begin{aligned} \mathbb{P} \left[\sum_{i=1}^t B_i \geq d \right] &\geq \mathbb{P}[Z_t = 1] \cdot \mathbb{P} \left[\sum_{i=1}^{t-1} Z_i \geq d-1 \right] + \mathbb{P}[Z_t = 0] \cdot \mathbb{P} \left[\sum_{i=1}^{t-1} Z_i \geq d \right] \\ &= \mathbb{P} \left[\sum_{i=1}^t Z_i \geq d \right] \end{aligned}$$

which gives the lemma. $\square$

We next prove a related theorem, but in a more specialized setting. We consider a set of neurons $\{u_1, \dots, u_s\}$ for which we can lower bound the probability of each u_i spiking at time $t+1$ given that it spiked at time t (i.e., given that $u_i^t = 1$). We show that, while the behavior of the neurons may be highly correlated, the number of neurons in the set that spike for t consecutive times can be lower bounded by comparing these neurons to a set of independent random variables with comparable spiking probabilities.

Lemma 2.4. Let $\mathcal{N} = \langle N, w, b, f \rangle$ be an SNN, and let $\{u_1, \dots, u_s\} \subseteq N$ be any set of neurons in the network. Let $Z_{i,t} \in \{0, 1\}$ be a set of independent random variables. Suppose that:

1. The initial configuration N^0 of $\mathcal{N}$ has $N^0(u_i) = 1$ for every $i \in \{1, \dots, s\}$.
2. For every $i \in \{1, \dots, s\}$, any configuration C of $\mathcal{N}$ with $C(u_i) = 1$, and any $t \geq 0$:

$$\mathbb{P}[u_i^{t+1} = 1 | N^t = C] \geq \mathbb{P}[Z_{i,t+1} = 1].$$

Let $\mathcal{I}_i(t) \in \{0, 1\}$ be an indicator variable for the event that $u_i^1 = \dots = u_i^t = 1$. Let $\bar{\mathcal{I}}_i(t) \in \{0, 1\}$ be an indicator variable for the event that $Z_{i,1} = \dots = Z_{i,t} = 1$. Then for every t and $d \in \mathbb{Z}^{\geq 0}$,

$$\mathbb{P} \left[\sum_{i=1}^s \mathcal{I}_i(t) \geq d \right] \geq \mathbb{P} \left[\sum_{i=1}^s \bar{\mathcal{I}}_i(t) \geq d \right]. \quad (3)$$

Proof. We prove the lemma via a coupling argument. At a high level, we define a set of auxiliary random variables $\hat{\mathcal{I}}_i(t)$ for $i \in \{1, \dots, s\}$. We construct these random variables such that their joint distribution is *identical* to that of the random variables $\mathcal{I}_i(t)$. Additionally, we correlate $\hat{\mathcal{I}}_i(t)$ with the variables $\{Z_{i,t}\}$ in such a way that we always have $\hat{\mathcal{I}}_i(t) \geq \bar{\mathcal{I}}_i(t)$. We thus have:

$$\mathbb{P} \left[\sum_{i=1}^s \mathcal{I}_i(t) \geq d \right] = \mathbb{P} \left[\sum_{i=1}^s \hat{\mathcal{I}}_i(t) \geq d \right] \geq \mathbb{P} \left[\sum_{i=1}^s \bar{\mathcal{I}}_i(t) \geq d \right], \quad (4)$$

which gives the lemma.

Definition of Coupled Random Variables $\hat{\mathcal{I}}_i(t)$.

Given $\mathcal{N}$, the distribution on executions of $\mathcal{N}$ with initial configuration N^0 and input configuration α_X , induced by the update rules described in Section 2.2 is given by $\mathcal{D}(\mathcal{N}, N^0, \alpha_X)$. We define the distribution $\hat{\mathcal{D}}(\mathcal{N}, N^0, \alpha_X)$, which is identical to $\mathcal{D}(\mathcal{N}, N^0, \alpha_X)$ except coupled to the auxiliary random variables $\{Z_{i,t}\}$ in the following way:

For any $t \geq 0$, execution $\alpha^t = C^0 \dots C^t$, and $i \in \{1, \dots, s\}$ with $C^0(u_i) = \dots = C^t(u_i) = 1$ let

$$\epsilon_{i,\alpha^t} = \mathbb{P}_{\mathcal{D}(\mathcal{N}, N^0, \alpha_X)} [u_i^{t+1} = 1 | N^0 \dots N^t = \alpha^t] - \mathbb{P}[Z_{i,t+1} = 1]. \quad (5)$$

By assumption (2) in the lemma statement and Lemma 2.2 we have $\epsilon_{i,\alpha^t} \geq 0$. Let $E_{i,\alpha^t} \in \{0, 1\}$ be a random variable which is independently set to 1 with probability $\frac{\epsilon_{i,\alpha^t}}{1 - \mathbb{P}[Z_{i,t+1} = 1]}$ and 0 otherwise. The distribution $\hat{\mathcal{D}}(\mathcal{N}, N^0, \alpha_X)$ is given by iteratively drawing a configuration N^{t+1} in the same way as in $\mathcal{D}(\mathcal{N}, N^0, \alpha_X)$, with spiking probabilities given by the potentials induced by N^t . However, if $i \in \{1, \dots, s\}$ and $N^0 \dots N^t = \alpha^t$ with $N^0(u_i) = \dots = N^t(u_i) = 1$, we set

$$u_i^{t+1} = \max(Z_{i,t+1}, E_{i,\alpha^t}). \quad (6)$$

Using (5) and the definition of E_{i,α^t} we can see that

$$\begin{aligned} \mathbb{P}_{\hat{\mathcal{D}}(\mathcal{N}, N^0, \alpha_X)} [u_i^{t+1} = 1 | N^0 \dots N^t = \alpha^t] &= 1 - \mathbb{P}_{\hat{\mathcal{D}}(\mathcal{N}, N^0, \alpha_X)} [u_i^{t+1} = 0 | N^0 \dots N^t = \alpha^t] \\ &= 1 - (1 - \mathbb{P}[Z_{i,t+1} = 1]) \cdot (1 - \mathbb{P}[E_{i,\alpha^t} = 1]) \\ &= 1 - (1 - \mathbb{P}[Z_{i,t+1} = 1]) \cdot \left(1 - \frac{\epsilon_{i,\alpha^t}}{1 - \mathbb{P}[Z_{i,t+1} = 1]} \right) \\ &= \mathbb{P}[Z_{i,t+1} = 1] + \epsilon_{i,\alpha^t} \\ &= \mathbb{P}_{\mathcal{D}(\mathcal{N}, N^0, \alpha_X)} [u_i^{t+1} = 1 | N^0 \dots N^t = \alpha^t]. \end{aligned} \quad (\text{By (5)})$$

Thus, the probability that any neuron spikes at time t conditioned on the network configuration at all times before t is identical in executions drawn from $\hat{\mathcal{D}}(\mathcal{N}, N^0, \alpha_X)$ and $\mathcal{D}(\mathcal{N}, N^0, \alpha_X)$. So for any t , if $N^0 \dots N^t$ is drawn from $\mathcal{D}(\mathcal{N}, N^0, \alpha_X)$ and $\hat{N}^0 \dots \hat{N}^t$ from $\hat{\mathcal{D}}(\mathcal{N}, N^0, \alpha_X)$, these two executions are identically distributed. In particular, if $\hat{\mathcal{I}}_i(t) \in \{0, 1\}$ is an indicator variable for the event that $\hat{N}^1(u_i) = \dots = \hat{N}^t(u_i) = 1$, then $\hat{\mathcal{I}}_i(t)$ and $\mathcal{I}_i(t)$ are identically distributed.

Proof that Coupled Random Variables Upper Bound Independent Variables.

We can see that $\hat{\mathcal{I}}_i(t) \geq \bar{\mathcal{I}}_i(t)$ via an inductive argument. In the base case, since we assume $N^0(u_i) = 1$ for all $i \in \{1, \dots, s\}$, we apply (6) to generate $\hat{N}^1(u_i)$. We set $\hat{N}^1(u_i) = \max(Z_{i,1}, E_{i,\alpha^0}) \geq Z_{i,1}$ which gives $\hat{\mathcal{I}}_i(1) \geq \bar{\mathcal{I}}_i(1)$. For $t \geq 1$, if $\bar{\mathcal{I}}_i(t) = 0$ the claim holds trivially since $\hat{\mathcal{I}}_i(t), \bar{\mathcal{I}}_i(t) \in \{0, 1\}$. Otherwise, we have $\bar{\mathcal{I}}_i(t) = 1$ which implies that $\bar{\mathcal{I}}_i(t-1) = 1$ and so $\hat{\mathcal{I}}_i(t-1) = 1$ by the inductive assumption. If $\hat{\mathcal{I}}_i(t-1) = 1$ then again we apply (6) to generate $\hat{N}^t(u_i)$ and so have $\hat{N}^t(u_i) = \max(Z_{i,t}, E_{i,\alpha^{t-1}}) \geq Z_{i,t}$, giving $\hat{\mathcal{I}}_i(t) \geq \bar{\mathcal{I}}_i(t)$.

Since $\hat{\mathcal{I}}_i(t)$ is identically distributed to $\bar{\mathcal{I}}_i(t)$ this completes the lemma by (4). $\square$

Our next lemma pertains specifically to networks with a sigmoid spike probability function, $f(x) = \frac{1}{1+e^{-x/\lambda}}$, which we use throughout this work. We show that given a network with temperature parameter $\lambda > 0$, we can construct a network with an identical execution distribution for any $\hat{\lambda} > 0$. Thus, we will always consider the case of $\lambda = 1$, which implies the existence of networks satisfying all bounds given for all $\lambda > 0$.

Lemma 2.5 (Equivalence of Temperature Parameters). *For $\lambda, \hat{\lambda} > 0$, let $f(x) = \frac{1}{1+e^{-x/\lambda}}$ and $\hat{f}(x) = \frac{1}{1+e^{-x/\hat{\lambda}}}$. Given $\mathcal{N} = \langle N, w, b, f \rangle$, let $\hat{\mathcal{N}} = \langle N, \hat{w}, \hat{b}, \hat{f} \rangle$ where for all $u, v \in N$, $\hat{w}(u, v) = w(u, v) \cdot \frac{\lambda}{\hat{\lambda}}$ and $\hat{b}(u) = b(u) \cdot \frac{\lambda}{\hat{\lambda}}$. For any length initial configuration N^0 and any infinite input execution α_X :*

$$\mathcal{D}(\mathcal{N}, N^0, \alpha_X) = \mathcal{D}(\hat{\mathcal{N}}, N^0, \alpha_X).$$

Proof. For any $t \geq 1$ and any configuration C , we can compute the probability that $\mathcal{N}$ is in this configuration at time t conditioned on all past configurations as:

$$\mathbb{P}_{\mathcal{D}(\mathcal{N}, N^0, \alpha_X)}[N^t = C | N^{t-1} \dots N^0] = \prod_{u \in N} [C(u) \cdot p(u, t) + (1 - C(u)) \cdot (1 - p(u, t))] \quad (7)$$

where $p(u, t) = f(\text{pot}(u, t))$. Fixing $N^{t-1} \dots N^0$, we can see that the potential computation (1) is a linear function of $w(u, v)$ and $b(u)$ for all $u, v \in N$. Thus, letting $\widehat{\text{pot}}(u, t)$ be the potential of u at time t in $\hat{\mathcal{N}}$ given $N^{t-1} \dots N^0$, since $\hat{w}(u, v) = w(u, v) \cdot \frac{\lambda}{\hat{\lambda}}$ and $\hat{b}(u) = b(u) \cdot \frac{\lambda}{\hat{\lambda}}$,

$$\widehat{\text{pot}}(u, t) = \text{pot}(u, t) \cdot \frac{\lambda}{\hat{\lambda}}.$$

This gives that the probability of u spiking at time t in $\hat{\mathcal{N}}$ given $N^{t-1} \dots N^0$ equals :

$$\hat{p}(u, t) = \hat{f}(\widehat{\text{pot}}(u, t)) = \hat{f}\left(\text{pot}(u, t) \cdot \frac{\lambda}{\hat{\lambda}}\right) = f(\text{pot}(u, t)) = p(u, t).$$

And since $\hat{p}(u, t) = p(u, t)$ for all $u \in N$ we have using (7), for any C ,

$$\mathbb{P}_{\mathcal{D}(\mathcal{N}, N^0, \alpha_X)}[N^t = C | N^{t-1} \dots N^0] = \mathbb{P}_{\mathcal{D}(\hat{\mathcal{N}}, N^0, \alpha_X)}[N^t = C | N^{t-1} \dots N^0]. \quad (8)$$

Finally inducting on t we can show that for any finite execution $C^0 \dots C^t$:

$$\mathbb{P}_{\mathcal{D}(\mathcal{N}, N^0, \alpha_X)}[N^0 \dots N^t = C^0 \dots C^t] = \mathbb{P}_{\mathcal{D}(\hat{\mathcal{N}}, N^0, \alpha_X)}[N^0 \dots N^t = C^0 \dots C^t]. \quad (9)$$

This holds trivially for $t = 0$ since

$$\mathbb{P}_{\mathcal{D}(\mathcal{N}, N^0, \alpha_X)}[N^0 = C^0] = \mathbb{P}_{\mathcal{D}(\hat{\mathcal{N}}, N^0, \alpha_X)}[N^0 = C^0] = 1$$

if $N^0 = C^0$. Both probabilities are zero otherwise. For $t \geq 1$, assume that (9) holds for all $t' < t$. Combined with (8) this gives:

$$\begin{aligned}
\mathbb{P}_{\mathcal{D}(\mathcal{N}, N^0, \alpha_X)}[N^0 \dots N^t = C^0 \dots C^t] &= \mathbb{P}_{\mathcal{D}(\mathcal{N}, N^0, \alpha_X)}[N^0 \dots N^{t-1} = C^0 \dots C^{t-1}] \\
&\quad \cdot \mathbb{P}_{\mathcal{D}(\mathcal{N}, N^0, \alpha_X)}[N^t = C^t | N^0 \dots N^{t-1} = C^0 \dots C^{t-1}] \\
&= \mathbb{P}_{\mathcal{D}(\hat{\mathcal{N}}, N^0, \alpha_X)}[N^0 \dots N^{t-1} = C^0 \dots C^{t-1}] \\
&\quad \cdot \mathbb{P}_{\mathcal{D}(\hat{\mathcal{N}}, N^0, \alpha_X)}[N^t = C^t | N^{t-1} \dots N^0 = C^0 \dots C^{t-1}] \\
&= \mathbb{P}_{\mathcal{D}(\hat{\mathcal{N}}, N^0, \alpha_X)}[N^0 \dots N^t = C^0 \dots C^t].
\end{aligned}$$

This completes the lemma. $\square$

2.5 Potential Modifications to the Basic Model

There are many potential modifications to the basic network model described in Sections 2.1 and 2.2 which may be interesting to consider in future work. We present some below.

- One interesting extension is to add a history period $h > 1$ to the network, so that the spiking probability of a neuron at time t depends on the configuration of the network at times $t - 1, \dots, t - h$. In Section 5, for example, we use a model with history period $h = 2$ to design very fast WTA networks.
- We could consider a very general history model, defining

$$\text{pot}(u, t) = f(u, N^{t-1}, \dots, N^{t-h})$$

where f is any function.

- A history period h can be thought of as giving each neuron access to a length h queue of firing patterns on which $\text{pot}(u, t)$ depends. It may be interesting to model such a queue as residing within the neuron's state. We could also consider neurons with other types of state, capturing various types of observed biological phenomena. For example, we could model a refractory period, in which a neuron cannot fire again for a certain number of time steps after firing [And03, Izh04].
- A history period may be used, for example, to model a universal decay in the influence of spikes over time. We may specify a non-increasing weight vector $= (c_1, c_2, \dots, c_h) \in \mathbb{R}^{\geq 0}$ and modify the potential computation of (1) such that for any $t \geq h$:

$$\text{pot}(u, t) = \left(\sum_{i=1}^h \sum_{v \in N} c_i \cdot w(v, u) \cdot v^{t-i} \right) - b(u).$$

- It may be interesting to consider networks with neurons of multiple types, with different firing dynamics. The human brain contains as many as 10,000 distinct neuron types [Stu]. Understanding how important neuron specialization is and for what reasons it arises is a very interesting question.
- Similarly, we note that there is evidence that Dale's principal can be violated and that some neurons do have both inhibitory and excitatory outgoing connections [Osb79]. Modeling such neurons to better understand their role and importance is an interesting direction.

2.6 The Winner-Take-All Problem

We now define the main problem that we consider in this work, the binary winner-take-all (WTA) problem. In this problem, given n input neurons, the goal is to converge to a configuration in which a single output neuron, corresponding to a firing input, fires. This neuron is referred to as the ‘winner’ of the computation. We first define a valid WTA output configuration for a given input configuration:

Definition 2.6 (Valid WTA Output Configuration). *Consider any network $\mathcal{N}$ with n input neurons, labeled $x_1, \dots, x_n$, and n output neurons, labeled $y_1, \dots, y_n$. For any input configuration C_X of $\mathcal{N}$, a valid WTA output configuration for C_X is any output configuration C_Y with $C_Y(y_i) \leq C_X(x_i)$ for all $i \in \{1, \dots, n\}$ and $\|C_Y\|_1 = \min(1, \|C_X\|_1)$.*

Interpreting the above definition, the restriction that $\|C_Y\|_1 = \min(1, \|C_X\|_1)$ requires that if at least one input fires, exactly one output fires. The condition $C_Y(y_i) \leq C_X(x_i)$ for all i requires that this firing output corresponds to a firing input. If no inputs fire (i.e., if $\|C_X\|_1 = 0$), then no outputs should fire. With this definition, we can define the WTA problem (see Section 2.3 for a description of how problems are defined in our SNN model):

Definition 2.7 (Winner-Take-All Problem). *Given input size $n \in \mathbb{Z}^{>0}$, convergence time $t_c \in \mathbb{Z}^{>0}$, stability time $t_s \in \mathbb{Z}^{>0}$, and failure probability $\delta > 0$, the winner-take-all problem $\text{WTA}(n, t_c, t_s, \delta)$ is defined as follows:*

- *If α_X is an input execution with X^t fixed for all t , the output distribution $\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X)$ can be any distribution on executions of n output neurons satisfying:*
 - *With probability $\geq 1 - \delta$, there exists some $t \leq t_c$ such that the output configuration is fixed at times $t, t+1, \dots, t+t_s$ and is a valid WTA output configuration for X^t (Def. 2.6).*
- *If α_X is any other input execution, the output distribution is unconstrained.*

Thus, to solve $\text{WTA}(n, t_c, t_s, \delta)$, with probability $\geq 1 - \delta$, the network must converge to a valid output configuration within t_c time steps and maintain this configuration for t_s time steps.

Due to the random firing behavior of our neurons, the network will eventually move to a different configuration with some probability. However, if the network solves $\text{WTA}(n, t_c, t_s, \delta)$, since convergence is required given any initial configuration N^0 , we can show that it must be self-stabilizing. That is, once it leaves a valid output configuration, it will converge again with probability $\geq 1 - \delta$ within t_c steps, and maintain the new valid configuration again for t_s steps. Formally:

Theorem 2.8 (Self Stabilization of Winner-Take-All Networks). *If $\mathcal{N}$ solves $\text{WTA}(n, t_c, t_s, \delta)$ for input execution α_X with X^t fixed for all t , given any finite execution $C^0 \dots C^t$ of $\mathcal{N}$, conditioned on $N^0 \dots N^t = C^0 \dots C^t$, with probability $\geq 1 - \delta$ there is some time $t' \leq t + t_c$ such that the output configuration for $\mathcal{N}$ is fixed at times $t', t'+1, \dots, t'+t_s$ and is a valid WTA output configuration for $X^{t'}$.*

Proof. Consider the distribution on infinite executions $N^{t+1}N^{t+2}\dots$ conditioned on $N^0 \dots N^t = C^0 \dots C^t$. Since the configuration at time $t' \geq t+1$ depends only on the configuration at time $t'-1$, this distribution is identical to $\mathcal{D}_Y(\mathcal{N}, C^t, \alpha_X)$.

Thus, if $\mathcal{N}$ solves $\text{WTA}(n, t_c, t_s, \delta)$ on α_X , conditioned on $N^0 \dots N^t = C^0 \dots C^t$, with probability $1 - \delta$, there is some time $t' \leq t + t_c$ in which $\mathcal{N}$ reaches a valid WTA output configuration for $X^{t'}$ and remains there for t_s steps, giving the lemma. □

We can also define an expected-time version of the winner-take-all problem as follows:

Definition 2.9 (Expected-Time Winner-Take-All Problem). *For any infinite input execution $\alpha_X = X^0 X^1 \dots$, stability time $t_s \in \mathbb{Z}^{>0}$, and infinite output execution $\alpha_Y = Y^0 Y^1 \dots$ define:*

$$t(\alpha_X, t_s, \alpha_Y) = \min \{t : Y^t \text{ is a valid WTA output configuration and } Y^t = \dots = Y^{t+t_s}\}.$$

Given input size $n \in \mathbb{Z}^{>0}$, convergence time $t_c \in \mathbb{Z}^{>0}$, stability time $t_s \in \mathbb{Z}^{>0}$, the expected-time winner-take-all problem WTA-EXP(n, t_c, t_s) is defined as follows:

- *If α_X is an input execution with X^t fixed for all t , the output distribution $\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X)$ can be any distribution on executions of n output neurons satisfying:*

$$\mathbb{E}_{\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X)} t(\alpha_X, t_s, \alpha_Y) \leq t_c.$$

- *If α_X is any other input execution, the output distribution is unconstrained.*

3 A Two-Inhibitor Solution to the WTA Problem

We now present a simple solution to the WTA problems in Definitions 2.7 and 2.9 in networks with spike probability given by a sigmoid function. We begin by defining a family of networks $\mathcal{T}_{n,\gamma}$ for any input size n and weight scaling parameter $\gamma \in \mathbb{R}^+$ that solve these problems.

3.1 Network Definition

We first give a full definition of our family of two-inhibitor WTA networks, before describing the intuition behind why these networks solve the WTA problem (Definitions 2.7 and 2.9).

Definition 3.1 (Two-Inhibitor WTA Network). *For any positive integer n and $\gamma \in \mathbb{R}^+$, let $\mathcal{T}_{n,\gamma} = \langle N, w, b, f \rangle$ where the spike probability, weight, and bias functions are defined as follows:*

- *The spike probability function f is defined to be the basic sigmoid function:*

$$f(x) \stackrel{\text{def}}{=} \frac{1}{1 + e^{-x}}. \quad (10)$$

- *The set of neurons N consists of a set of n input neurons X , labeled $x_1, \dots, x_n$, a set of n corresponding outputs Y , labeled $y_1, \dots, y_n$, and two auxiliary inhibitor neurons labeled a_s, a_c .*
- *The weight function w is given by:*
 - $w(x_i, y_i) = 3\gamma$, for all i .
 - $w(y_i, y_i) = 2\gamma$, for all i .
 - $w(a_s, y_i) = w(a_c, y_i) = -\gamma$, for all i .
 - $w(y_i, a_s) = w(y_i, a_c) = \gamma$, for all i .
 - $w(u, v) = 0$ for any u, v whose connection is not specified above.
- *The bias function b is given by:*
 - $b(y_i) = 3\gamma$ for all i .
 - $b(a_s) = \gamma/2$.
 - $b(a_c) = 3\gamma/2$.

A diagram of $\mathcal{T}_{n,\gamma}$ is shown in Figure 1. Note that the two inhibitors a_s and a_c have identical outgoing connections, and differ just in their bias.

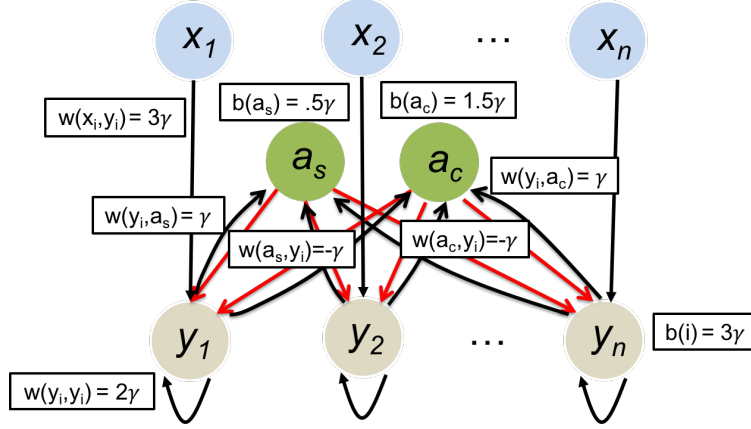


Figure 1: Our two-inhibitor WTA network $\mathcal{T}_{n,\gamma}$ as described in Definition 3.1.

3.1.1 Intuition Behind the Two-Inhibitor Network

Before giving a formal analysis of the behavior of $\mathcal{T}_{n,\gamma}$, we give some intuition behind why this family of two-inhibitor networks solves the WTA problem. In the description below, we informally refer to events that occur ‘with high probability’. We will quantify the meaning of such statements in our full analysis.

In $\mathcal{T}_{n,\gamma}$, each input is connected to its corresponding output with a positive weight. Thus, firing inputs will initially cause their corresponding outputs to fire with high probability. For the network to solve WTA, it must converge to a state in which just a single one of these outputs fires.

To ensure this convergence, $\mathcal{T}_{n,\gamma}$ has two inhibitors $A = \{a_s, a_c\}$. The neuron a_s is a *stability* inhibitor that maintains a valid WTA output configuration once it has been reached. It fires with high probability at time t whenever at least one output fires at time $t - 1$. The neuron a_c is a *convergence* inhibitor that fires with high probability whenever at least two outputs fire at time $t - 1$.

The weights connecting a_s and a_c to the outputs are set such that when both fire at time t , any output that fired at time t will fire with probability $1/2$ at time $t + 1$. Any output that did not fire at time t will not fire at time $t + 1$ with high probability. This distinguished behavior between previously firing and non-firing outputs is due to the self-loops on each output neuron, which allow firing outputs to partially overcome the strong inhibition from a_s and a_c .

In this way, if two or more outputs fire at time t , both inhibitors fire with high probability and the high level of inhibition causes outputs to ‘drop out of contention’ for the winning position with probability $1/2$. After $O(\log n)$ time steps, nearly all the outputs stop firing and, with constant probability, there is a time step in which exactly one output fires. Once this step occurs, with high probability, a_c ceases firing and just a_s fires. This decreased level of inhibition allows the winner to keep firing with high probability, as the inhibition is fully offset by the winner’s excitatory self-loop. However, with high probability, the inhibition prevents any other output whose excitatory self-loop is inactive from firing. Thus the network remains in the valid WTA output configuration for a large number of time steps with high probability.

In the event that a time in which a single output fires does not occur, then the network ‘resets’. No outputs fire at some time, causing the inhibitors to both cease firing. Thus, all outputs with firing inputs are able to fire, and convergence starts again. Within $O(\log 1/\delta)$ of these resets each reaching a valid WTA output state with constant probability, the network reaches a valid WTA output state with probability $\geq 1 - \delta$ and so solves the WTA problem of Definition 2.7. Similarly,

the network requires $O(1)$ resets in expectation to reach a valid WTA output state, giving a solution to the expected-time version of the problem in Definition 2.9. Formally, we will prove the following:

Theorem 3.2 (Two-Inhibitor WTA). *For $\gamma \geq 4 \ln((n+2)t_s/\delta) + 10$, $\mathcal{T}_{n,\gamma}$ solves WTA(n, t_c, t_s, δ) for any $t_c \geq 72(\log_2 n + 1) \cdot (\log_2(1/\delta) + 1)$.*

Theorem 3.3 (Two-Inhibitor Expected-Time WTA). *For $\gamma \geq 4 \ln((n+2)t_s) + 10$, $\mathcal{T}_{n,\gamma}$ solves WTA-EXP(n, t_c, t_s) for any $t_c \geq 108(\log_2 n + 3)$.*

Proof Roadmap. We prove Theorems 3.2 and 3.3 in Sections 3.2-3.7. The analysis is broken down as follows:

Section 3.2: Prove basic *one-step lemmas* which characterize single time step transitions of $\mathcal{T}_{n,\gamma}$, showing that the neurons behave as described in the above high-level description.

Section 3.3: Prove that, once in a valid WTA configuration, $\mathcal{T}_{n,\gamma}$ stays in this configuration with high probability (that is, valid WTA configurations are stable).

Section 3.4: Show that all configurations of $\mathcal{T}_{n,\gamma}$ transition with high probability within two time steps to a small set of *good configurations*, from which we will prove fast convergence.

Section 3.5 Show basic transition lemmas for this set of good configurations, characterizing the network's behavior at the times immediately following a good configuration.

Section 3.6 Use the above transition lemmas to show that the network converges, with constant probability, from any good configuration (and hence any configuration by Section 3.4) to a valid WTA configuration within $O(\log n)$ time steps.

Section 5.6 Complete the analysis, demonstrating with what parameters $\mathcal{T}_{n,\gamma}$ solves the winner-take-all problem (Definitions 2.7 and 2.9).

3.2 Basic Results and One-step Lemmas

We begin with some basic results that will be important throughout our analysis, including a few ‘one-step’ lemmas, which characterize the transition probabilities from a set of configurations at time t to a set of configurations at time $t + 1$.

We first show that, unless a neuron has potential 0, either it fires with high probability (i.e., except with probability that is inverse exponential in γ) or it does not fire with high probability.

Lemma 3.4 (Characterization of Firing Probabilities). *For any time $t \geq 1$ and any $u \in N$:*

If $\text{pot}(u, t) = 0$, then $p(u, t) = 1/2$.

If $\text{pot}(u, t) < 0$, then $p(u, t) \leq e^{-\gamma/2}$.

If $\text{pot}(u, t) > 0$, then $p(u, t) \geq 1 - e^{-\gamma/2}$.

Proof. If $\text{pot}(u, t) = 0$, then by (1) $p(u, t) = f(\text{pot}(u, t)) = \frac{1}{1+e^0} = 1/2$. Otherwise consider the potential calculation of (1) in the case when $h = 1$:

$$\text{pot}(u, t) = \sum_{v \in N} w(v, u) \cdot v^{t-1} - b(u).$$

By Definition 3.1, for all u, v , $w(v, u)$ and $b(u)$ are integer multiples of $\gamma/2$. Thus, since $v^{t-1} \in \{0, 1\}$, $\text{pot}(u, t)$ is also an integer multiple of $\gamma/2$. So, if $\text{pot}(u, t) < 0$, then $\text{pot}(u, t) \leq -\gamma/2$ and:

$$p(u, t) = f(\text{pot}(u, t)) \leq f(-\gamma/2) = \frac{1}{1 + e^{\gamma/2}} \leq e^{-\gamma/2}.$$

Similarly, if $\text{pot}(u, t) > 0$, then $\text{pot}(u, t) \geq \gamma/2$ and so:

$$p(u, t + 1) = f(\text{pot}(u, t)) \geq f(\gamma/2) = \frac{1}{1 + e^{-\gamma/2}} \geq 1 - e^{-\gamma/2}.$$

□

We next show that if output y_i does not correspond to a firing input (i.e., $x_i^t = 0$), then starting from any configuration of $\mathcal{T}_{n,\gamma}$ at time t , with high probability y_i does not fire at time $t + 1$. That is, with high probability, outputs that are not valid winners of the WTA computation do not fire.

Lemma 3.5 (Correct Output Behavior). *For any time t , any configuration C of $\mathcal{T}_{n,\gamma}$, and any i with $C(x_i) = 0$,*

$$\mathbb{P}[y_i^{t+1} = 1 | N^t = C] \leq e^{-\gamma/2}.$$

Proof. If $N^t = C$ then $x_i^t = C(x_i)$. We can compute y_i 's potential at time $t + 1$, assuming $x_i^t = 0$:

$$\begin{aligned} \text{pot}(y_i, t + 1) &= w(x_i, y_i)x_i^t + w(y_i, y_i)y_i^t + w(a_s, y_i)a_s^t + w(a_c, y_i)a_c^t - b(y_i) \\ &\leq 0 + 2\gamma + 0 + 0 - 3\gamma = -\gamma. \end{aligned}$$

Thus, by Lemma 3.4, since $\text{pot}(y_i, t + 1) < 0$, $p(y_i, t + 1) \leq e^{-\gamma/2}$. □

Applying Lemma 3.5 and a simple union bound over all n outputs yields the following corollary:

Corollary 3.6 (Correct Output Behavior, All Neurons). *For any time t and configuration C of $\mathcal{T}_{n,\gamma}$,*

$$\mathbb{P}[y_i^{t+1} \leq x_i^t \text{ for all } i | N^t = C] \geq 1 - ne^{-\gamma/2}.$$

Proof. If $C(x_i) = 1$ then conditioned on $N^t = C$, $x_i^t = 1$ and so $y_i^{t+1} \leq x_i^t$ always. Otherwise, by Lemma 3.5, if $C(x_i) = 0$, then $\mathbb{P}[y_i^{t+1} = 0 | N^t = C] \geq 1 - e^{-\gamma/2}$. Union bounding over all such inputs (of which there are at most n) gives the corollary. □

We next show that the inhibitors a_s and a_c behave as expected with high probability.

Lemma 3.7 (Correct Inhibitor Behavior). *For any time t and configuration C of $\mathcal{T}_{n,\gamma}$,*

1. *If $\|C(Y)\|_1 = 0$, then $\mathbb{P}[a_s^{t+1} = a_c^{t+1} = 0 | N^t = C] \geq 1 - 2e^{-\gamma/2}$.*
2. *If $\|C(Y)\|_1 = 1$, then $\mathbb{P}[a_s^{t+1} = 1 \text{ and } a_c^{t+1} = 0 | N^t = C] \geq 1 - 2e^{-\gamma/2}$.*
3. *If $\|C(Y)\|_1 \geq 2$, then $\mathbb{P}[a_s^{t+1} = 1 = a_c^{t+1} = 1 | N^t = C] \geq 1 - 2e^{-\gamma/2}$.*

Proof. We prove each case above separately. Note that, conditioned on $N^t = C$, $Y^t = C(Y)$.

Case 1: $\|C(Y)\|_1 = 0$.

In this case, the inhibitors receive no excitatory signal from the outputs so,

$$\text{pot}(a_s, t+1) = -b(a_s) < 0 \quad \text{and} \quad \text{pot}(a_c, t+1) = -b(a_c) < 0.$$

Thus by Lemma 3.4 and a union bound over the two inhibitors,

$$\mathbb{P}[a_s^{t+1} = a_c^{t+1} = 0 | N^t = C] \geq 1 - 2e^{-\gamma/2}.$$

Case 2: $\|C(Y)\|_1 = 1$.

In this case we have:

$$\begin{aligned} \text{pot}(a_s, t+1) &= \sum_{j=1}^n w(y_j, a_s) y_j^t - b(a_s) \\ &= \gamma - \gamma/2 = \gamma/2. \\ \text{pot}(a_c, t+1) &= \sum_{j=1}^n w(y_j, a_c) y_j^t - b(a_c) \\ &= \gamma - 3\gamma/2 = -\gamma/2. \end{aligned}$$

Again by Lemma 3.4 and a union bound, $\mathbb{P}[a_s^{t+1} = 1 \text{ and } a_c^{t+1} = 0 | N^t = C] \geq 1 - 2e^{-\gamma/2}$.

Case 3: $\|C(Y)\|_1 \geq 2$

Finally, in this case:

$$\begin{aligned} \text{pot}(a_s, t+1) &= \sum_{j=1}^n w(y_j, a_s) y_j^t - b(a_s) \\ &\geq 2\gamma - \gamma/2 = 3\gamma/2. \\ \text{pot}(a_c, t+1) &= \sum_{j=1}^n w(y_j, a_c) y_j^t - b(a_c) \\ &\geq 2\gamma - 3\gamma/2 = \gamma/2. \end{aligned}$$

So by Lemma 3.4 and a union bound, $\mathbb{P}[a_s^{t+1} = a_c^{t+1} = 1 | N^t = C] \geq 1 - 2e^{-\gamma/2}$, completing the lemma. $\square$

Combined with Corollary 3.6, Lemma 3.7 conclusion (1) gives:

Lemma 3.8 (Quiescent Behavior). *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t and $\|X^t\|_1 = 0$. For any time t and configuration C with $C(X) = X^t$,*

$$\mathbb{P}[\|N^{t+2}\|_1 = 0 | N^t = C] \geq 1 - 2(n+1)e^{-\gamma/2}.$$

Proof. Let $\mathcal{E}_{10}$ be the event that $N^t = C$ and $\|Y^{t+1}\|_1 = 0$. Let $\mathcal{E}_{20}$ be the event that $\|N^{t+2}\|_1 = 0$. That is, that no neurons fire at time $t+2$. Conditioned on $\mathcal{E}_{10}$, by Lemma 3.7 conclusion (1), with probability $\geq 1 - 2e^{-\gamma/2}$, $a_s^{t+2} = a_c^{t+2} = 0$. Again by Corollary 3.6, conditioned on $\mathcal{E}_{10}$, with probability $\geq 1 - ne^{-\gamma/2}$, $\|Y^{t+2}\|_1 = 0$. So, overall by a union bound,

$$\mathbb{P}[\mathcal{E}_{20} | \mathcal{E}_{10}] \geq 1 - (n+2)e^{-\gamma/2}.$$

By Corollary 3.6, since $\|X^t\|_1 = 0$, $\mathbb{P}[\mathcal{E}_{10}|N^t = C] \geq 1 - ne^{-\gamma/2}$. We can thus bound:

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{20}|N^t = C] &\geq \mathbb{P}[\mathcal{E}_{10}|N^t = C] \cdot \mathbb{P}[\mathcal{E}_{20}|\mathcal{E}_{10}, N^t = C] \\ &= \mathbb{P}[\mathcal{E}_{10}|N^t = C] \cdot \mathbb{P}[\mathcal{E}_{20}|\mathcal{E}_{10}] \quad (\text{Since, by definition, } \mathcal{E}_{10} \text{ implies } N^t = C.) \\ &\geq \left(1 - ne^{-\gamma/2}\right) \cdot \left(1 - (n+2)e^{-\gamma/2}\right) \\ &\geq 1 - 2(n+1)e^{-\gamma/2}, \end{aligned}$$

which gives the lemma. $\square$

We next show that the stability inhibitor, with high probability, induces exactly the outputs that fired at the previous time step to fire in the next step. We show the lemma in fact for any configuration in which exactly one inhibitor fires. Since a_s and a_c have identical outgoing edges, they have a symmetric effect on the firing probabilities of other neurons.

Lemma 3.9 (Stability Inhibitor Effect). *For any time t and configuration C of $\mathcal{T}_{n,\gamma}$ with $(C(a_s) = 1 \text{ and } C(a_c) = 0)$ or $(C(a_s) = 0 \text{ and } C(a_c) = 1)$ and $C(y_i) \leq C(x_i)$ for all i ,*

$$\mathbb{P}[Y^{t+1} = Y^t | N^t = C] \geq 1 - ne^{-\gamma/2}.$$

Proof. For any configuration C with $C(a_c) = 1$ and $C(a_s) = 0$, let $\bar{C}$ denote the configuration with $\bar{C}(a_c) = 0$, $\bar{C}(a_s) = 1$, and $\bar{C}(u) = C(u)$ for all other $u \in N \setminus \{a_c, a_s\}$. Since a_c and a_s have no self-loops and have identical outgoing connections, the distribution of N^{t+1} given $N^t = C$ is identical to its distribution given $N^t = \bar{C}$. Thus, we can assume without loss of generality in the proof of this lemma that $C(a_s) = 1$ and $C(a_c) = 0$.

