Computational Tradeoffs in Biological Neural Networks: Self-Stabilizing Winner-Take-All Networks*

Nancy Lynch[1], Cameron Musco[2], and Merav Parter[3]

1 **Massachusetts Institute of Technology, Cambridge, MA, USA**
`lynch@csail.mit.edu`
2 **Massachusetts Institute of Technology, Cambridge, MA, USA**
`cnmusco@mit.edu`
3 **Massachusetts Institute of Technology, Cambridge, MA, USA**
`parter@mit.edu`

Abstract

We initiate a line of investigation into biological neural networks from an algorithmic perspective. We develop a simplified but biologically plausible model for distributed computation in *stochastic spiking neural networks* and study tradeoffs between computation time and network complexity in this model. Our aim 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.

In this paper, we focus on the important 'winner-take-all' (WTA) problem, which is analogous to a neural leader election unit: a network consisting of n input neurons and n corresponding output neurons must converge to a state in which a single output corresponding to a firing input (the 'winner') fires, while all other outputs remain silent. Neural circuits for WTA rely on inhibitory neurons, which suppress the activity of competing outputs and drive the network towards a converged state with a single firing winner. We attempt to understand how the number of inhibitors used affects network convergence time.

We show that it is possible to significantly outperform naive WTA constructions through a more refined use of inhibition, solving the problem in $O(\theta)$ rounds in expectation with just $O(\log^{1/\theta} n)$ inhibitors for any θ. An alternative construction gives convergence in $O(\log^{1/\theta} n)$ rounds with $O(\theta)$ inhibitors. We complement these upper bounds with our main technical contribution, a nearly matching lower bound for networks using $\geq \log\log n$ inhibitors. Our lower bound uses familiar indistinguishability and locality arguments from distributed computing theory applied to the neural setting. It lets us derive a number of interesting conclusions about the structure of any network solving WTA with good probability, and the use of randomness and inhibition within such a network.

1998 ACM Subject Classification F.1.1 Models of Computation – Unbounded-action devices, C.1.3 Other Architecture Styles – Neural nets

Keywords and phrases biological distributed algorithms, neural networks, distributed lower bounds, winner-take-all networks

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

* This work was partially supported by NSF Graduate Research Fellowship No. 1122374, AFOSR grant FA9550-13-1-0042 and the NSF Center for Science of Information.

1 Introduction

In this paper, we study biological neural networks from an algorithmic perspective, focusing on understanding tradeoffs between computation time and network complexity. We use a biologically plausible yet simplified neural computational model. 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.

1.1 Model and Problem Statement

Model.

We work with *spiking neural networks* (SNNs) [26, 27, 12, 18, 15], 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 its membrane potential. In this respect, our networks are similar to the popular Boltzmann machine [1], 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. While a rich literature focuses on deterministic threshold circuits [31, 16] we employ a stochastic model as it is widely accepted that neural computation is inherently stochastic [3, 42, 10], and that while this can lead to a number of challenges, it also affords significant computational advantages [30].

The WTA Problem.

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.

WTA is widely applicable, including in circuits that implement visual attention via WTA competition between groups of neurons that process different input classes [21, 23, 17]. It is also the foundation of competitive learning [32, 20, 14], in which classifiers compete to respond to specific input types. More broadly, WTA is known to be a powerful computational primitive [28, 29] – a network equipped with WTA units can perform some tasks significantly more efficiently than with just linear threshold neurons (McCulloch-Pitts neurons or perceptrons).

Related Work.

Due to its importance, there has been significant work on WTA, including in biologically plausible spiking networks [22, 48, 43, 7, 47, 34, 33, 2]. 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, 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.

Within theoretical computer science, our work is most inspired by: (1) work on the computational power of spiking neural networks, including the power of WTA as a black-box primitive, most notably by Maass et al. [27, 28, 29] (2) the pioneering work of Les Valiant on the neuroidal model [44, 45, 46] and (3) self-stabilization algorithms in distributed networks [8, 25]. We survey this literature in more depth in Appendix A.1.

Basic WTA Networks.

We restrict our attention to a simple network structure that can implement WTA efficiently using a small number of auxiliary neurons. A network consists of three layers: n input neurons X, n output neurons Y, and α auxiliary neurons Z. We usually assume all auxiliary neurons are inhibitory, however in Appendix C give extensions to the more general case where we allow auxiliary neurons to also be excitatory. Similar to well-known feedforward networks, all synaptic connections are between layers[2] with the exception of an excitatory self-loop from each output y_i to itself. This basic structure is biologically plausible; in particular self-loops and reciprocal excitatory-inhibitory connections (as implemented in our networks) are used in many biological models of WTA computation [48, 7, 38].

It is well known that inhibition is crucial for solving WTA – outputs compete for activation via *lateral inhibition* or *recurrent inhibition* [7, 38]. In our network, outputs fire in response to stimulation by their corresponding inputs, thereby stimulating inhibitors which suppress the activity of other outputs. Once a single winner is selected, it must remain distinguished from the remainder of the outputs. This is achieved via positive feedback – a consistently firing output will tend to continue firing due to its excitatory self-loop.

1.2 Our Contribution

Computational Tradeoffs.

We explore the tradeoff between the number of inhibitors α 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). In artificial neural networks, inhibitory and excitatory connections are often treated equally, as connections with either positive or negative weights. However, in reality, neurons themselves are either inhibitory or excitatory and do not have outgoing connections of both types. There are many fewer inhibitors (around 15% of the neural population [39, 13]), and they typically have restricted connectivity structures, often inhibiting just neurons in their local vicinity [29]. This gives natural motivation to understanding how the number of inhibitors used in a network affects its computational power. We give two main results:

▶ **Theorem 1** (Upper bound). *(1) For any $\alpha \geq 2$ there exists a basic WTA network with α inhibitors that, from any* arbitrary *starting configuration, converges to a valid WTA state in $O(\alpha \log^{1/\alpha} n)$ expected time. (2) For any $\theta \geq 1$ there exists a basic WTA network with $\alpha = O(\theta \log^{1/\theta} n)$ inhibitors that converges in $O(\theta)$ expected time.*

[1] Aside from immediate bounds for deterministic circuits using many ($\Omega(n)$) auxiliary neurons [22, 29].
[2] Although, due to recurrent connections the network convergence time is not synonymous with the number of layers.