Conditioned on $N^t = C$, $y_i^t \leq x_i^t$ by assumption. So for any output with $y_i^t = 1$, we have $x_i^t = 1$. This gives:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i)x_i^t + w(y_i, y_i)y_i^t + w(a_s, y_i)a_s^t + w(a_c, y_i)a_c^t - b(y_i) \\ &= 3\gamma + 2\gamma - \gamma + 0 - 3\gamma \\ &= \gamma. \end{aligned}$$

In contrast, for any output with $y_i^t = 0$:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i)x_i^t + w(y_i, y_i)y_i^t + w(a_s, y_i)a_s^t + w(a_c, y_i)a_c^t - b(y_i) \\ &\leq 3\gamma + 0 - \gamma + 0 - 3\gamma \\ &= -\gamma. \end{aligned}$$

Thus, by Lemma 3.4, if $y_i^t = 1$, then $y_i^{t+1} = 1$ with probability $\geq 1 - e^{-\gamma/2}$. If $y_i^t = 0$, then $y_i^{t+1} = 0$ with probability $\geq 1 - e^{-\gamma/2}$. The lemma follows after union bounding over all n outputs. $\square$

Finally, we show that when both the stability and convergence inhibitors fire at time t , not only do outputs not firing at time t not fire at time $t+1$ with high probability, but also, all firing outputs at time t stop firing with probability $1/2$ at time $t+1$. Conditioned on the configuration at time t , these outputs fire independently, a property which will be useful in our eventual proof of progress towards a valid WTA configuration in Lemma 3.23.

Lemma 3.10 (Convergence Inhibitor Effect). *For any time t and configuration C of $\mathcal{T}_{n,\gamma}$ with $C(a_s) = C(a_c) = 1$ and $C(y_i) \leq C(x_i)$ for all i ,*

1. $\mathbb{P}[y_i^{t+1} \leq y_i^t \text{ for all } i | N^t = C] \geq 1 - ne^{-\gamma/2}$.
2. *If $y_i^t = 1$, $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] = 1/2$.*
3. *For $i \neq j$, y_i^{t+1} and y_j^{t+1} are independent conditioned on $N^t = C$.*

Proof. Conditioned on $N^t = C$, if $y_i^t = 1$, by assumption $x_i^t = 1$. We can thus compute:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i)x_i^t + w(y_i, y_i)y_i^t + w(a_s, y_i)a_s^t + w(a_c, y_i)a_c^t - b(y_i) \\ &= 3\gamma + 2\gamma - \gamma - \gamma - 3\gamma \\ &= 0. \end{aligned}$$

We thus have $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] = 1/2$ by Lemma 3.4. This gives conclusion (2). Conclusion (3) holds trivially since, with N^t fixed, u^{t+1} is independent of v^{t+1} for all $u \neq v$.

We can also bound if $y_i^t = 0$:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i)x_i^t + w(y_i, y_i)y_i^t + w(a_s, y_i)a_s^t + w(a_c, y_i)a_c^t - b(y_i) \\ &\leq 3\gamma + 0 - \gamma - \gamma - 3\gamma \\ &= -2\gamma. \end{aligned}$$

Thus, by Lemma 3.4, $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] \leq e^{-\gamma/2}$. By a union bound over at most n such outputs, we have, with probability $\geq 1 - ne^{-\gamma/2}$, $y_i^{t+1} \leq y_i^t$ for all i , completing the lemma. $\square$

3.3 Stability

In this section we extend our definition of a valid WTA output configuration (Definition 2.6), to give a more restrictive notion of a valid WTA configuration, which additionally requires that the auxiliary neurons a_s, a_c are in a good state. We show that once the network is in such a state at time t , it remains there with high probability at time $t+1$.

Definition 3.11 (Valid WTA Configuration). *A valid WTA configuration of $\mathcal{T}_{n,\gamma}$ is a configuration C with $C(y_i) \leq C(x_i)$ for all $i \in \{1, \dots, n\}$ and $\|C(Y)\|_1 = \min(1, \|C(X)\|_1)$ (i.e., the outputs satisfy Definition 2.6) and further, $C(a_c) = 0$ and $C(a_s) = \min(1, \|C(X)\|_1)$.*

In the above we require $C(a_s) = \min(\|C(X)\|_1, 1)$. That is, the stability inhibitor fires in a valid WTA configuration, unless no inputs fire. If no inputs fire, a valid WTA configuration requires that neither a_s nor a_c fire and additionally, that no outputs fire.

Lemma 3.12 (Stability of Valid Configurations). *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t . For any time t and valid WTA configuration C with $C(X) = X^t$,*

$$\mathbb{P}[N^{t+1} = N^t | N^t = C] \geq 1 - (n+2)e^{-\gamma/2}.$$

Proof. By Definition 3.11, since C is a valid WTA configuration, we have

$$\|C(Y)\|_1 = \min(1, \|C(X)\|_1) \in \{0, 1\}.$$

We prove the lemma via a case analysis on $\|C(Y)\|_1$.

Case 1 : $\|C(Y)\|_1 = 0$

In this case, since C is a valid WTA configuration, according to Definition 3.11, conditioned on $N^t = C$, we must have $\|X^t\|_1 = 0$ and $a_s^t = a_c^t = 0$. By Corollary 3.6, since $\|X^t\|_1 = 0$,

$$\mathbb{P}[\|Y^{t+1}\|_1 = 0 | N^t = C] \geq 1 - ne^{-\gamma/2}.$$

By Lemma 3.7 conclusion (1), since $\|C(Y)\|_1 = 0$, $\mathbb{P}[a_s^{t+1} = a_c^{t+1} = 0 | N^t = C] \geq 1 - 2e^{-\gamma/2}$. By a union bound, recalling that X^t is fixed for all t by assumption, $\mathbb{P}[N^{t+1} = N^t | N^t = C] \geq 1 - (n+2)e^{-\gamma/2}$.

Case 2 : $\|C(Y)\|_1 = 1$

In this case by Definition 3.11, we have $C(y_i) \leq C(x_i)$ for all i , $C(a_s) = 1$, and $C(a_c) = 0$. We can thus apply Lemma 3.9, giving that $\mathbb{P}[Y^{t+1} = Y^t | N^t = C] \geq 1 - ne^{-\gamma/2}$. Additionally, by Lemma 3.7 conclusion (2), since $\|C(Y)\|_1 = 1$, $\mathbb{P}[a_s^{t+1} = 1 \text{ and } a_c^{t+1} = 0 | N^t = C] \geq 1 - 2e^{-\gamma/2}$. So by a union bound,

$$\mathbb{P}[N^{t+1} = N^t | N^t = C] \geq 1 - (n+2)e^{-\gamma/2},$$

giving the result in this case and completing the lemma. $\square$

Lemma 3.12 immediately implies a bound on the probability that $\mathcal{T}_{n,\gamma}$ remains in a valid WTA configuration for t_s consecutive times.

Corollary 3.13 (Stability of Valid WTA Configurations). *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t . For any time t and valid WTA configuration C with $C(X) = X^t$,*

$$\mathbb{P}[N^t = N^{t+1} = \dots = N^{t+t_s} | N^t = C] \geq 1 - t_s(n+2)e^{-\gamma/2}.$$

Proof. Applying Lemma 3.12 for each time $t+1, \dots, t+t_s$ in succession gives the result. $\square$

3.4 Convergence to Good Configurations

Configuration Type	$C(y_i) \leq C(x_i) \ \forall i?$	$\ C(Y)\ _1$	$C(a_s)$	$C(a_c)$
Valid WTA (Def. 3.11)	$\checkmark$	$\min(1, \ C(X)\ _1)$	1	0
Near-Valid WTA (Def. 3.15)	$\checkmark$	$\min(1, \ C(X)\ _1)$	1	1
Valid k -WTA (Def. 3.14)	$\checkmark$	$k \geq 2$	1	1
Reset (Def. 3.16)	–	–	0	0

Table 1: Summary of good configuration types (Definition 3.17), from which we show rapid convergence to a valid WTA configuration. We refer to the configuration types shaded in gray as *active configurations* (Definition 3.18).

With the stability bound of Corollary 3.13 in place, it remains to prove that $\mathcal{T}_{n,\gamma}$ converges quickly to a valid WTA configuration. We do this in two main steps:

1. In this Section, we define three additional good configuration types and show that all other network configurations converge to a good configuration within just two time steps with high probability.

2. In Sections 3.5 and 3.6 we show that, in turn, each of these good configurations rapidly converges to a valid WTA configuration with constant probability.

A high level illustration of our proof is shown in Figure 2.

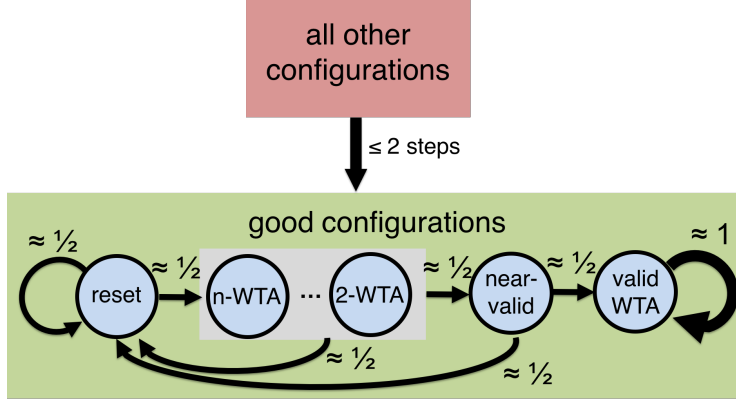


Figure 2: A high level illustration of our proof that $\mathcal{T}_{n,\gamma}$ solves the WTA problem. We show that all configurations converge to the set of good configurations. Once in a good configuration, $\mathcal{T}_{n,\gamma}$ converges to a valid WTA state very rapidly, with constant probability. In the illustration, arrow size corresponds to relative probability.

The first class of good configurations are *valid k -WTA configurations*. In such configurations the network behaves as expected before convergence. Multiple outputs corresponding to firing inputs fire and the inhibitors a_s and a_c fire, driving convergence towards a valid WTA configuration.

Definition 3.14 (Valid k -WTA Configuration). *For any $k \geq 2$, a valid k -WTA configuration of $\mathcal{T}_{n,\gamma}$ is a configuration C with $C(y_i) \leq C(x_i)$ for all $i \in \{1, \dots, n\}$, $\|C(Y)\|_1 = k$ and $C(a_s) = C(a_c) = 1$.*

We next define a class of *near-valid WTA configurations*, each of which is a small perturbation of a valid WTA configuration, with correct output but incorrect inhibitor behavior. We will show in Section 3.6 that the network rapidly converges to a near-valid WTA configuration from any configuration. In turn, it transitions with probability $\approx 1/2$ from a near-valid WTA configuration to a valid WTA configuration.

Definition 3.15 (Near-Valid WTA Configuration). *A near-valid WTA configuration of $\mathcal{T}_{n,\gamma}$ is a configuration C in which $C(Y)$ is a valid WTA output configuration for $C(X)$ (Definition 2.6) but $C(a_s) = C(a_c) = 1$.*

Finally, we define the class of *reset configurations* with $C(a_s) = C(a_c) = 0$. Since there is no inhibition in such a configuration, each output corresponding to a firing input will fire with probability $\geq 1/2$ at the next time. With probability $\approx 1/2$, the network will transition to either a valid k -WTA, a near-valid WTA, or a valid WTA configuration within three steps (Lemma 3.22).

Definition 3.16 (Reset Configuration). *A reset configuration of $\mathcal{T}_{n,\gamma}$ is a configuration C with $C(a_s) = C(a_c) = 0$.*

Definition 3.17 (Good Configuration). *A good configuration is any configuration that is either a valid WTA configuration (Definition 3.11), a valid k -WTA configuration (Definition 3.14), a reset configuration (Definition 3.16), or a near-valid WTA configuration (Definition 3.15).*

For conciseness, we also give a name to the good configurations excluding reset configurations:

Definition 3.18 (Active Configuration). *An active configuration is any configuration that is either a valid WTA configuration (Definition 3.11), a valid k -WTA configuration (Definition 3.14), or a near-valid WTA configuration (Definition 3.15).*

We first give a simple lemma building on Lemma 3.9, which characterizes the network's behavior when a single inhibitor and at least one output corresponding to a firing input fire at time t :

Lemma 3.19. *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t . For any time t and any configuration C with $C(X) = X^t$, $(C(a_s) = 1$ and $C(a_c) = 0)$ or $(C(a_s) = 0$ and $C(a_c) = 1)$, $\|C(Y)\|_1 \geq 1$, and $C(y_i) \leq C(x_i)$ for all i :*

$$\mathbb{P}[N^{t+1} \text{ is a valid WTA or valid } k\text{-WTA configuration} \mid N^t = C] \geq 1 - (n+2)e^{-\gamma/2}.$$

Proof. If $\|C(Y)\|_1 = 1$, then C is a valid WTA configuration. Thus, by Lemma 3.12, conditioned in $N^t = C$, N^{t+1} is a valid WTA configuration with probability $\geq 1 - (n+2)e^{-\gamma/2}$.

If $\|C(Y)\|_1 \geq 2$, then by Lemma 3.7 conclusion (3),

$$\mathbb{P}[a_s^{t+1} = a_c^{t+1} = 1 \mid N^t = C] \geq 1 - 2e^{-\gamma/2}.$$

Further, by Lemma 3.9, $\mathbb{P}[Y^{t+1} = Y^t \mid N^t = C] \geq 1 - ne^{-\gamma/2}$, which gives that N^{t+1} is a valid k -WTA configuration since, by conditioned on $N^t = C$, $\|Y^t\|_1 \geq 2$ and $y_i^t \leq x_i^t$ for all i . Thus, by a union bound, conditioned on $N^t = C$, N^{t+1} is a valid k -WTA configuration with probability $\geq 1 - (n+2)e^{-\gamma/2}$, giving the lemma. $\square$

Theorem 3.20 (Convergence to a Good Configuration). *Assume that the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t . For any time t and configuration C with $C(X) = X^t$,*

$$\mathbb{P}[\text{at least one of } \{N^{t+1}, N^{t+2}\} \text{ is a good configuration} \mid N^t = C] \geq 1 - 2(n+1)e^{-\gamma/2}.$$

Proof. Let $\mathcal{E}$ be the event that at least one of $\{N^{t+1}, N^{t+2}\}$ is a good configuration. Let $\mathcal{E}_1$ be the event that, $N^t = C$ and for all i , $y_i^{t+1} \leq x_i^{t+1}$. By Corollary 3.6 and the fact that X^t is fixed

$$\mathbb{P}[\mathcal{E}_1 \mid N^t = C] \geq 1 - ne^{-\gamma/2}. \quad (11)$$

We now give a simple case analysis, considering all values of a_c^{t+1} and a_s^{t+1} .

Case 1: $a_c^{t+1} = a_s^{t+1} = 0$.

In this case, N^{t+1} is a reset configuration (Definition 3.16). So we have:

$$\mathbb{P}[\mathcal{E} \mid \mathcal{E}_1, a_c^{t+1} = a_s^{t+1} = 0] = 1. \quad (12)$$

Case 2: $a_c^{t+1} = a_s^{t+1} = 1$.

If $\|Y^{t+1}\|_1 = 0$, then N^{t+2} is a reset configuration with probability $\geq 1 - 2e^{-\gamma/2}$ by Lemma 3.7 conclusion (1). If $\|Y^{t+1}\|_1 = 1$ and $\mathcal{E}_1$ occurs, then N^{t+1} is a near-valid WTA configuration (Definition 3.15). If $\|Y^{t+1}\|_1 \geq 2$ and $\mathcal{E}_1$ occurs, then N^{t+1} is a valid k -WTA configuration (Definition 3.14). Thus:

$$\mathbb{P}[\mathcal{E} \mid \mathcal{E}_1, a_c^{t+1} = a_s^{t+1} = 1] \geq 1 - 2e^{-\gamma/2}. \quad (13)$$

Case 3: $(a_s^{t+1} = 1 \text{ and } a_c^{t+1} = 0)$ or $(a_s^{t+1} = 0 \text{ and } a_c^{t+1} = 1)$.

Let $\mathcal{E}_2$ be the event that $(a_s^{t+1} = 1 \text{ and } a_c^{t+1} = 0)$ or $(a_s^{t+1} = 0 \text{ and } a_c^{t+1} = 1)$. Again, if $\|Y^{t+1}\|_1 = 0$, then N^{t+2} is a reset configuration with probability $\geq 1 - 2e^{-\gamma/2}$ by Lemma 3.7 conclusion (1).

If $\mathcal{E}_2$ and $\mathcal{E}_1$ occur and $\|Y^{t+1}\|_1 \geq 1$, we can apply Lemma 3.19, giving that N^{t+2} is either a valid WTA or valid k -WTA configuration with probability $\geq 1 - (n+2)e^{-\gamma/2}$. We thus have:

$$\mathbb{P}[\mathcal{E}|\mathcal{E}_1, \mathcal{E}_2] \geq 1 - (n+2)e^{-\gamma/2}. \quad (14)$$

Completing the lemma.

Combining (12), (13), and (14), by the law of total probability, $\mathbb{P}[\mathcal{E}|\mathcal{E}_1] \geq 1 - (n+2)e^{-\gamma/2}$. We can then use that $\mathbb{P}[\mathcal{E}_1|N^t = C] \geq 1 - ne^{-\gamma/2}$ by (11) to give:

$$\begin{aligned} \mathbb{P}[\mathcal{E}|N^t = C] &\geq \mathbb{P}[\mathcal{E}_1|N^t = C] \cdot \mathbb{P}[\mathcal{E}|\mathcal{E}_1, N^t = C] \\ &= \mathbb{P}[\mathcal{E}_1|N^t = C] \cdot \mathbb{P}[\mathcal{E}|\mathcal{E}_1] \quad (\text{Since, by definition, } \mathcal{E}_1 \text{ implies } N^t = C.) \\ &\geq 1 - 2(n+1)e^{-\gamma/2}, \end{aligned}$$

completing the lemma. □

3.5 Transition Lemmas for Good Configurations

We now give a set of lemmas that characterize the transitions of $\mathcal{T}_{n,\gamma}$ when starting from a good configuration. We first show that a near-valid WTA configuration transitions with probability $\approx 1/2$ to the adjacent valid WTA configuration (i.e., the configuration with the same output behavior, but correct inhibitor behavior).

Lemma 3.21 (From Near-Valid to Valid Configurations). *Assume that the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t . For any time t and near-valid WTA configuration C with $C(X) = X^t$,*

$$\mathbb{P}[N^{t+1} \text{ is a valid WTA configuration} | N^t = C] \geq 1/2 - (n+2)e^{-\gamma/2}.$$

Proof. We give a proof similar to that of Lemmas 3.9 and 3.12. We consider two cases:

Case 1: $\|C(Y)\|_1 = 0$.

In this case, for C to be a near-valid WTA configuration, according to Definition 3.15, we must also have $\|C(X)\|_1 = 0$. So by Corollary 3.6,

$$\mathbb{P}[\|Y^{t+1}\|_1 = 0 | N^t = C] \geq 1 - ne^{-\gamma/2}.$$

Additionally, by Lemma 3.7 conclusion (1), since $\|C(Y)\|_1 = 0$, $\mathbb{P}[a_s^{t+1} = a_c^{t+1} = 0 | N^t = C] \geq 1 - 2e^{-\gamma/2}$. By a union bound,

$$\mathbb{P}[\|N^{t+1}\|_1 = 0 | N^t = C] \geq 1 - (n+2)e^{-\gamma/2}.$$

By Definition 3.11, since $\|X^{t+1}\|_1 = 0$, if $\|N^{t+1}\|_1 = 0$, then N^{t+1} is a valid WTA configuration. So, conditioned on $N^t = C$, N^{t+1} is a valid WTA configuration with probability $\geq 1 - (n+2)e^{-\gamma/2}$, giving the lemma in this case.

Case 2: $\|C(Y)\|_1 = 1$.

In this case, for some i , $C(y_i) = 1$ and $C(y_j) = 0$ for all $j \neq i$. Further, we must have $C(x_i) = 1$ and $C(a_s) = C(a_c) = 1$ by the requirements of Definition 3.15. Define the event $\mathcal{E}_1$ by:

$$\mathcal{E}_1 \stackrel{\text{def}}{=} \left(y_i^{t+1} = 1, a_s^{t+1} = 1, a_c^{t+1} = 0 \text{ and } y_j^{t+1} = 0 \text{ for all } j \neq i \right).$$

By Lemma 3.10, $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] = 1/2$ and $\mathbb{P}[y_i^t \leq y_i^{t+1} \text{ for all } i | N^t = C] \geq 1 - ne^{-\gamma/2}$. By Lemma 3.7 conclusion (2), since $\|C(Y)\|_1 = 1$, $\mathbb{P}[a_s^{t+1} = 1 \text{ and } a_c^{t+1} = 0 | N^t = C] \geq 1 - 2e^{-\gamma/2}$. By a union bound this gives,

$$\mathbb{P}[\mathcal{E}_1 | N^t = C] \geq 1/2 - (n+2)e^{-\gamma/2}.$$

If $\mathcal{E}_1$ occurs, then N^{t+1} is a valid WTA configuration, giving the lemma in this case. $\square$

We next show that a reset configuration transitions to some active configuration (Definition 3.18) with probability $\approx 1/2$.

Lemma 3.22 (From Reset to Active Configurations). *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t . For any time t and reset configuration C with $C(X) = C^t$,*

$$\mathbb{P}[\text{at least one of } \{N^{t+1}, N^{t+2}, N^{t+3}\} \text{ is active} | N^t = C] \geq 1/2 - 3(n+2)e^{-\gamma/2}.$$

Proof. Let $\mathcal{E}$ be the event that at least one of $\{N^{t+1}, N^{t+2}, N^{t+3}\}$ is an active configuration. We consider two cases:

Case 1: $\|X^t\|_1 = 0$.

Let $\mathcal{E}_{20}$ be the event that $\|N^{t+2}\|_1 = 0$. That is, that no neurons fire at time $t+2$. By Lemma 3.8, since $\|X^t\|_1 = 0$, $\mathbb{P}[\mathcal{E}_{20} | N^t = C] \geq 1 - 2(n+1)e^{-\gamma/2}$.

$\mathcal{E}_{20}$ requires that no neurons fire in N^{t+2} , which makes this a valid WTA configuration since $\|X^t\|_1 = 0$ and the input is fixed for all t . Thus, $\mathbb{P}[\mathcal{E} | N^t = C] \geq \mathbb{P}[\mathcal{E}_{20} | N^t = C] \geq 1 - 2(n+1)e^{-\gamma/2}$, completing the lemma in this case.

Case 2: $\|X^t\|_1 \geq 1$.

Let $\mathcal{E}_{11}$ be the event that $N^t = C$ and $\|Y^{t+1}\| \geq 1$ (i.e., at least one output fires at time $t+1$) and $y_i^{t+1} \leq x_i^{t+1}$ for all i . For any y_i with $x_i^t = 1$, we have:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i)x_i^t + w(y_i, y_i)y_i^t + w(a_s, y_i)a_s^t + w(a_c, y_i)a_c^t - b(y_i) \\ &\geq 3\gamma + 0 + 0 + 0 - 3\gamma = 0. \end{aligned}$$

So, by Lemma 3.4, each output y_i with $x_i^t = 1$ fires with probability at least $1/2$ at time $t+1$. Each with $x_i^t = 0$ does not fire with probability $\geq 1 - e^{-\gamma/2}$ by Lemma 3.5. Since, by assumption, $\|X^t\|_1 \geq 1$, by a union bound, with probability $\geq 1/2 - ne^{-\gamma/2}$, at least one output with $x_i^t = 1$ fires in N^{t+1} , and no outputs with $x_i^t = 0$ fire. That is,

$$\mathbb{P}[\mathcal{E}_{11} | N^t = C] \geq 1/2 - ne^{-\gamma/2}. \tag{15}$$

We now proceed with a case analysis on the inhibitor behavior at time $t+1$:

Sub-case 1: $a_s^{t+1} = a_c^{t+1} = 1$.

In this case, assuming $\mathcal{E}_{11}$ occurs, N^{t+1} is either a valid k -WTA configuration for some k or a near-valid WTA configuration. We thus have:

$$\mathbb{P}[\mathcal{E}|\mathcal{E}_{11}, a_s^{t+1} = a_c^{t+1} = 1] = 1. \quad (16)$$

Sub-case 2: $a_s^{t+1} = a_c^{t+1} = 0$.

Let $\mathcal{E}_{21}$ be the event that $\|Y^{t+2}\|_1 \geq 1$, $y_i^{t+2} \leq x_i^{t+2}$ for all i , and $a_s^{t+2} = 1$. Assuming $\mathcal{E}_{11}$ occurs and $a_s^{t+1} = a_c^{t+1} = 0$, any output with $y_i^{t+1} = 1$ has

$$\begin{aligned} \text{pot}(y_i, t+2) &= w(x_i, y_i)x_i^{t+1} + w(y_i, y_i)y_i^{t+1} + w(a_s, y_i)a_s^{t+1} + w(a_c, y_i)a_c^{t+1} - b(y_i) \\ &= 3\gamma + 2\gamma + 0 + 0 - 3\gamma > 0. \end{aligned}$$

Thus, any such output has $y_i^{t+2} = 1$ with probability $\geq 1 - e^{-\gamma/2}$ by Lemma 3.4. Combined with Corollary 3.5, with probability $\geq 1 - ne^{-\gamma/2}$, at least one output fires in N^{t+2} , and no outputs with $x_i^{t+1} = 0$ fire. Further by Lemma 3.7 conclusions (2) and (3), since $\mathcal{E}_{11}$ requires that $\|Y^{t+1}\| \geq 1$, with probability $\geq 1 - 2e^{-\gamma/2}$, $a_s^{t+2} = 1$. Thus,

$$\mathbb{P}[\mathcal{E}_{21}|\mathcal{E}_{11}, a_s^{t+1} = a_c^{t+1} = 0] \geq 1 - (n+2)e^{-\gamma/2}. \quad (17)$$

Assume that both $\mathcal{E}_{11}$ and $\mathcal{E}_{21}$ occur. $\mathcal{E}_{21}$ requires that $a_s^{t+2} = 1$ and $\|Y^{t+2}\|_1 \geq 1$. If we also have $a_c^{t+2} = 1$, then N^{t+2} is a near-valid WTA or valid k -WTA configuration. If $a_c^{t+2} = 0$ and $\|Y^{t+2}\|_1 = 1$, then N^{t+2} is a valid WTA configuration. If $a_c^{t+2} = 0$ and $\|Y^{t+2}\|_1 \geq 2$, then by Lemma 3.7 conclusion (3), with probability $\geq 1 - 2e^{-\gamma/2}$, $a_s^{t+3} = a_c^{t+3} = 1$. Further, we can apply Lemma 3.9, giving that $Y^{t+3} = Y^{t+2}$ with probability $\geq 1 - ne^{-\gamma/2}$. This ensures that N^{t+3} is a valid k -WTA configuration. So we have by a union bound:

$$\mathbb{P}[\mathcal{E}|\mathcal{E}_{11}, \mathcal{E}_{21}, a_s^{t+1} = a_c^{t+1} = 0] \geq 1 - (n+2)e^{-\gamma/2}.$$

Combined with (17) the above gives:

$$\begin{aligned} \mathbb{P}[\mathcal{E}|\mathcal{E}_{11}, a_s^{t+1} = a_c^{t+1} = 0] &\geq \mathbb{P}[\mathcal{E}_{21}|\mathcal{E}_{11}, a_s^{t+1} = a_c^{t+1} = 0] \cdot \mathbb{P}[\mathcal{E}|\mathcal{E}_{11}, \mathcal{E}_{21}, a_s^{t+1} = a_c^{t+1} = 0] \\ &\geq \left(1 - (n+2)e^{-\gamma/2}\right) \cdot \left(1 - (n+2)e^{-\gamma/2}\right) \\ &\geq 1 - 2(n+2)e^{-\gamma/2}. \end{aligned} \quad (18)$$

Sub-case 3: $(a_s^{t+1} = 1 \text{ and } a_c^{t+1} = 0) \text{ or } (a_s^{t+1} = 0 \text{ and } a_c^{t+1} = 1)$.

Let $\mathcal{E}_{13}$ denote the event that $(a_s^{t+1} = 1 \text{ and } a_c^{t+1} = 0) \text{ or } (a_s^{t+1} = 0 \text{ and } a_c^{t+1} = 1)$. Assuming $\mathcal{E}_{13}$ and $\mathcal{E}_{11}$, we can apply Lemma 3.19, which gives that N^{t+2} is either a valid WTA or valid k -WTA configuration with probability $\geq 1 - (n+2)e^{-\gamma/2}$. I.e.,

$$\mathbb{P}[\mathcal{E}|\mathcal{E}_{11}, \mathcal{E}_{13}] \geq 1 - (n+2)e^{-\gamma/2}. \quad (19)$$

Completing Case 2.

Overall, combining (16), (18), and (19), by the law of total probability $\mathbb{P}[\mathcal{E}|\mathcal{E}_{11}] \geq 1 - 2(n+2)e^{-\gamma/2}$. Recalling that by (15) $\mathbb{P}[\mathcal{E}_{11}|N^t = C] \geq 1/2 - ne^{-\gamma/2}$:

$$\begin{aligned} \mathbb{P}[\mathcal{E}|N^t = C] &\geq \mathbb{P}[\mathcal{E}_{11}|N^t = C] \cdot \mathbb{P}[\mathcal{E}|\mathcal{E}_{11}, N^t = C] \\ &= \mathbb{P}[\mathcal{E}_{11}|N^t = C] \cdot \mathbb{P}[\mathcal{E}|\mathcal{E}_{11}] \quad (\text{Since, by assumption, } \mathcal{E}_{11} \text{ implies } N^t = C.) \\ &\geq 1/2 - 3(n+2)e^{-\gamma/2} \end{aligned}$$

which completes the lemma in this case. $\square$

We next show that, in each time step, with high probability, the number of firing outputs k does not increase. Further, with probability $\approx 1/2$, k is reduced by a factor of $1/2$. This ensures rapid convergence towards having just a single firing output (i.e., a near-valid WTA state). While there is some chance that the convergence will ‘overshoot’ the target and zero outputs will fire at some time step, we show that the probability of this event is upper bounded by the probability of the desired event – i.e., reaching a near-valid WTA configuration.

Lemma 3.23 (Progress from k -WTA Configurations). *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t . For any time t and any valid k -WTA configuration C with $C(X) = X^t$,*

1. *Letting $\mathcal{E}$ be the event that N^{t+1} is either a near-valid WTA configuration, a valid k -WTA configuration with $\|Y^{t+1}\|_1 \leq \|Y^t\|_1$, or has $\|Y^{t+1}\|_1 = 0$.*

$$\mathbb{P}[\mathcal{E}|N^t = C] \geq 1 - (n+2)e^{-\gamma/2}.$$

$$2. \mathbb{P}\left[\|Y^{t+1}\|_1 \leq \left\lceil \frac{\|Y^t\|_1}{2} \right\rceil \mid N^t = C\right] \geq 1/2 - (n+2)e^{-\gamma/2}.$$

$$3. \mathbb{P}[\|Y^{t+1}\|_1 = 0|N^t = C] - (n+2)e^{-\gamma/2} \leq \mathbb{P}[N^{t+1} \text{ is a near-valid} | N^t = C].$$

Proof. Since C is a valid k -WTA configuration, conditioned on $N^t = C$, we have $\|Y^t\|_1 \geq 2$ and $a_s^t = a_c^t = 1$. Event $\mathcal{E}$ in (1) above occurs if and only if $a_s^{t+1} = a_c^{t+1} = 1$, and $y_i^{t+1} \leq y_i^t$ for all i .

By Lemma 3.7 conclusion (3), since $\|Y^t\|_1 \geq 2$, both inhibitors remain firing at time $t+1$ with high probability. That is,

$$\mathbb{P}[a_s^{t+1} = a_c^{t+1} = 1|N^t = C] \geq 1 - 2e^{-\gamma/2}.$$

Further, by Lemma 3.10, $\mathbb{P}[y_i^{t+1} \leq y_i^t \text{ for all } i|N^t = C] \geq 1 - ne^{-\gamma/2}$. By a union bound, this gives that $\mathbb{P}[\mathcal{E}|N^t = C] \geq 1 - (n+2)e^{-\gamma/2}$, giving (1).

Let $\bar{Y} \subseteq Y$ denote the set of all output neurons with $y_i^t = 1$. Let $k = \|Y^t\|_1 = \|\bar{Y}^t\|_1$ and $k' = \|\bar{Y}^{t+1}\|_1$. By Lemma 3.10 properties (2) and (3), conditioned on $N^t = C$, k' is distributed according to the binomial distribution $B(k, 1/2)$. That is, it is the number of successes in k independent trials each with success probability $1/2$. Since $B(k, 1/2)$ is symmetric with mean $k/2$, its median is upper bounded by $\lceil k/2 \rceil$. Thus, $\mathbb{P}[k' \leq \lceil k/2 \rceil | N^t = C] \geq 1/2$. This gives by a union bound,

$$\mathbb{P}[k' \leq \lceil k/2 \rceil \text{ and } \mathcal{E}|N^t = C] \geq 1/2 - (n+2)e^{-\gamma/2}.$$

Note that if $\mathcal{E}$ holds, then $k' = \|\bar{Y}^{t+1}\|_1 = \|Y^{t+1}\|_1$, which thus gives conclusion (2).

Finally, we have $\mathbb{P}[k' = 1|N^t = C] = k \cdot \frac{1}{2^k}$ and $\mathbb{P}[k' = 0|N^t = C] = \frac{1}{2^k}$ and so

$$\mathbb{P}[k' = 0|N^t = C] \leq \mathbb{P}[k' = 1|N^t = C].$$

Let $\mathcal{E}_1$ be the event that N^{t+1} is a near-valid WTA configuration. Assuming $\mathcal{E}$ occurs, N^{t+1} is a near-valid WTA configuration if and only if $k' = 1$. That is,

$$\mathbb{P}[\mathcal{E}_1|\mathcal{E}, N^t = C] = \mathbb{P}[k' = 1|\mathcal{E}, N^t = C] = \mathbb{P}[k' = 1|N^t = C].$$

The second equality follows since $\mathcal{E}$ and k' are independent conditioned on N^t . k' only depends on

the firing of $y \in \bar{Y}$ while $\mathcal{E}$ only depends on the firing of $u \in \{a_s, a_c\} \cup (Y \setminus \bar{Y})$. Using the above:

$$\begin{aligned}
\mathbb{P}[\mathcal{E}_1 | N^t = C] &\geq \mathbb{P}[\mathcal{E} | N^t = C] \cdot \mathbb{P}[\mathcal{E}_1 | \mathcal{E}, N^t = C] \\
&\geq \left(1 - (n+2)e^{-\gamma/2}\right) \cdot \mathbb{P}[k' = 1 | N^t = C] \\
&\geq \mathbb{P}[k' = 0 | N^t = C] - (n+2)e^{-\gamma/2} \\
&\geq \mathbb{P}[\|Y^{t+1}\|_1 = 0 | N^t = C] - (n+2)e^{-\gamma/2}
\end{aligned}$$

where the last bound follows since $\|Y^{t+1}\|_1 \geq \|\bar{Y}_{t+1}\|_1 = k'$ so $\|Y^{t+1}\|_1 = 0$ at least requires $k' = 0$. This gives conclusion (3), completing the lemma. $\square$

3.6 Convergence to WTA

We now use the good configuration transition probabilities given in Section 3.5, along with the results of Section 3.4, to show that, if sufficiently large γ , starting from any configuration $\mathcal{T}_{n,\gamma}$ converges with probability $\geq 1/18$ to a valid WTA configuration within $O(\log n)$ steps (see Lemma 3.32). The proof is in four main parts, which we outline here. We first define:

Definition 3.24 (Terminal Configuration). *For $\mathcal{T}_{n,\gamma}$, a terminal configuration is any configuration C which is either a near-valid WTA configuration or has $\|C(Y)\|_1 = 0$ (i.e., no outputs fire).*

With this definition we can describe the general proof outline:

1. **Monotonicity** (Lemma 3.27). We prove that, starting from a k -WTA configuration, with high probability, $\mathcal{T}_{n,\gamma}$ remains in a k -WTA configuration, with the number of firing outputs consistently decreasing until it reaches a terminal configuration.
2. **Convergence** (Lemma 3.29). We prove that the number of firing outputs decreases rapidly. That is, starting from a k -WTA configuration, with high probability, a terminal configuration is reached within $O(\log n)$ steps.
3. **Probability of valid WTA** (Lemma 3.30, Corollary 3.31). We show that, starting from a valid k -WTA configuration, with constant probability, the terminal configuration reached is in fact a near-valid WTA configuration. By Lemma 3.21, with constant probability, this configuration transitions to a valid WTA configuration.
4. **Convergence from any starting configuration** (Theorem 3.32). We show that, starting in any configuration, with constant probability, the network reaches either a valid WTA configuration or a k -WTA configuration in few steps. Combined with our convergence results for k -WTA configurations, this proves fast convergence to a valid WTA state from any starting configuration.

We begin with a few definitions which we use to formalize the high level description above.

Definition 3.25 (Termination Step). *Given any infinite execution $\alpha = C^0 C^1 \dots$ let $\text{term}(\alpha, t, \Delta)$ be the minimum value in $\{t+1, \dots, t+\Delta\}$ for which $C^{\text{term}(\alpha, t, \Delta)}$ is a terminal configuration (Definition 3.24). If no such time exists let $\text{term}(\alpha, t, \Delta) = t + \Delta$.*

Definition 3.26 (Monotonicity Until Termination). *Let $\mathcal{E}_{\text{mono}}(t, \Delta)$ be the event that the execution of $\mathcal{T}_{n,\gamma}$ is in set of executions $\alpha = C^0 C^1 \dots$ satisfying:*

$$\begin{aligned}
&\{\alpha \mid \text{for all } t' \in \{t+1, \dots, \text{term}(\alpha, t, \Delta)\}, C^{t'} \text{ is a valid } k\text{-WTA configuration} \\
&\quad \text{with } \|Y^{t'}\|_1 \leq \|Y^{t'-1}\|_1\}.
\end{aligned}$$

We begin by showing that, starting from any k -WTA configuration, with high probability $\mathcal{T}_{n,\gamma}$ behaves monotonically as described above.

Lemma 3.27 (Monotonicity). *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t . For any time t , any valid k -WTA configuration C with $C(X) = X^t$, and any $\Delta \geq 1$,*

$$\mathbb{P}[\mathcal{E}_{mono}(t, \Delta) | N^t = C] \geq 1 - \Delta(n+2)e^{-\gamma/2}.$$

Proof. Consider any $\Delta \geq 2$. If $\mathcal{E}_{mono}(t, \Delta - 1)$ occurs, then either $N^{t+(\Delta-1)}$ is a valid k -WTA configuration, or, for some $t' \in \{t+1, \dots, t+\Delta-1\}$, $N^{t'}$ is a terminal configuration. Thus, by conclusion (1) of Lemma 3.23 we have:

$$\mathbb{P}[\mathcal{E}_{mono}(t, \Delta) | \mathcal{E}_{mono}(t, \Delta - 1), N^t = C] \geq 1 - (n+2)e^{-\gamma/2}. \quad (20)$$

Using (20) we can show by induction that for any $\Delta \geq 1$,

$$\mathbb{P}[\mathcal{E}_{mono}(t, \Delta) | N^t = C] \geq 1 - \Delta(n+2)e^{-\gamma/2}. \quad (21)$$

For any $\Delta \geq 2$, assume by way of induction that (21) holds for all $\Delta' < \Delta$. The assumption holds in the base case when $\Delta = 1$ again by conclusion (1) of Lemma 3.23, since C is a valid k -WTA configuration so $\mathbb{P}[\mathcal{E}_{mono}(t, 1) | N^t = C] \geq 1 - (n+2)e^{-\gamma/2}$. Applying (20) and the inductive assumption:

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{mono}(t, \Delta) | N^t = C] &\geq \mathbb{P}[\mathcal{E}_{mono}(t, \Delta - 1) | N^t = C] \cdot \mathbb{P}[\mathcal{E}_{mono}(t, \Delta) | \mathcal{E}_{mono}(t, \Delta - 1), N^t = C] \\ &\geq \left(1 - (\Delta - 1)(n+2)e^{-\gamma/2}\right) \cdot \left(1 - (n+2)e^{-\gamma/2}\right) \\ &\geq 1 - \Delta(n+2)e^{-\gamma/2}. \end{aligned}$$

which gives (21) for all $\Delta \geq 1$, and so the lemma. $\square$

We next show that, starting from a k -WTA configuration, with high probability, $\mathcal{T}_{n,\gamma}$ reaches a terminal configuration within $O(\log n)$ steps. This requires showing that for $\Delta = O(\log n)$ with high probability, $N^{\text{term}(\alpha, t, \Delta)}$ (where $\text{term}(\alpha, t, \Delta)$ is defined in Definition 3.25) is actually a terminal configuration. We note that if the network does not reach a terminal configuration within Δ steps after time t , then, by definition, $N^{\text{term}(\alpha, t, \Delta)} = N^{t+\Delta}$, which is some non-terminal configuration.

We first define a termination event:

Definition 3.28 (Termination by Δ). *Let $\mathcal{E}_{term}(t, \Delta)$ be the intersection of $\mathcal{E}_{mono}(t, \Delta)$ (Definition 3.26) and the event that $N^{\text{term}(\alpha, t, \Delta)}$ is a terminal configuration.*

Lemma 3.29 (Convergence from k -WTA Configurations). *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t and that $\gamma \geq 4\ln(n+2) + 10$. Let $\Delta = 12(\log_2 n + 2)$. For any time t and valid k -WTA configuration C with $C(X) = X^t$,*

$$\mathbb{P}[\mathcal{E}_{term}(t, \Delta) | N^t = C] \geq 1 - \Delta(n+2)e^{-\gamma/2} - \frac{1}{7n}$$

Proof. Let $\hat{\mathcal{E}}_{term}(t, \Delta)$ be the event that $\mathcal{E}_{mono}(t, \Delta)$ occurs but $\mathcal{E}_{term}(t, \Delta)$ does not. For any $t' \in \{t+1, \dots, t+\Delta\}$, define the indicator $I_{t'} \in \{0, 1\}$ with $I_{t'} = 1$ if and only if either:

- $N^{t'-1}$ is a valid k -WTA configuration and $\|Y^{t'}\|_1 \leq \lceil k/2 \rceil$.

- $N^{t'-1}$ is not a valid k -WTA configuration for any $k \geq 2$.

$\widehat{\mathcal{E}}_{term}(t, \Delta)$ requires that each of $N^t, \dots, N^{t+\Delta}$ is a valid k -WTA configuration and that

$$\|Y^t\|_1 \geq \|Y^{t+1}\|_1 \geq \dots \geq \|Y^{t+\Delta}\|_1 \geq 2.$$

Otherwise, a terminal configuration with $\|Y^{t'}\|_1 = 0$ would be reached and $\mathcal{E}_{term}(t, \Delta)$ would occur.

Initially $\|Y^t\|_1 \leq n$. Since each time $I_{t'} = 1$, either $\|Y^t\|_1$ is cut in half or a configuration other than a valid k -WTA configuration occurs, $\widehat{\mathcal{E}}_{term}(t)$ can only occur if $\sum_{t'=t+1}^{t+\Delta} I_{t'} < \log_2 n + 1$. Thus we can bound:

$$\mathbb{P}[\widehat{\mathcal{E}}_{term}(t, \Delta) | N^t = C] \leq \mathbb{P}\left[\sum_{t'=t+1}^{t+\Delta} I_{t'} < (\log_2 n + 1) | N^t = C\right]. \quad (22)$$

We will show that this probability is low since $I_{t'} = 1$ with good probability. Specifically, if $N^{t'-1}$ is not a valid k -WTA configuration, then $I_{t'} = 1$ deterministically. If $N^{t'-1}$ is a valid k -WTA configuration, then by conclusion (2) of Lemma 3.23, $I_{t'} = 1$ with probability $\geq 1/2 - (n+2)e^{-\gamma/2}$. Overall, we have: $\mathbb{P}[I_{t'} = 1 | N^{t'-1}] \geq 1/2 - (n+2)e^{-\gamma/2}$. In fact, by Lemma 2.2, we can also condition on all past configurations and have:

$$\mathbb{P}[I_{t'} = 1 | N^{t'-1} N^{t'-2} \dots N^t, N^t = C] \geq 1/2 - (n+2)e^{-\gamma/2}.$$

The above bound lets us use Lemma 2.3 to upper bound the probability that $\sum_{t'=t+1}^{t+\Delta} I_{t'}$ is below any value d by the probability that a sum of Δ independent coin flips, each with success probability $1/2 - (n+2)e^{-\gamma/2}$, is below d . Specifically, let $Z_{t+1}, \dots, Z_{t+\Delta}$ be i.i.d. random variables with $Z_{t'} = 1$ with probability $1/2 - (n+2)e^{-\gamma/2}$ and $Z_{t'} = 0$ otherwise. Invoking (22) and Lemma 2.3,

$$\mathbb{P}[\widehat{\mathcal{E}}_{term}(t, \Delta) | N^t = C] \leq \mathbb{P}\left[\sum_{t'=t+1}^{t+\Delta} I_{t'} < (\log_2 n + 1) | N^t = C\right] \quad (23)$$

$$\leq \mathbb{P}\left[\sum_{t'=t+1}^{t+\Delta} Z_{t'} < (\log_2 n + 1)\right]. \quad (24)$$

By our assumption that $\gamma \geq 4 \ln(n+2) + 10$ and our setting of $\Delta = 12(\log_2 n + 2) \leq 14n$:

$$\mathbb{E}\left[\sum_{t'=t+1}^{t+\Delta} Z_{t'}\right] = \Delta/2 - \Delta(n+2)e^{-\gamma/2} \geq \Delta/3 = 4(\log_2 n + 2).$$

By a standard Chernoff bound [MU05],

$$\mathbb{P}\left[\sum_{t'=t+1}^{t+\Delta} Z_{t'} \leq (\log_2 n + 1)\right] \leq e^{-\frac{(3/4)^2 \cdot 4(\log_2 n + 2)}{2}} \leq e^{-(\log_2 n + 2)} \leq \frac{1}{7n}.$$

We thus have, by (23), $\mathbb{P}[\widehat{\mathcal{E}}_{term}(t, \Delta) | N^t = C] \leq \frac{1}{7n}$. Combined with Lemma 3.27 this gives:

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{term}(t, \Delta) | N^t = C] &= \mathbb{P}[\mathcal{E}_{mono}(t, \Delta) | N^t = C] - \mathbb{P}[\widehat{\mathcal{E}}_{term}(t, \Delta) | N^t = C] \\ &\geq 1 - \Delta(n+2)e^{-\gamma/2} - \frac{1}{7n}. \end{aligned}$$

□

We next combine Lemma 3.29 with conclusion (3) of Lemma 3.23 and Lemma 3.21 to show that, starting from a valid k -WTA configuration, not only does $\mathcal{T}_{n,\gamma}$ reach a terminal configuration quickly, but also, if γ is large enough, this terminal configuration is a near-valid WTA configuration with probability $\approx 1/2$.

Lemma 3.30 (Constant Probability of Near-Valid WTA, from k -WTA Configurations). *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t and that $\gamma \geq 4\ln(n+2) + 10$. Let $\Delta = 12(\log_2 n + 2)$ and $\mathcal{E}_1(t)$ be the event that there is some $t' \in \{t+1, \dots, t+\Delta\}$, such that $N^{t'}$ is a near-valid WTA configuration. For any t and valid k -WTA configuration C with $C(X) = X^t$,*

$$\mathbb{P}[\mathcal{E}_1(t)|N^t = C] \geq \frac{1}{2} - \frac{\Delta+1}{2}(n+2)e^{-\gamma/2} - \frac{1}{14n}.$$

Proof. $\mathcal{E}_1(t)$ is equivalent to the event that $\mathcal{E}_{term}(t, \Delta)$ occurs and $N^{\text{term}(\alpha, t, \Delta)}$ is a near-valid WTA configuration. Let $\mathcal{E}_0(t)$ be the event that $\mathcal{E}_{term}(t, \Delta)$ occurs and $\|Y^{\text{term}(\alpha, t, \Delta)}\|_1 = 0$. $\mathcal{E}_0(t)$ and $\mathcal{E}_1(t)$ are disjoint with $\mathcal{E}_0(t) \cup \mathcal{E}_1(t) = \mathcal{E}_{term}(t, \Delta)$. So by Lemma 3.29,

$$\begin{aligned} \mathbb{P}[\mathcal{E}_0(t)|N^t = C] + \mathbb{P}[\mathcal{E}_1(t)|N^t = C] &= \mathbb{P}[\mathcal{E}_{term}(t, \Delta)|N^t = C] \\ &\geq 1 - \Delta(n+2)e^{-\gamma/2} - \frac{1}{7n}. \end{aligned} \quad (25)$$

We will use conclusion (3) of Lemma 3.23 to show that

$$\mathbb{P}[\mathcal{E}_1(t)|N^t = C] \geq \mathbb{P}[\mathcal{E}_0(t)|N^t = C] - (n+2)e^{-\gamma/2}, \quad (26)$$

which combined with (25) gives the conclusion of the lemma, that

$$\mathbb{P}[\mathcal{E}_1(t)|N^t = C] \geq \frac{1}{2} - \frac{\Delta+1}{2}(n+2)e^{-\gamma/2} - \frac{1}{14n}. \quad (27)$$

For each $t' \in \{t+1, \dots, t+\Delta\}$, let $\mathcal{E}_{term}(t, t', \Delta)$ be the event that $\mathcal{E}_{term}(t, \Delta)$ occurs and $\text{term}(\alpha, t, \Delta) = t'$. Define $\mathcal{E}_0(t, t')$ and $\mathcal{E}_1(t, t')$ analogously. Let $\mathcal{E}_2(t, t')$ be the event that $N^t, \dots, N^{t'-1}$ are all valid k -WTA configurations with $\|Y^t\|_1 \geq \dots \geq \|Y^{t'-1}\|_1 \geq 2$. Let $\bar{\mathcal{E}}_2(t, t')$ be its complement.

$$\mathbb{P}[\mathcal{E}_1(t, t')|\bar{\mathcal{E}}_2(t, t')] = \mathbb{P}[\mathcal{E}_0(t, t')|\bar{\mathcal{E}}_2(t, t')] = 0 \quad (28)$$

since both $\mathcal{E}_1(t, t')$ and $\mathcal{E}_0(t, t')$ require $\mathcal{E}_{mono}(t, \Delta)$ to hold, which requires $\mathcal{E}_2(t, t')$ to hold if $\text{term}(\alpha, t, \Delta) = t'$. Further, by conclusion (3) of Lemma 3.23, since $\mathcal{E}_2(t, t')$ requires that $N^{t'-1}$ is a valid k -WTA configuration,

$$\mathbb{P}[\mathcal{E}_1(t, t')|\mathcal{E}_2(t, t'), N^t = C] \geq \mathbb{P}[\mathcal{E}_0(t, t')|\mathcal{E}_2(t, t'), N^t = C] - (n+2)e^{-\gamma/2} \quad (29)$$

By the law of total probability, (28) and (29) give

$$\mathbb{P}[\mathcal{E}_1(t, t')|N^t = C] \geq \mathbb{P}[\mathcal{E}_0(t, t')|N^t = C] - (n+2)e^{-\gamma/2}.$$

Again by the law of total probability, this gives

$$\mathbb{P}[\mathcal{E}_1(t)|N^t = C] \geq \mathbb{P}[\mathcal{E}_0(t)|N^t = C] - (n+2)e^{-\gamma/2},$$

yielding (26) and thus (27) and the lemma. $\square$

We next combine Lemma 3.30 with Lemma 3.21, which shows that any near-valid WTA configuration transitions with probability $\approx 1/2$ to a valid WTA configuration. This gives fast convergence to a valid WTA configuration starting from any valid k -WTA configuration, with probability $\geq 1/8$.

Corollary 3.31 (Constant Probability of Success, from k -WTA Configurations). *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t and that $\gamma \geq 4\ln(n+2) + 10$. Let $\mathcal{E}(t)$ be the event that there is some $t' \in \{t+1, \dots, t+12\log_2 n + 25\}$, such that $N^{t'}$ is a valid WTA configuration. For any t and valid k -WTA configuration C with $C(X) = X^t$,*

$$\mathbb{P}[\mathcal{E}(t)|N^t = C] \geq 1/8.$$

Proof. As in Lemma 3.30, let $\Delta = 12(\log_2 n + 2)$ and $\mathcal{E}_1(t)$ be the event that there is some $t' \in \{t+1, \dots, t+\Delta\}$, such that $N^{t'}$ is a near-valid WTA configuration. Let $\mathcal{E}_{val}(t)$ be the event that $N^{t'+1}$ is a valid WTA configuration. We have $\mathcal{E}(t) \subseteq \mathcal{E}_{val}(t)$ (since $t'+1 \in \{t+2, \dots, t+\Delta+1\}$ where $t+\Delta+1 = t+(12\log_2 n + 25)$). Thus it suffices to show that $\mathbb{P}[\mathcal{E}_{val}(t)|N^t = C] \geq 1/8$.

By Lemmas 3.30 and 3.21, for any configuration C :

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{val}(t)|N^t = C] &\geq \mathbb{P}[\mathcal{E}_1(t)|N^t = C] \cdot \mathbb{P}[\mathcal{E}_{val}(t)|\mathcal{E}_1(t), N^t = C] \\ &\geq \left(\frac{1}{2} - \frac{\Delta+1}{2}(n+2)e^{-\gamma/2} - \frac{1}{14n}\right) \cdot \left(1/2 - (n+2)e^{-\gamma/2}\right) \\ &\geq \frac{1}{4} - \frac{\Delta+3}{4}(n+2)e^{-\gamma/2} - \frac{1}{28n}. \end{aligned}$$

We can loosely bound $\frac{\Delta+3}{4} = \frac{12(\log_2 n + 2) + 3}{4} \leq \frac{12(n+2n) + 3n}{4} \leq 10n$. Further, by our assumption that $\gamma \geq 4\ln(n+2) + 10$ we have:

$$\mathbb{P}[\mathcal{E}_{val}(t)|N^t = C] \geq \frac{1}{4} - \frac{10n(n+2)}{(n+2)^2 \cdot e^5} - \frac{1}{28} \geq \frac{1}{8}$$

which gives the corollary. $\square$

Finally, we show that, starting from *any configuration*, with constant probability, $\mathcal{T}_{n,\gamma}$ converges to a valid WTA configuration in $O(\log n)$ steps. Our proof combines Theorem 3.20 and Lemma 3.22 which show that any configuration transitions to an active configuration (Definition 3.18) in few steps with constant probability. We then apply Corollary 3.31 to show convergence from such a configuration.

Theorem 3.32. *Assume the input execution α_X of $\mathcal{T}_{n,\gamma}$ has X^t fixed for all t and that $\gamma \geq 4\ln(n+2) + 10$. Let $\mathcal{E}(t)$ be the event that there is some $t' \in \{t+1, \dots, t+(12\log_2 n + 30)\}$ such that $N^{t'}$ is a valid WTA configuration. For any time t and configuration C with $C(X) = X^t$,*

$$\mathbb{P}[\mathcal{E}(t)|N^t = C] \geq 1/18.$$

Proof. Let $\mathcal{E}_{act}(t)$ be the event that $N^t = C$ and that at least one of $\{N^{t+1}, \dots, N^{t+5}\}$ is an active configuration (Definition 3.18). Let $N^{\hat{t}}$ be the first active configuration in this set, or $N^{\hat{t}} = N^{t+5}$ if there is no such configuration.

By Theorem 3.20, conditioned on $N^t = C$, with probability $\geq 1 - 2(n+1)e^{-\gamma/2}$ one of $\{N^{t+1}, N^{t+2}\}$ is a good configuration. Let $N^{\hat{t}}$ be the first good configuration in this set or $N^{\hat{t}} = N^{t+2}$ if neither are good. If $N^{\hat{t}}$ is also an active configuration then $\mathcal{E}_{act}(t)$ holds.

If not, then $N^{\hat{t}}$ is a reset configuration. Let C' be any reset configuration. By Lemma 3.22, conditioned on $N^{\hat{t}} = C'$, with probability $\geq 1/2 - 3(n+2)e^{-\gamma/2}$ at least one of $\{N^{\hat{t}+1}, N^{\hat{t}+2}, N^{\hat{t}+3}\}$, is an active configuration. Thus, overall we have:

$$\begin{aligned}\mathbb{P}[\mathcal{E}_{act}|N^{\hat{t}} = C] &\geq \left(1 - 2(n+1)e^{-\gamma/2}\right) \cdot \left(1/2 - 3(n+2)e^{-\gamma/2}\right) \\ &\geq \frac{1}{2} - 5(n+2)e^{-\gamma/2}.\end{aligned}\tag{30}$$

We can define three disjoint events:

$$\begin{aligned}\mathcal{E}_{act,1}(t) &\stackrel{\text{def}}{=} (N^{\hat{t}} \text{ is a valid } k\text{-WTA configuration}) \\ \mathcal{E}_{act,2}(t) &\stackrel{\text{def}}{=} (N^{\hat{t}} \text{ is a near-valid WTA configuration}) \\ \mathcal{E}_{act,3}(t) &\stackrel{\text{def}}{=} (N^{\hat{t}} \text{ is a valid WTA configuration})\end{aligned}$$

We have $\mathcal{E}_{act}(t) = \bigcup_{i=1}^3 \mathcal{E}_{act,i}(t)$ and so by the law of total probability:

$$\begin{aligned}\mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act}(t)] &= \sum_{i=1}^3 \mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act,i}(t)] \cdot \mathbb{P}[\mathcal{E}_{act,i}(t)|\mathcal{E}_{act}(t)] \\ &\geq \sum_{i=1}^3 \min_{i \in \{1,2,3\}} \mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act,i}(t)] \cdot \mathbb{P}[\mathcal{E}_{act,i}(t)|\mathcal{E}_{act}(t)] \\ &\geq \min_{i \in \{1,2,3\}} \cdot \mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act,i}(t)]\end{aligned}\tag{31}$$

where the last bound follows since the $\mathcal{E}_{act,i}(t)$ events are disjoint and $\sum_{i=1}^3 \mathbb{P}[\mathcal{E}_{act,i}(t)|\mathcal{E}_{act}(t)] = 1$. We now bound this minimum via a case analysis:

Case 1: $\mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act,1}(t)]$

In this case, $N^{\hat{t}}$ is a valid k -WTA configuration, so applying Corollary 3.31, conditioned on $\mathcal{E}_{act,1}(t)$, with probability $1/8$ there is some time $t' \in \{\hat{t}+1, \dots, \hat{t} + (12 \log_2 n + 25)\}$ such that $N^{t'}$ is a valid WTA configuration. Note that $\hat{t} \leq t + 5$ giving $t' \leq (12 \log_2 n + 30)$. We thus have:

$$\mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act,1}(t)] \geq 1/8.\tag{32}$$

Case 2: $\mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act,2}(t)]$

In this case, $N^{\hat{t}}$ is a near-valid WTA configuration, so applying Lemma 3.21, conditioned on $\mathcal{E}_{act,2}(t)$, $N^{\hat{t}+1}$ is a valid WTA configuration with probability $\geq 1/2 - (n+2)e^{-\gamma/2}$. By our assumption that $\gamma \geq 4 \ln(n+2) + 10$:

$$\mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act,2}(t)] \geq 1/2 - (n+2)e^{-\gamma/2} \geq 1/3.\tag{33}$$

Case 3: $\mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act,3}(t)]$

If $\mathcal{E}_{act,3}(t)$ holds then $N^{\hat{t}}$ is a valid WTA configuration by definition, so trivially

$$\mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act,3}(t)] = 1.\tag{34}$$

Completing the theorem:

Combining (32), (33), (34), and (31) we have $\mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act}(t)] \geq 1/8$. Using (30) we then have:

$$\begin{aligned} \mathbb{P}[\mathcal{E}(t)|N^t = C] &\geq \mathbb{P}[\mathcal{E}_{act}(t)|N^t = C] \cdot \mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act}(t), N^t = C] \\ &= \mathbb{P}[\mathcal{E}_{act}(t)|N^t = C] \cdot \mathbb{P}[\mathcal{E}(t)|\mathcal{E}_{act}(t)] \quad (\text{Since } \mathcal{E}_{act}(t) \subseteq (N^t = C) \text{ by definition.}) \\ &\geq \frac{1}{8} \cdot \left(\frac{1}{2} - 5(n+2)e^{-\gamma/2} \right) \geq \frac{1}{18} \end{aligned}$$

where the last bound follows from our assumption that $\gamma \geq 4\ln(n+2) + 10$. $\square$

3.7 Completing the Bounds

Given Theorem 3.32, it is easy to show that, with γ set large enough, $\mathcal{T}_{n,\gamma}$ solves the WTA problem (Definitions 2.7 and 2.9), giving Theorem 3.2 and 3.3. We start with the basic WTA problem of Definition 2.7. By Theorem 3.32, starting from any configuration, the network converges to a valid WTA configuration in $O(\log n)$ steps. By applying this analysis in sequence to $O(\log 1/\delta)$ sets of $O(\log n)$ steps, we show that the network converges to a valid WTA state with probability $\geq 1 - \delta$ within $O(\log n \cdot \log(1/\delta))$ steps. Further, if γ is large enough, by Lemma 3.13, it remains in this state for t_s steps with high probability.

Theorem 3.2 (Two-Inhibitor WTA). *For $\gamma \geq 4\ln((n+2)t_s/\delta) + 10$, $\mathcal{T}_{n,\gamma}$ solves WTA(n, t_c, t_s, δ) for any $t_c \geq 72(\log_2 n + 1) \cdot (\log_2(1/\delta) + 1)$.*

Proof. Consider $\mathcal{T}_{n,\gamma}$ starting from any initial configuration N^0 and given an infinite input execution α_X with X^t fixed for all t . Let $\Delta = (12\log_2 n + 30)$ and $r = 6(\log_2(1/\delta) + 1)$. Let $\mathcal{E}$ be the event that there is some time $t \leq t_c$ where N^t is a valid WTA configuration.

For any $i \geq 0$, let $\mathcal{E}_i$ be the event that there is some time $t \in \{i\Delta + 1, \dots, (i+1)\Delta\}$ where N^t is a valid WTA configuration. By Theorem 3.32 and Lemma 2.2 we have:

$$\mathbb{P}[\mathcal{E}_i|N^{i\Delta}] = \mathbb{P}[\mathcal{E}_i|N^{i\Delta}, N^{i\Delta-1}, \dots, N^1] \geq 1/8.$$

Let $Z_0, \dots, Z_{r-1} \in \{0, 1\}$ be independent coin flips, with $\mathbb{P}[Z_i = 1] = 1/8$. Applying Lemma 2.3:

$$\mathbb{P}[\mathcal{E}] = \mathbb{P}\left[\bigcap_{i=0}^{r-1} \mathcal{E}_i\right] \geq \mathbb{P}\left[\sum_{i=0}^{r-1} Z_i \geq 1\right] = 1 - \left(\frac{7}{8}\right)^r.$$

Using that $r = 6(\log_2(1/\delta) + 1)$:

$$\mathbb{P}[\mathcal{E}] \geq 1 - \left(\frac{7}{8}\right)^{6(\log_2(1/\delta)+1)} \geq 1 - \frac{\delta}{2}.$$

Thus, with probability $\geq 1 - \frac{\delta}{2}$ there is some time $t \leq r \cdot \Delta \leq 72(\log_2 n + 1) \cdot (\log_2(1/\delta) + 1) \leq t_c$ in which N^t is a valid WTA configuration. By Corollary 3.13, if C is a valid WTA configuration then

$$\mathbb{P}[N^t = N^{t+1} = \dots = N^{t+t_s}|N^t = C] \geq 1 - t_s(n+2)e^{-\gamma/2} \geq 1 - \frac{\delta}{e^5},$$

where the bound holds by our assumption that $\gamma \geq 4\ln((n+2)t_s/\delta) + 10$. We thus have that the network reaches a valid WTA configuration within time t_c and remains in it for time t_s with probability $\geq (1 - \frac{\delta}{2}) \cdot (1 - \frac{\delta}{e^5}) \geq 1 - \delta$, yielding the theorem. $\square$

We conclude by showing with what parameters $\mathcal{T}_{n,\gamma}$ solves the expected-time WTA problem of Definition 2.9.

Theorem 3.3 (Two-Inhibitor Expected-Time WTA). *For $\gamma \geq 4 \ln((n+2)t_s) + 10$, $\mathcal{T}_{n,\gamma}$ solves WTA-EXP(n, t_c, t_s) for any $t_c \geq 108(\log_2 n + 3)$.*

Proof. Recall that in Definition 2.9 we defined the convergence time for any infinite input execution α_X and output execution α_Y :

$$t(\alpha_X, t_s, \alpha_Y) = \min \{t : Y^t \text{ is a valid WTA output configuration for } X^t \text{ and } Y^t = \dots = Y^{t+t_s}\}.$$

Define the worst case expected convergence time of $\mathcal{T}_{n,\gamma}$ on input α_X by:

$$t_{max}(\alpha_X) = \max_{N^0} \left(\mathbb{E}_{\alpha_Y \sim \mathcal{D}_Y(\mathcal{T}_{n,\gamma}, N^0, \alpha_X)} t(\alpha_X, t_s, \alpha_Y) \right).$$

To prove the lemma we must prove that for any α_X with X^t fixed for all t , $t_{max}(\alpha_X) \leq 108(\log_2 n + 3)$. Fixing such an α_X , for any starting configuration N^0 , let $\mathbb{E}_{N^0}$ and $\mathbb{P}_{N^0}$ denote the expectation and probability of an event taken over executions drawn from $\mathcal{D}(\mathcal{T}_{n,\gamma}, N^0, \alpha_X)$.

Let $\Delta = (12 \log_2 n + 30)$ and let $\mathcal{E}_1$ be the event that there is some $t \in \{1, \dots, \Delta\}$ where N^t is a valid WTA configuration. Let $\mathcal{E}_{stab}$ be the event that there is some $t \in \{1, \dots, \Delta\}$ where N^t is a valid WTA configuration and additionally, where $N^t = \dots = N^{t+t_s}$. Let $\bar{\mathcal{E}}_1$ and $\bar{\mathcal{E}}_{stab}$ be the complements of these two events. By Theorem 3.32, for any initial configuration N^0

$$\mathbb{P}_{N^0}[\mathcal{E}_1] \geq 1/8. \quad (35)$$

Further, by Corollary 3.13, if C is a valid WTA configuration then,

$$\mathbb{P}_{N^0}[N^t = N^{t+1} = \dots = N^{t+t_s} | N^t = C] \geq 1 - t_s(n+2)e^{-\gamma/2} \geq 1 - \frac{1}{t_s \cdot e^5} \quad (36)$$

where the bound holds since $\gamma \geq 4 \ln((n+2)t_s) + 10 \geq 2 \ln((n+2)t_s^2) + 10$ and so $e^{-\gamma/2} \leq \frac{1}{(n+2)t_s^2 \cdot e^5}$. Together (35) and (36) give that:

$$\mathbb{P}_{N^0}[\mathcal{E}_{stab}] \geq \mathbb{P}_{N^0}[\mathcal{E}_{stab} | \mathcal{E}_1] \cdot \mathbb{P}_{N^0}[\mathcal{E}_1] \geq \frac{1}{8} \cdot \left(1 - \frac{1}{t_s \cdot e^5} \right) \geq \frac{1}{8} - \frac{1}{t_s \cdot e^5}.$$

We can write:

$$\begin{aligned} \mathbb{E}_{N^0}[t(\alpha_X, t_s, \alpha_Y)] &= \mathbb{E}_{N^0}[t(\alpha_X, t_s, \alpha_Y) | \mathcal{E}_{stab}] \cdot \mathbb{P}_{N^0}[\mathcal{E}_{stab}] \\ &\quad + \mathbb{E}_{N^0}[t(\alpha_X, t_s, \alpha_Y) | \mathcal{E}_1, \bar{\mathcal{E}}_{stab}] \cdot \mathbb{P}_{N^0}[\mathcal{E}_1, \bar{\mathcal{E}}_{stab}] \\ &\quad + \mathbb{E}_{N^0}[t(\alpha_X, t_s, \alpha_Y) | \bar{\mathcal{E}}_1] \cdot \mathbb{P}_{N^0}[\bar{\mathcal{E}}_1] \end{aligned} \quad (37)$$

Conditioned on $\mathcal{E}_{stab}$ (which also requires that $\mathcal{E}_1$ occurs), the network converges within Δ steps and stabilizes for t_s steps. Thus, we have:

$$\mathbb{E}_{N^0}[t(\alpha_X, t_s, \alpha_Y) | \mathcal{E}_{stab}] \leq \Delta.$$

Conditioned on $\mathcal{E}_1, \bar{\mathcal{E}}_{stab}$ the network converges, but does not stabilize. We can bound

$$\begin{aligned}\mathbb{E}_{N_0}[t(\alpha_X, t_s, \alpha_Y)|\mathcal{E}_1, \bar{\mathcal{E}}_{stab}] &\leq (\Delta + t_s) + \mathbb{E}_{N^{\Delta+t_s}}[t(\alpha_X, t_s, \alpha_Y)] \\ &\leq \Delta + t_s + t_{max}(\alpha_X).\end{aligned}$$

Finally, conditioned on $\bar{\mathcal{E}}_1$, the network does not converge within Δ steps. We have:

$$\begin{aligned}\mathbb{E}_{N_0}[t(\alpha_X, t_s, \alpha_Y)|\bar{\mathcal{E}}_1] &\leq \Delta + \mathbb{E}_{N^\Delta}[t(\alpha_X, t_s, \alpha_Y)] \\ &\leq \Delta + t_{max}(\alpha_X).\end{aligned}$$

We can plug these bounds along with the probability bounds of (35) and (36) into (37) to obtain:

$$\begin{aligned}\mathbb{E}_{N_0}[t(\alpha_X, t_s, \alpha_Y)] &\leq \Delta \cdot \left(\frac{1}{8} - \frac{1}{t_s \cdot e^5}\right) + (\Delta + t_{max}(\alpha_X) + t_s) \cdot \frac{1}{t_s \cdot e^5} + (\Delta + t_{max}(\alpha_X)) \cdot \frac{7}{8} \\ &\leq \Delta + t_{max}(\alpha_X) \left(\frac{7}{8} + \frac{1}{t_s \cdot e^5}\right) + \frac{t_s}{t_s \cdot e^5} \\ &\leq \Delta + t_{max}(\alpha_X) \cdot \frac{8}{9} + \frac{1}{e^5}.\end{aligned}$$

Since this bound holds for all N^0 we have:

$$t_{max}(\alpha_X) \leq \Delta + t_{max}(\alpha_X) \cdot \frac{8}{9} + \frac{1}{e^5}$$

which gives $t_{max}(\alpha_X) \leq 9\Delta + \frac{9}{e^5} \leq 9\Delta + 1$. This bound holds for all α_X and so gives the lemma, after recalling that $\Delta = (12 \log_2 n + 30)$ so $9\Delta + 1 \leq 108(\log_2 n + 3)$. $\square$

4 WTA Lower Bounds

The simple family of two-inhibitor networks presented in Section 3 gives convergence to a valid WTA output configuration in $O(\log n \cdot \log(1/\delta))$ steps with probability $\geq 1 - \delta$, as long as the weight scaling parameter γ is set large enough. Specifically, by Theorem 3.2, these networks solve $\text{WTA}(n, t_c, t_s, \delta)$ with $t_c = O(\log n \cdot \log(1/\delta))$ and t_s exponentially large in γ . In this section we ask whether this is optimal, considering two questions:

1. Are there networks that achieve comparable convergence speed with just a single auxiliary neuron?
2. Are there networks using two auxiliary neurons that converge faster?

We answer these questions for somewhat restricted classes of *simple SNNs* and *symmetric SNNs*, described in Definitions 4.1 and 4.2 below. These classes of networks include, in particular, the construction studied in Section 3.

We show that a simple SNN with just a single auxiliary neuron cannot solve $\text{WTA}(n, t_c, t_s, \delta)$ with $t_s = \tilde{\Omega}\left(\frac{t_c}{\log n}\right)$. That is, the network cannot effectively converge to a valid WTA output state and remain in this state for a significant time compared to its convergence time. Additionally, we show that no symmetric SNN with two auxiliary neurons can improve on the convergence time of the two-inhibitor network $\mathcal{T}_{n,\gamma}$ proven in Theorem 3.2 by more than a $O(\log \log n)$ factor.

We define the restricted network classes we consider in our lower bounds below.

Definition 4.1 (Simple SNN). A spiking neural network $\mathcal{N} = \langle N, w, b, f \rangle$ is a simple SNN if it contains n input neurons labeled $x_1, \dots, x_n$ and n output neurons labels $y_1, \dots, y_n$ and satisfies:

- $w(x_i, y_j) = 0$ and $w(y_i, y_j) = 0$ for all $j \neq i$. I.e., each input does not connect to outputs, other than its corresponding output, and outputs do not connect to each other.

Note that in a simple SNN, auxiliary neurons may connect to each other, may have incoming edges from the input neurons, and may form unrestricted connections with the output neurons. In our two-auxiliary neuron lower bound we consider a further restricted class of networks:

Definition 4.2 (Symmetric SNN). A simple SNN $\mathcal{N} = \langle N, w, b, f \rangle$ is a symmetric SNN if it contains n input neurons labeled $x_1, \dots, x_n$ and n output neurons labels $y_1, \dots, y_n$ and satisfies:

- For all $u, v \in A$, $w(u, v) = w(v, u) = 0$. I.e., there are no connections between auxiliary neurons.
- For all $u \in A$, $w(y_i, u) = w(y_j, u)$. I.e., each auxiliary neuron is affected in the same way by each output.
- For all i, j , $w(x_i, y_i) = w(x_j, y_j)$, $w(y_i, y_i) = w(y_j, y_j)$, $w(u, y_i) = w(u, y_j)$ for all $u \in A$, and $b(y_i) = b(y_j)$. I.e., all outputs have identical incoming connections from their corresponding inputs, themselves, and the auxiliary neurons, and have identical biases.

4.1 Single Auxiliary Neuron Lower Bound

We begin with our lower bound for simple SNNs with just a single auxiliary neuron.

Theorem 4.3 (One Neuron Lower Bound). For any $n \geq 20$, $\delta \leq 1/2$, and any spike probability function $f : \mathbb{R} \rightarrow [0, 1]$ (satisfying the restrictions in Section 2.1), there is no simple SNN $\mathcal{N} = \langle N, w, b, f \rangle$ with just a single auxiliary neuron that solves WTA(n, t_c, t_s, δ) with $t_s > 10t_c \cdot \frac{\ln(2t_c)}{\ln n}$.

We note that Theorem 4.3 applies to simple SNNs (Definition 4.1). However, we conjecture that the result holds for any SNN, without structural restrictions. We leave proving a more general bound as an open question, for now.

Proof Outline. We will prove Theorem 4.3 assuming that the single auxiliary neuron is an inhibitor. It will be easy to see that a nearly identical proof goes through when the auxiliary neuron is excitatory. At a high level, our proof shows that the convergence and stability inhibitors employed by the two-inhibitor networks described Section 3 are necessary. A single auxiliary neuron is not able to both drive fast convergence to a valid WTA output configuration and to maintain stability once the network is in such a state. In more detail, our proof breaks into three steps:

1. We show in Lemma 4.4 that the network must be relatively ‘active’ if it solves WTA(n, t_c, t_s, δ). Specifically, if the single inhibitor a does not fire, then any output corresponding to a firing input must fire with probability $\geq 1 - \delta^{1/t_c}$. Otherwise, starting from a configuration in which no outputs fire, the network would take longer than t_c steps to reach a valid WTA output configuration with probability $1 - \delta$.
2. Conversely, we show in Lemma 4.8 that if the inhibitor a does fire, then any output with a firing input must *cease firing* at the next time with probability $\geq 1 - n^{-\frac{1}{10t_c}}$. Otherwise, starting from a configuration in which $\Omega(n)$ outputs fire, with probability $\geq 1/2$, the network would take longer than t_c steps to converge to a valid WTA output configuration (with a single firing output).

3. We combine these results a network with just one inhibitor cannot maintain a valid WTA output configuration for $t_s = \Omega\left(t_c \cdot \frac{\ln t_c}{\ln n}\right)$ consecutive steps with probability $\geq 1/2$ (i.e., the network cannot achieve sufficient stability).

Consider any time t in which $\mathcal{N}$ is in a valid WTA output configuration. If a does not fire at time t , then by (1), if there are $\Omega(n)$ active inputs, at least one output which did not fire at time t fires with probability $\geq 1 - \delta^{\Omega(n/t_c)}$ at time $t+1$. If a does fire, then by (2) the winning output stops firing at time $t+1$ with probability $\geq 1 - n^{-\frac{1}{10t_c}}$. In any case, if $\delta \leq 1/2$, with probability $\geq 1 - n^{-\frac{1}{10t_c}}$, the output configuration changes, and stability is broken.

Since this relatively high probability of breaking stability holds at any time in which $\mathcal{N}$ is in a valid WTA output configuration, it is enough to show that stability cannot be maintained with probability $\geq \epsilon$ for $\Omega\left(t_c \cdot \frac{\ln 1/\epsilon}{\ln n}\right)$ steps. By setting $\epsilon = O(t_c)$ and applying a union bound, we can show that, with probability $\geq 1/2$, in the first t_c time steps, $\mathcal{N}$ never reaches a valid WTA output configuration and remains in this configuration for t_s consecutive steps. Thus, $\mathcal{N}$ does not solve $\text{WTA}(n, t_c, t_s, \delta)$ for $\delta \leq 1/2$.

We start by showing that if the inhibitor a does not fire at time t , then any output corresponding to a firing input must fire with reasonably high probability at time $t+1$. We in fact prove a more general lemma, for networks containing any number of inhibitors, since this result will be useful in our lower bound for two-inhibitor networks presented in Section 4.2. The proof is not complicated by adding more inhibitors.