For $\alpha \geq \log\log n$ the above gives runtime $\tilde{O}\left(\frac{\log\log n}{\log \alpha}\right)$. We give a near matching lower bound in this case, which holds even if we allow both excitatory and inhibitory auxiliary neurons.

▶ **Theorem 2** (Lower bound). *Any basic WTA network with α inhibitors requires $\Omega(\log\log n/\log\alpha)$ rounds to solve WTA in expectation.*

Upper Bound Techniques.

Our upper bounds are based on random competition between outputs that fire in response to stimulation from their firing inputs. One "stability" inhibitor is responsible for maintaining a WTA steady-state: as soon as just a single output fires in a round it becomes the winner of the network. Its positive feedback self-loop allows it to keep firing in subsequent rounds, while all other outputs do not fire due to inhibition from the stability inhibitor.

In order to reach a round in which just a single output fires, we employ a number of "convergence inhibitors". Ideally, if k competing outputs fire in a round, each would fire in the next round with probability $1/k$ and we would have just a single firing output with constant probability. We can approximate this behavior using $\lfloor \log n \rfloor$ convergence inhibitors, each of which acts as a threshold circuit and fires whenever $\geq 2^i$ outputs fire for $i \in 1, ..., \lfloor \log n \rfloor$. Thus when k outputs fire, approximately $\log(k)$ inhibitors fire, the inhibition causes outputs to continue firing with probability $\Theta(1/k)$, and convergence is achieved in constant rounds in expectation. This technique implicitly splits the possible number of firing outputs into $\log n$ *density classes* and uses one inhibitor to ensure fast convergence from each class. To obtain more general runtime tradeoffs, we will use density classes of increasing coarseness, with the inhibitors assigned to each density classes ensuring that the number of firing outputs decreases in few rounds until it falls into a finer density class, and eventually until just a single output fires.

Lower Bound Techniques.

Our lower bound shows that *any* network which solves WTA must have a similar structure to the network described above. The inhibitory neurons can always be roughly be divided into two classes: stability and convergence inhibitors. Further, while randomness is important in breaking symmetry between competing inputs, we show that in any efficient network, the inhibitors behave in a *nearly deterministic* manner, matching behavior seen in our upper bounds. After significantly constraining inhibitor behavior, we are able to analyze how any network which solves WTA behaves on inputs with varying numbers of firing neurons. Specifically, we consider $\Theta(\log n)$ different inputs configurations, with geometrically increasing numbers of firing input neurons, ranging from $O(1)$ to $O(n)$. We show that, after t rounds, with good probability, the network *does not distinguish between* (i.e. behaves identically for) $\Theta(\log n/\alpha^t)$ inputs.

As long as $\log n/\alpha^t > 2$, after t rounds, there are at least two inputs not distinguished by the network, and so on which the network cannot achieve WTA with good probability. This yields our lower bound of $t = \Omega(\log\log n/\log\alpha)$ rounds in expectation. Our argument uses techniques familiar in distributed computing theory [24], showing that limited local information prevents outputs from behaving in distinct manners for a large number of density classes in each round.

We obtain a corresponding lower bound for the number of rounds required to solve WTA *with high probability* by showing that in general, the high probability runtime is $\Omega(\log n/\log\log\log n)$ times the expected runtime. This nearly matches the $O(\log n)$ gap which can be achieved by noting that in $O(\log n)$ runs, any network will converge within its

expected runtime at least once with high probability. Our conversion result shows that, in our setting, expected runtime is a more natural metric – it is controlled by the number of inhibitors used, whereas the high probability runtime is just a function of expected runtime, independent of the number of inhibitors

Inhibitors	Lower Bound (Expected Time)	Upper Bound (Expected Time)
Unbounded	$\Omega(1)$ ($\Omega(\log n)$ high probability time)	$O(1)$ with $\alpha = \Theta(\log^{1/c} n)$
1	$\Omega(n^c)$	$O(n^c)$
2	$\Omega(\log n / \log\log n)$	$O(\log n)$
α	$\Omega(\log\log n / \log \alpha)$	$O(\alpha \cdot \log^{1/\alpha} n)$, for $\alpha = O(\log\log n)$ $\tilde{O}\left(\frac{\log\log n}{\log \alpha}\right)$, for $\alpha = \Omega(\log\log n)$

Table 1 Expected Time vs. Number of Inhibitors Tradeoff in Basic WTA Networks.

1.3 Biological Insights in Our Results

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 [29]. Our work shows that WTA can be achieved and maintained efficiently using very few inhibitors and with a very simple connectivity structure.

Our upper and lower bound constructions have a common take home message that may shed some light into the biological implementations of WTA networks. For instance, the division of inhibitors into "task preservers" (stability inhibitors) and "task solvers" (convergence inhibitors) seems fundamental. Further, while randomness is crucial as it allows for symmetry breaking amongst competing outputs, it appears (both in the upper bounds and the corresponding lower bound) that in optimal networks the inhibitors behave almost as deterministic threshold circuits, firing with high probability whenever the number of firing outputs is above a certain level. This presents an interesting dichotomy – while randomness is necessary computationally, it also has a cost in leading to unpredictable behavior amongst the inhibitors which 'control' the network.