Lemma 4.4 (Output Firing Probability When No Inhibitors Are Active). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be any simple SNN which solves $\text{WTA}(n, t_c, t_s, \delta)$ and whose auxiliary neurons $a_1, \dots, a_m$ are all inhibitory. For any i , any configuration C with $C(x_i) = 1$ and $C(a_j) = 0$ for all j , and any t ,*

$$\mathbb{P}[y_i^{t+1} = 1 | N^t = C] \geq 1 - \delta^{1/t_c}.$$

Proof. Consider any $i \in \{1, \dots, n\}$. Assume for the sake of contradiction that there is some configuration C with $C(x_i) = 1$, $C(a_j) = 0$ for all j , and

$$\mathbb{P}[y_i^{t+1} = 1 | N^t = C] < 1 - \delta^{1/t_c}.$$

This assumption additionally implies three claims. First we can see that, since y_i is excitatory,

Claim 4.5. *There exists some configuration C with $C(x_i) = 1$, $C(a_j) = 0$ for all j and $C(y_i) = 0$ with $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] < 1 - \delta^{1/t_c}$.*

From Claim 4.5 we can additionally conclude that, since, by Definition 4.1, y_i can only have connections from itself, x_i , and $a_1, \dots, a_m$:

Claim 4.6. *For any configuration C with $C(x_i) = 1$, $C(a_j) = 0$ for all j and $C(y_i) = 0$, $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] < 1 - \delta^{1/t_c}$.*

Finally, Claim 4.6 implies that, since $a_1, \dots, a_m$ are all inhibitors,

Claim 4.7. *For any configuration with $C(x_i) = 1$ and $C(y_i) = 0$, $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] < 1 - \delta^{1/t_c}$.*

Consider input execution α_X with X^t fixed for all t (i.e., $X^t = X^{t'}$ for any t, t'), $x_i^t = 1$, and $x_j^t = 0$ for all $j \neq i$. Let N^0 be any initial configuration of $\mathcal{N}$ consistent with α_X (i.e., $N^0(X) = X^0$) and with $y_i^0 = 0$.

Since $\mathcal{N}$ solves $\text{WTA}(n, t_c, t_s, \delta)$, an infinite output execution drawn from $\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X)$ must, with probability $\geq 1 - \delta$, reach a valid WTA output configuration (Definition 2.6) for some $t \leq t^c$. In particular, with probability $\geq 1 - \delta$, there must be some $t \leq t_c$ in which $y_i^t = 1$. So, letting $\mathcal{E}_0(t)$ be the event that $y_i^{t'} = 0$ for all $t' \leq t$, we have $\mathbb{P}[\mathcal{E}_0(t_c)] \leq \delta$.

Additionally, $\mathbb{P}[\mathcal{E}_0(0)] = 1$ and using Claim 4.7 above and inducting on t , for any $t \geq 1$:

$$\begin{aligned} \mathbb{P}[\mathcal{E}_0(t)] &= \mathbb{P}[\mathcal{E}_0(t) | \mathcal{E}_0(t-1)] \cdot \mathbb{P}[\mathcal{E}_0(t-1)] && (\text{Since } \mathcal{E}_0(t) \subseteq \mathcal{E}_0(t-1).) \\ &> (1 - (1 - \delta^{1/t_c}))^t = \delta^{t/t_c}. \end{aligned}$$

We thus have

$$\mathbb{P}[\mathcal{E}_0(t_c)] > \delta^{t_c/t_c} = \delta$$

which contradicts that fact that $\mathbb{P}[\mathcal{E}_0(t_c)] \leq \delta$, giving the lemma. $\square$

We next show that if the inhibitor a does fire at time t , then any output must cease firing at time $t + 1$ with reasonably large probability. Formally, we show this statement for roughly half of the n outputs. It may be possible that some outputs fire with large probability at time $t + 1$ whenever their corresponding inputs fire at time t . However, for convergence to occur rapidly give an input execution in which all inputs, this cannot be the case for most outputs.

Again, since it will be useful in the two-inhibitor lower bound proven in Section 4.2, we show a more general result which pertains to networks with any number of auxiliary inhibitors.

Lemma 4.8 (Output Firing Probability When All Inhibitors Are Active). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be any simple SNN which solves $\text{WTA}(n, t_c, t_s, \delta)$ for $n \geq 20$ and $\delta \leq 1/2$ and whose auxiliary neurons $a_1, \dots, a_m$ are all inhibitory. There is some set $\mathcal{S} \subseteq \{1, \dots, n\}$ with $|\mathcal{S}| \geq \lceil n/2 \rceil$ such that, for any $i \in \mathcal{S}$, any configuration C with $C(x_i) = 1$ and $C(a_j) = 1$ for all j , and any t ,*

$$\mathbb{P}[y_i^{t+1} = 0 | N^t = C] \geq 1 - n^{-\frac{1}{10t_c}}.$$

Proof. Assume for the sake of contradiction that there is some set $\mathcal{R} \subseteq \{1, \dots, n\}$ with $|\mathcal{R}| = \lceil n/2 \rceil + 1$ such that, for each $i \in \mathcal{R}$, there exists some configuration C with $C(x_i) = 1$, $C(a_j) = 1$ for all j and

$$\mathbb{P}[y_i^{t+1} = 1 | N^t = C] > n^{-\frac{1}{10t_c}}.$$

From this assumption we can deduce three claims. First, since each y_i is excitatory,

Claim 4.9. *For each $i \in \mathcal{R}$ there exists some configuration C with $C(x_i) = 1$, $C(a_j) = 1$ for all j and $C(y_i) = 1$ with $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] > n^{-\frac{1}{10t_c}}$.*

Claim 4.9 further implies, since y_i can only have connections from itself, x_i , and $a_1, \dots, a_m$ (see Definition 4.1):

Claim 4.10. *For any configuration C with $C(x_i) = 1$, $C(a_j) = 1$ for all j and $C(y_i) = 1$, $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] > n^{-\frac{1}{10t_c}}$.*

Finally, from Claim 4.10 and the fact that $a_1, \dots, a_m$ are all inhibitors we can conclude:

Claim 4.11. *For any configuration with $C(x_i) = 1$ and $C(y_i) = 1$, $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] > n^{-\frac{1}{10t_c}}$.*

Let α_X be any infinite input execution with X^t fixed for all t and $X^t(x_i) = 1$ for all i . Let N^0 be any initial configuration of $\mathcal{N}$ consistent with α_X (i.e., $N^0(X) = X^0$) and with $Y^0(y_i) = 1$ for all i .

Since $\mathcal{N}$ solves $\text{WTA}(n, t_c, t_s, \delta)$, an infinite output execution drawn from $\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X)$ must, with probability $\geq 1 - \delta$, reach a valid WTA output configuration (Definition 2.6) for some $t \leq t^c$. In particular, with probability $\geq 1 - \delta$, there must be some $t \leq t_c$ in which $y_i^t = 0$ for at most one $i \in \mathcal{R}$.

Let $\mathcal{I}_{act}(t, i) \in \{0, 1\}$ be an indicator of the event that $y_i^{t'} = 1$ for all $t' \leq t$. If $\mathcal{I}_{act}(t_c, i) = 1$ for more than one $i \in \mathcal{R}$, then the network has not reached a valid WTA output state within time t_c . Thus, since $\mathcal{N}$ solves $\text{WTA}(n, t_c, t_s, \delta)$ we have:

$$\mathbb{P} \left[\sum_{i \in \mathcal{R}} \mathcal{I}_{act}(t_c, i) \geq 2 \right] \leq \delta. \quad (38)$$

We can bound $\mathbb{P} \left[\sum_{i \in \mathcal{R}} \mathcal{I}_{act}(t_c, i) \geq 2 \right]$ from below using Claim 4.11 above and Lemma 2.4. Define for all i and t a random variable $Z_{i,t} \in \{0, 1\}$ which is set to 1 independently with probability $n^{-\frac{1}{10t_c}}$. Let $\tilde{\mathcal{I}}(t_c, i)$ be an indicator of the event that $Z_{i,1} = \dots = Z_{i,t_c} = 1$. Clearly the $\tilde{\mathcal{I}}(t_c, i)$ variables are independent and $\mathbb{P}[\tilde{\mathcal{I}}(t_c, i) = 1] = \left(n^{-\frac{1}{10t_c}}\right)^{t_c} = n^{-\frac{1}{10}}$.

By (38) and Lemma 2.4 we thus have:

$$\begin{aligned} \delta &\geq \mathbb{P} \left[\sum_{i \in \mathcal{R}} \mathcal{I}_{act}(t_c, i) \geq 2 \right] \geq \mathbb{P} \left[\sum_{i \in \mathcal{R}} \tilde{\mathcal{I}}(t_c, i) \geq 2 \right] \\ &= 1 - \left(1 - n^{-\frac{1}{10}}\right)^{|\mathcal{R}|} - |\mathcal{R}| \left(1 - n^{-\frac{1}{10}}\right)^{|\mathcal{R}|-1} \\ &\geq 1 - n \left(1 - n^{-\frac{1}{10}}\right)^{n/4} \end{aligned}$$

where in the last step we bound $|\mathcal{R}| - 1 \geq \lfloor n/2 \rfloor \geq n/4$. Rearranging:

$$\frac{1 - \delta}{n} \leq \left(1 - n^{-\frac{1}{10}}\right)^{n/4} \leq e^{-\frac{n^{9/10}}{4}}. \quad (39)$$

We can check that whenever $n \geq 20$, $e^{-\frac{n^{9/10}}{4}} < \frac{1}{2n} < \frac{1-\delta}{n}$, by our assumption that $\delta \leq 1/2$. This contradicts (39), thus giving the lemma. $\square$

We conclude by combining Lemmas 4.4 and 4.8 to prove Theorem 4.3. We first show a simple auxiliary lemma which lower bounds the probability that a valid WTA output configuration at time t remains fixed at time $t + 1$ in terms of the network convergence time t_c and stability time t_s . The smaller t_c and the larger t_s , the larger the lower bound on this probability is.

Note that the theorem only lower bounds the probability that the output Y^t stays fixed for *some configuration* C where $C(Y)$ is a valid WTA output. It does not bound this probability for all such configurations, and in fact such a bound cannot be shown for all such configurations. Recall, for example, that for our two-inhibitor networks in Section 3, both near-valid WTA configurations (Definition 3.15) and valid WTA configurations (Definition 3.11) have valid WTA output configurations. However, the output of a near-valid configuration only remains fixed with constant probability (Lemma 3.21).

Lemma 4.12 (Single Step Stability Probability). *If $\mathcal{N} = \langle N, w, b, f \rangle$ is a simple SNN which solves $WTA(n, t_c, t_s, \delta)$ for $\delta \leq 1/2$, then there must exist some configuration C with $C(Y)$ a valid WTA output configuration such that:*

$$\mathbb{P}[Y^{t+1} = Y^t | N^t = C] \geq \frac{1}{(2t_c)^{1/t_s}} \geq 1 - \frac{\ln 2t_c}{t_s}. \quad (40)$$

Proof. The second inequality simply follows since $e^{-x} \geq 1 - x$ for all x . So we focus on proving the first inequality. Assume for the sake of contradiction that for every configuration C with $C(Y)$ a valid WTA output configuration,

$$\mathbb{P}[Y^{t+1} = Y^t | N^t = C] < \frac{1}{(2t_c)^{1/t_s}}.$$

For any t, i , let $\mathcal{E}_{stab}(t, i)$ be the event that $Y^t = Y^{t+1} = \dots = Y^{t+i}$. Trivially $\mathbb{P}[\mathcal{E}_{stab}(t, 0)] = 1$ for all t . Using the assumption of (40) and induction we can bound for any $i \geq 1$:

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{stab}(t, i) | N^t = C] &= \mathbb{P}[\mathcal{E}_{stab}(t, i) | \mathcal{E}_{stab}(t, i-1), N^t = C] \cdot \mathbb{P}[\mathcal{E}_{stab}(t, i-1) | N^t = C] \\ &\quad \text{(Since } \mathcal{E}_{stab}(t, i-1) \subseteq \mathcal{E}_{stab}(t, i) \text{.)} \\ &< \frac{1}{(2t_c)^{1/t_s}} \cdot \mathbb{P}[\mathcal{E}_{stab}(t, i-1) | N^t = C] \\ &\leq \left(\frac{1}{(2t_c)^{1/t_s}} \right)^i. \end{aligned}$$

Thus, for any C with $C(Y)$ a valid WTA output configuration,

$$\mathbb{P}[\mathcal{E}_{stab}(t, t_s) | N^t = C] < \left(\frac{1}{(2t_c)^{1/t_s}} \right)^{t_s} < \frac{1}{2t_c}. \quad (41)$$

Let $\mathcal{E}$ be the event that $\mathcal{N}$ reaches a valid WTA output configuration within $t \leq t_c$ steps and remains in this output configuration for t_s consecutive steps. Let $Z_1, \dots, Z_{t_c}$ be i.i.d. random variables with $Z_t = 1$ with probability $\frac{1}{2t_c}$ and $Z_t = 0$ otherwise. Invoking (41) and Lemma 2.3,

$$\begin{aligned} \mathbb{P}[\mathcal{E}] &< \mathbb{P} \left[\sum_{t=1}^{t_c} Z_t \geq 1 \right] = 1 - \left(1 - \frac{1}{2t_c} \right)^{t_c} \\ &< 1/2 \\ &< 1 - \delta \end{aligned} \quad (42)$$

for $\delta \leq 1/2$. This contradicts the fact that $\mathcal{N}$ solves $WTA(n, t_c, t_s, \delta)$, giving the lemma. $\square$

We can now combine Lemmas 4.4 and 4.8 to show that in a single-inhibitor network, for any configuration C with $C(Y)$ a valid WTA output configuration $\mathbb{P}[Y^{t+1} = Y^t | N^t = C]$ cannot be too large. This contradicts Lemma 4.12, giving our lower bound.

Proof of Theorem 4.3.

Consider any simple SNN $\mathcal{N} = \langle N, w, b, f \rangle$ with just a single auxiliary neuron a , which solves $WTA(n, t_c, t_s, \delta)$ for $n \geq 20$, $\delta \leq 1/2$, $t_s > 10t_c \cdot \frac{\ln(2t_c)}{\ln n}$. Let $\mathcal{S}$ be the set of indices shown to exist in Lemma 4.8, which do not fire with too high probability when all inhibitors in the network fire.

Let α_X be the infinite input execution with X^t fixed for all t , $X^t(x_i) = 1$ for all $i \in \mathcal{S}$ and $X^t(x_i) = 0$ for all $i \notin \mathcal{S}$. Let N^0 be any initial configuration of $\mathcal{N}$ consistent with α_X .

Since $\mathcal{N}$ solves $\text{WTA}(n, t_c, t_s, \delta)$, an infinite output execution drawn from $\mathcal{D}_Y(\mathcal{N}, N^0, \alpha_X)$ must, with probability $\geq 1 - \delta$, reach a valid WTA output state (Definition 2.6) for some $t \leq t_c$ and remain in this state for t_s consecutive steps. Let C be any configuration where $C(Y)$ is a valid WTA output configuration for α_X . We must have for exactly one $i \in \mathcal{S}$, $C(y_i) = 1$. Let $\mathcal{E}_{fail}(t)$ be the event that $Y^{t+1} \neq Y^t$. We consider two cases:

Case 1: $C(a) = 0$.

In this case, by Lemma 4.4, for all $j \in \mathcal{S}$, y_j fires with probability $\geq 1 - \delta^{1/t_c}$ at time $t + 1$. So, with probability $\geq 1 - \delta^{\frac{|\mathcal{S}|-1}{t_c}} \geq 1 - \delta^{\frac{n}{4t_c}}$ as least one output other than the winner fires at time $t + 1$. This gives:

$$\mathbb{P}[\mathcal{E}_{fail}(t) | N^t = C] \geq 1 - \delta^{\frac{n}{4t_c}} \geq 1 - n^{-\frac{1}{4t_c}}, \quad (43)$$

where the second inequality follows from the assumption that $n \geq 20$ and $\delta \leq 1/2$.

Case 2: $C(a) = 1$.

In this case, Lemma 4.8 gives that, conditioned on $N^t = C$, y_i does not fire at time $t + 1$ with probability $\geq 1 - n^{-\frac{1}{10t_c}}$. This gives:

$$\mathbb{P}[\mathcal{E}_{fail}(t) | N^t = C] \geq 1 - n^{-\frac{1}{10t_c}}. \quad (44)$$

So overall, combining (43) and (44) we have for C with $C(Y)$ a valid WTA output configuration,

$$\mathbb{P}[\mathcal{E}_{fail}(t) | N^t = C] \geq 1 - n^{-\frac{1}{10t_c}}$$

and so

$$\mathbb{P}[Y^{t+1} = Y^t | N^t = C] \leq n^{-\frac{1}{10t_c}}.$$

For $t_s > 10t_c \cdot \frac{\ln(2t_c)}{\ln n}$ this gives

$$\mathbb{P}[Y^{t+1} = Y^t | N^t = C] < n^{-\frac{\ln 2t_c}{t_s \cdot \ln n}} < \frac{1}{(2t_c)^{1/t_s}}. \quad (45)$$

This contradicts Lemma 4.12, giving the theorem. $\square$

Remark on the Tightness of Theorem 4.3. We note that our proof of Theorem 4.3 is loose. In bounding $\mathbb{P}[\mathcal{E}]$ in (42), we do not consider the time required to converge to a valid WTA output state, or the time spent in this converged state before convergence is broken. We conjecture that if this time were taken into account, it would be possible improve the $\ln(2t_c)$ term, as well as add a dependence on the failure probability δ , giving a lower bound of $t_s = O\left(\frac{t_c}{\log n \cdot \log(1/\delta)}\right)$.

We note that by simply removing the stability inhibitor from our two-inhibitor network family presented in Section 3, we obtain a family of single inhibitor WTA networks with $t_s = 1$ and $t_c = O(\log n \cdot \log(1/\delta))$. This matches the conjectured stronger lower bound above up to a constant factor since for $t_c = O(\log n \cdot \log(1/\delta))$ we have $O\left(\frac{t_c}{\log n \cdot \log(1/\delta)}\right) = O(1)$.

Theorem 4.13 (Single-Inhibitor WTA Network). *There exists a simple SNN with a single inhibitory auxiliary neuron which solves $WTA(n, t_c, t_s, \delta)$ for $t_s = 1$ and $t_c = O(\log n \cdot \log(1/\delta))$.*

Proof Sketch. Let $\mathcal{T}'_{n,\gamma}$ be identical to $\mathcal{T}_{n,\gamma}$ as described in Section 3.1, but with a_s removed from the network and with $w(a_c, y_i) = -2\gamma$ for all i . In $\mathcal{T}_{n,\gamma}$, $w(a_c, y_i) = w(a_s, y_i) = -\gamma$ so $w(a_c, y_i) + w(a_s, y_i) = -2\gamma$. Thus, in $\mathcal{T}'_{n,\gamma}$, when $a_c^t = 1$, the firing probabilities of all neurons at time $t + 1$ are identical to what they would be in $\mathcal{T}_{n,\gamma}$ assuming the same configuration at time t but with $a_c^t = a_s^t = 1$.

Using this equivalence, it is tedious but easy to check that an analogous result to Theorem 3.32 holds, where a valid WTA configuration is redefined to be any configuration C where $C(Y)$ is a valid WTA output configuration and $C(a_c) = 1$. By a similar result to Lemma 3.21, with probability $1/2 - ne^{-\gamma/2}$, if $\mathcal{T}'_{n,\gamma}$ is in such a configuration at time t , it is also in a valid WTA output configuration at time $t + 1$. Thus, with $\Theta(1)$ probability, a valid WTA output configuration is reached and maintained for $t_s = 1$ steps within $O(\log n)$ steps. We can then use an argument similar to that of Theorem 3.2 to argue that in $O(\log n \cdot \log(1/\delta))$ steps a valid WTA output configuration is reached and maintained for 1 step with probability $\geq 1 - \delta$. $\square$

4.2 Two Auxiliary Neuron Lower Bound

We next give a convergence time lower bound for SNNs with two auxiliary neurons, showing that the rate given by the family of two-inhibitor networks $\mathcal{T}_{n,\gamma}$ presented in Section 3 is optimal up to a $O(\log \log n)$ factor. To simplify our argument, we focus the further restricted class of symmetric SNNs described in Definition 4.2, proving:

Theorem 4.14 (Two Neuron Lower Bound). *For any $n \geq 341$, $\delta \leq 1/2$, and any spike probability function $f : \mathbb{R} \rightarrow [0, 1]$ (satisfying the restrictions in Section 2.1), there is no symmetric SNN $\mathcal{N} = \langle N, w, b, f \rangle$ using two auxiliary neurons that solves $WTA(n, t_c, t_s, \delta)$ with $t_c \leq \frac{\ln n}{30 \ln \ln n}$ and $t_s \geq 32 \ln n \cdot \ln 2t_c$.*

As with Theorem 4.3, we conjecture that this result holds more generally, for any SNN with two auxiliary neurons (i.e., without making the assumptions of Definition 4.2). In the theorem we require $t_s \geq 32 \ln n \cdot \ln 2t_c$. However, we conjecture that the result holds even for $t_s = O(1)$.

Proof Outline. Our proof of Theorem 4.14 is similar in spirit to that of Theorem 4.3. We consider the case when both auxiliary neurons are inhibitors and note that a similar proof applies when one or both of the neurons are excitatory.

With two inhibitors we have to not only consider the cases when neither inhibitor fires (analyzed in Lemma 4.4) and when both inhibitors fire (analyzed in Lemma 4.8), but also the cases in which one of the inhibitors fires. Our analysis breaks down as follows:

1. By Lemma 4.4, if neither inhibitor fires at time t , then at time $t + 1$ any output corresponding to a firing input must fire with probability $\geq 1 - \delta^{1/t_c} = \Omega(\frac{1}{\ln n})$ when $\delta \leq 1/2$ and $t_c \leq \frac{\ln n}{30 \ln \ln n}$. By Lemma 4.8, if both inhibitors fire at time t , then at time $t + 1$, any output with a firing input must *not* fire at time $t + 1$ with probability $\geq 1 - n^{-\frac{1}{10t_c}} \geq 1 - \frac{1}{\ln^3 n}$ when $t_c \leq \frac{\ln n}{30 \ln \ln n}$ and n is sufficiently large.
2. In Lemma 4.16 we show that, by the above bounds, if we consider an input with $\Theta(\ln^2 n)$ firing inputs, then if either both inhibitors fire or neither inhibitor fires at time t , except with probability $\leq \frac{1}{\ln n}$, at time $t + 1$, either 0 or ≥ 2 outputs will fire, and so $\mathcal{N}$ will not be in a valid WTA output configuration.

3. In Lemma 4.18 we show that, in order for $\mathcal{N}$ to stabilize to a valid WTA output configuration for t_s steps, if Y^t is a valid WTA output configuration, then *exactly one inhibitor* must fire with good probability at time $t + 1$. Otherwise, if neither or both inhibitors fire at time $t + 1$, by Lemma 4.16, Y^{t+2} is unlikely to be a valid WTA output configuration. So if t_s is large, over t_s steps, stability is likely to be broken at some point.
4. In Lemma 4.20 we prove that, since the inhibitors fire independently at time $t + 1$ conditioned on N^t , Lemma 4.18 in fact requires that one of the inhibitors fires with high probability at time $t + 1$ when Y^t is a valid WTA output configuration and that the other inhibitor is silent with high probability.

By our symmetry assumption (Definition 4.2), given a fixed input firing pattern, the probability that an inhibitor fires at time $t + 1$ depends only on the number of outputs that fire at time t . So we can in fact show a stronger result: one inhibitor (assume without loss of generality a_1) fires with high probability at time $t + 1$ when $\|Y^t\|_1 = 1$ while the other (assume without loss of generality a_2) remains silent with high probability at time $t + 1$ when $\|Y^t\|_1 = 1$. Since all outputs are excitatory, this implies that a_1 fires with high probability whenever $\|Y^t\|_1 \geq 1$, while a_2 does not fire with high probability whenever $\|Y^t\|_1 \leq 1$.

This result shows that, in any two-inhibitor network with fast convergence and a reasonably long stability period t_s , the inhibitors must exhibit separate behaviors. Neuron a_1 fires with high probability at time $t + 1$ whenever one output fires at time t , maintaining stability of valid WTA configurations. Neuron a_2 is silent with high probability at time $t + 1$, except possibly when $\|Y^t\|_1 \geq 2$. We can see this separation, for example, in the behavior of a_s and a_c in the two-inhibitor $\mathcal{T}_{n,\gamma}$ networks analyzed in Section 3 (see Lemma 3.7). The stability inhibitor, a_s , fires with high probability at time $t + 1$ when at least one output fires at time t (e.g., when the network is in a valid WTA configuration). The convergence inhibitor, a_c , only fires with high probability at time $t + 1$ when at least two outputs fire at time t .

5. In Corollary 4.25 we use Lemma 4.18 to show that it is unlikely that a_2 ever fires at a time in which a_1 does not, which will be useful in our eventual case analysis (see Step 7).
6. In Lemma 4.26 we show that, since when a_1 fires alone it must maintain stability of a valid WTA output configuration, when a_1 fires at time t , any output with an active input which fired at time t is likely to continue firing at time $t + 1$. Any output that did not fire at time t is unlikely to fire at time $t + 1$. This result shows that a_1 *must act as a stability inhibitor*, reflecting the role of a_s in the two-inhibitor network family presented in Section 3.

In Corollary 4.27 we show that Lemma 4.26 implies that, if Y^t is *not a valid WTA output configuration* and just a_1 fires at time t , since the output firing states are maintained with high probability at time $t + 1$, Y^{t+1} is unlikely to be a valid WTA output configuration.

7. We finally prove Theorem 4.14 via a case analysis. By Lemma 4.16, if 0 or 2 inhibitors fire at time t , Y^{t+1} is unlikely to be a valid WTA output configuration (see Step 2 above). By Corollary 4.27, if $a_1^t = 1$, $a_2^t = 0$, and Y^t is not a valid WTA output configuration, then Y^{t+1} is unlikely to be a valid WTA output configuration (see Step 6 above). Finally, it is unlikely that we ever have $a_1^t = 0$, $a_2^t = 1$ (see Step 5 above). Thus, convergence to a valid WTA output configuration is relatively unlikely at all times, letting us prove a lower bound on convergence time.

We first define a *hard input execution*, based on Lemma 4.8, which shows that at least half of the output neurons must not fire with too high probability when all inhibitors in the network fire.

Specifically, Lemma 4.8 guarantees the existence of some set $\mathcal{S}$ with $|\mathcal{S}| \geq \lceil n/2 \rceil$ such that for any $i \in \mathcal{S}$, any configuration C of $\mathcal{N}$ with $C(x_i) = 1$ and $C(a_1) = C(a_2) = 1$, and any time t ,

$$\mathbb{P}[y_i^{t+1} = 0 | N^t = C] \geq 1 - n^{-\frac{1}{10t_c}}.$$

Since in Theorem 4.14 we require $n \geq 341$ we easily have $\lceil n/2 \rceil \geq \lceil \ln^2 n \rceil$ and so can construct a hard input execution as follows:

Definition 4.15 (Hard Input Execution). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be any simple SNN with two auxiliary inhibitory neurons a_1, a_2 which solves $\text{WTA}(n, t_c, t_s, \delta)$ for $n \geq 341$ and $\delta \leq 1/2$. Fix any set $\mathcal{R} \subseteq \{1, \dots, n\}$ with $|\mathcal{R}| = \lceil \ln^2 n \rceil$ such that for any $i \in \mathcal{R}$, any C with $C(x_i) = 1$ and $C(a_1) = C(a_2) = 1$, and any time t ,*

$$\mathbb{P}[y_i^{t+1} = 0 | N^t = C] \geq 1 - n^{-\frac{1}{10t_c}}.$$

Let $X_{\text{hard}, \mathcal{N}}$ be the input configuration with $X_{\text{hard}, \mathcal{N}}(x_i) = 1$ for all $i \in \mathcal{R}$, and $X_{\text{hard}, \mathcal{N}}(x_i) = 0$ for all $i \notin \mathcal{R}$. Let $\alpha_{\text{hard}, \mathcal{N}}$ be the infinite input execution with $X^t = X_{\text{hard}, \mathcal{N}}$ for all t .

Lemma 4.16 (Valid WTA is Unlikely After Both or Neither Inhibitors Fire). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be any simple SNN with two auxiliary inhibitory neurons a_1, a_2 which solves $\text{WTA}(n, t_c, t_s, \delta)$ for $n \geq 341$, $\delta \leq 1/2$, and $t_c \leq \frac{\ln n}{30 \ln \ln n}$. For any configuration C with $C(X) = X_{\text{hard}, \mathcal{N}}$ (Definition 4.15) and with $C(a_1) = C(a_2) = 1$ or $C(a_1) = C(a_2) = 0$,*

$$\mathbb{P}[\|Y^{t+1}\| = 1 | N^t = C] \leq \frac{1}{\ln n}.$$

Proof. We prove the lemma in two cases depending on the inhibitor behavior.

Case 1: $C(a_1) = C(a_2) = 0$.

In this case, by Lemma 4.4, since $X^t(x_i) = X_{\text{hard}, \mathcal{N}}(x_i) = 1$ for any $i \in \mathcal{R}$ by Definition 4.15,

$$\mathbb{P}[y_i^{t+1} = 1 | N^t = C] \geq 1 - \delta^{1/t_c} \geq 1 - \frac{1}{2^{\frac{\ln \ln n}{\ln n}}}$$

where the second inequality follows from our assumption that $\delta \leq 1/2$ and $t_c \leq \frac{\ln n}{30 \ln \ln n} \leq \frac{\ln n}{\ln \ln n}$. For any $x \in [1, 2]$, $(1 - 1/x) \geq \frac{\log_2 x}{2}$. Since $2^{\frac{\ln \ln n}{\ln n}} \in [1, 2]$ this gives:

$$\mathbb{P}[y_i^{t+1} = 1 | N^t = C] \geq \frac{\ln \ln n}{2 \ln n}. \quad (46)$$

Using (46) we can bound the probability that $\|Y^{t+1}\|_1 \neq 1$ by:

$$\mathbb{P}[\|Y^{t+1}\|_1 \neq 1 | N^t = C] \geq 1 - |\mathcal{R}| \left(1 - \frac{\ln \ln n}{2 \ln n}\right)^{|\mathcal{R}|-1}.$$

We can check numerically that since $|\mathcal{R}| = \lceil \ln^2 n \rceil$ and by assumption $n \geq 341$, the above can be lower bounded to give:

$$\mathbb{P}[\|Y^{t+1}\|_1 \neq 1 | N^t = C] \geq 1 - \frac{1}{\ln n}.$$

This gives $\mathbb{P}[\|Y^{t+1}\|_1 = 1 | N^t = C] \leq \frac{1}{\ln n}$ and thus the lemma in this case.

Case 2: $C(a_1) = C(a_2) = 1$.

In this case, by Lemma 4.8 and the assumption that $t_c \leq \frac{\ln n}{30 \ln \ln n}$, for any $i \in \mathcal{R}$:

$$\mathbb{P}[y_i^{t+1} = 1 | N^t = C] \leq n^{-\frac{1}{10t_c}} \leq e^{-\frac{30 \ln \ln n}{10}} \leq \frac{1}{\ln^3 n}. \quad (47)$$

Using (47) we can bound:

$$\begin{aligned} \mathbb{P}[\|Y^{t+1}\|_1 = 1 | N^t = C] &\leq \frac{|\mathcal{R}|}{\ln^3 n} \\ &= \frac{\lfloor \ln^2 n \rfloor}{\ln^3 n} \\ &\leq \frac{1}{\ln n}, \end{aligned}$$

which gives the lemma in this case. $\square$

By Lemma 4.16, if $a_1^t = a_2^t = 0$ or $a_1^t = a_2^t = 1$, then Y^{t+1} is unlikely to have $\|Y^{t+1}\|_1 = 1$ and thus is unlikely to be a valid WTA output configuration. Thus, to stabilize to any valid WTA output configuration C , if $N^t = C$ exactly one of a_1 or a_2 must fire at time $t + 1$. Otherwise convergence will likely be broken at time $t + 2$. To prove this, we first show an auxiliary lemma, very similar to Lemma 4.12, which bounds the probability of maintaining stability over two time steps in terms of the convergence and stability times t_c and t_s .

Lemma 4.17 (Lower Bound on the Two-Step Stability Probability). *If $\mathcal{N} = \langle N, w, b, f \rangle$ is a simple SNN which solves WTA(n, t_c, t_s, δ) for $\delta \leq 1/2$, then there exists some configuration C with $C(Y)$ a valid WTA output configuration such that:*

$$\mathbb{P}[Y^{t+2} = Y^{t+1} = Y^t | N^t = C] \geq 1 - \frac{4 \ln 2t_c}{t_s}. \quad (48)$$

Proof. Our proof mirrors that of Lemma 4.12. Assume for the sake of contradiction that for any C with $C(Y)$ a valid WTA output configuration,

$$\mathbb{P}[Y^{t+2} = Y^{t+1} = Y^t | N^t = C] < 1 - \frac{4 \ln 2t_c}{t_s}.$$

For any t, i , let $\mathcal{E}_{stab}(t, i)$ be the event that $Y^t = Y^{t+1} = \dots = Y^{t+i}$. Trivially, $\mathbb{P}[\mathcal{E}_{stab}(t, 0)] = 1$ for all t . Using the assumption if (48) and induction we can bound for any $i \geq 2$:

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{stab}(t, i) | N^t = C] &= \mathbb{P}[\mathcal{E}_{stab}(t, i) | \mathcal{E}_{stab}(t, t-2), N^t = C] \cdot \mathbb{P}[\mathcal{E}_{stab}(t, i-2) | N^t = C] \\ &\quad (\text{Since } \mathcal{E}_{stab}(t, i-2) \subseteq \mathcal{E}_{stab}(t, i).) \\ &< \left(1 - \frac{4 \ln 2t_c}{t_s}\right) \cdot \mathbb{P}[\mathcal{E}_{stab}(t, i-2) | N^t = C] \\ &\leq \left(1 - \frac{4 \ln 2t_c}{t_s}\right)^{\lfloor i/2 \rfloor}. \end{aligned}$$

Thus, for any C with $C(Y)$ a valid WTA output configuration,

$$\begin{aligned}\mathbb{P}[\mathcal{E}_{stab}(t, t_s) | N^t = C] &< \left(1 - \frac{4 \cdot \ln 2t_c}{t_s}\right)^{\lfloor t_s/2 \rfloor} \\ &< \left(1 - \frac{4 \cdot \ln 2t_c}{t_s}\right)^{t_s/4} \\ &< \frac{1}{2t_c}.\end{aligned}\tag{49}$$

Let $\mathcal{E}$ be the event that $\mathcal{N}$ reaches a valid WTA output configuration within $t \leq t_c$ steps and remains in this output configuration for t_s consecutive steps. Let $Z_1, \dots, Z_{t_c}$ be i.i.d. random variables with $Z_t = 1$ with probability $\frac{1}{2t_c}$ and $Z_t = 0$ otherwise. Invoking (49) and Lemma 2.3,

$$\begin{aligned}\mathbb{P}[\mathcal{E}] &< \mathbb{P}\left[\sum_{t=1}^{t_c} Z_t \geq 1\right] = 1 - \left(1 - \frac{1}{2t_c}\right)^{t_c} \\ &< 1/2 \\ &< 1 - \delta\end{aligned}\tag{50}$$

for $\delta \leq 1/2$. This contradicts the fact that $\mathcal{N}$ solves $\text{WTA}(n, t_c, t_s, \delta)$, giving the lemma. $\square$

We can now use Lemmas 4.16 and 4.17 to show that, for $\mathcal{N}$ to stabilize to a valid WTA output configuration for t_s steps with good probability, if Y^t is a valid WTA output configuration, then, with high probability, exactly one inhibitor must fire at time $t + 1$.

Lemma 4.18 (A Single Inhibitor is Likely to Fire After a Valid WTA Configuration). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be any simple SNN with two auxiliary inhibitory neurons $A = \{a_1, a_2\}$ which solves $\text{WTA}(n, t_c, t_s, \delta)$ for $n \geq 341$, $\delta \leq 1/2$, and $t_c \leq \frac{\ln n}{30 \ln \ln n}$. For any configuration C with $C(X) = X_{hard, \mathcal{N}}$ (Definition 4.15) and with $C(Y)$ a valid WTA output configuration,*

$$\mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] \geq 1 - \frac{8 \cdot \ln 2t_c}{t_s}.$$

Proof. Assume for the sake of contradiction that there exists C with $C(Y)$ a valid WTA output configuration and

$$\mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] < 1 - \frac{8 \cdot \ln 2t_c}{t_s}.$$

From this assumption, since in a symmetric SNN (Definition 4.2), $w(y_i, u) = w(y_j, u)$ for all i, j and $u \in A$, and since each auxiliary neuron may only have connections from the inputs and outputs (not to the other auxiliary neurons), we can deduce the stronger claim:

Claim 4.19. *For every C with $C(Y)$ a valid WTA output configuration,*

$$\mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] < 1 - \frac{8 \cdot \ln 2t_c}{t_s}.$$

Using Claim 4.19 we can bound the probability that $\mathcal{N}$ remains in a valid WTA output configuration at time $t + 2$ if it is in a valid configuration at time t . Specifically, for any C with $C(Y)$ a valid WTA output configuration:

$$\begin{aligned}\mathbb{P}[Y^{t+2} = Y^t | N^t = C] &= \mathbb{P}[Y^{t+2} = Y^t | \|A^{t+1}\|_1 = 1, N^t = C] \cdot \mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] \\ &\quad + \mathbb{P}[Y^{t+2} = Y^t | \|A^{t+1}\|_1 \neq 1, N^t = C] \cdot \mathbb{P}[\|A^{t+1}\|_1 \neq 1 | N^t = C] \\ &< \left(1 - \frac{8 \cdot \ln 2t_c}{t_s}\right) + \left(\frac{8 \cdot \ln 2t_c}{t_s}\right) \cdot \mathbb{P}[Y^{t+2} = Y^t | \|A^{t+1}\|_1 \neq 1, N^t = C].\end{aligned}\tag{51}$$

Since $C(Y)$ is a valid WTA output configuration, having $Y^{t+2} = Y^t$ requires that $\|Y^{t+2}\|_1 = 1$. Thus by Lemma 4.16 we can bound:

$$\mathbb{P}[Y^{t+2} = Y^t | \|A^{t+1}\|_1 \neq 1, N^t = C] \leq \frac{1}{\ln n}.$$

Plugging back into (51), for any C with $C(Y)$ a valid WTA output configuration:

$$\begin{aligned} \mathbb{P}[Y^{t+2} = Y^t | N^t = C] &< \left(1 - \frac{8 \cdot \ln 2t_c}{t_s}\right) + \left(\frac{8 \cdot \ln 2t_c}{t_s}\right) \cdot \left(\frac{1}{\ln n}\right) \\ &< 1 - \frac{4 \cdot \ln 2t_c}{t_s} \end{aligned} \quad (52)$$

where the last inequality holds easily since we require $n \geq 341$ and so $\ln n \geq 2$. Using (52) we can bound:

$$\mathbb{P}[Y^{t+2} = Y^{t+1} = Y^t | N^t = C] \leq \mathbb{P}[Y^{t+2} = Y^t | N^t = C] < 1 - \frac{4 \cdot \ln 2t_c}{t_s}.$$

This contradicts Lemma 4.17, giving the lemma. $\square$

Since at time $t + 1$, conditioned on the configuration at time t , a_1 and a_2 fire independently, we can in fact strengthen Lemma 4.18 to show that one of a_1, a_2 fires with high probability at time $t + 1$ when $\|Y^t\|_1 \geq 1$ and that the other remains silent with high probability when $\|Y^t\|_1 \leq 1$.

Lemma 4.20 (Separation of Inhibitor Behaviors). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be any symmetric SNN with two auxiliary inhibitory neurons $A = \{a_1, a_2\}$ which solves $WTA(n, t_c, t_s, \delta)$ for $n \geq 341$, $\delta \leq 1/2$, $t_c \leq \frac{\ln n}{30 \ln \ln n}$, and $\frac{8 \ln 2t_c}{t_s} \leq \frac{1}{4}$. There exists some $i \in \{1, 2\}$ such that, for any configuration C with $C(X) = X_{hard, \mathcal{N}}$ (Definition 4.15),*

1. *If $\|C(Y)\|_1 \geq 1$, then*

$$\mathbb{P}[a_i^{t+1} = 1 | N^t = C] \geq 1 - \frac{8 \cdot \ln 2t_c}{t_s}.$$

2. *If $\|C(Y)\|_1 \leq 1$, then for $j \neq i$,*

$$\mathbb{P}[a_j^{t+1} = 1 | N^t = C] \leq \frac{12 \cdot \ln 2t_c}{t_s}.$$

We can assume without loss of generality that $i = 1$.

Proof. We prove the two conclusions in sequence. We first show that there exists some i such that conclusion (1) holds. Fixing i , we then show that for $j \neq i$, conclusion (2) holds.

Conclusion 1:

Assume for the sake of contradiction that for both $i \in \{1, 2\}$ there exists some configuration C with $C(X) = X_{hard, \mathcal{N}}$, with $\|C(Y)\|_1 \geq 1$, and with

$$\mathbb{P}[a_i^{t+1} = 1 | N^t = C] < 1 - \frac{8 \cdot \ln 2t_c}{t_s}.$$

From this assumption we can deduce two claims. First, since all y_i are excitatory:

Claim 4.21. For both $i \in \{1, 2\}$, there exists some configuration C with $C(X) = X_{hard, \mathcal{N}}$, with $\|C(Y)\|_1 = 1$, and with $\mathbb{P}[a_i^{t+1} = 1 | N^t = C] < 1 - \frac{8 \cdot \ln 2t_c}{t_s}$.

Further, since by Definition 4.2, a symmetric SNN has $w(y_i, u) = w(y_j, u)$ for all i, j and $u \in A$, and since each inhibitor can only have incoming connections from the inputs and outputs,

Claim 4.22. For both $i \in \{1, 2\}$, for every C with $C(X) = X_{hard, \mathcal{N}}$ and with $C(Y)$ a valid WTA output configuration,

$$\mathbb{P}[a_i^{t+1} = 1 | N^t = C] < 1 - \frac{8 \cdot \ln 2t_c}{t_s}.$$

Since conditioned on N^t , a_1^{t+1} and a_2^{t+2} are independent, we can compute, for any C with $C(Y)$ a valid WTA output configuration,

$$\begin{aligned} \mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] &= \mathbb{P}[a_1^{t+1} = 1 | N^t = C] \cdot \mathbb{P}[a_2^{t+1} = 0 | N^t = C] \\ &\quad + \mathbb{P}[a_1^{t+1} = 0 | N^t = C] \cdot \mathbb{P}[a_2^{t+1} = 1 | N^t = C] \\ &= \mathbb{P}[a_1^{t+1} = 1 | N^t = C] \cdot (1 - \mathbb{P}[a_2^{t+1} = 1 | N^t = C]) \\ &\quad + (1 - \mathbb{P}[a_1^{t+1} = 1 | N^t = C]) \cdot \mathbb{P}[a_2^{t+1} = 1 | N^t = C]. \end{aligned} \quad (53)$$

If $\mathbb{P}[a_1^{t+1} = 1 | N^t = C] \geq \frac{1}{2}$, then (53) is maximized by setting $\mathbb{P}[a_2^{t+1} = 1 | N^t = C] = 0$, giving via Claim 4.22:

$$\mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] = \mathbb{P}[a_1^{t+1} = 1 | N^t = C] < 1 - \frac{8 \ln 2t_c}{t_s}.$$

If $\mathbb{P}[a_1^{t+1} = 1 | N^t = C] \leq \frac{1}{2}$, (53) is maximized by setting $\Pr[a_2^{t+1} = 1 | N^t = C] = 1 - \frac{8 \ln 2t_c}{t_s}$, giving:

$$\begin{aligned} \mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] &< \mathbb{P}[a_1^{t+1} = 1 | N^t = C] \cdot \frac{8 \ln 2t_c}{t_s} \\ &\quad + (1 - \mathbb{P}[a_1^{t+1} = 1 | N^t = C]) \cdot \left(1 - \frac{8 \ln 2t_c}{t_s}\right). \end{aligned} \quad (54)$$

Using our requirement that $\frac{8 \ln 2t_c}{t_s} \leq \frac{1}{4}$, (54) is maximized by setting $\mathbb{P}[a_1^{t+1} = 1 | N^t = C] = 0$, again giving:

$$\mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] = \mathbb{P}[a_1^{t+1} = 1 | N^t = C] < 1 - \frac{8 \ln 2t_c}{t_s}.$$

In either case, we have a contradiction of Lemma 4.18, giving the result.

Conclusion 2:

From conclusion (1) proven above, we can assume without loss of generality that, for any configuration C with $C(X) = X_{hard, \mathcal{N}}$ and with $\|C(Y)\|_1 \geq 1$,

$$\mathbb{P}[a_1^{t+1} = 1 | N^t = C] \geq 1 - \frac{8 \cdot \ln 2t_c}{t_s}. \quad (55)$$

I.e., we assume that the index i in the lemma statement satisfies $i = 1$. To prove conclusion (2) Assume for the sake of contradiction that there exists some configuration C with $C(X) = X_{hard, \mathcal{N}}$ and with $\|C(Y)\| \leq 1$ such that $\mathbb{P}[a_2^{t+1} = 1 | N^t = C] > \frac{12 \cdot \ln 2t_c}{t_s}$. Since all outputs are excitatory, from this assumption we can conclude:

Claim 4.23. *There exists some configuration C with $C(X) = X_{hard, \mathcal{N}}$ and with $\|C(Y)\|_1 = 1$ such that $\mathbb{P}[a_2^{t+1} = 1 | N^t = C] > \frac{12 \cdot \ln 2t_c}{t_s}$.*

Further, since, by Definition 4.2, a symmetric SNN has $w(y_i, u) = w(y_j, u)$ for all i, j and $u \in A$, and since each inhibitor can only have incoming connections from the inputs and outputs,

Claim 4.24. *For every C with $C(X) = X_{hard, \mathcal{N}}$ and with $C(Y)$ a valid WTA output configuration,*

$$\mathbb{P}[a_2^{t+1} = 1 | N^t = C] > \frac{12 \cdot \ln 2t_c}{t_s}.$$

Using Claim 4.24 we can prove conclusion (2) by considering two cases.

Case 1: $\mathbb{P}[a_2^{t+1} = 1 | N^t = C] > 1/2$.

In this case, by (53) and (55), we have for C with $C(Y)$ a valid WTA output configuration:

$$\begin{aligned} \mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] &= \mathbb{P}[a_1^{t+1} = 1 | N^t = C] \cdot (1 - \mathbb{P}[a_2^{t+1} = 1 | N^t = C]) \\ &\quad + (1 - \mathbb{P}[a_1^{t+1} = 1 | N^t = C]) \cdot \mathbb{P}[a_2^{t+1} = 1 | N^t = C] \\ &< \frac{1}{2} + \frac{8 \cdot \ln 2t_c}{t_s} \\ &< 1 - \frac{8 \cdot \ln 2t_c}{t_s} \end{aligned} \tag{56}$$

where the last inequality follows from our requirement that $\frac{8 \cdot \ln 2t_c}{t_s} \leq \frac{1}{4}$. (56) contradicts Lemma 4.18, giving the result in this case.

Case 2: $\mathbb{P}[a_2^{t+1} = 1 | N^t = C] \leq 1/2$.

In this case, by (55), Claim 4.24, and (53) we have for any C with $C(Y)$ a valid WTA output configuration:

$$\begin{aligned} \mathbb{P}[\|A^{t+1}\|_1 = 1 | N^t = C] &= \mathbb{P}[a_1^{t+1} = 1 | N^t = C] \cdot (1 - \mathbb{P}[a_2^{t+1} = 1 | N^t = C]) \\ &\quad + (1 - \mathbb{P}[a_1^{t+1} = 1 | N^t = C]) \cdot \mathbb{P}[a_2^{t+1} = 1 | N^t = C] \\ &< \left(1 - \frac{12 \ln 2t_c}{t_s}\right) + \frac{4 \cdot \ln 2t_c}{t_s} \\ &< 1 - \frac{8 \cdot \ln 2t_c}{t_s}. \end{aligned} \tag{57}$$

Again, (57) contradicts Lemma 4.18, giving the result in this case. $\square$

From Lemma 4.20 we can easily show that it is unlikely that a_2 ever fires when a_1 does not.

Corollary 4.25 (Neuron a_2 Rarely Fires Alone). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be any symmetric SNN with two auxiliary inhibitory neurons a_1, a_2 which solves $WTA(n, t_c, t_s, \delta)$ for $n \geq 341$, $\delta \leq 1/2$, $t_c \leq \frac{\ln n}{30 \ln \ln n}$, and $\frac{8 \ln 2t_c}{t_s} \leq \frac{1}{4}$. For any configuration C with $C(X) = X_{hard, \mathcal{N}}$ (Definition 4.15):*

$$\mathbb{P}[a_1^{t+1} = 0 \text{ and } a_2^{t+1} = 1 | N^t = C] \leq \frac{12 \cdot \ln 2t_c}{t_s}.$$

Proof. We prove this result in two cases.

Case 1: $\|C(Y)\|_1 \leq 1$.

In this case, using Lemma 4.20 conclusion (2):

$$\mathbb{P}[a_1^{t+1} = 0 \text{ and } a_2^{t+1} = 1 | N^t = C] \leq \mathbb{P}[a_2^{t+1} = 1 | N^t = C] \leq \frac{12 \cdot \ln 2t_c}{t_s}.$$

Case 2: $\|C(Y)\|_1 > 1$.

In this case, applying Lemma 4.20 conclusion (1):

$$\begin{aligned} \mathbb{P}[a_1^{t+1} = 0 \text{ and } a_2^{t+1} = 1 | N^t = C] &\leq \mathbb{P}[a_1^{t+1} = 0 | N^t = C] \\ &= 1 - \mathbb{P}[a_1^{t+1} = 1 | N^t = C] \\ &\leq \frac{8 \cdot \ln 2t_c}{t_s}. \end{aligned}$$

□

In combination, Lemma 4.16 and Corollary 4.25 show that, in order for a valid WTA output state to be maintained with high probability, some inhibitor, (which we assume without loss of generality is a_1) must fire alone. Using this fact, we can characterize the firing behavior of the outputs when a_1 fires alone. Since a_1 maintains stability, if it fires alone at time t , all outputs maintain the same firing state at time $t + 1$ as at time t with high probability.

Lemma 4.26 (Neuron a_1 Enforces Stability). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be any symmetric SNN with two auxiliary inhibitory neurons a_1, a_2 which solves $WTA(n, t_c, t_s, \delta)$ for $n \geq 341$, $\delta \leq 1/2$, $t_c \leq \frac{\ln n}{30 \ln \ln n}$, and $\frac{8 \ln 2t_c}{t_s} \leq \frac{1}{4}$. For any configuration C with $C(X) = X_{hard, \mathcal{N}}$ (Definition 4.15) and with $C(a_1) = 1$ and $C(a_2) = 0$:*

1. For all $i \in \mathcal{R}$, if $C(y_i) = 1$ then $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] \geq 1 - \frac{16 \ln 2t_c}{t_s}$.
2. For all $i \in \mathcal{R}$, if $C(y_i) = 0$ then $\mathbb{P}[y_i^{t+1} = 0 | N^t = C] \geq \left(1 - \frac{16 \ln 2t_c}{t_s}\right)^{1/(|\mathcal{R}|-1)}$

Proof. In a symmetric SNN (Definition 4.2) all outputs have identical incoming connections and biases. Thus, to prove the lemma it suffices to show that *there exists a configuration C with $C(X) = X_{hard, \mathcal{N}}$ and with $C(a_1) = 1$ and $C(a_2) = 0$, such that:*

1. There exists $i \in \mathcal{R}$ with $C(y_i) = 1$ and $\mathbb{P}[y_i^{t+1} = 1 | N^t = C] \geq 1 - \frac{16 \ln 2t_c}{t_s}$.
2. There exists $i \in \mathcal{R}$ with $C(y_i) = 0$ and $\mathbb{P}[y_i^{t+1} = 0 | N^t = C] \geq \left(1 - \frac{16 \ln 2t_c}{t_s}\right)^{1/(|\mathcal{R}|-1)}$.

We will show that these bounds must hold for at least one configuration in order for the network to reach a valid WTA output configuration in t_c steps and remain in this configuration for t_s consecutive steps. Let $\mathcal{E}(t)$ denote the event that $Y^t = Y^{t+1} = Y^{t+2}$. For any C we have:

$$\begin{aligned} \mathbb{P}[\mathcal{E}(t) | N^t = C] &= \mathbb{P}[\mathcal{E}(t) | N^t = C, a_1^{t+1} = a_2^{t+1}] \cdot \mathbb{P}[a_1^{t+1} = a_2^{t+1} | N^t = C] \\ &\quad + \mathbb{P}[\mathcal{E}(t) | N^t = C, a_1^{t+1} = 0, a_2^{t+1} = 1] \cdot \mathbb{P}[a_1^{t+1} = 0, a_2^{t+1} = 1 | N^t = C] \\ &\quad + \mathbb{P}[\mathcal{E}(t) | N^t = C, a_1^{t+1} = 1, a_2^{t+1} = 0] \cdot \mathbb{P}[a_1^{t+1} = 1, a_2^{t+1} = 0 | N^t = C] \\ &\leq \mathbb{P}[a_1^{t+1} = 0, a_2^{t+1} = 1 | N^t = C] \\ &\quad + \max(\mathbb{P}[\mathcal{E}(t) | N^t = C, a_1^{t+1} = a_2^{t+1}], \mathbb{P}[\mathcal{E}(t) | Y^t = C, a_1^{t+1} = 1, a_2^{t+1} = 0]). \end{aligned} \quad (58)$$

We can upper bound the first term of (58) by $\frac{12 \ln 2t_c}{t_s}$ using Corollary 4.25. Additionally, using Lemma 4.16, we can bound $\mathbb{P}[\mathcal{E}(t)|N^t = C, a_1^{t+1} = a_2^{t+1}] \leq \frac{1}{\ln n}$. This gives:

$$\mathbb{P}[\mathcal{E}(t)|N^t = C] \leq \frac{12 \ln 2t_c}{t_s} + \max \left(\frac{1}{\ln n}, \mathbb{P}[\mathcal{E}(t)|N^t = C, a_1^{t+1} = 1, a_2^{t+1} = 0] \right). \quad (59)$$

Applying Lemma 4.17 gives that there must exist some C with $C(Y)$ a valid WTA output configuration such that $\mathbb{P}[\mathcal{E}(t)|N^t = C] \geq 1 - \frac{4 \ln 2t_c}{t_s}$. So combining with (59):

$$\max \left(\frac{1}{\ln n}, \mathbb{P}[\mathcal{E}(t)|N^t = C, a_1^{t+1} = 1, a_2^{t+1} = 0] \right) \geq 1 - \frac{16 \ln 2t_c}{t_s}$$

By our assumption that $\frac{8 \ln 2t_c}{t_s} \leq \frac{1}{4}$ and the fact that $n \geq 341$ we easily have that $\frac{1}{\ln n} \leq \frac{1}{2} \leq 1 - \frac{16 \ln 2t_c}{t_s}$, meaning that in fact, there must exist some C with $C(Y)$ a valid WTA output configuration such that:

$$\mathbb{P}[\mathcal{E}(t)|N^t = C, a_1^{t+1} = 1, a_2^{t+1} = 0] \geq 1 - \frac{16 \ln 2t_c}{t_s}. \quad (60)$$

We can now easily prove the two conclusions of the lemma using similar arguments.

Conclusion 1:

Assume for the sake of contradiction that for every configuration C with $C(X) = X_{hard, \mathcal{N}}$ and with $C(a_1) = 1$ and $C(a_2) = 0$, for all $i \in \mathcal{R}$ with $C(y_i) = 1$,

$$\mathbb{P}[y_i^{t+1} = 1|N^t = C] < 1 - \frac{16 \ln 2t_c}{t_s}.$$

Then, for every C with $C(Y)$ a valid WTA output configuration we have:

$$\begin{aligned} \mathbb{P}[\mathcal{E}(t)|N^t = C, a_1^{t+1} = 1, a_2^{t+1} = 0] &\leq \mathbb{P}[Y^{t+2} = Y^{t+1}|N^{t+1} = C, a_1^{t+1} = 1, a_2^{t+1} = 0] \\ &< 1 - \frac{16 \ln 2t_c}{t_s}. \end{aligned}$$

contradicting (60) and giving the lemma.

Conclusion 2:

Assume for the sake of contradiction that for every configuration C with $C(X) = X_{hard, \mathcal{N}}$ and with $C(a_1) = 1$ and $C(a_2) = 0$, for all $i \in \mathcal{R}$ with $C(y_i) = 0$,

$$\mathbb{P}[y_i^{t+1} = 0|N^t = C] < \left(1 - \frac{16 \ln 2t_c}{t_s} \right)^{1/(|\mathcal{R}|-1)}.$$

Then, again for every C with $C(Y)$ a valid WTA output configuration we have:

$$\begin{aligned} \mathbb{P}[\mathcal{E}(t)|N^t = C, a_1^{t+1} = 1, a_2^{t+1} = 0] &\leq \mathbb{P}[Y^{t+2} = Y^{t+1}|N^{t+1} = C, a_1^{t+1} = 1, a_2^{t+1} = 0] \\ &< \left(\left(1 - \frac{16 \ln 2t_c}{t_s} \right)^{1/(|\mathcal{R}|-1)} \right)^{|\mathcal{R}|-1} \\ &< 1 - \frac{16 \ln 2t_c}{t_s} \end{aligned}$$

contradicting (60) and giving the lemma. □

Lemma 4.26 implies that when just a_1 fires, since this leads to stability, it is unlikely to lead to a valid WTA output configuration when starting from an invalid output configuration. That is, a_1 does not drive convergence to a valid WTA output configuration.

Corollary 4.27 (Neuron a_1 Does Not Drive Convergence). *Let $\mathcal{N} = \langle N, w, b, f \rangle$ be any symmetric SNN with two auxiliary inhibitory neurons a_1, a_2 which solves $WTA(n, t_c, t_s, \delta)$ for $n \geq 341$, $\delta \leq 1/2$, $t_c \leq \frac{\ln n}{30 \ln \ln n}$, and $\frac{8 \ln 2t_c}{t_s} \leq \frac{1}{4}$. For any configuration C with $C(X) = X_{hard, \mathcal{N}}$ (Definition 4.15), with $C(a_1) = 1$ and $C(a_2) = 0$, and in which $C(Y)$ is not a valid WTA output configuration:*

$$\mathbb{P}[Y^{t+1} \text{ is a valid WTA output configuration} \mid N^t = C] \leq \frac{32 \ln 2t_c}{t_s}.$$

Proof. The proof can be split into two cases, depending on if $C(Y)$ is not a valid WTA output configuration because too many valid outputs are firing or because no outputs are firing. Let $\mathcal{R} \subseteq \{1, \dots, \}$ be the set of indices with $X_{hard, \mathcal{N}}(x_i) = 1$.

Case 1: $|\{i \in \mathcal{R} \mid C(y_i) = 1\}| \geq 2$.

Let $Y(\mathcal{R})$ denote the set of outputs restricted to the indices in $\mathcal{R}$. In order for Y^{t+1} to be a valid WTA output configuration for $X_{hard, \mathcal{N}}$, we must have $\|Y(\mathcal{R})^t\|_1 = 1$. We can apply Lemma 4.26 conclusion (1) to bound:

$$\mathbb{P}[\|Y(\mathcal{R})^{t+1}\|_1 \geq 2 \mid N^t = C] \geq \left(1 - \frac{16 \ln 2t_c}{t_s}\right)^2 \geq 1 - \frac{32 \ln 2t_c}{t_s}$$

and thus $\mathbb{P}[Y^{t+1} \text{ is a valid WTA output configuration} \mid N^t = C] \leq \frac{32 \ln 2t_c}{t_s}$.

Case 2: $|\{i \in \mathcal{R} \mid C(y_i) = 1\}| = 0$.

In this case, we can apply Lemma 4.26 conclusion (2) to bound:

$$\mathbb{P}[\|Y(\mathcal{R})^{t+1}\|_1 = 0 \mid N^t = C] \geq \left(\left(1 - \frac{16 \ln 2t_c}{t_s}\right)^{1/(|\mathcal{R}|-1)} \right)^{|\mathcal{R}|}$$

By our assumption that $n \geq 341$, $\frac{|\mathcal{R}|}{|\mathcal{R}|-1} = \frac{\lfloor \ln^2 n \rfloor}{\lfloor \ln^2 n \rfloor - 1} \leq \frac{34}{33} < 2$. So we have:

$$\mathbb{P}[\|Y(\mathcal{R})^{t+1}\|_1 = 0 \mid N^t = C] \geq \left(1 - \frac{16 \ln 2t_c}{t_s}\right)^2 \geq 1 - \frac{32 \ln 2t_c}{t_s}.$$

Thus, $\mathbb{P}[Y^{t+1} \text{ is a valid WTA output configuration} \mid N^t = C] \leq \frac{32 \ln 2t_c}{t_s}$, giving the lemma. $\square$

We can now prove the main two-inhibitor lower bound Theorem 4.14 via a case analysis which combines Lemma 4.16, Corollary 4.25, and Corollary 4.27.

Proof of Theorem 4.14.

Assume that $\mathcal{N} = \langle N, w, b, f \rangle$ is a symmetric SNN with two auxiliary inhibitory neurons a_1, a_2 which solves $WTA(n, t_c, t_s, \delta)$ for $n \geq 341$, $\delta \leq 1/2$, $t_c \leq \frac{\ln n}{30 \ln \ln n}$, and $\frac{\ln 2t_c}{t_s} \leq \frac{1}{32 \ln n}$. Assume that the network is given the hard input execution $\alpha_{hard, \mathcal{N}}$ (Definition 4.15) and starts with initial state N^0 in which Y^0 is not a valid WTA output configuration and $a_1^0 = a_2^0 = 0$.

Let $\mathcal{E}_{fail}(t)$ be the event that for every $t' \leq t$, $Y^{t'}$ is not a valid WTA output configuration. Let $\mathcal{E}_{01}(t)$ be the event that for every $t' \leq t$, we *do not have* $a_1^{t'} = 0$ and $a_2^{t'} = 1$.

We can bound $\mathbb{P}[\mathcal{E}_{fail}(t)] \geq \mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t)]$ and write, using that $\mathcal{E}_{fail}(t-1) \subseteq \mathcal{E}_{fail}(t)$ and $\mathcal{E}_{01}(t-1) \subseteq \mathcal{E}_{01}(t)$,

$$\mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t)] = \mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t) | \mathcal{E}_{fail}(t-1), \mathcal{E}_{01}(t-1)] \cdot \mathbb{P}[\mathcal{E}_{fail}(t-1), \mathcal{E}_{01}(t-1)]. \quad (61)$$

Conditioned on $\mathcal{E}_{fail}(t-1)$ and $\mathcal{E}_{01}(t-1)$, Y^{t-1} is not a valid WTA output configuration and $(a_1^{t-1}, a_2^{t-1}) \neq (0, 1)$. Let $\mathcal{C}$ be the set of all configurations which are not valid WTA output configurations and which have $(C(a_1), C(a_2)) \neq (0, 1)$. We can expand (61) as:

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t)] &= \mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t) | \mathcal{E}_{fail}(t-1), \mathcal{E}_{01}(t-1)] \cdot \mathbb{P}[\mathcal{E}_{fail}(t-1), \mathcal{E}_{01}(t-1)] \\ &= \sum_{C \in \mathcal{C}} \mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t) | N^{t-1} = C] \cdot \mathbb{P}[N^{t-1} = C, \mathcal{E}_{fail}(t-1), \mathcal{E}_{01}(t-1)]. \end{aligned} \quad (62)$$

Note that by the law of total probability since $\mathcal{E}_{fail}(t-1)$ and $\mathcal{E}_{01}(t-1)$ requires that $C \in \mathcal{C}$:

$$\sum_{C \in \mathcal{C}} \mathbb{P}[N^{t-1} = C, \mathcal{E}_{fail}(t-1), \mathcal{E}_{01}(t-1)] = \mathbb{P}[\mathcal{E}_{fail}(t-1), \mathcal{E}_{01}(t-1)]. \quad (63)$$

We now bound $\mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t) | N^{t-1} = C]$ in (62) in two cases:

Case 1: $C(a_1) = C(a_2) = 1$ or $C(a_1) = C(a_2) = 0$.

In this case, for any $C \in \mathcal{C}$ with $C(a_1) = C(a_2)$, by Lemma 4.16,

$$\mathbb{P}[\mathcal{E}_{fail}(t) | N^{t-1} = C] \geq 1 - \frac{1}{\ln n}.$$

By Corollary 4.25,

$$\mathbb{P}[\mathcal{E}_{01}(t) | N^{t-1} = C] \geq 1 - \frac{12 \cdot \ln 2t_c}{t_s} \geq 1 - \frac{1}{\ln n},$$

where the second inequality follows from our requirement that $\frac{\ln 2t_c}{t_s} \leq \frac{1}{32 \ln n}$. Additionally, conditioned on N^{t-1} , $\mathcal{E}_{fail}(t)$ and $\mathcal{E}_{01}(t)$ are independent since they involve disjoint sets of neurons. Thus we can bound:

$$\mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t) | N^{t-1} = C] \geq \left(1 - \frac{1}{\ln n}\right)^2. \quad (64)$$

Case 2: $C(a_1) = 1$ and $C(a_2) = 0$.

In this case, for any $C \in \mathcal{C}$, since $C(Y)$ is not a valid WTA output configuration, by Corollary 4.27,

$$\mathbb{P}[\mathcal{E}_{fail}(t) | N^{t-1} = C] \geq 1 - \frac{32 \cdot \ln 2t_c}{t_s} \geq 1 - \frac{1}{\ln n}$$

where the second inequality follows from our requirement that $\frac{\ln 2t_c}{t_s} \leq \frac{1}{32 \ln n}$. As above we also have $\mathbb{P}[\mathcal{E}_{01}(t) | N^{t-1} = C] \geq 1 - \frac{1}{\ln n}$ and since $\mathcal{E}_{fail}(t)$ and $\mathcal{E}_{01}(t)$ are independent conditioned on N^{t-1} :

$$\mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t) | N^{t-1} = C] \geq \left(1 - \frac{1}{\ln n}\right)^2. \quad (65)$$

Together (64) and (65) and the definition of $\mathcal{C}$ give us:

Claim 4.28. *For any C which is not a valid WTA configuration and with $(C(a_1), C(a_2)) \neq (0, 1)$,*

$$\mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t) | N^{t-1} = C] \geq \left(1 - \frac{1}{\ln n}\right)^2.$$

Completing the Theorem.

We conclude by using Claim 4.28 to lower bound the probability of converging to a valid WTA configuration within t_c steps. Substituting the bound of Claim 4.28 into (62) and (63) we have:

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t)] &\geq \sum_{C \in \mathcal{C}} \left(1 - \frac{1}{\ln n}\right)^2 \cdot \mathbb{P}[N^{t-1} = C, \mathcal{E}_{fail}(t-1), \mathcal{E}_{01}(t-1)] \\ &\geq \left(1 - \frac{1}{\ln n}\right)^2 \cdot \mathbb{P}[\mathcal{E}_{fail}(t-1), \mathcal{E}_{01}(t-1)] \end{aligned}$$

Since we choose N^0 with Y^0 not a valid WTA output configuration and $a_1^0 = a_2^0 = 0$, $\mathbb{P}[\mathcal{E}_{fail}(0), \mathcal{E}_{01}(0)] = 1$. Thus, via induction, we have for any $t \geq 0$,

$$\mathbb{P}[\mathcal{E}_{fail}(t), \mathcal{E}_{01}(t)] \geq \left(1 - \frac{1}{\ln n}\right)^{2t}.$$

Plugging in $t_c \leq \frac{\ln n}{30 \ln \ln n}$ we have:

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{fail}(t_c)] &\geq \mathbb{P}[\mathcal{E}_{fail}(t_c), \mathcal{E}_{01}(t_c)] \geq \left(1 - \frac{1}{\ln n}\right)^{\frac{\ln n}{15 \ln \ln n}} \\ &\geq \frac{1}{e^{1/15}} > 1/2. \end{aligned}$$

This contradicts the fact that $\mathcal{N}$ solves $WTA(n, t_c, t_s, \delta)$ for $\delta \geq 1/2$, giving the lower bound. $\square$

5 Faster Convergence With More Inhibitors

In this section we show how to speed up the two-inhibitor construction of Section 3 by using $\alpha > 2$ inhibitors. We give a formal convergence proof for a construction which uses $\lceil \log_2 n \rceil + 1$ inhibitors and converges with constant probability in $O(1)$ time (and with probability $\geq 1 - \delta$ in $O(\log 1/\delta)$ time). We then describe the high level idea behind two constructions that give a tradeoff between the number of inhibitors used and the convergence time.

5.1 Use of History Period

Our $\lceil \log_2 n \rceil + 1$ -inhibitor construction (as well as the sketched constructions which give an inhibitor-convergence time tradeoff) requires using a history period of $h = 2$, as suggested in Section 2.5. At a high level, to achieve fast convergence, our networks use the larger number of inhibitors available to create higher levels of inhibition at time t corresponding to higher number of firing outputs at time $t - 1$. This strategy however, leads to ‘race conditions’. If many outputs fire at time t and exactly one output fires at time $t + 1$, there will still be a high level of inhibition at time $t + 1$. Thus, at time $t + 2$ it is likely that the single firing output at time $t + 1$ will stop firing and so the network will not stabilize to a valid WTA output state.

To avoid this race condition, we use history so that an output's self-loop excites the output for two time steps. The stability inhibitor a_s will also be excited for two time steps by the outputs. In this way, if a single output y_i fires at time t , while at time $t + 1$ no outputs may fire due to high levels of inhibition, at time $t + 2$ the history will cause both a_s and y_i to fire, and the network to stabilize to a valid WTA state.

5.1.1 Generalized Model with History Period

Following the basic SNN model of Section 2 we define an SNN model with history period h for any $h \geq 1$. The changes to the definition are highlighted in gray.

An SNN $\mathcal{N} = \langle N, w, b, f, h \rangle$ with history period h consists of:

- N , a set of neurons, partitioned into a set of input neurons X , a set of output neurons Y , and a set of auxiliary neurons A . N is also partitioned into a set of excitatory and inhibitory neurons E and I . All input and output neurons are excitatory.
- $h \in \mathbb{Z}^{\geq 1}$, a positive integer indicating the neural response history period.
- $w : N \times N \times \{1, \dots, h\} \rightarrow \mathbb{R}$, a weight function describing the weighted synaptic connections between the neurons in the network. w is restricted in a few notable ways:
 - $w(u, x, l) = 0$ for all $u \in N$, $x \in X$, and $l \in \{1, \dots, h\}$.
 - Each excitatory neuron $v \in E$ has $w(v, u, l) \geq 0$ for every u and $l \in \{1, \dots, h\}$. Each inhibitory neuron $v \in I$ has $w(v, u, l) \leq 0$ for every u and $l \in \{1, \dots, h\}$.
- $b : N \rightarrow \mathbb{R}$, a bias function, assigning an activation bias to each neuron.
- $f : \mathbb{R} \rightarrow [0, 1]$, a spike probability function, satisfying a few restrictions:
 - f is continuous and monotonically increasing.
 - $\lim_{x \rightarrow \infty} f(x) = 1$ and $\lim_{x \rightarrow -\infty} f(x) = 0$.

Remark on the Time Dependent Weight Function: The only difference between the above model and our basic SNN model of Section 2.1 is in the specification of the weight function w . In the model with history period h , w describes the strength of the synaptic connections between neurons in N , as a function of the time difference between a spike and the current time (from 1 up to h). $w(u, v, 1)$ is the weight corresponding to the most recent time and $w(u, v, h)$ corresponds to the most distant time within the history period. The weight function, for example, can be used to model the decaying effect of a spike over time, if we set $|w(u, v, 1)| \geq |w(u, v, 2)| \geq \dots \geq |w(u, v, h)|$.

5.1.2 Network Dynamics With History Period

In our SNN with history period model, configurations and executions are defined as in our basic model (see Section 2.2). The behavior of the SNN is determined as follows:

- **Input Neurons:** As in our basic model, we specify how the infinite input execution $X^0 X^1 \dots$ is determined. Through this work, we will fix the input so that for each $u \in X$, u^t is constant for all $t \geq 0$.

- **Initial Firing States:** For each non-input $u \in N \setminus X$, the firing states in the first h time slots $u^0, u^1, \dots, u^{h-1}$ are arbitrary, where h is this history length.
- **Firing Dynamics:** For each non-input neuron $u \in N \setminus X$ and every time $t \geq h$, let $\text{pot}(u, t)$ denote the membrane potential at time t and $p(u, t)$ denote the corresponding firing probability. These values are calculated as:

$$\text{pot}(u, t) = \left(\sum_{i=1}^h \sum_{v \in N} w(v, u, i) \cdot v^{t-i} \right) - b(u) \text{ and } p(u, t) = f(\text{pot}(u, t)) \quad (66)$$

where f is the spike probability function. At time t , each non-input neuron u fires independently with probability $p(u, t)$. Note that equation (66) is defined only for $t \geq h$, in which case $t - i \geq 0$ for all $i \in \{1, \dots, h\}$. It is analogous to the potential calculation (1), used for our basic model, except that the summation of spikes is over h time steps.

An SSN $\mathcal{N} = \langle N, w, b, f, h \rangle$ with history period h , length h execution $\alpha^h = N^0 N^1 \dots N^{h-1}$, and infinite input execution α_X define a probability distribution over infinite executions, $\mathcal{D}(\mathcal{N}, \alpha^h, \alpha_X)$. This distribution is the natural distribution that follows from applying the stochastic firing dynamics of (66). As with our basic model (see Section 2.2), we can also define a corresponding distribution $\mathcal{D}_Y(\mathcal{N}, \alpha^h, \alpha_X)$ on infinite output executions.

5.1.3 Solving Problems in Networks with History

As in our basic model, a problem P is a mapping from an infinite input execution α_X to a set of output distributions. A network $\mathcal{N}$ is said to *solve problem P on input α_X* if, for any length h initial execution α^h , the output distribution $\mathcal{D}_Y(\mathcal{N}, \alpha^h, \alpha_X)$ is an element of $P(\alpha_X)$. A network $\mathcal{N}$ is said to *solve problem P* if it solves P on every infinite input execution α_X .

5.2 $O(1)$ Convergence Time with $O(\log n)$ Inhibitors

In this section we describe and analyze a family of networks that converge to a valid WTA state with constant probability in constant time and uses $O(\log n)$ inhibitors.

We have one stability inhibitor a_s that functions similarly to the stability inhibitor in our two-inhibitor construction (Definition 3.1), ensuring that, once the network reaches a valid WTA output configuration, it remains in this configuration for t_s consecutive time steps with high probability (see Corollary 5.19). This stability inhibitor employs a history period of length 2. We prove in Lemma 5.7 that it fires with high probability at time $t + 1$ whenever at least one output fires at time t or $t - 1$.

We additionally have $\lceil \log_2 n \rceil$ convergence inhibitors, labeled $a_1, \dots, a_{\lceil \log_2 n \rceil}$. For each i , a_i fires with high probability at time $t + 1$ whenever $k \geq 2^i$ outputs fire at time t . In this way, with high probability, $a_1, \dots, a_i$ fire and all other inhibitors do not fire at time $t + 1$ whenever the number of firing outputs k is in the range $[2^i, 2^{i+1})$ (see Lemma 5.8). Note that we define our network so $a_1, \dots, a_{\lceil \log_2 n \rceil}$ have no incoming connections that use the length 2 history period (i.e., $w(u, a_i, 2) = 0$ for all u and all $i \in \{1, \dots, \lceil \log_2 n \rceil\}$). Thus, the firing probabilities of these inhibitors at time $t + 1$ depend only on the firing pattern at time t .

We set the inhibitory weights such that when a_i fires at time t (along with a_j for all $j \leq i$), each firing output fires at time $t + 1$ with probability $p_i \approx \frac{1}{2^i}$ (see Lemma 5.12). For $k \in [2^i, 2^{i+1})$ we thus have $p_i \cdot k \in [1, 2)$. We will show that this ensures that, with constant probability, exactly one output fires at time $t + 1$ (see Corollary 5.13). Once a single output fires, using the length-two

history mechanism described in Section 5.1, the network stabilizes to a valid WTA state with high probability.

We begin with a formal definition of the network below:

Definition 5.1 (Constant Time WTA Network). *For any $n \in \mathbb{Z}^{\geq 2}$ and $\gamma \in \mathbb{R}^+$, let $\mathcal{L}_{n,\gamma} = \langle N, w, b, f, 2 \rangle$ where the spike probability, weight, and bias functions are defined as follows:*

- *The spike probability function f is defined to be the basic sigmoid function:*

$$f(x) \stackrel{\text{def}}{=} \frac{1}{1 + e^{-x}}.$$

- *The set of neurons N consists of a set of n input neurons X , labeled $x_1, \dots, x_n$, a set of n corresponding outputs Y , labeled $y_1, \dots, y_n$, and $\lceil \log_2 n \rceil + 1$ auxiliary inhibitory neurons labeled a_s and $a_1, \dots, a_{\lceil \log_2 n \rceil}$.*
- *The weight function w is given by:*
 - $w(x_i, y_i, 1) = 6\gamma$, for all i .
 - $w(y_i, y_i, 1) = w(y_i, y_i, 2) = 2\gamma$, for all i .
 - $w(a_s, y_i, 1) = -\gamma$, for all i
 - $w(a_1, y_i, 1) = -7\gamma/2 - \ln(2)$, for all i
 - $w(a_j, y_i, 1) = -\ln(2)$, for all i and $j \in \{2, \dots, \lceil \log_2 n \rceil\}$.
 - $w(y_i, a_s, 1) = w(y_i, a_s, 2) = \gamma$, for all i .
 - $w(y_i, a_j, 1) = \gamma$, for all i and $j \in \{1, \dots, \lceil \log_2 n \rceil\}$.
 - $w(u, v, i) = 0$ for any u, v and $i \in \{1, 2\}$ whose connection is not specified above.
- *The bias function b is given by:*
 - $b(y_i) = 11\gamma/2$ for all i .
 - $b(a_s) = \gamma/2$.
 - $b(a_i) = 2^i \cdot \gamma - \gamma/2$ for all $i \in \{1, \dots, \lceil \log_2 n \rceil\}$.

Use of History. Note that in our network definition above, we use the history mechanism only in two places. We set $w(y_i, y_i, 1) = w(y_i, y_i, 2) = 2\gamma$, for all i , meaning that each output's self-loop affects its potential for two time steps. We also set $w(y_i, a_s, 1) = w(y_i, a_s, 2) = \gamma$ for all i , meaning that the stability inhibitor is affected by the outputs for two steps.

Due to the high level of inhibition induced by the convergence inhibitors $a_1, \dots, a_{\lceil \log_2 n \rceil}$, after the network reaches a configuration with just a single firing output, it will likely transition to a state with no firing outputs, since the number of firing inhibitors will still reflect the number of firing outputs in the previous time step. The length-two output self-loop and output to inhibitor connections allow the network to recover from this state. Specifically, in Lemma 5.18 and Corollary 5.19, we show that if the network reaches a valid WTA output state at time t , with good probability it will return to this state at time $t + 2$ and remain in this state for t_s consecutive steps.