1.4 Road Map

In Sec. 2 we describe our spiking neural network model and specify the WTA problem. In Sec. 3 we give two warm up examples of WTA networks to illustrate the tradeoff between convergence time and network size. The first has two inhibitors and converges to the WTA state within $O(\log n)$ rounds in expectation. The second has $O(\log n)$ inhibitors and $O(1)$ expected runtime. In Sec. 4.1, we provide more delicate constructions for any number of inhibitors α. Our key technical result appears in Sec. 4.2 where we provide a runtime lower bound (both for expected and high probability time) for circuits using α inhibitors, for any α. Our lower bound nearly matches our upper bounds for $\alpha = \Omega(\log\log n)$. Missing proofs are deferred to the Appendix.

2 Neural Network Model

A *Spiking Neural Network* (SNN) $\mathrm{N} = \langle \mathrm{X}, \mathrm{Y}, \mathrm{Z}, w, b \rangle$ consists of n input neurons $\mathrm{X} = \{x_1, \ldots, x_n\}$, n output neurons $\mathrm{Y} = \{y_1, \ldots, y_n\}$, and α auxiliary neurons $\mathrm{Z} = \{z_1, ..., z_\alpha\}$. The directed, weighted synaptic connections between X, Y, and Z are described by the weight function $w : [\mathrm{X} \cup \mathrm{Y} \cup \mathrm{Z}] \times [\mathrm{X} \cup \mathrm{Y} \cup \mathrm{Z}] \to \mathbb{R}$. The in-degree of every input neuron x_i is zero.

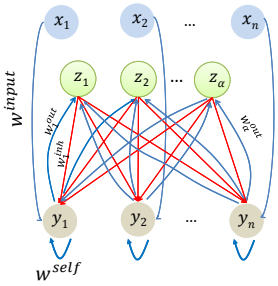

Figure 1 Basic WTA Network structure.

Each neuron is either inhibitory or excitatory: if v is inhibitory, then $w(v,u) \leq 0$ for every u, and if v is excitatory, then $w(v,u) \geq 0$ for every u. Finally, for any neuron v, $b(v) \in \mathbb{R}_{\geq 0}$ is the activation bias – as we will see, roughly, v's membrane potential must reach $b(v)$ for a spike to occur with good probability.

The Basic WTA Network and its Dynamics:

We focus on a restricted class of *basic SNNs*, in which all auxiliary neurons are inhibitory, inputs connect only to their corresponding outputs, and there are no connections within the inhibitory or output layers, aside from an excitatory self-loop from each output to itself. All outputs have identical parameters, i.e., bias values and edge weights.

We introduce some more concise notation to describe basic SNNs. Let $\mathrm{w}^{\mathrm{input}} > 0$ be the synaptic weight from each input x_j to its corresponding output y_j. Let $\mathrm{w}^{\mathrm{self}} > 0$ be the weight of the excitatory self-loop from output y_j to itself. Let $\mathrm{w}^{\mathrm{inh}}{}_j \leq 0$ be the weight of each inhibitory synapse from inhibitor z_j to an output neuron. Conversely, let $\mathrm{w}^{\mathrm{out}}{}_j \geq 0$ be the weight of each excitatory synapse from an output in Y to inhibitor z_j. Finally, let $\mathrm{b}^{\mathrm{out}}$ be the bias value for each output neuron. For an diagram of the basic architecture, see Fig. 1.

The network evolves in discrete, synchronous *rounds* as a Markov chain, with an alternating dynamic between the neurons in X, Y and Z. We give in-depth biological motivation in Appendix A.2. Each round t consists of three sub-rounds denoted by $(t,1),(t,2)$ and $(t,3)$ where the three layers inputs, outputs and inhibitors are scheduled to fire: In the first sub-round $(t,1)$ of each round t, the input layer fires. We consider static inputs so each x_i either fires in every round or does not fire in any round. After that, in sub-round $(t,2)$ the output neurons in Y spike with probabilities dependent on their membrane potentials. Finally, in sub-round $(t,3)$ the inhibitors in Z spike in response to their potentials. The firing probability of every neuron depends on the firing status of its neighboring neurons in the preceding three sub-rounds (i.e., a length of one round). This probabilistic firing is modeled using a standard sigmoid function. For each neuron u, and each round $t \geq 1$, let $u^{(t,k)} = 1$ if u fires (i.e., generates a spike) in sub-round (t,k) for $k \in \{1,2,3\}$. Let $u^{0,k}$ denote the initial firing state of the neuron – we will discuss how this is determined below.

Since each neuron is always scheduled to fire in one of $(t,1)$, $(t,2)$ or $(t,3)$ depending on whether it is in layer X, Y, or Z, for convenience we will often omit the sub-round notation, writing $u^t = 1$ if u fires in *one* of the sub-rounds (t,k). We call u^t, the *firing state* of u in round t. Informally, we say that u *fires in round* t if $u^t = 1$. For each output $y_j \in \mathrm{Y}$ and every $t \geq 1$, let $pot(y_j,t)$ denote the membrane potential at sub-round $(t,2)$ and $p(y_j,t)$

denote the corresponding firing probability. These values are calculated as:

$$pot(y_j,t) = (x_j^{(t,1)}\, \mathrm{w}^{\mathrm{input}}) + (y_j^{(t-1,2)} \cdot \mathrm{w}^{\mathrm{self}}) + \left[\sum_{z_i \in \mathrm{Z}} z_i^{(t-1,3)} \cdot \mathrm{w}^{\mathrm{inh}}{}_i\right] - \mathrm{b}^{\mathrm{out}}$$
$$\text{and } p(y_j,t) = \frac{1}{1+e^{-pot(y_j,t)/\lambda}} \tag{1}$$

where $\lambda > 0$ is a *temperature parameter*, which determines the steepness of the sigmoid. Note that (1) incorporates excitatory and inhibitory effects from any spikes occurring within the three sub-rounds before the outputs spike in sub-round $(t,2)$. Specifically, this includes input spikes in sub-round $(t,1)$ along with output and inhibitory spikes in sub-rounds $(t-1,2),(t-1,3)$ respectively. Also note that when $t=1$, the firing probability depends on the initial firing states $x_j^{(0,1)}$, $y_j^{(0,2)}$ and $z_i^{(0,3)}$. We will discuss how these are determined below. Applying the same rules, in sub-round $(t,3)$, each inhibitor in Z fires with probability $p(z_j,t)$ calculated as:

$$pot(z_j,t) = \left[\sum_{y_i \in \mathrm{Y}} y_i^{(t,2)} \cdot \mathrm{w}^{\mathrm{out}}{}_j\right] - b(z_j) \text{ and } p(z_j,t) = \frac{1}{1+e^{-pot(z_j,t)/\lambda}}. \tag{2}$$

Again (2) incorporates effects from relevant spikes within three sub-rounds $(t-1,3),(t,1)$ and $(t,2)$. However, since inhibitors are connected only to outputs, the only sub-round that affects them is $(t,2)$. After the inhibitors fire, we proceed to round $t+1$, beginning with the firing of the inputs.

We finally specify how the initial firing states are determined. As inputs are static, $x_j^{(0,1)}$ is 1 for firing inputs and 0 for non-firing inputs. $y_j^{(0,2)}$ is arbitrary, while $z_i^{(0,3)}$ is determined as in any regular round according to (2) below with $t=0$ (and so depends on each $y_j^{(0,2)}$). It is not hard to see that is equivalent to just allowing all initial firing states to be arbitrary. This would lead to arbitrary $y_i^{(1,2)}$ and $z_i^{(1,3)}$ determined according to equation (2), which matches our model if we relabel the states in round 1 to be the initial states.

Temperature and Background Noise.

It is clear that the temperature λ does not affect the computational power of the network as we can simply adjust all synapse weights and neuron biases by a factor of λ/λ' to simulate a network with temperature λ'. Hence, we can fix λ to make exposition easier. In our proofs we will always set $\lambda = 1/\Theta(\log n)$. We assume that neurons in Z, Y have bias $b(v) = \Omega(\lambda \log n)$, so they do not fire with probability $1 - 1/(1+e^{-c\cdot\log n}) = 1 - 1/n^c$ when they receive no external stimulation. We call this the *no-background noise* assumption: the network is quiet when no input is introduced. This assumption is used only for technical reasons in our general α inhibitor lower bound. We are hopeful that it could be removed.

System Configuration.

For any $t \geq 1$, the configuration $\mathcal{C}^t = (\mathrm{X}^t, \mathrm{Y}^t, \mathrm{Z}^t)$ in round t is defined by the firing states[3] of the corresponding neurons in round t where $\mathrm{X}^t = [x_1^t, ..., x_n^t]$ and Y^t and Z^t are defined analogously. Recall that $x_i^t = 1, y_i^t = 1, z_i^t = 1$ if the input x_i (output y_i, inhibitor z_i) fires in

[3] The firing state of a neuron is a binary number indicating if it is firing or not.

sub-round $(t, 1)$ (resp., $(t, 2), (t, 3)$). We consider a static input setting where $\mathrm{X}^t = \mathrm{X}$ for all t.[4] We abuse notation slightly, thinking of X as a vector of binary input values where $x_j = 1$ indicates that x_j fires in every round ($x_j^t = 1$ for all t) and $x_j = 0$ implies that x_j never fires ($x_j^t = 0$ for all t). We denote the the initial configuration $\mathcal{C}^0$. As discussed we have $\mathrm{X}^0 = \mathrm{X}$, Y^0 arbitrary, and Z^0 determined as in any round according to equation (2).

The WTA Problem.

A binary winner-take-all network given n inputs should converge to having a single firing output corresponding to a firing input (the 'winner'), if one exists. Formally, given $\mathrm{X} \in \{0,1\}^n$, let $f(\mathrm{X}) = \{\mathrm{Y} \in \{0,1\}^n \mid y_i \leq x_i \ \forall i \text{ and } \| \mathrm{Y} \|_1 = \min(1, \| \mathrm{X} \|_1)\}$ where $\| \cdot \|_1$ is the standard 1-norm, used to denote the number of firing neurons in a set.

We say N *satisfies WTA* in round t if $\mathrm{Y}^t \in f(\mathrm{X})$. We say N *converges to WTA* in t rounds with probability $1 - \delta$ if for every input $\mathrm{X} \in \{0,1\}^n$ and every initial output configuration Y^0, with probability at least $1 - \delta$, $\mathrm{Y}^t \in f(x)$ and $\mathrm{Y}^{t'} = \mathrm{Y}^t$ for all $t' \in [t+1, t+n^c]$ where c is a positive constant.[5] That is, the network satisfies WTA in round t and maintains the satisfying configuration for polynomial in n subsequent rounds. As our neurons are inherently probabilistic, our definition of convergence is as well – we will never be able to avoid occasional random deviations from a correct output state and so just demand that the state is maintained a large number of rounds.

We let $\mathcal{ET}(\mathrm{N})$ denote the maximum expected time required to converge to WTA, taken over all possible inputs X and initial output configurations Y^0. In the same manner, $\mathcal{HT}(\mathrm{N})$ denotes the maximum time required for convergence to WTA with high probability.[6]

3 Warm Up: Two Simple Networks for WTA

We begin by presenting two WTA networks that represent two extremes of the inhibitor-time tradeoff. They also illustrate the rough intuition that will appear in our later network constructions and lower bound strategies.

WTA with Two Inhibitors.

In our two inhibitor network we have $\mathrm{Z} = \{z_s, z_c\}$. The neuron z_s is a *stability* inhibitor that maintains the WTA state once it has been reached. It fires w.h.p. in sub-round $(t, 3)$ whenever at least one output fires in sub-round $(t, 2)$. The neuron z_c is a *convergence* inhibitor that fires w.h.p. whenever WTA has not yet been reached – i.e. whenever ≥ 2 outputs fire in sub-round $(t, 2)$.