In our two-inhibitor construction, no history was necessary. The inhibition in the network was always low enough such that, after reaching a near-valid WTA configuration with a single firing output, with constant probability the network would transition to a valid-WTA configuration and stabilize in this configuration (see Lemma 3.21).

We prove the following theorems on the performance of $\mathcal{L}_{n,\gamma}$:

Theorem 5.2 ($O(\log n)$ -Inhibitor WTA). *For $\gamma \geq \ln((n+2)t_s/\delta)$, $\mathcal{L}_{n,\gamma}$ solves WTA(n, t_c, t_s, δ) for any $t_c \geq \log_2(1/\delta) + 1$. $\mathcal{L}_{n,\gamma}$ contains $\lceil \log_2 n \rceil + 1$ auxiliary inhibitors.*

Theorem 5.3 ($O(\log n)$ -Inhibitor Expected-Time WTA). *For $\gamma \geq \ln((n+2)t_s/\delta)$, $\mathcal{L}_{n,\gamma}$ solves WTA-EXP(n, t_c, t_s) for any $t_c \geq 4$. $\mathcal{L}_{n,\gamma}$ contains $\lceil \log_2 n \rceil + 1$ auxiliary inhibitors.*

Proof Roadmap. We prove Theorems 5.2 and 5.3 in Section 5.5. The analysis is broken down as follows:

Section 5.3: Prove basic *two-step lemmas* which characterize single time step transitions of $\mathcal{L}_{n,\gamma}$, showing that the neurons behave as described in the above high-level description.

Section 5.4: Prove that, once in a valid WTA output configuration, as long as certain other stability conditions are satisfied, $\mathcal{L}_{n,\gamma}$ stays in this configuration with high probability.

Section 5.5: Show that all configurations of $\mathcal{L}_{n,\gamma}$ transition with constant probability to a stable and valid WTA configuration within $O(1)$ time steps.

Section 5.6: Complete the analysis, demonstrating with what parameter values $\mathcal{L}_{n,\gamma}$ solves the winner-take-all problem (Definitions 2.7 and 2.9).

5.3 Two-Step Lemmas

As in our analysis of our two-inhibitor network family in Section 3.2, we begin with a series of lemmas which characterize the basic transition probabilities of $\mathcal{L}_{n,\gamma}$. Since we employ history period $h = 2$, these lemmas consider the behavior of the network at time $t + 1$ conditioned on its configuration at times t and $t - 1$.

We first note that an analog to Lemma 3.4 still holds for all inhibitory neurons.

Lemma 5.4 (Characterization of Firing Probabilities). *For any time $t \geq 1$ and any $a \in A$:*

If $\text{pot}(u, t) = 0$, then $p(u, t) = 1/2$.

If $\text{pot}(u, t) < 0$, then $p(u, t) \leq e^{-\gamma/2}$.

If $\text{pot}(u, t) > 0$, then $p(u, t) \geq 1 - e^{-\gamma/2}$.

Proof. The proof is essentially identical to that of Lemma 3.4. We just use that for each inhibitor $a \in A$ and each i and $h \in \{1, 2\}$, $w(y_i, a, h)$ and $b(a)$ are all multiples of $\gamma/2$. $\square$

Analogous to Lemma 3.5 and Corollary 3.6 also still hold, ensuring that, with high probability, outputs that do not correspond to firing inputs do not fire.

Lemma 5.5 (Correct Output Behavior). *For any time $t \geq 1$, any configurations C, C' of $\mathcal{L}_{n,\gamma}$, and any i with $C(x_i) = 0$,*

$$\mathbb{P}[y_i^{t+1} = 1 | N^t = C, N^{t-1} = C'] \leq e^{-3\gamma/2}.$$

Proof. If $N^t = C$ then $x_i^t = C(x_i)$. We can compute y_i 's potential at time $t + 1$, assuming $x_i^t = 0$:

$$\begin{aligned} \text{pot}(y_i, t + 1) &= w(x_i, y_i, 1)x_i^t + w(y_i, y_i, 1)y_i^t + w(y_i, y_i, 2)y_i^{t-1} + w(a_s, y_i, 1)a_s^t \\ &\quad + \sum_{j=1}^{\lceil \log_2 n \rceil} w(a_c, y_i, 1)a_c^t - b(y_i) \\ &\leq 0 + 2\gamma + 2\gamma + 0 + 0 - 11\gamma/2 = -3\gamma/2. \end{aligned}$$

We thus have $p(y_i, t + 1) \leq \frac{1}{1+e^{3\gamma/2}} \leq e^{-3\gamma/2}$. $\square$

Corollary 5.6 (Correct Output Behavior, All Neurons). *For any time $t \geq 1$ and configurations C, C' of $\mathcal{L}_{n,\gamma}$,*

$$\mathbb{P}[y_i^{t+1} \leq x_i^t \text{ for all } i | N^t = C, N^{t-1} = C'] \geq 1 - ne^{-3\gamma/2}.$$

Proof. The proof is essentially identical to that of Corollary 3.6. If $C(x_i) = 1$ then conditioned on $N^t = C$, $x_i^t = 1$ and so $y_i^{t+1} \leq x_i^t$ always. Otherwise, by Lemma 5.5, if $C(x_i) = 0$, then $\mathbb{P}[y_i^{t+1} = 0 | N^t = C, N^{t-1} = C'] \geq 1 - e^{-3\gamma/2}$. Union bounding over all such inputs (of which there are at most n) gives the corollary. $\square$

We next show that the inhibitors $a_s, a_1, \dots, a_{\lceil \log_2 n \rceil}$ behave as expected. The following lemmas can be viewed as a generalization of Lemma 3.7. We first show that, due to our use of history, a_s fires with high probability at time $t + 1$ whenever *at least one output fires at time t or $t - 1$* .

Lemma 5.7 (Correct Stability Inhibitor Behavior). *For any time $t \geq 1$ and configurations C, C' of $\mathcal{L}_{n,\gamma}$,*

1. *If $\|C(Y)\|_1 = \|C'(Y)\|_1 = 0$, then $\mathbb{P}[a_s^{t+1} = 0 | N^t = C, N^{t-1} = C'] \geq 1 - e^{-\gamma/2}$.*
2. *If $\|C(Y)\|_1 \geq 1$ or $\|C'(Y)\|_1 \geq 1$, then $\mathbb{P}[a_s^{t+1} = 1 | N^t = C, N^{t-1} = C'] \geq 1 - e^{-\gamma/2}$.*

Proof. We prove the two conclusions separately.

Conclusion 1: $\|C(Y)\|_1 = \|C'(Y)\|_1 = 0$.

In this case, a_s receives no excitatory signal from the outputs so,

$$\text{pot}(a_s, t + 1) = -b(a_s) < 0.$$

Thus by Lemma 5.4,

$$\mathbb{P}[a_s^{t+1} = 0 | N^t = C] \geq 1 - e^{-\gamma/2}.$$

Conclusion 2: $\|C(Y)\|_1 \geq 1$ or $\|C'(Y)\|_1 \geq 1$.

In this case we have:

$$\begin{aligned} \text{pot}(a_s, t + 1) &= \sum_{j=1}^n [w(y_j, a_s, 1)y_j^t + w(y_j, a_s, 2)y_j^{t-1}] - b(a_s) \\ &\geq \gamma - \gamma/2 = \gamma/2. \end{aligned}$$

We thus have by Lemma 5.4:

$$\mathbb{P}[a_s^{t+1} = 1 | N^t = C] \geq 1 - e^{-\gamma/2},$$

which gives the lemma. $\square$

As described in Section 5.2, the convergence inhibitors $a_1, \dots, a_{\lceil \log_2 n \rceil}$ fire at time $t+1$ depending on the number of firing outputs at time t . They have no incoming connections which affect them for two rounds, and thus their firing probabilities at time $t+1$ do not depend on the firing pattern at time $t-1$. We prove that a_j for all $j \leq i$ fire with high probability at time $t+1$ whenever the number of firing outputs at time t falls in the range $[2^i, 2^{i+1})$. Further, all a_j for $j > i$ do not fire with high probability.

Lemma 5.8 (Correct Convergence Inhibitor Behaviors). *For any time $t \geq 1$ and configurations C, C' of $\mathcal{L}_{n,\gamma}$,*

1. *If $\|C(Y)\|_1 \leq 1$, then*

$$\mathbb{P}[a_i^{t+1} = 0 \text{ for all } i \in \{1, \dots, \lceil \log_2 n \rceil\} | N^t = C, N^{t-1} = C'] \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}.$$

2. *For any $i \in \{1, \dots, \lceil \log_2 n \rceil\}$, if $\|C(Y)\|_1 = k$ for $k \in [2^i, 2^{i+1})$, then:*

$$\begin{aligned} \mathbb{P}[a_1^{t+1} = \dots = a_i^{t+1} = 1 \text{ and } a_{i+1}^{t+1} = \dots = a_{\lceil \log_2 n \rceil}^{t+1} = 0 | N^t = C, N^{t-1} = C'] \\ \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}. \end{aligned}$$

Proof. We prove the two conclusions of the lemma separately.

Conclusion 1: $\|C(Y)\|_1 \leq 1$.

In this case we have for all i :

$$\begin{aligned} \text{pot}(a_i, t+1) &= \sum_{j=1}^n w(y_j, a_i, 1) y_j^t - b(a_i) \\ &\leq \gamma - 3\gamma/2 = -\gamma/2. \end{aligned}$$

Again by Lemma 5.4 and a union bound,

$$\mathbb{P}\left[\sum_{i=1}^{\lceil \log_2 n \rceil} a_i^{t+1} = 0 | N^t = C, N^{t-1} = C'\right] \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}.$$

Conclusion 2: $\|C(Y)\|_1 = k$ for $k \in [2^i, 2^{i+1})$.

In this case, for any $j \leq i$ we have:

$$\begin{aligned} \text{pot}(a_j, t+1) &= \sum_{l=1}^n w(y_l, a_j, 1) y_l^t - b(a_j) \\ &= k \cdot \gamma - 2^j \gamma + \gamma/2 \geq \gamma/2 \end{aligned}$$

where the last inequality follows since $k \geq 2^j \geq 2^j$. In contrast for $j > i$:

$$\begin{aligned} \text{pot}(a_j, t+1) &= \sum_{l=1}^n w(y_l, a_s, 1) y_l^t - b(a_j) \\ &= k \cdot \gamma - 2^j \gamma + \gamma/2 \leq -\gamma/2 \end{aligned}$$

where the last inequality follows from the fact that $k < 2^{i+1} \leq 2^j$. Overall, by Lemma 5.4 and a union bound,

$$\begin{aligned} \mathbb{P}[a_1^{t+1} = \dots = a_i^{t+1} = 1 \text{ and } a_{i+1}^{t+1} = \dots = a_{\lceil \log_2 n \rceil}^{t+1} = 0 | N^t = C, N^{t-1} = C'] \\ \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}, \end{aligned}$$

which gives the lemma □

It will be useful in our future bounds to consider the class of configurations in which all outputs and inhibitors behave as expected. Such a configuration is analogous to the good configurations of our two-inhibitor networks (Definition 3.17), from which we were able to show convergence.

Definition 5.9 (Typical Configuration). *A typical configuration is any configuration C with $C(y_i) \leq C(x_i)$ for all i and $C(a_s) \geq C(a_1) \geq \dots \geq C(a_{\lceil \log_2 n \rceil})$.*

In combination, Corollary 5.6, Lemma 5.7 and Lemma 5.8 give:

Corollary 5.10 (Correct Behavior, All Neurons). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and consider configurations C, C' with $C(X) = C'(X) = X^t$. For any time $t \geq 1$:*

$$\mathbb{P}[N^{t+1} \text{ is a typical configuration} | N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}.$$

Proof. Since the input is fixed, by Corollary 5.6,

$$\mathbb{P}[y_i^{t+1} \leq x_i^{t+1} | N^t = C, N^{t-1} = C'] \geq 1 - n \cdot e^{-3\gamma/2}. \quad (67)$$

If $\|C(Y)\|_1 \leq 1$ then by Lemma 5.8 conclusion (2),

$$\mathbb{P}[a_1^{t+1} = \dots = a_{\lceil \log_2 n \rceil}^{t+1} = 0 | N^t = C, N^{t-1} = C'] \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}.$$

If $\|C(Y)\|_1 > 1$ then by Lemma 5.7 conclusion (2), Lemma 5.8 conclusion (2), and a union bound,

$$\begin{aligned} \mathbb{P}[a_s^{t+1} = a_1^{t+1} = \dots = a_i^{t+1} = 1 \text{ and } a_{i+1}^{t+1} = \dots = a_{\lceil \log_2 n \rceil}^{t+1} = 0 | N^t = C, N^{t-1} = C'] \\ \geq 1 - (\lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \end{aligned}$$

Combining these two cases, we have

$$\mathbb{P}[a_s^{t+1} \geq a_1^{t+1} \geq \dots \geq a_{\lceil \log_2 n \rceil}^{t+1} | N^t = C, N^{t-1} = C'] \geq 1 - (\lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2},$$

which gives the lemma after a union bound with (67) □

We next show that the stability inhibitor firing alone, with high probability, induces exactly the outputs that fired at one of the previous two time steps to fire in the next step. This Lemma is analogous to Lemma 3.9 for our two-inhibitor networks.

Lemma 5.11 (Stability Inhibitor Effect). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and consider configurations C, C' with $C(X) = C'(X) = X^t$, $C(a_s) = 1$, $C(a_i) = 0$ for all $i \in \{1, \dots, \lceil \log_2 n \rceil\}$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,*

$$\mathbb{P}[y_i^{t+1} = \max(y_i^t, y_i^{t-1}) \text{ for all } i | N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}$$

Proof. Conditioned on $N^t = C, N^{t-1} = C'$, $y_i^t \leq x_i^t$ and $y_i^{t-1} \leq x_i^{t-1}$ by assumption. So for any output with $\max(y_i^t, y_i^{t-1}) = 1$ we must have $x_i^t = 1$. This gives:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i, 1)x_i^t + w(y_i, y_i, 1)y_i^t + w(y_i, y_i, 2)y_i^{t-1} + w(a_s, y_i, 1)a_s^t \\ &\quad + \sum_{j=1}^{\lceil \log_2 n \rceil} w(a_j, y_i, 1)a_j^t - b(y_i) \\ &\geq 6\gamma + 2\gamma \cdot \max(y_i^t, y_i^{t-1}) - \gamma + 0 - 11\gamma/2s \\ &\geq 3\gamma/2. \end{aligned}$$

In contrast, for any output with $\max(y_i^t, y_i^{t-1}) = 0$:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i, 1)x_i^t + w(y_i, y_i, 1)y_i^t + w(y_i, y_i, 2)y_i^{t-1} + w(a_s, y_i, 1)a_s^t \\ &\quad + \sum_{j=1}^{\lceil \log_2 n \rceil} w(a_j, y_i, 1)a_j^t - b(y_i) \\ &\leq 6\gamma + 0 + 0 - \gamma + 0 - 11\gamma/2 \\ &= -\gamma/2. \end{aligned}$$

So, if $\max(y_i^t, y_i^{t-1}) = 1$, then $y_i^{t+1} = 1$ with probability $\geq 1 - e^{-3\gamma/2}$. If $\max(y_i^t, y_i^{t-1}) = 0$, then $y_i^{t+1} = 0$ with probability $\geq 1 - e^{-\gamma/2}$. The lemma follows after union bounding over all n outputs. $\square$

We next characterize the effect of the convergence inhibitors. We show that when l inhibitors fire, any firing output that also fired in the previous time step, fires with probability $\Theta(1/l)$. We will show in Corollary 5.13 that this implies that in the next step, with constant probability, exactly one output fires, and in fact the configuration is a valid WTA output configuration.

Lemma 5.12 below is analogous to Lemma 3.10 for our two-inhibitor networks, except that the firing probability is $\Theta(1/l)$ rather than $1/2$. Since this firing probability is smaller when a larger number of outputs fire at time t , convergence to a single firing output with constant probability occurs in 1 step, rather than $O(\log n)$ steps.

Lemma 5.12 (Convergence Inhibitor Effect). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and consider configurations C, C' with $C(X) = C'(X) = X^t$, $C(a_s) = C(a_1) = \dots C(a_l) = 1$, $C(a_{l+1}) = \dots = C(a_{\lceil \log_2 n \rceil}) = 0$ for some $l \geq 1$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,*

1. $\mathbb{P}[y_i^{t+1} \leq \min(y_i^t, y_i^{t-1}) \text{ for all } i | N^t = C, N^{t-1} = C'] \geq 1 - ne^{-2\gamma}.$
2. *If $\min(y_i^t, y_i^{t-1}) = 1$, $\mathbb{P}[y_i^{t+1} = 1 | N^t = C, N^{t-1} = C'] = \frac{1}{1+2^l}.$*
3. *For $i \neq j$, y_i^{t+1} and y_j^{t+1} are independent conditioned on $N^t = C, N^{t-1} = C'.$*

Proof. Conditioned on $N^t = C, N^{t-1} = C'$, if $\min(y_i^t, y_i^{t-1}) = 1$, then $y_i^t = y_i^{t-1} = 1$ and by assumption $x_i^t = 1$. We can thus compute:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i, 1)x_i^t + w(y_i, y_i, 1)y_i^t + w(y_i, y_i, 2)y_i^{t-1} + w(a_s, y_i, 1)a_s^t + w(a_1, y_i, 1)a_1^t \\ &\quad + \sum_{j=2}^{\lceil \log_2 n \rceil} w(a_j, y_i, 1)a_j^t - b(y_i) \\ &= 6\gamma + 4\gamma - \gamma - 7\gamma/2 - \ln 2 - (l-1) \cdot \ln 2 - 11\gamma/2 \\ &= -l \cdot \ln 2. \end{aligned}$$

We thus have

$$\mathbb{P}[y_i^{t+1} = 1 | N^t = C] = f(-l \cdot \ln 2) = \frac{1}{1+2^l}.$$

This gives conclusion (2). Conclusion (3) holds since, with N^t and N^{t-1} fixed with history length 2, u^{t+1} is independent of v^{t+1} for all $u \neq v$. We can also bound if $\min(y_i^t, y_i^{t-1}) = 0$:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i, 1)x_i^t + w(y_i, y_i, 1)y_i^t + w(y_i, y_i, 2)y_i^{t-1} + w(a_s, y_i, 1)a_s^t + w(a_1, y_i, 1)a_1^t \\ &\quad + \sum_{j=2}^{\lceil \log_2 n \rceil} w(a_j, y_i, 1)a_j^t - b(y_i) \\ &\leq 6\gamma + 2\gamma - \gamma - 7\gamma/2 - \ln 2 - (l-1) \cdot \ln 2 - 11\gamma/2 \\ &\leq -2\gamma. \end{aligned}$$

Thus, $\mathbb{P}[y_i^{t+1} = 1 | N^t = C, N^{t-1} = C'] \leq e^{-2\gamma}$. By a union bound over at most n such outputs, we have, with probability $\geq 1 - ne^{-2\gamma}$, $y_i^{t+1} \leq \min(y_i^t, y_i^{t-1})$ for all i , giving conclusion (1) and completing the lemma. $\square$

We now formalize the fact that the network converges to a valid WTA output configuration in just a single step with constant probability, as long as the number of inhibitors matches the minimum number of firing outputs in the preceding two steps. Corollary 5.13 can be viewed as an analog to Lemma 3.23 for our two inhibitor networks, except that the number of outputs is reduced to 1, rather than just cut in half, with constant probability.

Corollary 5.13 (Constant Probability of a Valid WTA Configuration). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and consider configurations C, C' with $C(X) = C'(X) = X^t$, $C(a_s) = C(a_1) = \dots C(a_l) = 1$, $C(a_{l+1}) = \dots = C(a_{\lceil \log_2 n \rceil}) = 0$ for some $l \geq 1$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$, if $\|\min(C', C)\|_1 \in [2^l, 2^{l+1})$, then*

$$\mathbb{P}[Y^{t+1} \text{ is a valid WTA output configuration} \mid N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - ne^{-2\gamma}.$$

Proof. Let $\bar{Y}$ be the set of outputs who fire in both in C, C' . So $\|\bar{Y}\|_1 = \|\min(C', C)\|_1$. By conclusions (2) and (3) of Lemma 5.12, and the assumption that $\|\min(C', C)\|_1 \in [2^l, 2^{l+1})$:

$$\begin{aligned} \mathbb{P}[\|\bar{Y}^{t+1}\|_1 = 1 \mid N^t = C, N^{t-1} = C'] &= \frac{1}{1+2^l} \cdot \left(1 - \frac{1}{1+2^l}\right)^{\|\bar{Y}^t\|_1} \cdot \|\bar{Y}^t\|_1 \\ &\geq \frac{1}{1+2^l} \left(1 - \frac{1}{1+2^l}\right)^{2^{l+1}} \cdot 2^l \\ &\geq \frac{2^l}{1+2^l} \cdot \frac{1}{8} \geq \frac{1}{16}. \end{aligned}$$

Further, by conclusion (1) of Lemma 5.12, no output outside if $\bar{Y}$ fires at time $t+1$ with probability $\geq 1 - ne^{-2\gamma}$. Additionally, by assumption, for all $y_i \in \bar{Y}$, $x_i^t = 1$. So since if exactly one output in $\bar{Y}$ fires, N^{t+1} is a valid WTA output configuration. This gives the corollary by a union bound. $\square$

We also show a related corollary – if the number of firing inhibitors *exceeds* the appropriate amount for the number of firing outputs, then with good probability, no outputs fire in the next time step.

Corollary 5.14 (Constant Probability of Zero Firing Outputs). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and consider configurations C, C' with $C(X) = C'(X) = X^t$, $C(a_s) = C(a_1) = \dots C(a_l) = 1$, $C(a_{l+1}) = \dots = C(a_{\lceil \log_2 n \rceil}) = 0$ for some $l \geq 1$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$, if $\|\min(C', C)\|_1 \in [0, 2^{l+1})$, then*

$$\mathbb{P}[\|Y^{t+1}\|_1 = 0 \mid N^t = C, N^{t-1} = C'] \geq \frac{1}{8} - ne^{-2\gamma}$$

Proof. Let $\bar{Y}$ be the set of outputs who fire in both in C, C' . So $\|\bar{Y}\|_1 = \|\min(C', C)\|_1 \in [0, 2^{l+1})$. By conclusions (2) and (3) of Lemma 5.12:

$$\begin{aligned} \mathbb{P}[\|\bar{Y}^{t+1}\|_1 = 0 \mid N^t = C, N^{t-1} = C'] &= \left(1 - \frac{1}{1+2^l}\right)^{\|\bar{Y}^t\|_1} \\ &\geq \left(1 - \frac{1}{1+2^l}\right)^{2^{l+1}} \\ &\geq \frac{1}{8}. \end{aligned}$$

Further, by conclusion (1) of Lemma 5.12, no output outside if $\bar{Y}$ fires at time $t+1$ with probability $\geq 1 - ne^{-2\gamma}$. This gives the corollary by a union bound. $\square$

Finally, we show that if there is no inhibition in the network, all outputs corresponding to firing inputs are likely to fire at the next time step.

Lemma 5.15 (No Inhibitor Effect). *For any time $t \geq 1$ and configurations C, C' of $\mathcal{L}_{n,\gamma}$, if $\|C(A)\|_1 = 0$, then*

$$\mathbb{P}[y_i^{t+1} = x_i^t \text{ for all } i | N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}.$$

Proof. We consider two cases:

Case 1: $x_i^t = 0$.

In this case:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i, 1)x_i^t + w(y_i, y_i, 1)y_i^t + w(y_i, y_i, 2)y_i^{t-1} + w(a_s, y_i, 1)a_s^t + \\ &\quad + \sum_{j=1}^{\lceil \log_2 n \rceil} w(a_j, y_i, 1)a_j^t - b(y_i) \\ &\leq 0 + 4\gamma + 0 + 0 - 11\gamma/2 \\ &\leq -3\gamma/2. \end{aligned}$$

This gives

$$\mathbb{P}[y_i^{t+1} = 0 = x_i^t | N^t = C, N^{t-1} = C'] \geq 1 - e^{-2\gamma}.$$

Case 2: $x_i^t = 1$.

In this case:

$$\begin{aligned} \text{pot}(y_i, t+1) &= w(x_i, y_i, 1)x_i^t + w(y_i, y_i, 1)y_i^t + w(y_i, y_i, 2)y_i^{t-1} + w(a_s, y_i, 1)a_s^t + \\ &\quad + \sum_{j=1}^{\lceil \log_2 n \rceil} w(a_j, y_i, 1)a_j^t - b(y_i) \\ &\geq 6\gamma + 0 + 0 + 0 - 11\gamma/2 \\ &\geq \gamma/2. \end{aligned}$$

This gives

$$\mathbb{P}[y_i^{t+1} = 1 = x_i^t | N^t = C, N^{t-1} = C'] \geq 1 - e^{-\gamma/2}.$$

The lemma then follows after union bounding over all n outputs. $\square$

5.4 Stability

In this section we show that once in a valid WTA output configuration (Definition 2.6), the network remains in this configuration with high probability. Due to our use of a length-two history period, our stability proof requires certain conditions on the firing states at both times t and times $t-1$. We will focus on the case when there is at least one firing input (i.e., when $\|X^t\| \geq 1$.) In the case $\|X^t\| = 0$, convergence to a valid WTA output configuration and stability of this configuration follow easily from Lemma 5.5.

Definition 5.16 below can be viewed as a two-step generalization of a near-valid WTA configuration of our two-inhibitor networks (Definition 3.15).

Definition 5.16 (Near-Stable Pair of Configurations). Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and $\|X^t\|_1 \geq 1$. Consider configurations C, C' with $C(X) = C'(X) = X^t$. The ordered pair (C', C) is near-stable if:

1. $\|\max(C'(Y), C(Y))\|_1 = 1$, where $\max(C'(Y), C(Y))$ is the entrywise maximum of $C'(Y), C(Y)$. This condition requires that exactly one output fires in configurations C', C . It may fire in one or both configurations.
2. $C(a_s) = C'(a_s) = 1$.
3. $C(a_i) = 0$ for all $i \in \{1, \lceil \log_2 n \rceil\}$.
4. $C(y_i) \leq C(x_i), C'(y_i) \leq C'(x_i)$ for all i .

Note that by conditions (1) and (4), at least one of $C(Y), C'(Y)$ is a valid WTA output configuration (Definition 2.6). In our proofs, it will be useful to refer to the output whose existence is guaranteed by condition (1). Thus we define:

Definition 5.17. Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and $\|X^t\|_1 \geq 1$. For any near-stable pair of configurations (C', C) with $C(X) = C'(X) = X^t$, let $\text{out}(C', C) \in \{1, \dots, n\}$ be equal to the index of the unique output that fires in C', C (whose existence is guaranteed by condition (1) of Definition 5.16).

We next show that if the configurations (N^{t-1}, N^t) are near-stable, then with high probability, N^{t+1} will be a valid WTA output configuration. Further, the network will stabilize for t_s steps. That is, with high probability, we will have $N^{t+1} = \dots = N^{t+t_s+1}$. Lemma 5.16 is analogous to Lemma 3.21 for our two-inhibitor networks.

Lemma 5.18 (Reaching Stability From Near-Stable Configurations). Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any near-stable pair of configurations (C', C) with $C(X) = C'(X) = X^t$. For any time $t \geq 1$, conditioned on $N^t = C, N^{t-1} = C'$, with probability $\geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}$,

1. N^{t+1} is a valid WTA output configuration (Definition 2.6).
2. $y_{\text{out}(N^{t-1}, N^t)}^{t+1} = 1$. That is, the winner at time $t+1$ is the output firing in N^{t-1} , and/or N^t .
3. (N^t, N^{t+1}) is also a near-stable pair of configurations.

Proof. By condition (1) of Definition 5.16, for all $j \neq \text{out}(N^{t-1}, N^t)$, $y_j^t = y_j^{t-1} = 0$. Additionally, by conditions (2) and (3), a_s is the only inhibitor that fires at time t . So by Lemma 5.11,

$$\mathbb{P}[Y^{t+1} = \max(Y^t, Y^{t-1}) | N^t = C] \geq 1 - ne^{-\gamma/2}. \quad (68)$$

This gives that $y_{\text{out}(N^{t-1}, N^t)}^{t+1} = 1$ while $y_j^{t+1} = 0$ for all $j \neq \text{out}(N^{t-1}, N^t)$. By condition (4) of Definition 5.16 we must have also have $x_{\text{out}(N^{t-1}, N^t)}^{t+1} = 1$. This implies conclusions (1) and (2) of the lemma. It remains to show conclusion (3).

Condition (1) of Definition 5.16 holds if $Y^{t+1} = \max(Y^t, Y^{t-1})$ (see (68)) since $y_{\text{out}(N^{t-1}, N^t)}^{t+1} = 1$ and further, $y_{\text{out}(N^{t-1}, N^t)}$ is the only output that may fire at time t . Conditions (2) and (3) with probability $\geq 1 - (\lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}$ conditioned on $N^t = C, N^{t-1} = C'$ by Lemmas 5.7 and 5.8 and a union bound. Finally, condition (4) holds if $Y^{t+1} = \max(Y^t, Y^{t-1})$ (see (68)). Overall, by a

union bound, all three conclusions hold with probability $\geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}$, giving the lemma. $\square$

We can use Lemma 5.18 to show that $\mathcal{L}_{n,\gamma}$ remains in a valid WTA configuration for t_s consecutive time steps with good probability.

Corollary 5.19 (Stability of Valid WTA Configurations). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any near-stable pair of configurations (C', C) with $C(X) = C'(X) = X^t$. For any time $t \geq 1$, . For any time t conditioned on $N^t = C, N^{t-1} = C'$, with probability $\geq 1 - 3t_s n \cdot e^{-\gamma/2}$, Y^{t+1} is a valid WTA output configuration and further,*

$$Y^{t+1} = Y^{t+2} = \dots = Y^{t+t_s+1}.$$

Proof. We apply Lemma 5.18 for each time $t+1, \dots, t+t_s+1$ in succession. This is possible since by conclusion (3), if (N^{t-1}, N^{t-1}) is a near-stable pair of configurations, then with high probability (N^t, N^{t+1}) is as well. Conclusion (1) gives that $N^{t+1}, \dots, N^{t+t_s+1}$ are all valid WTA configurations and conclusion (2) gives that $Y^{t+1} = Y^{t+2} = \dots = Y^{t+t_s+1}$. By a union bound over these t_s steps, the conclusion holds with probability $\geq 1 - t_s(n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2} \geq 1 - 3t_s n \cdot e^{-\gamma/2}$. $\square$

5.5 Convergence in $O(1)$ Steps

We now use the transition lemmas of Section 5.3 to show that, starting from any configuration, the network converges to a near-stable pair of configurations (Definition 5.16) with constant probability in $O(1)$ time steps. Combined with Corollary 5.19 this shows convergence to a valid WTA configuration (and stability within this configuration for t_s steps) in $O(1)$ steps with constant probability.

Our analysis is tedious by straightforward. It breaks down into nine cases, based on the initial output and inhibitor behavior. These cases are summarized in Table 2. Since some cases depend on our bounds for others, we do not prove them in the order listed.

Output Count $\ Y^t\ _1$	Inhibitor Count $\ A^t\ _1$	Lemma
0	0	Lemma 5.20
0	1	Lemma 5.27
0	any $a > 1$	Lemma 5.28
1	0	Lemma 5.24
1	1	Lemma 5.25
1	any $a > 1$	Lemma 5.26
any $k > 1$	0	Lemma 5.23
any $k > 1$	1	Lemma 5.22
any $k > 1$	any $a > 1$	Lemma 5.21

Table 2: Summary of cases from which we show convergence in $O(1)$ steps to a near-stable pair of configurations (Definition 5.16) with constant probability.

Lemma 5.20 ($\|Y^t\|_1 = \|A^t\|_1 = 0$). Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$ and $\|C(Y)\|_1 = \|C(A)\|_1 = 0$. For any time $t \geq 1$,

$$\mathbb{P}[(N^{t+3}, N^{t+4}) \text{ is near-stable} \mid N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - 12n \cdot e^{-\gamma/2}.$$

Proof. The proof follows from a series of four steps, arguing about the state of $\mathcal{L}_{n,\gamma}$ at times $t+1, t+2, t+3, t+4$.

Step 1:

Let $\mathcal{E}_1$ be the event that $y_i^{t+1} = x_i^{t+1}$ for all i and that $a_i^{t+1} = 0$ for all $i \in \{1, \dots, \lceil \log_2 n \rceil\}$. By Lemma 5.15, since $\|C(A)\|_1 = 0$,

$$\mathbb{P}[y_i^{t+1} = x_i^{t+1} \text{ for all } i \mid N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}.$$

Additionally, since $\|C(Y)\|_1 = 0$, by Lemma 5.8 conclusion (1),

$$\mathbb{P}[a_i^{t+1} = 0 \text{ for all } i \in \{1, \dots, \lceil \log_2 n \rceil\} \mid N^t = C, N^{t-1} = C'] \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}.$$

Thus, by a union bound we have:

$$\mathbb{P}[\mathcal{E}_1 \mid N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil) \cdot e^{-\gamma/2}. \quad (69)$$

Step 2:

Let $\mathcal{E}_2$ be the event that $y_i^{t+2} = x_i^{t+2}$ for all i and that for $l = \lfloor \log_2 (\|X^t\|_1) \rfloor$, $a_s^{t+2} = a_1^{t+2} = \dots = a_l^{t+2} = 1$ and $a_{l+1}^{t+2} = \dots = a_{\lceil \log_2 n \rceil}^{t+2} = 0$ (if $l = 0$, just $a_s^{t+2} = 1$). Conditioned on $\mathcal{E}_1$, the only inhibitor that possibly fires at time $t+1$ is a_s . We can separately consider the cases when $a_s^{t+1} = 0$ and when $a_s^{t+1} = 1$. By Lemma 5.11,

$$\mathbb{P}[y_i^{t+2} = x_i^{t+2} \text{ for all } i \mid \mathcal{E}_1, a_s^{t+1} = 1, N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}.$$

By Lemma 5.15 we also have

$$\mathbb{P}[y_i^{t+2} = x_i^{t+2} \text{ for all } i \mid \mathcal{E}_1, a_s^{t+1} = 0, N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}.$$

By the law of total probability this gives:

$$\mathbb{P}[y_i^{t+2} = x_i^{t+2} \text{ for all } i \mid \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}. \quad (70)$$

We also apply Lemma 5.8. Conditioned on $\mathcal{E}_1$, for $l = \lfloor \log_2 (\|X^t\|_1) \rfloor$, we have $\|Y^{t+1}\|_1 = \|X^t\|_1 \in [2^l, 2^{l+1})$, which gives that,

$$\begin{aligned} \mathbb{P}[a_1^{t+2} = \dots = a_i^{t+2} = 1 \text{ and } a_{i+1}^{t+2} = \dots = a_{\lceil \log_2 n \rceil}^{t+2} = 0 \mid \mathcal{E}_1, N^t = C, N^{t-1} = C'] \\ \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}. \end{aligned} \quad (71)$$

Similarly, applying Lemma 5.7, since conditioned on $\mathcal{E}_1$, $\|Y^{t+1}\|_1 = \|X^{t+1}\|_1 \geq 1$:

$$\mathbb{P}[a_s^{t+2} = 1 \mid \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq 1 - e^{-\gamma/2}. \quad (72)$$

Combining (70), (71), and (72) we have:

$$\mathbb{P}[\mathcal{E}_2 | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \quad (73)$$

Step 3:

Let $\mathcal{E}_3$ be the event that Y^{t+3} is a valid WTA configuration, and that for $l = \lfloor \log_2 (\|X^t\|_1) \rfloor$, $a_s^{t+3} = a_1^{t+3} = \dots = a_l^{t+3} = 1$ and $a_{l+1}^{t+3} = \dots = a_{\lceil \log_2 n \rceil}^{t+3} = 0$.

If $l = 0$, conditioned on $\mathcal{E}_1, \mathcal{E}_2$, we have $Y^{t+1} = Y^{t+2} = X^t$ and so $\|Y^{t+1}\|_1 = \|Y^{t+2}\|_1 = \|X^t\|_1 = 1$. By the stability property of Lemma 5.11 we thus have:

$$\mathbb{P}[Y^{t+3} \text{ is a valid WTA output configuration} | \mathcal{E}_1, \mathcal{E}_2, N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}.$$

Otherwise, for $l \geq 1$, by Corollary 5.13, since conditioned on $\mathcal{E}_1$ and $\mathcal{E}_2$,

$$\|\min(Y^{t+1}, Y^{t+2})\|_1 = \|X^t\|_1 \in [2^l, 2^{l+1})$$

and $a_s^{t+2} = a_1^{t+2} = \dots = a_l^{t+2} = 1$ and $a_{l+1}^{t+2} = \dots = a_{\lceil \log_2 n \rceil}^{t+2} = 0$,

$$\mathbb{P}[Y^{t+3} \text{ is a valid WTA output configuration} | \mathcal{E}_1, \mathcal{E}_2, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - ne^{-2\gamma}.$$

We can easily bound the probability of $a_s^{t+3} = a_1^{t+3} = \dots = a_l^{t+3} = 1$ and $a_{l+1}^{t+3} = \dots = a_{\lceil \log_2 n \rceil}^{t+3} = 0$ using the same arguments as in (71) and (72), giving, via a union bound:

$$\mathbb{P}[\mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_2, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \quad (74)$$

Step 4:

Finally, let $\mathcal{E}_4$ be the event that $\max(Y^{t+3}, Y^{t+4}) = 1$, $a_s^{t+4} = 1$, $\sum_{j=1}^{\lceil \log_2 n \rceil} a_j^{t+4} = 0$ and $y_i^{t+4} \leq x_i^{t+4}$ for all i . We can check via Definition 5.16 that if $\mathcal{E}_3$ and $\mathcal{E}_4$ occur, then (N^{t+3}, N^{t+4}) is a near-stable pair.

Since conditioned on $\mathcal{E}_3$, $a_s^{t+3} = a_1^{t+3} = \dots = a_l^{t+3} = 1$ and $a_{l+1}^{t+3} = \dots = a_{\lceil \log_2 n \rceil}^{t+3} = 0$, if $l \geq 1$, by Lemma 5.12 conclusion (1),

$$\mathbb{P}[\|Y^{t+4}\|_1 \leq \|Y^{t+3}\|_1 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}. \quad (75)$$

Since conditioned on $\mathcal{E}_3$, $\|Y^{t+3}\|_1 = 1$, this gives $\max(Y^{t+3}, Y^{t+4}) = 1$. If $l = 0$, then we have an identical bound via the stability property of Lemma 5.11.

Again, since conditioned on $\mathcal{E}_3$, $\|Y^{t+3}\|_1 = 1$, by Lemma 5.7,

$$\mathbb{P}[a_s^{t+4} = 1 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'] \geq 1 - e^{-\gamma/2}. \quad (76)$$

By Lemma 5.8, this also gives

$$\mathbb{P}\left[\sum_{j=1}^{\lceil \log_2 n \rceil} a_j^{t+4} = 0 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'\right] \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}. \quad (77)$$

By a union bound using (75), (76), and (77),

$$\mathbb{P}[\mathcal{E}_4 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \quad (78)$$

Completing the proof:

Let $\mathcal{E}$ be the event that (N^{t+3}, N^{t+4}) is near-stable. We can complete the proof by bounding:

$$\begin{aligned} \mathbb{P}[\mathcal{E} | N^t = C, N^{t-1} = C'] &\geq \mathbb{P}[\mathcal{E}_3, \mathcal{E}_4 | N^t = C, N^{t-1} = C'] \\ &\geq \mathbb{P}[\mathcal{E}_4 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'] \\ &\quad \cdot \mathbb{P}[\mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_2, N^t = C, N^{t-1} = C'] \\ &\quad \cdot \mathbb{P}[\mathcal{E}_2 | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \\ &\quad \cdot \mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C']. \end{aligned}$$

We can bound the above terms using (69), (73), (74), and (78) giving:

$$\begin{aligned} \mathbb{P}[\mathcal{E} | N^t = C, N^{t-1} = C'] &\geq \left(1 - (n + \lceil \log_2 n \rceil) \cdot e^{-\gamma/2}\right) \cdot \left(1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}\right) \\ &\quad \cdot \left(\frac{1}{16} - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}\right) \cdot \left(1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}\right) \\ &\geq \frac{1}{16} - 4(n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2} \\ &\geq \frac{1}{16} - 12n \cdot e^{-\gamma/2}. \end{aligned}$$

□

Lemma 5.21 ($\|Y^t\|_1 > 1, \|A^t\|_1 > 1$). Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$, $\|C(Y)\|_1 > 1$, $\|C(A)\|_1 > 1$, $C(a_s) = 1$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,

$$\mathbb{P}[(N^{t+i}, N^{t+i+1}) \text{ is near-stable for some } i \leq 7 | N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 24n \cdot e^{-\gamma/2}.$$

Proof. Again the proof follows from a series of steps, arguing about the state of $\mathcal{L}_{n,\gamma}$ at times $t+1, t+2, \dots$. Let $\mathcal{E}$ be the event that (N^{t+i}, N^{t+i+1}) is near-stable for some $i \leq 7$.

Step 1:

Let $\mathcal{E}_1$ be the event that $y_i^{t+1} \leq y_i^t$ for all i and that for $l = \lfloor \log_2 (\|Y^t\|_1) \rfloor$, $a_s^{t+2} = a_1^{t+2} = \dots = a_l^{t+2} = 1$ and $a_{l+1}^{t+2} = \dots = a_{\lceil \log_2 n \rceil}^{t+2} = 0$ (note that $l \geq 1$ since $\|Y^t\|_1 > 1$). Let $\mathcal{E}_{1,0}$ be the event that $\mathcal{E}_1$ holds and $\|Y^{t+1}\|_1 = 0$. Let $\mathcal{E}_{1,1}$ be the event that $\mathcal{E}_1$ holds and $\|Y^{t+1}\|_1 = 1$. Finally, let $\mathcal{E}_{1,>1}$ be the event that $\mathcal{E}_1$ holds and $\|Y^{t+1}\|_1 > 1$. Applying Lemmas 5.7, 5.8, and 5.12 and a union bound:

$$\mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \quad (79)$$

Step 2:

Let $\mathcal{E}_2$ be the event that $\|Y^{t+2}\|_1 = 0$, that $a_s^{t+2} = 1$ if $\|Y^{t+1}\|_1 > 0$, and that for $l' = \lfloor \log_2(\|Y^{t+1}\|_1) \rfloor$, $a_1^{t+2} = \dots = a_{l'}^{t+2} = 1$ and $a_{l'+1}^{t+2} = \dots = a_{\lfloor \log_2 n \rfloor}^{t+2} = 0$. Since, conditioned on $\mathcal{E}_1$, $\|Y^{t+1}\|_1 \leq \|Y^t\|_1 \leq 2^{l+1}$, $a_s^{t+2} = a_1^{t+2} = \dots = a_l^{t+2} = 1$ and $a_{l+1}^{t+2} = \dots = a_{\lfloor \log_2 n \rfloor}^{t+2} = 0$, by Corollary 5.14 combined with Lemmas 5.7, 5.8:

$$\mathbb{P}[\mathcal{E}_2 | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq \frac{1}{8} - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \quad (80)$$

We now write:

$$\begin{aligned} \mathbb{P}[\mathcal{E} | N^t = C, N^{t-1} = C'] &\geq \min \left(\mathbb{P}[\mathcal{E} | \mathcal{E}_2, \mathcal{E}_{1,0}, N^t = C, N^{t-1} = C'], \right. \\ &\quad \mathbb{P}[\mathcal{E} | \mathcal{E}_2, \mathcal{E}_{1,1}, N^t = C, N^{t-1} = C'], \\ &\quad \left. \mathbb{P}[\mathcal{E} | \mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \right) \\ &\quad \cdot \mathbb{P}[\mathcal{E}_2 | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \cdot \mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \end{aligned}$$

Using (79) and (80) we can bound the above by:

$$\begin{aligned} \mathbb{P}[\mathcal{E} | N^t = C, N^{t-1} = C'] &\geq \min \left(\mathbb{P}[\mathcal{E} | \mathcal{E}_2, \mathcal{E}_{1,0}, N^t = C, N^{t-1} = C'], \right. \\ &\quad \mathbb{P}[\mathcal{E} | \mathcal{E}_2, \mathcal{E}_{1,1}, N^t = C, N^{t-1} = C'], \\ &\quad \left. \mathbb{P}[\mathcal{E} | \mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \right) \cdot \left(\frac{1}{8} - 6n \cdot e^{-\gamma/2} \right). \quad (81) \end{aligned}$$

We bound the minimum above by considering each of the three cases separately.

Case 1: $\mathbb{P}[\mathcal{E} | \mathcal{E}_2, \mathcal{E}_{1,0}, N^t = C, N^{t-1} = C']$.

Conditioned on $\mathcal{E}_2$, $\|Y^{t+2}\|_1 = 0$. So by Lemma 5.20,

$$\mathbb{P}[(N^{t+5}, N^{t+6}) \text{ is near-stable} | \mathcal{E}_2, \mathcal{E}_{1,0}, \|A^{t+2}\|_1 = 0, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - 12n \cdot e^{-\gamma/2}. \quad (82)$$

If $\|A^{t+2}\|_1 \geq 1$ then, conditioned on $\mathcal{E}_2$, we must have $a_s^{t+2} = 1$. Let $\mathcal{E}_3$ be the event that $\|Y^{t+3}\|_1 = \|A^{t+3}\|_1 = 0$. By Lemmas 5.7, 5.8, and 5.11 and the fact that conditioned on $\mathcal{E}_{1,0}$, $\|Y^{t+1}\|_1 = 0$,

$$\mathbb{P}[\mathcal{E}_3 | \mathcal{E}_2, \mathcal{E}_{1,0}, \|A^{t+2}\|_1 \geq 1, N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2} \quad (83)$$

Again by Lemma 5.20,

$$\mathbb{P}[(N^{t+6}, N^{t+7}) \text{ is near-stable} | \mathcal{E}_3, \mathcal{E}_2, \mathcal{E}_{1,0}, \|A^{t+2}\|_1 \geq 1, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - 12n \cdot e^{-\gamma/2}.$$

Combined with (83) this gives

$$\mathbb{P}[(N^{t+6}, N^{t+7}) \text{ is near-stable} | \mathcal{E}_2, \mathcal{E}_{1,0}, \|A^{t+2}\|_1 \geq 1, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - 15n \cdot e^{-\gamma/2}.$$

By the law of total probability, combined with (82) we have

$$\mathbb{P}[\mathcal{E} | \mathcal{E}_2, \mathcal{E}_{1,0}, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - 15n \cdot e^{-\gamma/2} \quad (84)$$

which completes this case.

Case 2: $\mathbb{P}[\mathcal{E}|\mathcal{E}_2, \mathcal{E}_{1,1}, N^t = C, N^{t-1} = C']$.

In this case, conditioned on $\mathcal{E}_{1,1}$ and $\mathcal{E}_2$, $a_s^{t+1} = a_s^{t+2} = 1$, a_s is the only inhibitor that fires at time $t+2$, $\|Y^{t+1}\|_1 = 1$, and $\|Y^{t+2}\|_1 = 0$. Thus, (N^{t+1}, N^{t+2}) is a near-stable pair of configurations, and so vacuously,

$$\mathbb{P}[\mathcal{E}|\mathcal{E}_2, \mathcal{E}_{1,0}, N^t = C, N^{t-1} = C'] = 1 \quad (85)$$

which completes this case.

Case 3: $\mathbb{P}[\mathcal{E}|\mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C']$.

In this case, conditioned on $\mathcal{E}_{1,>1}$ and $\mathcal{E}_2$, $\|A^{t+2}\|_1 > 1$ and $a_s^{t+2} = 1$. Let $\mathcal{E}_3$ be the event that $\|Y^{t+3}\|_1 = 0$ and that if $\|A^{t+3}\|_1 \geq 1$, $a_s^{t+3} = 1$. By Lemmas 5.7, 5.8, and 5.12,

$$\mathbb{P}[\mathcal{E}_3|\mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1)e^{-\gamma/2}. \quad (86)$$

In the case that $\|A^{t+3}\|_1 = 0$, we can again apply Lemma 5.20 to give:

$$\begin{aligned} \mathbb{P}[(N^{t+6}, N^{t+7}) \text{ is near-stable} | \mathcal{E}_3, \|A^{t+3}\|_1 = 0, \mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \\ \geq \frac{1}{16} - 12n \cdot e^{-\gamma/2}. \end{aligned} \quad (87)$$

In the case that $\|A^{t+3}\|_1 \geq 1$, let $\mathcal{E}_4$ be the event that $\|Y^{t+4}\|_1 = \|A^{t+4}\|_1 = 0$. We have by Lemmas 5.7, 5.8 and 5.11,

$$\mathbb{P}[\mathcal{E}_4|\mathcal{E}_3, \|A^{t+3}\|_1 \geq 1, \mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1)e^{-\gamma/2}. \quad (88)$$

Further, again by Lemma 5.20 we have:

$$\begin{aligned} \mathbb{P}[(N^{t+7}, N^{t+8}) \text{ is near-stable} | \mathcal{E}_4, \mathcal{E}_3, \|A^{t+3}\|_1 = 0, \mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \\ \geq \frac{1}{16} - 12n \cdot e^{-\gamma/2}. \end{aligned}$$

Combined with (88) this gives:

$$\begin{aligned} \mathbb{P}[(N^{t+7}, N^{t+8}) \text{ is near-stable} | \mathcal{E}_3, \|A^{t+3}\|_1 = 0, \mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \\ \geq \frac{1}{16} - 15n \cdot e^{-\gamma/2}. \end{aligned}$$

Further, combined with (87), by the law of total probability,

$$\mathbb{P}[\mathcal{E}|\mathcal{E}_3, \mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - 15n \cdot e^{-\gamma/2}. \quad (89)$$

Finally, combining (89) with (86) we have:

$$\mathbb{P}[\mathcal{E}|\mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - 18n \cdot e^{-\gamma/2} \quad (90)$$

completing this case.

Completing the proof.

Using equations (84), (85), and (90) in the three cases above along (81):

$$\begin{aligned}
\mathbb{P}[\mathcal{E}|N^t = C, N^{t-1} = C'] &\geq \min \left(\mathbb{P}[\mathcal{E}|\mathcal{E}_2, \mathcal{E}_{1,0}, N^t = C, N^{t-1} = C'], \right. \\
&\quad \mathbb{P}[\mathcal{E}|\mathcal{E}_2, \mathcal{E}_{1,1}, N^t = C, N^{t-1} = C'], \\
&\quad \left. \mathbb{P}[\mathcal{E}|\mathcal{E}_2, \mathcal{E}_{1,>1}, N^t = C, N^{t-1} = C'] \right) \cdot \left(\frac{1}{8} - 6n \cdot e^{-\gamma/2} \right) \\
&\geq \left(\frac{1}{16} - 18n \cdot e^{-\gamma/2} \right) \cdot \left(\frac{1}{8} - 6n \cdot e^{-\gamma/2} \right) \\
&\geq \frac{1}{128} - 24n \cdot e^{-\gamma/2},
\end{aligned}$$

completing the lemma. $\square$

Using Lemma 5.21 it is not hard to complete the cases when $\|Y^t\|_1 > 1$ and $\|A^t\|_1 \leq 1$.

Lemma 5.22 ($\|Y^t\|_1 > 1, \|A^t\|_1 = 1$). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$, $\|C(Y)\|_1 > 1$, $\|C(A)\|_1 = 1$, $C(a_s) = 1$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,*

$$\mathbb{P}[(N^{t+i}, N^{t+i+1}) \text{ is near-stable for some } i \leq 8 | N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 28n \cdot e^{-\gamma/2}.$$

Proof. Let $\mathcal{E}$ be the event that (N^{t+i}, N^{t+i+1}) is near-stable for some $i \leq 8$. Let $\mathcal{E}_1$ be the event that $y_i^{t+1} \geq y_i^t$ and $y_i^{t+1} \leq x_i^{t+1}$ for all i , that $a_s^{t+1} = 1$, and that $\|A^{t+1}\|_1 > 1$. By Corollary 5.6 and Lemmas 5.7, 5.8, and 5.11, we have:

$$\mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \geq 1 - (2n + \lceil \log_2 n \rceil + 1)e^{-\gamma/2}.$$

Further, by Lemma 5.22 since conditioned on $\mathcal{E}_1$, $\|Y^{t+1}\|_1 > 1$ and $\|A^{t+1}\|_1 > 1$ with $a_s^{t+1} = 1$:

$$\mathbb{P}[\mathcal{E} | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 24n \cdot e^{-\gamma/2}.$$

This gives the lemma since:

$$\begin{aligned}
\mathbb{P}[\mathcal{E} | N^t = C, N^{t-1} = C'] &\geq \mathbb{P}[\mathcal{E} | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \cdot \mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \\
&\geq \frac{1}{128} - 28n \cdot e^{-\gamma/2}.
\end{aligned}$$

$\square$

Lemma 5.23 ($\|Y^t\|_1 > 1, \|A^t\|_1 = 0$). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$, $\|C(Y)\|_1 > 1$, $\|C(A)\|_1 = 0$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,*

$$\mathbb{P}[(N^{t+i}, N^{t+i+1}) \text{ is near-stable for some } i \leq 8 | N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 27n \cdot e^{-\gamma/2}.$$

Proof. Let $\mathcal{E}$ be the event that (N^{t+i}, N^{t+i+1}) is near-stable for some $i \leq 8$. Let $\mathcal{E}_1$ be the event that $y_i^{t+1} = x_i^{t+1}$ for all i , that $a_s^{t+1} = 1$, and that $\|A^{t+1}\|_1 > 1$. Note that since $\|C(Y)\|_1 > 1$ and $C(y_i) \leq C(x_i)$ for all i , $\mathcal{E}_1$ implies that $\|Y^{t+1}\|_1 = \|X^{t+1}\|_1 > 1$. By Lemmas 5.7, 5.8, and 5.15:

$$\mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1)e^{-\gamma/2}.$$

Further, by Lemma 5.22 since conditioned on $\mathcal{E}_1$, $\|Y^{t+1}\|_1 > 1$ and $\|A^{t+1}\|_1 > 1$ with $a_s^{t+1} = 1$:

$$\mathbb{P}[\mathcal{E} | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 24n \cdot e^{-\gamma/2}.$$

This gives the lemma since:

$$\begin{aligned} \mathbb{P}[\mathcal{E} | N^t = C, N^{t-1} = C'] &\geq \mathbb{P}[\mathcal{E} | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \cdot \mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \\ &\geq \frac{1}{128} - 27n \cdot e^{-\gamma/2}. \end{aligned}$$

□

We next complete the remaining cases when $\|Y^t\|_1 \leq 1$.

Lemma 5.24 ($\|Y^t\|_1 = 1, \|A^t\|_1 = 0$). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$, $\|C(Y)\|_1 = 1$, $\|C(A)\|_1 = 0$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,*

$$\mathbb{P}[(N^{t+3}, N^{t+4}) \text{ is near-stable} | N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - 12n \cdot e^{-\gamma/2}.$$

Proof. The proof follows from a series of four steps, arguing about the state of $\mathcal{L}_{n,\gamma}$ at times $t+1, t+2, t+3, t+4$. The analysis closely mirrors that of Lemma 5.20, for the case when $\|Y^t\|_1 = 0$ and $\|A^t\|_1 = 0$.

Step 1:

Let $\mathcal{E}_1$ be the event that $y_i^{t+1} = x_i^{t+1}$ for all i , that $a_i^{t+1} = 0$ for all $i \in \{1, \dots, \lceil \log_2 n \rceil\}$, and that $a_s^{t+1} = 1$. By Lemma 5.15, since $\|C(A)\|_1 = 0$,

$$\mathbb{P}[y_i^{t+1} = x_i^{t+1} \text{ for all } i | N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}.$$

Additionally, since $\|C(Y)\|_1 = 1$, by Lemma 5.8 conclusion (1),

$$\mathbb{P}[a_i^{t+1} = 0 \text{ for all } i \in \{1, \dots, \lceil \log_2 n \rceil\} | N^t = C, N^{t-1} = C'] \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}.$$

Finally, by Lemma 5.7,

$$\mathbb{P}[a_s^{t+1} = 1 | N^t = C, N^{t-1} = C'] \geq 1 - e^{-\gamma/2}.$$

Thus, by a union bound we have:

$$\mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \quad (91)$$

Step 2:

Let $\mathcal{E}_2$ be the event that $y_i^{t+2} = x_i^{t+2}$ for all i and that for $l = \lfloor \log_2(\|X^t\|_1) \rfloor$, $a_s^{t+2} = a_1^{t+2} = \dots = a_l^{t+2} = 1$ and $a_{l+1}^{t+2} = \dots = a_{\lceil \log_2 n \rceil}^{t+2} = 0$ (if $l = 0$, just $a_s^{t+2} = 1$). Conditioned on $\mathcal{E}_1$, the only inhibitor that fires at time $t+1$ is a_s . We can separately consider the cases when $a_s^{t+1} = 0$ and when $a_s^{t+1} = 1$. By Lemma 5.11 and the fact that $y_i^{t+1} = x_i^{t+1}$ for all i ,

$$\mathbb{P}[y_i^{t+2} = x_i^{t+2} \text{ for all } i | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}. \quad (92)$$

We also apply Lemma 5.8. Conditioned on $\mathcal{E}_1$, for $l = \lfloor \log_2(\|X^t\|_1) \rfloor$, we have $\|Y^{t+1}\|_1 = \|X^t\|_1 \in [2^l, 2^{l+1})$, which gives that,

$$\begin{aligned} \mathbb{P}[a_1^{t+2} = \dots = a_i^{t+2} = 1 \text{ and } a_{i+1}^{t+2} = \dots = a_{\lceil \log_2 n \rceil}^{t+2} = 0 | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \\ \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}. \end{aligned} \quad (93)$$

Similarly, applying Lemma 5.7, since conditioned on $\mathcal{E}_1$, $\|Y^{t+1}\|_1 = \|X^{t+1}\|_1 \geq 1$:

$$\mathbb{P}[a_s^{t+2} = 1 | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq 1 - e^{-\gamma/2}. \quad (94)$$

Combining (92), (93), and (94) we have:

$$\mathbb{P}[\mathcal{E}_2 | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \quad (95)$$

Step 3:

Let $\mathcal{E}_3$ be the event that Y^{t+3} is a valid WTA configuration, and that for $l = \lfloor \log_2(\|X^t\|_1) \rfloor$, $a_s^{t+3} = a_1^{t+3} = \dots = a_l^{t+3} = 1$ and $a_{l+1}^{t+3} = \dots = a_{\lceil \log_2 n \rceil}^{t+3} = 0$.

If $l = 0$, conditioned on $\mathcal{E}_1, \mathcal{E}_2$, we have $\|Y^{t+1}\|_1 = \|Y^{t+2}\|_1 = 1$ and $Y^{t+1} = Y^{t+2} = X^t$. By the stability property of Lemma 5.11 we thus have:

$$\mathbb{P}[Y^{t+3} \text{ is a valid WTA output configuration} | \mathcal{E}_1, \mathcal{E}_2, N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}.$$

Otherwise, for $l \geq 1$, by Corollary 5.13, since conditioned on $\mathcal{E}_1$ and $\mathcal{E}_2$,

$$\|\min(Y^{t+1}, Y^{t+2})\|_1 = \|X^t\|_1 \in [2^l, 2^{l+1})$$

and $a_s^{t+2} = a_1^{t+2} = \dots = a_l^{t+2} = 1$ and $a_{l+1}^{t+2} = \dots = a_{\lceil \log_2 n \rceil}^{t+2} = 0$,

$$\mathbb{P}[Y^{t+3} \text{ is a valid WTA output configuration} | \mathcal{E}_1, \mathcal{E}_2, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - ne^{-2\gamma}.$$

We can easily bound the probability of $a_s^{t+3} = a_1^{t+3} = \dots = a_l^{t+3} = 1$ and $a_{l+1}^{t+3} = \dots = a_{\lceil \log_2 n \rceil}^{t+3} = 0$ using the same arguments as in (93) and (94), giving, via a union bound:

$$\mathbb{P}[\mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_2, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \quad (96)$$

Step 4:

Finally, let $\mathcal{E}_4$ be the event that $\max(Y^{t+3}, Y^{t+4}) = 1$, $a_s^{t+4} = 1$, $\sum_{j=1}^{\lceil \log_2 n \rceil} a_j^{t+4} = 0$ and $y_i^{t+4} \leq x_i^{t+4}$ for all i . We can check via Definition 5.16 that if $\mathcal{E}_3$ and $\mathcal{E}_4$ occur, then (N^{t+3}, N^{t+4}) is a near-stable pair.

Since conditioned on $\mathcal{E}_3$, $a_s^{t+3} = a_1^{t+3} = \dots = a_l^{t+3} = 1$ and $a_{l+1}^{t+3} = \dots = a_{\lceil \log_2 n \rceil}^{t+3} = 0$, if $l \geq 1$, by Lemma 5.12 conclusion (1),

$$\mathbb{P}[\|Y^{t+4}\|_1 \leq \|Y^{t+3}\|_1 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}. \quad (97)$$

Since conditioned on $\mathcal{E}_3$, $\|Y^{t+3}\|_1 = 1$, this gives $\max(Y^{t+3}, Y^{t+4}) = 1$. If $l = 0$, then we have an identical bound via the stability property of Lemma 5.11.

Again, since conditioned on $\mathcal{E}_3$, $\|Y^{t+3}\|_1 = 1$, by Lemma 5.7,

$$\mathbb{P}[a_s^{t+4} = 1 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'] \geq 1 - e^{-\gamma/2}. \quad (98)$$

By Lemma 5.8, this also gives

$$\mathbb{P}\left[\sum_{j=1}^{\lceil \log_2 n \rceil} a_j^{t+4} = 0 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'\right] \geq 1 - \lceil \log_2 n \rceil \cdot e^{-\gamma/2}. \quad (99)$$

By a union bound using (97), (98), and (99),

$$\mathbb{P}[\mathcal{E}_4 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}. \quad (100)$$

Completing the proof:

Let $\mathcal{E}$ be the event that (N^{t+3}, N^{t+4}) is near-stable. We can complete the proof by bounding:

$$\begin{aligned} \mathbb{P}[\mathcal{E} | N^t = C, N^{t-1} = C'] &\geq \mathbb{P}[\mathcal{E}_3, \mathcal{E}_4 | N^t = C, N^{t-1} = C'] \\ &\geq \mathbb{P}[\mathcal{E}_4 | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, N^t = C, N^{t-1} = C'] \\ &\quad \cdot \mathbb{P}[\mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_2, N^t = C, N^{t-1} = C'] \\ &\quad \cdot \mathbb{P}[\mathcal{E}_2 | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \\ &\quad \cdot \mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \end{aligned}$$

We can bound the above terms using (91), (95), (96), and (100) giving:

$$\begin{aligned} \mathbb{P}[\mathcal{E} | N^t = C, N^{t-1} = C'] &\geq \left(1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}\right) \cdot \left(1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}\right) \\ &\quad \cdot \left(\frac{1}{16} - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}\right) \cdot \left(1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}\right) \\ &\geq \frac{1}{16} - 4(n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2} \\ &\geq \frac{1}{16} - 12n \cdot e^{-\gamma/2}. \end{aligned}$$

□

The remaining cases follow relatively straightforwardly from the previous lemmas.

Lemma 5.25 ($\|Y^t\|_1 = 1$, $\|A^t\|_1 = 1$). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$, $\|C(Y)\|_1 = 1$, $\|C(A)\|_1 = 1$, $C(a_s) = 1$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,*

$$\mathbb{P}[(N^{t+i}, N^{t+i+1}) \text{ is near-stable for some } i \leq 9 | N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 28n \cdot e^{-\gamma/2}.$$

Proof. Let $\mathcal{E}$ be the event that (N^{t+i}, N^{t+i+1}) is near-stable for some $i \leq 9$. Let $\mathcal{E}_1$ be the event that $y_i^{t+1} \geq y_i^t$ and $y_i^{t+1} \leq x_i^{t+1}$ for all i , that $a_s^{t+1} = 1$, and that $\|A^{t+1}\|_1 > 1$. By Corollary 5.6 and Lemmas 5.7, 5.8, and 5.11, we have:

$$\mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \geq 1 - (2n + \lceil \log_2 n \rceil + 1)e^{-\gamma/2}. \quad (101)$$

We consider two cases, dependent on the number of firing outputs at time $t+1$. Conditioned on $\mathcal{E}_1$, $\|Y^{t+1}\|_1 \geq \|Y^t\|_1 = 1$. By Lemma 5.22 since conditioned on $\mathcal{E}_1$, $\|A^{t+1}\|_1 > 1$ with $a_s^{t+1} = 1$:

$$\mathbb{P}[\mathcal{E} | \mathcal{E}_1, \|Y^{t+1}\| > 1, N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 24n \cdot e^{-\gamma/2}.$$

Alternatively, if $\|Y^{t+1}\| = 1$, then (N^t, N^{t+1}) is already a near-stable pair of configurations (Definition 5.16) so we vacuously have

$$\mathbb{P}[\mathcal{E} | \mathcal{E}_1, \|Y^{t+1}\| = 1, N^t = C, N^{t-1} = C'] = 1.$$

By the law of total probability this gives:

$$\mathbb{P}[\mathcal{E} | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 24n \cdot e^{-\gamma/2}.$$

Finally, we obtain the lemma by combining this bound with (101). $\square$

Lemma 5.26 ($\|Y^t\|_1 = 1, \|A^t\|_1 > 1$). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$, $\|C(Y)\|_1 = 1$, $\|C(A)\|_1 > 1$, $C(a_s) = 1$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,*

$$\mathbb{P}[(N^t, N^{t+1}) \text{ is near-stable} | N^t = C, N^{t-1} = C'] \geq 1 - 3n \cdot e^{-\gamma/2}.$$

Proof. Let $\mathcal{E}_1$ be the event that $\|A^{t+1}\|_1 = 1$ and $a_s^{t+1} = 1$. Again Lemmas 5.7 and 5.8 we have:

$$\mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \geq 1 - (\lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2}.$$

Let $\mathcal{E}_2$ be the event that $\|Y^{t+1}\| \leq 1$ and that $y_i^{t+1} \leq x_i^{t+1}$ for all i . By Lemma 5.12,

$$\mathbb{P}[\mathcal{E}_2 | N^t = C, N^{t-1} = C'] \geq 1 - ne^{-\gamma/2}.$$

If $\mathcal{E}_1$ and $\mathcal{E}_2$ both occur, (N^t, N^{t+1}) is a near-stable pair of configurations, giving the lemma by a union bound and the fact that $(n + \lceil \log_2 n \rceil + 1) \leq 3n$. $\square$

Lemma 5.27 ($\|Y^t\|_1 = 0, \|A^t\|_1 = 1$). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$, $\|C(Y)\|_1 = 0$, $\|C(A)\|_1 = 1$, $C(a_s) = 1$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,*

$$\mathbb{P}[(N^{t+i}, N^{t+i+1}) \text{ is near-stable for some } i \leq 10 | N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 31ne^{-\gamma/2}.$$

Proof. Let $\mathcal{E}$ be the event that (N^{t+i}, N^{t+i+1}) is near-stable for some $i \leq 10$. Let $\mathcal{E}_1$ be the event that $a_s^{t+1} = 1$ if $\|Y^{t-1}\|_1 \geq 1$ and 0 otherwise, that $a_i^{t+1} = 0$ for all $i \in \{1, \dots, \lceil \log_2 n \rceil\}$, and that $y_i^{t+1} = \max(y_i^{t-1}, y_i^t)$ for all i . By Lemmas 5.7, 5.8, and 5.11

$$\mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2} \quad (102)$$

We next consider two cases, based off the number of firing outputs at time $t - 1$.

Case 1: $\|C'(Y)\|_1 = 0$.

In this case, conditioned on $\mathcal{E}_1$, $\|Y^{t+1}\|_1 = \|A^{t+1}\|_1 = 0$. Thus, by Lemma 5.20,

$$\mathbb{P}[(N^{t+4}, N^{t+5}) \text{ is near-stable} | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq \frac{1}{16} - 12n \cdot e^{-\gamma/2} \quad (103)$$

Case 2: $\|C'(Y)\|_1 \geq 1$.

In this case, conditioned on $\mathcal{E}_1$, $a_s^{t+1} = 1$, $\|A^{t+1}\|_1 = 1$ and $\|Y^{t+1}\|_1 \geq 1$. Thus by Lemmas 5.22 and 5.25,

$$\mathbb{P}[\mathcal{E} | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 28n \cdot e^{-\gamma/2} \quad (104)$$

Overall by (103) and (104) we have

$$\mathbb{P}[\mathcal{E} | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 28n \cdot e^{-\gamma/2},$$

which gives the lemma when combined with (102). $\square$

Lemma 5.28 ($\|Y^t\|_1 = 0$, $\|A^t\|_1 > 1$). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$, $\|C(Y)\|_1 = 0$, $\|C(A)\|_1 > 1$, $C(a_s) = 1$, and $C(y_i) \leq C(x_i)$, $C'(y_i) \leq C'(x_i)$ for all i . For any time $t \geq 1$,*

$$\mathbb{P}[(N^{t+i}, N^{t+i+1}) \text{ is near-stable for some } i \leq 11 | N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 33ne^{-\gamma/2}.$$

Proof. Let $\mathcal{E}$ be the event that (N^{t+i}, N^{t+i+1}) is near-stable for some $i \leq 11$. Let $\mathcal{E}_1$ be the event that $a_i^{t+1} = 0$ for all $i \in \{1, \dots, \lceil \log_2 n \rceil\}$, and that $\|Y^{t+1}\|_1 = 0$. By Lemmas 5.8 and 5.12,

$$\mathbb{P}[\mathcal{E}_1 | N^t = C, N^{t-1} = C'] \geq 1 - (n + \lceil \log_2 n \rceil) \cdot e^{-\gamma/2}. \quad (105)$$

Further, by Lemmas 5.20 and 5.27,

$$\mathbb{P}[\mathcal{E} | \mathcal{E}_1, N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 31n \cdot e^{-\gamma/2}.$$

This gives the lemma when combined with (105). $\square$

5.6 Completing the Analysis

By combining the nine cases of Lemmas 5.20, 5.21, 5.22, 5.23, 5.24, 5.25, 5.26, 5.27, and 5.28 we can conclude the following general statement:

Lemma 5.29 (*$O(1)$ Step Convergence From Typical Configurations*). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of typical configurations (Definition 5.9) C', C with $C(X) = C'(X) = X^t$. For any time $t \geq 1$,*

$$\mathbb{P}[(N^{t+i}, N^{t+i+1}) \text{ is near-stable for some } i \leq 11 | N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 33ne^{-\gamma/2}.$$

Proof. The nine cases of the above listed lemmas cover all possible pairs of typical configurations C, C' . So the lemma follows immediately. $\square$

From Lemma 5.29 we can conclude an even more general result:

Lemma 5.30 (*General $O(1)$ Step Convergence*). *Assume the input execution α_X of $\mathcal{L}_{n,\gamma}$ has X^t fixed for all t and that $\|X^t\|_1 \geq 1$. Consider any pair of configurations C', C with $C(X) = C'(X) = X^t$. For any time $t \geq 1$,*

$$\mathbb{P}[(N^{t+i}, N^{t+i+1}) \text{ is near-stable for some } i \leq 13 | N^t = C, N^{t-1} = C'] \geq \frac{1}{128} - 39ne^{-\gamma/2}.$$

Proof. By Corollary 5.10, for any C', C ,

$$\begin{aligned} \mathbb{P}[N^{t+1}, N^{t+2} \text{ are typical} | N^t = C, N^{t-1} = C'] &\geq 1 - 2(n + \lceil \log_2 n \rceil + 1) \cdot e^{-\gamma/2} \\ &\geq 1 - 6n \cdot e^{-\gamma/2}. \end{aligned}$$

The lemma then follows by combining this bound with Lemma 5.29. $\square$

Finally, using Lemma 5.30, we can prove the two main theorems of this section. The proofs are similar to those of the main theorems for the two-inhibitor network in Section 3.7.

Theorem 5.2 (*$O(\log n)$ -Inhibitor WTA*). *For $\gamma \geq 12 \ln(39t_s n/\delta)$, $\mathcal{L}_{n,\gamma}$ solves WTA(n, t_c, t_s, δ) for any $t_c \geq 2086(\log_2(1/\delta) + 1)$. $\mathcal{L}_{n,\gamma}$ contains $\lceil \log_2 n \rceil + 1$ auxiliary inhibitors.*

Proof. Consider $\mathcal{L}_{n,\gamma}$ starting with any initial configurations N^0, N^1 and given an infinite input execution α_X with X^t fixed for all t . We consider two cases:

Case 1: $\|X^t\|_1 \geq 1$.

Let $\Delta = 15$ and $r = 139(\log_2(1/\delta) + 1)$. Let $\mathcal{E}$ be the event that there is some time $t \leq t_c$ where (N^t, N^{t+1}) is a near-stable pair of configurations.

For any $i \geq 0$, let $\mathcal{E}_i$ be the event that there is some time $t \in \{i\Delta + 1, \dots, (i+1)\Delta - 1\}$ where (N^t, N^{t+1}) is a near-stable pair of configurations. By Lemma 5.30 we have:

$$\mathbb{P}[\mathcal{E}_i | N^{i\Delta}] \geq \frac{1}{128} - 39n \cdot e^{-\gamma/2} \geq \frac{1}{200}$$

by the requirement that $\gamma > 12 \ln(39t_s n / \delta)$. Let $Z_0, \dots, Z_{r-1} \in \{0, 1\}$ be independent coin flips, with $\mathbb{P}[Z_i = 1] = 1/200$. Applying Lemma 2.3:

$$\mathbb{P}[\mathcal{E}] = \mathbb{P}\left[\bigcap_{i=0}^{r-1} \mathcal{E}_i\right] \geq \mathbb{P}\left[\sum_{i=0}^{r-1} Z_i \geq 1\right] = 1 - \left(\frac{199}{200}\right)^r.$$

Using that $r = 139(\log_2(1/\delta) + 1)$ and that $\left(\frac{199}{200}\right)^{139} \leq \frac{1}{2}$:

$$\mathbb{P}[\mathcal{E}] \geq 1 - \left(\frac{199}{200}\right)^{139(\log_2(1/\delta)+1)} \geq 1 - \frac{\delta}{2}. \quad (106)$$

Thus, with probability $\geq 1 - \frac{\delta}{2}$ there is some time $t \leq r \cdot \Delta - 1 \leq 2085 \cdot (\log_2(1/\delta) + 1) - 1 \leq t_c - 2$ in which (N^t, N^{t+1}) is a near-stable pair of configurations. By Corollary 5.19, if (N^t, N^{t+1}) is a near-stable pair, with probability $\geq 1 - 3t_s n \cdot e^{-\gamma/2}$, N^{t+2} is a valid WTA configuration and:

$$Y^{t+2} = Y^{t+3} = \dots = Y^{t+2+t_s}.$$

By the requirement that $\gamma \geq 12 \ln(39t_s n / \delta)$ and (106) we thus have that the network reaches a valid WTA configuration within time t_c and remains in it for time t_s with probability at least:

$$\left(1 - \frac{\delta}{2}\right) \cdot \left(1 - \frac{\delta}{e^6}\right) \geq 1 - \delta,$$

yielding the theorem in this case.

Case 0: $\|X^t\|_1 = 0$

In this case, by Corollary 5.6 and a union bound, for any configurations C, C' with $C(X) = C'(X) = X^t$:

$$\begin{aligned} \mathbb{P}[\|Y^2\|_1 = \dots = \|Y^{2+t_s}\|_1 = 0 | N^1 = C, N^0 = C'] &\geq 1 - 3t_s n \cdot e^{-\gamma/2} \\ &\geq 1 - \delta \end{aligned}$$

by the requirement that $\gamma \geq 12 \ln(39t_s n / \delta)$. This easily gives the theorem in this case since when $\|X^t\|_1 = 0$, $\|Y^t\|_1 = 0$ is a valid WTA output configuration. □

We now use a similar argument to show with what parameters $\mathcal{L}_{n,\gamma}$ solves the expected-time WTA problem of Definition 2.9.

Theorem 5.3 (*$O(\log n)$ -Inhibitor Expected-Time WTA*). *For $\gamma \geq 12 \ln(39t_s n)$, $\mathcal{L}_{n,\gamma}$ solves WTA-EXP(n, t_c, t_s) for any $t_c \geq 4001$. $\mathcal{L}_{n,\gamma}$ contains $\lceil \log_2 n \rceil + 1$ auxiliary inhibitors.*

Proof. Our proof closely follows that of Theorem 3.3. Recall that in Definition 2.9 we defined the convergence time for any infinite input execution α_X and output execution α_Y :

$$t(\alpha_X, t_s, \alpha_Y) = \min \{t : Y^t \text{ is a valid WTA output configuration for } X^t \text{ and } Y^t = \dots = Y^{t+t_s}\}.$$

Define the worst case expected convergence time of $\mathcal{L}_{n,\gamma}$ on input α_X by:

$$t_{max}(\alpha_X) = \max_{\alpha^1 = N^0 N^1} \left(\mathbb{E}_{\alpha_Y \sim \mathcal{D}_Y(\mathcal{L}_{n,\gamma}, \alpha^1, \alpha_X)} t(\alpha_X, t_s, \alpha_Y) \right).$$

To prove the lemma we must prove that for any α_X with X^t fixed for all t , $t_{max}(\alpha_X) \leq 16$. Fixing such an α_X , for any initial configuration $\alpha^1 = N^0 N^1$, let $\mathbb{E}_{\alpha^1}$ and $\mathbb{P}_{\alpha^1}$ denote the expectation and probability of an event taken over executions drawn from $\mathcal{D}(\mathcal{L}_{n,\gamma}, \alpha^1, \alpha_X)$.

We consider the case when α_X has $\|X^t\|_1 \geq 1$. The case when $\|X^t\|_1 = 0$ follows easily from a similar proof using Corollary 5.6.