We set the weights connecting z_s and z_c to the outputs such that when both fire in round t, any output that fired in round t will fire with probability $1/2$ in round $t+1$. Any output that *did not fire* in round t will not fire in round $t+1$ w.h.p. as it will not have an active excitatory self-loop and so its membrane potential will be too low to overcome the inhibition.

---

[4] Note however that our model can easily handle non-static inputs. All algorithms given will converge from an arbitrary initial configuration and so will converge if X changes.

[5] Formally, a family if networks $\mathcal{N} = \{N(n)\}$ for all integers $n \geq 1$ converges to WTA in $t(n)$ rounds with probability $1 - \delta$ if there exists $c > 0$ such that for all n, for all $X \in \{0,1\}^n$ and for all $Y^0 \in \{0,1\}^n$, with probability at least $1 - \delta$, $N(n)$ satisfies WTA in rounds $[t(n), ..., t(n) + n^c]$.

[6] Throughout, *with high probability* (w.h.p.) refers to events occuring with probability $\geq 1 - 1/n^c$ for constant c. Formally, a family of events $\mathcal{E} = \{E(n)\}$ for all integers $n \geq 1$ occurs w.h.p. if there exists $c > 0$ such that for all n, $\Pr[E(n)] \geq 1 - 1/n^c]$.

In this way, as long as ≥ 2 outputs fire in round t, both inhibitors fire w.h.p. and the high level of inhibition causes outputs to 'drop out of contention' for the winning position with probability $1/2$. After $O(\log n)$ rounds, nearly all the outputs stop firing and with constant probability there is a round in which exactly 1 output fires. Once this round occurs, z_c ceases firing w.h.p. and just z_s fires. This decreased level of inhibition allows the winner to keep firing, as it is offset by the winner's excitatory self-loop. However, it prevents any other output, whose excitatory self-loop is inactive, from firing w.h.p. See Fig. 2 in Appendix B.1 for illustration of the network with its edge weights. We analyze the network in depth in B.1, showing convergence given any input X and initial output configuration Y^0, and yielding:

▶ **Theorem 3.** *There exists a basic WTA network* N *with* $\alpha = 2$ *inhibitors and* $\mathcal{ET}(\mathrm{N}) = O(\log n)$ *and* $\mathcal{HT}(\mathrm{N}) = O(\log^2 n)$.

In Appendix B.1, we show that the network is optimal up to a $\log \log n$ factor and in Appendix B.2 we show that it represents a critical point in the inhibitor-time tradeoff: any network with just one inhibitor requires $\Omega(n^c)$ rounds to solve WTA. Essentially, it is not possible for a single inhibitor to implement the two opposing tasks of stability and convergence.

WTA with $O(\log n)$ Inhibitors.

Our second network represents another extreme point of the inhibitor-time tradeoff, using $\alpha = O(\log n)$ inhibitors to achieve $O(1)$ expected convergence time.

The idea is to approximate the ideal behavior in which outputs fire with probability $1/k_t$ in round $t+1$ if k_t outputs fired in round t. As in our two inhibitor algorithm, we have a single stability inhibitor z_s that fires w.h.p. whenever at least one output fires and insures that as soon as a single output fires in a round, the network converges to WTA. We then have $\lceil \log n \rceil - 1$ convergence inhibitors $z_1, ..., z_{\alpha-1}$. We set the bias of the z_i to $b(z_i) = 2^i - .5$ and set $\mathrm{w^{out}}_i = 1$ for all i. In this way, z_i fires w.h.p. in round t whenever $\geq 2^i$ outputs fire. We set the inhibitor to output weights to $\mathrm{w^{inh}}_i = \Theta(\lambda)$ for all i. Thus, when $k_t \in [2^i, 2^{i+1})$, w.h.p. inhibitors $z_1, ..., z_i$ all fire (while $z_{i+1}, ..., z_{\alpha-1}$ do not). The total inhibition from the inhibitors is thus $\Theta(i\lambda)$ and hence each of the k_t outputs fire with probability $1/(1 + e^{\Theta(i)}) \approx 1/2^i \approx 1/k_t$ in round $t+1$. In expectation (and with constant probability) there will be exactly one firing output, giving an expected runtime of just $O(1)$ rounds to reach WTA. In Appendix B.3, we give a full analysis, yielding:

▶ **Theorem 4.** *There exists a basic WTA network* N *with* $\alpha = O(\log n)$ *inhibitors,* $\mathcal{ET}(\mathrm{N}) = O(1)$ *and* $\mathcal{HT}(\mathrm{N}) = O(\log n)$.

Vacuously, no network can beat this expected runtime. We also show in Appendix B.3 that no network can do better with high probability: even with an unlimited number of inhibitors, $\Theta(\log n)$ rounds are requires to solve WTA w.h.p. Intuitively, as long as WTA has not yet been reached in round t, there is no single distinguished output. All outputs have identical connections to X, Z so each active output fires with the *same* probability p in round $t+1$. Hence the probability that a single output becomes distinguished (is the only one to fire) is $k_t \cdot p(1-p)^{k_t - 1}$, which is bounded by a constant for all k_t, p. Thus, converging to the WTA state w.h.p. takes at least $\Omega(\log n)$ rounds.

4 WTA with $\alpha \geq 2$ Inhibitors

The above results give a rough outline of the tradeoff between the number of inhibitors and the runtime for WTA. We now explore this tradeoff in more depth for general $\alpha \in (2, \log n]$

4.1 Upper Bound Networks

We first show that both our two inhibitor and $\lceil \log n \rceil$ inhibitor networks can be improved significantly with modest increases in the number of inhibitors or runtime used. We can (up to constant factors) match the runtime of the $\lceil \log n \rceil$ inhibitor network with just $O(\log^{1/c} n)$ inhibitors for any c. Additionally, for any $\alpha \geq \log\log n$ we can achieve expected runtime $O\left(\frac{\log\log n \log\log\log n}{\log \alpha}\right)$, nearly matching our main lower bound of Section 4.2.

▶ **Theorem 5.** *For any integer θ, there is a basic WTA network* N *with $\alpha = O(\theta \log^{1/\theta} n)$ inhibitors, $\mathcal{ET}(\mathrm{N}) = O(\theta)$, and $\mathcal{HT}(\mathrm{N}) = O(\theta \log n)$.*

For $\alpha \geq \log\log n$, writing $\alpha = \log\log^x n$ for $x \geq 1$ if we set $\theta = \frac{c_1 \log\log n \log\log\log n}{\log \alpha} = \frac{c_1 \log\log n}{x}$ then the number of inhibitors required is: $\frac{c_1 \log\log n}{x} \cdot e^{x/c_1} \leq \log\log^x n \leq \alpha$ for small enough c_1.

Proof Sketch. To see the high level idea, consider the case of $\theta = 2$. We will $2\sqrt{\log n}$ inhibitors which are divided into two classes: $\sqrt{\log n}$ coarse inhibitors and $\sqrt{\log n}$ fine inhibitors. The edges from the fine inhibitors to outputs have weight -1 and the edges from coarse inhibitors to outputs have weight $-\sqrt{\log n}$. All the edges from the outputs to the inhibitors have weight 1. We set the bias values of the inhibitors such that: (1) the i^{th} coarse inhibitor fires if the number of active outputs is at least $2^i\sqrt{\log n}$ and (2) the i^{th} fine inhibitor fires if the number of active outputs is at least 2^i. Consider any output density 2^d and let $d' = \lfloor d/\sqrt{\log n} \rfloor$. When 2^d outputs fire in round t, this will excite the first d' coarse inhibitors. As a result, the firing probability for the outputs in round $t+1$ will be approximately $2^{-d' \cdot \sqrt{logn}}$ (ignoring negligible effects from the fine inhibitors). In other words, within a single round the density will be reduced from 2^d to $2^{d - d'\sqrt{\log n}}$ which is a new density in the range $1, 2, 4, ..., 2^{\sqrt{\log n}}$. After this initial round, since at most $2^{\sqrt{\log n}}$ outputs fire, the circuit converges in constant rounds in expectation as the $\sqrt{\log n}$ fine inhibitors can induce probabilities roughly equal to $1/k_t$ just as is done in the $O(\log n)$ inhibitor circuit.

Generalization to larger θ is by repeating the above construction: we have θ levels of increasing coarseness: $[1, 2^{\log^{1/\theta} n}], [2^{\log^{1/\theta} n}, 2^{\log^{2/\theta} n}], ..., [2^{\log^{(\theta-1)/\theta} n}, 2^{\log n}]$. The $\log^{1/\theta} n$ inhibitors at each level ensure that if the number of firing outputs is at level i in round t, it is reduced to level $i-1$ in round $t+1$, yielding $O(\theta)$ expected runtime. We give a full analysis in Appendix B.4. ◀

Our second construction uses similar techniques, but uses just one convergence inhibitor per density class, balancing the time required to move through each density class and the number of classes used. It significantly improves on our two inhibitor algorithm, achieving runtime $O(\log^{1/c} n)$ for any constant c with $O(1)$ inhibitors and $O(\log\log n)$ runtime with $O(\log\log n)$ inhibitors.

▶ **Theorem 6.** *For any $\alpha \geq 2$, there is a basic WTA network* N *with α inhibitors, $\mathcal{ET}(\mathrm{N}) = O\left(\alpha \log^{1/(\alpha-1)} n\right)$ and $\mathcal{HT}(\mathrm{N}) = O\left(\alpha \log^{1+1/(\alpha-1)} n\right)$.*

Proof Sketch. Consider $\alpha = 3$. We have 2 convergence inhibitors: a fine inhibitor z_f and a coarse inhibitor z_c. The inhibitor z_c fires whenever the number of active outputs is at least $2^{\sqrt{\log n}}$, and induces outputs to fire with probability $1/2^{\sqrt{\log n}}$ in the next round. In this way, starting with any density of firing inputs $k_t \in [2^{\sqrt{\log n}}, n]$, within $\sqrt{\log n}$ rounds the density will be reduced to $\leq 2^{\sqrt{\log n}}$. The inhibitor z_f fires whenever at least 2 outputs fire, and induces outputs to fire with probability $1/2$ in the next round. So, within $\sqrt{\log n}$

additional rounds, with constant probability just a single output will remain firing. Again, a full network description for general α and proof is given in Appendix B.4. ◀

4.2 Lower Bound: The Tradeoff Between Inhibitors and Time

We now present our main lower bound which matches Theorem 5 up to $\log\log\log n$ factors.

▶ **Theorem 7.** *For any basic WTA network* N *with* α *inhibitors,* $\mathcal{ET}(\mathrm{N}) = \Omega\left(\frac{\log n \log n}{\log \alpha}\right)$ *and* $\mathcal{HT}(\mathrm{N}) = \Omega\left(\frac{\log\log n}{\log \alpha} \cdot \frac{\log n}{\log\log\log n}\right)$.

Lower Bound Overview.

We focus on initial output configuration $Y^0 = \vec{0}$ (i.e., no output fires in the sub-round $(0, 2)$) which we call the *reset configuration*. We show that for any network N with α inhibitors there exists at least one input X for which the expected time to reach WTA starting from the reset configuration is $\Omega(\log\log n/\log\alpha)$. If suffices to consider the case where $\alpha = O(\log^{1/c} n)$ for some constant c since for $\alpha = \Theta(\log^{1/c} n)$, the expected runtime is $O(1)$. Throughout this section, we say an event happens with *good probability* if its probability is at least $1 - O(\log^4 n)$.