Let $\Delta = 15$ and let $\mathcal{E}_1$ be the event that there is some $t \in \{1, \dots, \Delta - 1\}$ where (N^t, N^{t+1}) is a near-stable pair of configurations. Let $\mathcal{E}_{stab}$ be the event that there is some $t \in \{1, \dots, \Delta - 1\}$ where (N^t, N^{t+1}) is a near-stable pair of configurations and additionally, where N^{t+2} is a valid WTA output configuration with $N^{t+2} = \dots = N^{t+2+t_s}$. Let $\bar{\mathcal{E}}_1$ and $\bar{\mathcal{E}}_{stab}$ be the complements of these two events. By Lemma 5.30, for any initial $\alpha^1 = N^0 N^1$:

$$\mathbb{P}_{\alpha^1}[\mathcal{E}_1] \geq \frac{1}{128} - 39ne^{-\gamma/2} \geq \frac{1}{200}. \quad (107)$$

by the requirement that $\gamma \geq 12 \ln(39t_s n)$. Further, by Corollary 5.19, if (C', C) is a near-stable pair of configurations, then

$$\begin{aligned} \mathbb{P}_{\alpha^1}[N^{t+1} \text{ is a valid WTA output config. and } N^{t+2} = \dots = N^{t+2+t_s} | N^t = C', N^{t+1} = C] \\ \geq 1 - 3t_s n e^{-\gamma/2} \\ \geq 1 - \frac{1}{1000t_s} \end{aligned} \quad (108)$$

where the bound holds since $\gamma \geq 12 \ln(39t_s n) \geq 6 \ln(39t_s^2 n)$ and so $e^{-\gamma/2} \leq \frac{1}{(39nt_s^2)^6} \leq \frac{1}{1000nt_s^2}$. Together (107) and (108) give that:

$$\mathbb{P}_{\alpha^1}[\mathcal{E}_{stab}] \geq \mathbb{P}_{\alpha^1}[\mathcal{E}_{stab} | \mathcal{E}_1] \cdot \mathbb{P}_{\alpha^1}[\mathcal{E}_1] \geq \frac{1}{200} \cdot \left(1 - \frac{1}{1000t_s}\right) \geq \frac{1}{200} - \frac{1}{1000t_s}.$$

We can write:

$$\begin{aligned} \mathbb{E}_{\alpha^1}[t(\alpha_X, t_s, \alpha_Y)] &= \mathbb{E}_{\alpha^1}[t(\alpha_X, t_s, \alpha_Y) | \mathcal{E}_{stab}] \cdot \mathbb{P}_{\alpha^1}[\mathcal{E}_{stab}] \\ &\quad + \mathbb{E}_{\alpha^1}[t(\alpha_X, t_s, \alpha_Y) | \mathcal{E}_1, \bar{\mathcal{E}}_{stab}] \cdot \mathbb{P}_{\alpha^1}[\mathcal{E}_1, \bar{\mathcal{E}}_{stab}] \\ &\quad + \mathbb{E}_{\alpha^1}[t(\alpha_X, t_s, \alpha_Y) | \bar{\mathcal{E}}_1] \cdot \mathbb{P}_{\alpha^1}[\bar{\mathcal{E}}_1] \end{aligned} \quad (109)$$

Conditioned on $\mathcal{E}_{stab}$ (which also requires that $\mathcal{E}_1$ occurs), the network reaches a near-stable pair of configuration within Δ steps, reaches a valid WTA output configuration within $\Delta + 1$ steps, and stabilizes for t_s steps. Thus, we have:

$$\mathbb{E}_{\alpha^1}[t(\alpha_X, t_s, \alpha_Y) | \mathcal{E}_{stab}] \leq \Delta + 1.$$

Conditioned on $\mathcal{E}_1, \bar{\mathcal{E}}_{stab}$ the network reaches a near-stable pair of configurations, but does not stabilize. We can bound

$$\begin{aligned} \mathbb{E}_{\alpha^1}[t(\alpha_X, t_s, \alpha_Y) | \mathcal{E}_1, \bar{\mathcal{E}}_{stab}] &\leq (\Delta + 1 + t_s) + \mathbb{E}_{N^{\Delta+t_s} N^{\Delta+1+t_s}}[t(\alpha_X, t_s, \alpha_Y)] \\ &\leq \Delta + 1 + t_s + t_{max}(\alpha_X). \end{aligned}$$

Finally, conditioned on $\bar{\mathcal{E}}_1$, the network does not reach a pair of near-stable configurations within Δ steps. We have:

$$\begin{aligned}\mathbb{E}_{\alpha^1}[t(\alpha_X, t_s, \alpha_Y) | \bar{\mathcal{E}}_1] &\leq \Delta + \mathbb{E}_{N^{\Delta-1}N^\Delta}[t(\alpha_X, t_s, \alpha_Y)] \\ &\leq \Delta + t_{\max}(\alpha_X).\end{aligned}$$

We can plug these bounds along with the probability bounds of (107) and (108) into (109) to obtain:

$$\begin{aligned}\mathbb{E}_{\alpha^1}[t(\alpha_X, t_s, \alpha_Y)] &\leq (\Delta + 1) \cdot \left(\frac{1}{200} - \frac{1}{1000t_s} \right) + (\Delta + 1 + t_{\max}(\alpha_X) + t_s) \cdot \frac{1}{1000t_s} \\ &\quad + (\Delta + t_{\max}(\alpha_X)) \cdot \frac{199}{200} \\ &\leq \Delta + 1 + t_{\max} \left(\frac{199}{200} + \frac{1}{1000t_s} \right) + \frac{t_s}{1000t_s} \\ &\leq \Delta + 1 + t_{\max}(\alpha_X) \cdot \frac{249}{250} + \frac{1}{1000}.\end{aligned}$$

Since this bound holds for all α^1 we have:

$$t_{\max}(\alpha_X) \leq \Delta + 1 + t_{\max}(\alpha_X) \cdot \frac{249}{250} + \frac{1}{100}$$

which gives $t_{\max}(\alpha_X) \leq 250(\Delta + 1) + \frac{250}{1000} \leq 250\Delta + 251$. This bound holds for all α_X and so gives the lemma, after recalling that $\Delta = 15$. $\square$

5.7 Constructions With Runtime Tradeoffs

The family of two-inhibitor networks of Section 3 and the family of $O(\log n)$ -inhibitor networks of Section 5.2 represent two extremes of a tradeoff between number of inhibitors and runtime. As shown in Theorem 4.3, the two-inhibitor network uses the minimum number of neurons required to solve WTA with a reasonably long stability period. At least when considering a constant probability of success, the $O(\log n)$ -inhibitor networks gives optimal convergence time of $O(1)$ steps. In this section we outline, at a high level, two families of networks that allow a tradeoff between these two extremes.

5.7.1 Fixed Convergence Time Construction

The first construction lets us to achieve any desired convergence time θ if sufficiently many inhibitors are used. Specifically, we describe a family of networks which converge to a valid WTA output configuration with constant probability in $O(\theta)$ steps (and with probability $\geq 1 - \delta$ in $O(\theta \cdot \log 1/\delta)$ steps), using $O(\theta \log^{1/\theta} n)$ inhibitors. Note that setting $\theta = 1$ recovers the runtime-inhibitor tradeoff of our $O(\log n)$ -inhibitor networks.

We have one stability inhibitor a_s that functions in the same way as the stability inhibitor in our $O(\log n)$ -inhibitor construction (Definition 5.1), ensuring that, once the network reaches a valid WTA output configuration, it remains in this configuration for t_s consecutive time steps with high probability (as long as some other stability conditions, similar to the near-stable condition of Definition 5.16 are satisfied).

The remaining inhibitors are split into θ groups each containing $O(\log^{1/\theta} n)$ inhibitors. The i^{th} group for $i \in \{2, \dots, \theta\}$ is responsible for reducing the number of competing outputs from any number

$$k \in \left(2^{\log^{(i-1)/\theta} n}, 2^{\log^{i/\theta} n}\right]$$

to some $k' \leq 2^{\log^{(i-1)/\theta} n}$. With $O(\log^{1/\theta} n)$ inhibitors this can be done with high probability in $O(1)$ time steps via a method described below. Thus, in $O(\theta)$ time steps, the number of firing outputs (corresponding to firing inputs will reduce from at most $2^{\log^{\theta/\theta} n} = n$ down to $2^{\log^{1/\theta} n}$. Once the number of computing outputs is this low, the final group of $O(\log^{1/\theta} n)$ inhibitors can drive convergence to WTA with constant probability in $O(1)$ steps using an identical strategy to that employed in our $O(\log n)$ -inhibitor network. n is just replaced by $2^{\log^{1/\theta} n}$.

It remains to explain how the number of competing outputs is reduced from $k \in [2^{\log^{(i-1)/\theta} n}, 2^{\log^{i/\theta} n}]$ to $k' \leq 2^{\log^{(i-1)/\theta} n}$ in a single time step using $O(\log^{1/\theta} n)$ inhibitors. Again, we use a strategy similar to our $O(\log n)$ -inhibitor construction. depending on the number of firing outputs at time t , the $O(\log^{1/\theta} n)$ inhibitors in group i induce $O(\log^{1/\theta} n)$ different firing probabilities. Letting $\Delta(j) = j \cdot \log^{(i-1)/\theta} n$, these firing probabilities are:

$$\frac{1}{2^{\Delta(1)}}, \frac{1}{2^{\Delta(2)}}, \dots, \frac{1}{2^{\Delta((\log^{1/\theta} n - 1))}}, \frac{1}{2^{\Delta(\log^{1/\theta} n)}}.$$

Note that $\Delta(1) = \log^{(i-1)/\theta} n$ and $\Delta(\log^{1/\theta} n) = \log^{i/\theta} n$. Additionally, the ratio between adjacent probabilities is $2^{\log^{(i-1)/\theta} n}$. Thus, for any number $k \in \left(2^{\log^{(i-1)/\theta} n}, 2^{\log^{i/\theta} n}\right]$ of firing outputs at time t , the inhibitors can induce a firing probability between $\frac{1}{k}$ and $\frac{2^{\log^{(i-1)/\theta} n}}{k}$. Thus, with good probability, the number of firing outputs will reduce to some value $k' \leq 2^{\log^{(i-1)/\theta} n}$.

While we do not analyze this construction in detail, using similar proof techniques to those used for our $O(\log n)$ -inhibitor construction, it is possible to show:

Theorem 5.31 (θ -Step WTA). *For any $n \in \mathbb{Z}^{\geq 2}$ and $\theta \in \mathbb{Z}^{\geq 1}$ there is an a family of SNNs containing $O(\theta \cdot \log^{1/\theta} n)$ auxiliary inhibitory neurons which solve $\text{WTA}(n, t_c, t_s, \delta)$ for any t_s, δ and any $t_c \geq c_1 \cdot \theta(\log_2(1/\delta) + 1)$, where c_1 is some fixed constant.*

5.7.2 Fixed Inhibitor Budget Construction

Our second construction is a family of networks using α inhibitors for any $\alpha \geq 2$, which converges to a valid WTA state with constant probability in $O(\alpha \cdot (\log n)^{1/(\alpha-1)})$ steps (and with probability $\geq 1 - \delta$ in $O(\alpha \cdot (\log n)^{1/(\alpha-1)} \cdot \log 1/\delta)$ steps. Note that for $\alpha = 2$, this gives convergence in $O(\log n)$ steps, matching the performance of the two-inhibitor network construction of Section 3.

As in our two-inhibitor construction (Definition 3.1) and $O(\log n)$ -inhibitor construction (Definition 5.1) we employ one stability inhibitor, which ensures that, once the network reaches a valid WTA output configuration, it remains in such a configuration with good probability for t_s consecutive steps (again, as long as some other stability conditions, similar to the near-stable condition of Definition 5.16 are satisfied).

We label the remaining $\alpha - 1$ inhibitors $a_1, \dots, a_{\alpha-1}$. For each i , a_i fires with high probability at time $t + 1$ whenever $k \geq 2^{(\log n)^{(i-1)/(\alpha-1)}}$ outputs fire at time t . In this way, with high probability, $a_1, \dots, a_i$ fire (and all other inhibitors do not fire) at time $t + 1$ whenever the number of firing outputs k is in the range:

$$R_i \stackrel{\text{def}}{=} \left[2^{(\log n)^{(i-1)/(\alpha-1)}}, 2^{(\log n)^{i/(\alpha-1)}}\right).$$

We set the inhibitory weights such that when a_i fires (along with a_j for all $j \leq i$), each firing output fires in the next step with probability $p_i = \frac{1}{2^{(\log n)^{(i-1)/(\alpha-1)}}$. For $k \in R_i$ we have:

$$p_i \cdot k \in \left[1, 2^{(\log n)^{i/(\alpha-1)} \cdot (1 - 1/(\log n)^{1/(\alpha-1)})}\right).$$

With each output continuing to fire with probability $\frac{1}{2^{(\log n)^{(i-1)/(\alpha-1)}}$ in each step, starting from at most $2^{(\log n)^{i/(\alpha-1)}}$ competing outputs, we thus reach a configuration with $k \leq 2^{(\log n)^{(i-1)/(\alpha-1)}}$ firing outputs with good probability within $O((\log n)^{1/(\alpha-1)})$ steps. Overall, over α such levels, the network reaches a valid WTA configuration with constant probability in $O(\alpha \cdot (\log n)^{1/(\alpha-1)})$ steps.

Again, while we do not analyze this construction here, it is possible to prove the following:

Theorem 5.32 (α -Inhibitor WTA). *For any $n, \alpha \in \mathbb{Z}^{\geq 2}$ there is a family of SNNs containing α auxiliary inhibitory neurons which solve $\text{WTA}(n, t_c, t_s, \delta)$ for any t_s, δ and any $t_c \geq c_1 \cdot \alpha (\log n)^{1/(\alpha-1)} \cdot (\log_2(1/\delta) + 1)$, where c_1 is some fixed constant.*

6 Discussion and Future Work

We have presented an exploration of the WTA problem in stochastic spiking neural networks, giving network constructions, runtime analysis, and lower bounds. Our work leaves open a number of questions, both in regards to the WTA problem, and more broadly, in exploring spiking neural networks from an algorithmic perspective. We discuss some of these directions below. See also Section 2.5 in which we discuss possible extensions two and modifications of our basic spiking neural network model.

6.1 Winner-Take-All Extensions and Open Questions

We first overview future work directly related to the WTA problem studied in this work.

6.1.1 Lower Bounds

In Section 4 we present lower bounds for one and two-inhibitor WTA networks (Theorems 4.3 and 4.14). In [LMP17a] we also give nearly tight lower bounds on the convergence time for networks using $\alpha > 2$ inhibitors, in a similar model to the one presented here. A few interesting open questions remain:

- Can our lower bound for two-inhibitor networks (Theorem 4.14) be tightened to match our upper bound (Theorem 3.2) up to a constant factor, rather than a $O(\log \log n)$ factor?
- Our lower bounds apply to somewhat restricted classes of simple and symmetric SNNs (Definitions 4.1 and 4.2 respectively.) We conjecture, however, that these lower bounds can be shown for general SNNs, with no restrictions on the network structure.
- It would be interesting to obtain lower bounds which help explain the relationship between history and convergence time. Our $O(\log n)$ -inhibitor network of Section 5.2 achieves constant probability, $O(1)$ convergence time. However, it requires a history period of two steps. Is it possible to show that, with no history period (i.e., in the basic SNN model of Section 2), it is not possible to achieve this convergence time? Is the $O(\log n)$ time for our two-inhibitor network optimal for historyless networks, regardless of the number of inhibitors used? In a network with history, is it possible to solve WTA with a single auxiliary neuron, or can the lower bound of Theorem 4.3 be extended to this case?

6.1.2 Variations on the WTA Problem

Our work focuses on a simplified WTA problem in which all inputs either fire at every time step or never fire (see Definition 2.7). The challenge in solving this problem is in breaking symmetry between the firing inputs. In future work we plan to extend this work to consider more general variants of WTA, discussed below.

- **Non-Binary WTA:** We plan to extend our WTA work to a *non-binary* setting, in which inputs have different firing rates and selection is based on those rates. The most common requirement is for the network to select an input neuron with the highest or near-highest firing rate [YG89, CGL92, OL06].

In the most basic setting, we can consider n inputs $x_1, \dots, x_n$ with firing rates $r_1, \dots, r_n \in [0, 1]$ as neurons that fire independently at random at each time step, each time with probability r_i . In [LMP17c], we presented an initial exploration of the non-binary WTA problem with inputs of this type, giving a solution using $O(n \log n)$ auxiliary neurons to select an input with firing rate within a constant factor of the maximum. It seems that in our basic SNN model, $\Omega(n \log n)$ auxiliary neurons may in fact be necessary: essentially, for each input neuron, $\Omega(\log n)$ auxiliary neurons are needed to record a short firing history, from which the firing probability r_i can be (implicitly) estimated to within a constant factor, with high probability. Thus, networks using a sublinear number of auxiliary neurons, like those we gave for binary WTA, may be ruled out.

However, more efficient solutions may be possible if we use an SNN model with a history period. Even without history, efficient solutions may be possible if we relax or modify the WTA problem. For example, we may just require selecting each neuron with probability $\frac{r_i}{\sum_{i=1}^n r_i}$, or some other function of the firing probabilities.

- **WTA with Different Output Conditions:** It would be interesting to consider variations on the basic output condition of the WTA problem, both in the binary and non-binary input setting. For example, k -WTA is a common variant of WTA, in which the goal is to select k inputs with the highest or near-highest firing rates [MEAM89, Maa96, WS03, HW08], rather than just as single input. k -WTA, for example, has been used in recent work in modeling sparse coding in the fly olfactory via random projection methods [DSN17]. It would be interesting to understand the fundamental complexity (in terms of network complexity and convergence time) of implementing this primitive in spiking neural networks.

6.1.3 Applications of WTA

Finally, it would be interesting to explore how our WTA constructions can be integrated into solutions to higher-level neural tasks. Some problems which may be worth considering are:

- Attention in sensory processing systems, which is thought to be implemented via WTA competition [LIKB99, IK01]. In this setting, a specific set of neurons is activated while others are suppressed, allowing computation to happen on the selected neurons without interference. Formulating and studying simple tasks which model the use of WTA as an attention mechanism would be an interesting direction.
- The use of WTA in neural sparse coding algorithms [DSN17]. In this setting, a stimulus, such as a specific odor, triggers the firing of a large set of neurons, which is then “sparsified” via non-binary k -WTA competition to k neurons that encode the odor – those with the strongest response to the stimulus.

- Clustering problems, in a which the network is set up such that, when given any input, the neuron which corresponds to the best cluster assignment for this input fires at the highest rate. Non-binary WTA can be used to select this neuron [AF94, WR94].

6.2 Other Neural Computational Primitives

WTA is a basic primitive that seems to be useful in a broad range of neural tasks, discussed above. It is also useful in exploring the computational power of SNNs and studying tradeoffs among such costs as runtime, network size, and the use of randomness. In future work, we hope to identify other such basic primitives, as a means of building a general algorithmic theory for spiking neural networks. We discuss some possible directions below:

6.2.1 Neural Indexing

In [LMP17b] we define and study the *Neuro-RAM* problem: the network is given a set of n binary inputs X_1 , along with a smaller set of $\log n$ inputs X_2 whose firing pattern represents an index into X_1 . The goal is for the output to fire if and only if the selected neuron in X_1 is firing. In [LMP17b], we applied our Neuro-RAM primitive to the similarity testing problem: given two binary inputs X_1, X_2 , determine with high probability whether their Hamming distance is greater than some threshold. We solved this problem efficiently by using a Neuro-RAM model to select and compare random positions in X_1 and X_2 .

It would be interesting to formulate other basic neural tasks that may employ a Neuro-RAM primitive. For example, one can consider estimating the differences between firing patterns (e.g., representing sensory inputs at successive times), estimating the average firing activity of a group of neurons responding to some stimulus, or randomly “exploring” different sets of neurons whose firing may trigger a desired response (such as the recall of a memory).

In our work, aside from efficient Neuro-RAM constructions, we give lower bounds demonstrating that these constructions give a near-optimal tradeoff between network size and runtime in our basic SNN model with a sigmoidal spike probability function. Our proofs are based on reducing an SNN to a distribution over deterministic networks, and then bounding the VC dimension [Vap98] of these networks.

Generally, while VC dimension has been studied for spiking networks [ZP96, MS99], we are the first to apply it to prove computational lower bounds. We hope to continue our work, greatly expanding the toolkit for proving lower bounds in biologically plausible networks, including those with different spike probability functions, with history, and even with refractory periods or spike propagation delays.

Further, Neuro-RAM lower bounds let us compare our SNN models with artificial neural network models that use continuous-output gates, e.g., sigmoidal gates [Bar93, Bar97] or rectified linear units (ReLU) [JKLR09, NH10]. The Neuro-RAM problem can be solved very efficiently in such networks [Koi96], contrasting with its apparent difficulty in SNNs. However, the continuous-output solutions do not seem very robust: they do not appear to tolerate small variations in edge weights, and they seem to rely on precise (infinite precision) gate outputs. We hope that comparing these models with our SNNs could lead to a better understanding of continuous-output circuits in noisy settings. We conjecture that in these settings, these networks behave more like their spiking network counterparts than like ideal continuous-output networks.

6.2.2 Binary Vector Problems

Beyond the WTA and Neuro-RAM problems it would also be interesting to consider other basic problems with fixed input firing patterns, that is, binary vector problems. For example, the problem of estimating the total firing strength of a population of neurons corresponds to determining the Hamming weight of a binary input vector. More complex problems include string matching, in which we test whether the input firing pattern contains a copy of a specified substring. In the brain, this may be used for pattern recognition, or for alignment and comparison of patterns triggered by multiple stimuli.

6.2.3 Problems with Non-Binary Inputs

It would also be valuable to consider problems in which the input is non-binary, for example, problems that require computing some function of the firing rates of the input neurons. For example, we may consider the problem of approximating the sum of input firing probabilities, or more generally, some norm of the firing rate vector $r_1, \dots, r_n$. We may also consider similarity testing under different norms (generalizing the Hamming distance similarity testing problem in [LMP17b]).

Many algorithms proposed in computational neuroscience and computer science apply operations such as addition and multiplication to continuous neuron output values [DR89, AZGMS14, AGMM15, DSN17]. These continuous output values can be thought of as abstractions of firing rates in SNNs. We would like to understand how such basic operations may be emulated in more biologically plausible SNNs.

6.2.4 Synchronization Problems

Finally, it would be interesting to study algorithmic primitives that are not expressed in terms of simple input-output mappings. Notably, we would like to consider synchronization problems in which neurons, starting from arbitrary firing states, cooperate to align their firing in some way. Synchronization of spiking patterns and the emergence of neural rhythms are widely studied in both empirical and computational neuroscience [VVAE94, Buz06, RGDA97, WSO⁺07]. Synchronization is also an important technique used in distributed algorithms [BL85, KO87, Cri89, LSW09], and thus is a natural problem to include in an algorithmic theory of neural computation.

6.3 Learning Problems and Dynamic Networks

The work presented in this work, as well as the work proposed in Section 6.2 focuses on computation in *static* networks, with fixed edge weights. An important direction is studying algorithms for learning in *dynamic spiking networks*, in which synapse weights change throughout the computation via, for example, a Hebbian update rule.

In classical Hebbian learning [Heb05, CD08], the weight of a synapse is continuously updated by a factor that depends on the product of the firing strengths of its two endpoints. The more the firing of the endpoints correlates, the stronger the synapse becomes. We hope to define a Hebbian-style rule for our synchronous SNN models similar to, for example, the simple rule used in [LMPV18]. When both endpoints fire at the same time, the synapse weight should increase by a small factor, and when just one fires, the weight may decrease. We then hope to use our dynamic network model to study the costs of many learning problems such as memory formation and concept association [Ama77, Val05, LMPV18], linear classification [HNGS⁺06], principal component analysis [San89, HO97, HO00], and sparse coding [OF04, AGMM15].

6.4 Neural Linear Algebraic Computation

Finally, we are interested in understanding how basic linear algebraic computation may be performed in spiking neural networks, possibly with the use of randomization. Randomized algorithms have recently led to a number of breakthroughs in fundamental linear-algebraic and geometric problems such as linear regression [CW13, CLM⁺15], low-rank approximation [Sar06, NN13], clustering [BZD10], and locality-sensitive hashing [DIIM04, AI06]. Many mechanisms used in neural computation have close analogs to randomized linear algebraic techniques; e.g., Oja’s Hebbian-style rule for neural principal component analysis (PCA) [HO97, HO00] is a common technique for low-memory PCA and eigenvector approximation [Sha15, JJK⁺16]. Fast linear-algebraic methods based on *random projections* [Ach03, Sar06] are also conjectured to play a role in neural computation [GS12, AZGMS14], with random synaptic connectivity providing a natural implementation of randomized dimensionality-reduction. We hope to study the application of these techniques in stochastic spiking neural networks, forging connections with work on new algorithms for fast linear algebraic computation.

References

- [Ach03] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003.
- [AF94] Hazem M Abbas and Moustafa M Fahmy. Neural networks for maximum likelihood clustering. *Signal Processing*, 36(1):111–126, 1994.
- [AGMM15] Sanjeev Arora, Rong Ge, Tengyu Ma, and Ankur Moitra. Simple, efficient, and neural algorithms for sparse coding. In *Proceedings of the 28th Annual Conference on Computational Learning Theory (COLT)*, pages 113–149, 2015.
- [AHS85] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9(1):147–169, 1985.
- [AI06] Alexandr Andoni and Piotr Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 459–468, 2006.
- [Ama77] S-I Amari. Neural theory of association and concept-formation. *Biological Cybernetics*, 26(3):175–185, 1977.
- [And03] Alex M Andrew. Spiking neuron models: Single neurons, populations, plasticity. *Kybernetes*, 32(7/8), 2003.
- [AS94] Christina Allen and Charles F Stevens. An evaluation of causes for unreliability of synaptic transmission. *Proceedings of the National Academy of Sciences*, 91(22):10380–10383, 1994.
- [ASNN⁺15] Maruan Al-Shedivat, Rawan Naous, Emre Neftci, Gert Cauwenberghs, and Khaled N Salama. Inherently stochastic spiking neurons for probabilistic neural computation. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 356–359. IEEE, 2015.

- [AZGMS14] Zeyuan Allen-Zhu, Rati Gelashvili, Silvio Micali, and Nir Shavit. Sparse sign-consistent Johnson–Lindenstrauss matrices: Compression with neuroscience-based constraints. *Proceedings of the National Academy of the Sciences*, 2014.
- [Bar93] Andrew R Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):930–945, 1993.
- [Bar97] Peter L Bartlett. For valid generalization the size of the weights is more important than the size of the network. In *Advances in Neural Information Processing Systems 10 (NIPS)*, pages 134–140, 1997.
- [BL85] James E Burns and Nancy A Lynch. The Byzantine firing squad problem. Technical report, Massachusetts Institute of Technology, 1985.
- [Buz06] Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford University Press, 2006.
- [BZD10] Christos Boutsidis, Anastasios Zouzias, and Petros Drineas. Random projections for k -means clustering. In *Advances in Neural Information Processing Systems 23 (NIPS)*, 2010.
- [CD08] Natalia Caporale and Yang Dan. Spike timing–dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience*, 31:25–46, 2008.
- [CGL92] Robert Coultrip, Richard Granger, and Gary Lynch. A cortical model of winner-take-all competition via lateral inhibition. *Neural Networks*, 5(1):47–54, 1992.
- [CLM⁺15] Michael B. Cohen, Yin Tat Lee, Cameron Musco, Christopher Musco, Richard Peng, and Aaron Sidford. Uniform sampling for matrix approximation. In *Proceedings of the 6th Conference on Innovations in Theoretical Computer Science (ITCS)*, 2015.
- [Cri89] Flaviu Cristian. Probabilistic clock synchronization. *Distributed Computing*, 3(3):146–158, 1989.
- [CW13] Kenneth L. Clarkson and David P. Woodruff. Low rank approximation and regression in input sparsity time. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing (STOC)*, 2013.
- [DIIM04] Mayur Datar, Nicole Immorlica, Piotr Indyk, and Vahab S Mirrokni. Locality-sensitive hashing scheme based on p -stable distributions. In *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, pages 253–262. ACM, 2004.
- [DR89] Richard Durbin and David E Rumelhart. Product units: A computationally powerful and biologically plausible extension to backpropagation networks. *Neural Computation*, 1(1):133–142, 1989.
- [DSN17] Sanjoy Dasgupta, Charles F Stevens, and Saket Navlakha. A neural algorithm for a fundamental computing problem. *Science*, 358(6364):793–796, 2017.
- [FSW08] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4):292–303, 2008.
- [GK02] Wulfram Gerstner and Werner M Kistler. *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press, 2002.

- [GL09] Ankur Gupta and Lyle N Long. Hebbian learning with winner take all for spiking neural networks. In *2009 International Joint Conference on Neural Networks*, pages 1054–1060. IEEE, 2009.
- [GS12] Surya Ganguli and Haim Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*, 35:485–508, 2012.
- [Hay09] Simon S Haykin. *Neural Networks and Learning Machines*, volume 3. Pearson, 2009.
- [Heb05] Donald Olding Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Psychology Press, 2005.
- [HJM13] Stefan Habenschuss, Zeno Jonke, and Wolfgang Maass. Stochastic computations in cortical microcircuit models. *PLoS Computational Biology*, 9(11):e1003311, 2013.
- [HNGS⁺06] Ramón Huerta, Thomas Nowotny, Marta García-Sánchez, Henry DI Abarbanel, and Mikhail I Rabinovich. Learning classification in the olfactory system of insects. *Learning*, 16(8), 2006.
- [HO97] Aapo Hyvärinen and Erkki Oja. One-unit learning rules for independent component analysis. In *Advances in Neural Information Processing Systems 10 (NIPS)*, pages 480–486, 1997.
- [HO00] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411–430, 2000.
- [HSW89] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.
- [HT⁺86] John J Hopfield, David W Tank, et al. Computing with neural circuits – a model. *Science*, 233(4764):625–633, 1986.
- [HW08] Xiaolin Hu and Jun Wang. An improved dual neural network for solving a class of quadratic programming problems and its k -winners-take-all application. *IEEE Transactions on Neural Networks*, 19(12):2022–2031, 2008.
- [IK01] Laurent Itti and Christof Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001.
- [Izh04] Eugene M Izhikevich. Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5):1063–1070, 2004.
- [JJK⁺16] Prateek Jain, Chi Jin, Sham M Kakade, Praneeth Netrapalli, and Aaron Sidford. Streaming PCA: Matching matrix bernstein and near-optimal finite sample guarantees for Ojaś algorithm. In *Proceedings of the 29th Annual Conference on Computational Learning Theory (COLT)*, pages 1147–1164, 2016.
- [JKLR09] Kevin Jarrett, Koray Kavukcuoglu, Yann LeCun, and Marc’Aurelio Ranzato. What is the best multi-stage architecture for object recognition? In *12th International Conference on Computer Vision*, pages 2146–2153. IEEE, 2009.
- [KK94] Samuel Kaski and Teuvo Kohonen. Winner-take-all networks for physiological models of competitive learning. *Neural Networks*, 7(6-7):973–984, 1994.

- [KKKL11] Majid Khabbazzian, Dariusz Kowalski, Fabian Kuhn, and Nancy Lynch. Decomposing broadcast algorithms using abstract MAC layers, 2011. <http://groups.csail.mit.edu/tds/papers/Lynch/MIT-CSAIL-TR-2011-010.pdf>.
- [KO87] Hermann Kopetz and Wilhelm Ochsenreiter. Clock synchronization in distributed real-time systems. *IEEE Transactions on Computers*, 100(8):933–940, 1987.
- [Koi96] Pascal Koiran. VC dimension in circuit complexity. In *Proceedings of the Eleventh Annual IEEE Conference on Computational Complexity*, pages 81–85. IEEE, 1996.
- [KU87] Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of Intelligence*, pages 115–141. Springer, 1987.
- [LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [LIK99] Dale K Lee, Laurent Itti, Christof Koch, and Jochen Braun. Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, 2(4):375–381, 1999.
- [LMP17a] Nancy Lynch, Cameron Musco, and Merav Parter. Computational tradeoffs in biological neural networks: Self-stabilizing winner-take-all networks. In *Proceedings of the 8th Conference on Innovations in Theoretical Computer Science (ITCS)*, 2017.
- [LMP17b] Nancy Lynch, Cameron Musco, and Merav Parter. Neuro-RAM unit with applications to similarity testing and compression in spiking neural networks. In *Proceedings of the 31st International Symposium on Distributed Computing (DISC)*, 2017. Full version available at <https://arxiv.org/abs/1706.01382>.
- [LMP17c] Nancy Lynch, Cameron Musco, and Merav Parter. Spiking neural networks: An algorithmic perspective. In *5th Workshop on Biological Distributed Algorithms (BDA 2017)*, July 2017.
- [LMPV18] Robert Legenstein, Wolfgang Maass, Christos H Papadimitriou, and Santosh S Vempala. Long term memory and the densest k-subgraph problem. In *Proceedings of the 9th Conference on Innovations in Theoretical Computer Science (ITCS)*, 2018.
- [LRMM88] John Lazzaro, Sylvie Ryckebusch, Misha Anne Mahowald, and Caver A Mead. Winner-take-all networks of $O(n)$ complexity. Technical report, DTIC Document, 1988.
- [LSW09] Christoph Lenzen, Philipp Sommer, and Roger Wattenhofer. Optimal clock synchronization in networks. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 225–238. ACM, 2009.
- [Maa96] Wolfgang Maass. On the computational power of noisy spiking neurons. In *Advances in Neural Information Processing Systems 9 (NIPS)*, pages 211–217, 1996.
- [Maa97] Wolfgang Maass. Networks of spiking neurons: the third generation of neural network models. *Neural Networks*, 10(9):1659–1671, 1997.
- [Maa99] Wolfgang Maass. Neural computation with winner-take-all as the only nonlinear operation. In *Advances in Neural Information Processing Systems 12 (NIPS)*, pages 293–299, 1999.

- [Maa00] Wolfgang Maass. On the computational power of winner-take-all. *Neural Computation*, 12(11):2519–2535, 2000.
- [Maa14] Wolfgang Maass. Noise as a resource for computation and learning in networks of spiking neurons. *Proceedings of the IEEE*, 102(5):860–880, 2014.
- [MEAM89] E Majani, Ruth Erlanson, and Yaser S Abu-Mostafa. On the k -winners-take-all network. In *Advances in Neural Information Processing Systems 2 (NIPS)*, pages 634–642, 1989.
- [MP69] Marvin Minsky and Seymour Papert. *Perceptrons*. MIT Press, 1969.
- [MS99] Wolfgang Maass and Michael Schmitt. On the complexity of learning for spiking neurons with temporal coding. *Information and Computation*, 153(1):26–46, 1999.
- [MSS91] Wolfgang Maass, Georg Schnitger, and Eduardo D Sontag. On the computational power of sigmoid versus boolean threshold circuits. In *Proceedings of the 32nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 767–776, 1991.
- [MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.
- [Mus18] Cameron Musco. *The Power of Randomized Algorithms: From Numerical Linear Algebra to Biological Systems*. PhD thesis, Massachusetts Institute of Technology, 2018.
- [NH10] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pages 807–814, 2010.
- [NN13] Jelani Nelson and Huy L Nguyễn. OSNAP: Faster numerical linear algebra algorithms via sparser subspace embeddings. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2013.
- [Now89] Steven J Nowlan. Maximum likelihood competitive learning. In *Advances in Neural Information Processing Systems 2 (NIPS)*, pages 574–582, 1989.
- [ODL09] Matthias Oster, Rodney Douglas, and Shih-Chii Liu. Computation with spikes in a winner-take-all network. *Neural Computation*, 21(9):2437–2465, 2009.
- [OF04] Bruno A Olshausen and David J Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, 2004.
- [OL06] Matthias Oster and Shih-Chii Liu. Spiking inputs to a winner-take-all network. *Advances in Neural Information Processing Systems 19 (NIPS)*, 2006.
- [Osb79] Neville N Osborne. Is Dale’s principle valid? *Trends in Neurosciences*, 2:73–75, 1979.
- [Osb13] Neville N Osborne. *Dale’s Principle and Communication Between Neurons*. Elsevier, 2013. Based on a Colloquium of the Neurochemical Group of the Biochemical Society, Held at Oxford University, July 1982.
- [RB15] Lisa Roux and György Buzsáki. Tasks for inhibitory interneurons in intact brain circuits. *Neuropharmacology*, 88:10–23, 2015.

- [RFLHL11] Bernardo Rudy, Gordon Fishell, SooHyun Lee, and Jens Hjerling-Leffler. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Developmental Neurobiology*, 71(1):45–61, 2011.
- [RGDA97] Alexa Riehle, Sonja Grün, Markus Diesmann, and Ad Aertsen. Spike synchronization and rate modulation differentially involved in motor cortical function. *Science*, 278(5345):1950–1953, 1997.
- [San89] Terence D Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2(6):459–473, 1989.
- [Sar06] Tamas Sarlós. Improved approximation algorithms for large matrices via random projections. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 143–152, 2006.
- [Sha15] Ohad Shamir. A stochastic PCA and SVD algorithm with an exponential convergence rate. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 144–152, 2015.
- [SN94] Michael N Shadlen and William T Newsome. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4(4):569–579, 1994.
- [SS95] Hava T Siegelmann and Eduardo D Sontag. On the computational power of neural nets. *Journal of Computer and System Sciences*, 50(1):132–150, 1995.
- [Stu] Robert Stufflebeam. Introduction to neurons, synapses, action potentials, and neurotransmission. In *The Mind Project*. <http://www.mind.ilstu.edu/curriculum/modOverview.php?modGUI=232>.
- [Tho90] Simon J Thorpe. Spike arrival times: A highly efficient coding scheme for neural networks. *Parallel Processing in Neural Systems*, pages 91–94, 1990.
- [Tra09] Thomas Trappenberg. *Fundamentals of Computational Neuroscience*. OUP Oxford, 2009.
- [Val00a] Leslie G Valiant. *Circuits of the Mind*. Oxford University Press, 2000.
- [Val00b] Leslie G Valiant. A neuroidal architecture for cognitive computation. *Journal of the ACM (JACM)*, 47(5):854–882, 2000.
- [Val05] Leslie G Valiant. Memorization and association on a realistic neural model. *Neural Computation*, 17(3):527–555, 2005.
- [Vap98] Vladimir Vapnik. *Statistical Learning Theory*, volume 1. Wiley New York, 1998.
- [VVAE94] Carl Van Vreeswijk, LF Abbott, and G Bard Ermentrout. When inhibition not excitation synchronizes neural firing. *Journal of Computational Neuroscience*, 1(4):313–321, 1994.
- [WMK⁺02] Masahito Watanabe, Kentaro Maemura, Kiyoto Kanbara, Takumi Tamayama, and Hana Hayasaki. GABA and GABA receptors in the central nervous system and other organs. In *International Review of Cytology*, volume 213, pages 1–47. Elsevier, 2002.

- [WR94] Steve R Waterhouse and Anthony J Robinson. Classification using hierarchical mixtures of experts. In *Proceedings of the IEEE Workshop Neural Networks for Signal Processing*, pages 177–186, 1994.
- [WS03] Wei Wang and Jean-Jacques E Slotine. k -winners-take-all computation with neural oscillators. [arXiv:q-bio/0401001](https://arxiv.org/abs/q-bio/0401001), 2003.
- [WSO⁺07] Thilo Womelsdorf, Jan-Mathijs Schoffelen, Robert Oostenveld, Wolf Singer, Robert Desimone, Andreas K Engel, and Pascal Fries. Modulation of neuronal interactions through neuronal synchronization. *Science*, 316(5831):1609–1612, 2007.
- [YG89] Alan L Yuille and Norberto M Grzywacz. A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Computation*, 1(3):334–347, 1989.
- [ZP96] Anthony M Zador and Barak A Pearlmutter. VC dimension of an integrate-and-fire neuron model. In *Proceedings of the 9th Annual Conference on Computational Learning Theory (COLT)*, pages 10–18. ACM, 1996.