Our argument contains two main parts. First, we show that the inhibitors fire in a *nearly* deterministic manner and hence we can treat them (up to some slack) as *threshold circuits*. Equipped with this property, we then consider $\Theta(\log n)$ *density classes* each covering a constant multiplicative range of firing outputs. The predictable behavior of the inhibitors is used to show that even after $\Omega(\log n \log n/\log\alpha)$ rounds, the network cannot distinguish between at least two different density classes, which yields our claim as it does not converge to WTA for at least one class.

(1) Inhibitor classification: inhibitors are nearly deterministic for most density classes.

To address the first challenge (i.e., showing that inhibitors are predictable), we divide the set of inhibitors Z into three classes and show the predictability property for each class separately. The "stability" class (or "WTA preservers") S contains inhibitors whose *goal* is to maintain the WTA steady state. The "convergence" class (or "progress inhibitors") C contains the inhibitors that are responsible for driving fast convergence to a WTA state. Finally, the third class R contains the remaining inhibitors whose contribution to both stability and convergence is negligible.

Formally, for any inhibitor $z_i \in \mathrm{Z}$ and $j \in [1, n]$ let $pot_j(z) = j \cdot \mathrm{w^{out}}_i - b(z_i)$ be the potential of z_i when exactly j outputs fire (I.e., if in sub-round $(t, 2)$ the number of firing outputs is j, then the potential of z_i in sub-round $(t, 3)$ is $pot_j(z)$ and it fires in sub-round $(t, 3)$ with probability $1/(1 + e^{-pot_j(z)})$). The set S contains all inhibitors that fire in steady state (i.e., when exactly one output is firing) with reasonably high probability. Fixing some constant $c \geq 1$, $S = \{z_i \in \mathrm{Z} \mid 1/(1 + e^{-pot_1(z_i)}) \geq 1/\log^{3c} n\}$. The set C is comprised of all inhibitors $z_i \notin S$ whose firing probability is least $1/\log^c n$ when all n outputs fire in the previous sub-round: $C = \{z_i \in \mathrm{Z} \mid z_i \notin S \text{ and } 1/(1 + e^{-pot_n(z_i)}) \geq 1/\log^c n\}$[7]. Finally, R contains all remaining inhibitors not in S or C.

[7] The difference between $1/\log^{3c} n$ when defining the threshold for the inhibitors in S and $1/\log^c n$ when defining the threshold for the inhibitors C, is crucial in the analysis.

We show that the firing states of the inhibitors can *in certain cases* be predicated with good probability. The argument for each of the three classes S, C and R is different and is presented in Appendix B.5.1. Since the inhibitors in S fire with good probability when just one output fires, we can show that they fire w.h.p. when at least two outputs fire:

▶ **Lemma 8** (S is predictable)**.** *Let $(t,2)$ be a sub-round in which at least two outputs fire, then sub-round $(t,3)$, all inhibitors of S fire with probability at least $1-1/n$.*

Since the firing probability of the R inhibitors is small compared to the $O(\log\log n/\log\alpha)$ execution length that we care about, we have:

▶ **Lemma 9** (R is predictable)**.** *Given any input* X *and any initial configuration, with probability at least $1-1/\log^{c-3} n$, none of the inhibitors in R fire in $O(\log^2 n)$ rounds of execution of* N*.*

Perhaps the most surprising claim concerns the predictability of the convergence inhibitors:

▶ **Lemma 10** (C is almost predictable)**.** *For every $z \in C$, there exists an integer $k(z) \in [1,n]$, such that for $c \geq 4$:*

(1) Low Density: When there are at most $k(z)/2$ firing outputs in sub-round $(t,2)$, the probability that z fires in sub-round $(t,3)$ is at most $1/\log^c n$ (i.e., with good probability, z does not fire);

(2) High Density: When there are at least $2k(z)$ firing outputs in sub-round $(t,2)$, the probability that z fires in sub-round $(t,3)$ is at least $1-1/\log^c n$ (i.e., with good probability, z fires).

Overall, except for the case where the number of firing outputs in sub-round $(t,2)$ is in the density class $K(z) = [k(z)/2, k(z)]$, z behaves in sub-round $(t,3)$ in an almost deterministic manner. Roughly speaking, this is shown by exploiting the *gap* in the firing probabilities of these inhibitors between the steady state rounds (when they fire with probability $\leq 1/\log^{3c} n$) and the rounds in which there are sufficiently many firing outputs (where they fire with probability $\geq 1/\log^c n$). The proof of Lemma 10 shows that this gap implies that the sigmoid function which converts the number of firing inputs to z's firing probability must be steep enough such that z has predictable behavior outside a small range around $k(z)$.

(2) Network prediction for nearly deterministic inhibitors.

Using the predictable nature of the inhibitors, we now show that there is at least one *density class* of competing inputs for which we can predict (with good probability) the behavior of N for $\Omega(\log\log n/\log\alpha)$ rounds, at the end of which the WTA state has not been reached. We consider a set of $\ell = \lfloor \log n \rfloor$ inputs $\mathcal{X} = \{\mathrm{X}_1, ..., \mathrm{X}_\ell\}$ where X_i contains exactly 2^i firing inputs (i.e. $\| \mathrm{X}_i \|_1 = 2^i$). Thus, $\mathcal{X}$ contains a representative input from each density class of input vectors whose number of firing inputs is within a factor two of each other.

For any $\mathrm{X} \in \mathcal{X}$ let $\widehat{R}_t(\mathrm{X}) \in \{1, \ldots, n\}$ be the random variable indicting the number of firing outputs in sub-round $(t,2)$ starting from the initial configuration $Y_0 = \vec{0}$. Let $\widehat{F}_t(\mathrm{X}) \in \{0,1\}^\alpha$ be the random variable indicating the firing status of the inhibitors in sub-round $(t,3)$. For each $\mathrm{X} \in \mathcal{X}$ we will attempt to maintain a *predicted* range $R_t(\mathrm{X})$ of the number of firing outputs in sub-round $(t,2)$ along with a *predicted* inhibitor configuration in sub-round $(t,3)$, $F_t(\mathrm{X})$. We will let $\mathcal{X}_t \subseteq \mathcal{X}$ denote the subset of inputs whose behavior we can predict well in (all sub-rounds of) round t – specifically, for which we know $\widehat{R}_t(\mathrm{X}) \in R_t(\mathrm{X})$ and $\widehat{F}_t(\mathrm{X}) = F_t(\mathrm{X})$ with good probability (at least $1-1/\log n$).

For any inhibitor $z \in C$, we call the range $K(z) = [k(z)/2, 2k(z)]$– the *critical range* of z (see Lemma 10 for the definition of $k(z)$). If the number of firing outputs enters this range,

we will not be able to predict the behavior of z in the next sub-round with good probability. On the other hand, as long as the number of firing outputs in sub-round $(t, 2)$ is not in the critical range of any $z \in C$, then the firing behavior of the inhibitors in sub-round $(t, 3)$ can be predicted with good probability.

We will progress through rounds, predicting the behavior of N in round t for each input in $\mathcal{X}_{t-1}$ based off the predictions in round $t-1$. We will ensure that in any round, not too may inputs have predicted ranges overlapping critical regions by ensuring that these predicted ranges remain separated by constant factors and hence, at most $|C|$ of them can overlap $K(z)$ for some $z \in C$.

Predicting the number of firing outputs given inhibitor states.

We now describe how to predict the range $R_t(X)$ given the prediction $F_{t-1}(X)$. Our main goal is to preserve the separation between the predicted ranges $R_t(\mathrm{X})$ for sufficiently many inputs $\mathrm{X} \in \mathcal{X}_{t-1}$.

To maintain the separation, we consider only the largest subset $\mathcal{X}_t^{same} \subseteq \mathcal{X}_{t-1}$ of inputs whose predicted firing configuration for the inhibitors in the previous sub-round $(t-1, 3)$ is exactly the *same* (i.e., inputs X with the same $F_{t-1}(\mathrm{X})$ vector). By doing this, we guarantee that the firing probabilities of all the outputs in sub-round $(t, 2)$ is the same. Letting this probability be p, the expected number of firing outputs in sub-round $(t, 2)$ is in the range $p \cdot R_{t-1}(\mathrm{X})$ for each $\mathrm{X} \in \mathcal{X}_t^{same}$ and the separation between these ranges is preserved in expectation. To show that the ranges are also separated with good probability, we omit from $\mathcal{X}_t^{same}$ at most $\Theta(\log \log n)$ inputs with ranges $R_t(\mathrm{X})$ containing values $\leq \log^c n$ for some constant c. They remaining inputs thus have output ranges concentrated around their expectation. The key point to observe is that because the inhibitors behave almost as threshold circuits, the number of different firing configurations in sub-round $(t-1, 3)$ is at most α (i.e., there are at most α different $F_{t-1}(\mathrm{X})$ vectors for $\mathrm{X} \in \mathcal{X}_{t-1}$) and hence the cardinality of the set $\mathcal{X}_t^{same}$ for which we predict the range of firing outputs in sub-round $(t, 2)$ is at least $|\mathcal{X}_{t-1}|/\alpha$.

Predicting the inhibitor states given the number of firing outputs.

We next describe how to predict the inhibitor firings $F_t(\mathrm{X})$ given the prediction $R_t(\mathrm{X})$. Since the convergence inhibitors are predictable when the number of firing outputs is not in any critical range $K(z)$, we first omit from $\mathcal{X}_t^{same}$ all inputs X whose predicted range $R_t(\mathrm{X})$ intersects the critical range of some $z \in C$ (i.e. $R_t(\mathrm{X}) \cap K(z) \neq \emptyset$ for some z). We call the resulting set $\mathcal{X}_t$. Since the ranges of $\mathcal{X}_t^{same}$ are separated by some constant, we do not discard more than $|C| = O(\alpha)$ inputs.

Overall, we predict the circuit behavior in sub-rounds $(t, 2), (t, 3)$ with good probability for all inputs $\mathrm{X} \in \mathcal{X}_t$ where $|\mathcal{X}_t| \geq |\mathcal{X}_{t-1}|/\alpha - \alpha$. Since $\alpha = O(\log^{1/c} n)$, we get that after t rounds, there are $|\mathcal{X}_t| = \Omega(\log n/\alpha^t)$ inputs for which the network behaves *exactly* the same in each of the t rounds with good probability. This argument proceeds as long as $\log n/\alpha^t \geq 2$, leading to the lower bound of expected time $\Omega(\log \log n/\log \alpha)$ since we can show if two inputs are not distinguished, at least one will not have reached WTA. In Appendix B.5.2, we describe the prediction process in detail and complete the proof of Theorem 7.

High Probability Lower Bound.

Finally, we show that our lower bound for expected runtime extends to a lower bound on the high probability runtime. Our lower bound implies that “repeating" the execution of a

network that converges with constant probability $\Theta(\log n)$ times to achieve a high probability guarantee is essentially the best one can do (up to a $\log\log\log n$ factor).

▶ **Lemma 11.** *For any basic WTA network* N *with* α *inhibitors* $\mathcal{HT}(\mathrm{N}) = \Omega(\frac{\log n \cdot \log\log n}{\log\alpha \log\log\log n})$.

Proof Sketch. Let $DC = \Theta\left(\frac{\log\log n}{\log\alpha}\right)$ and $DH = DC \cdot \left(\frac{\log n}{\log\log\log n}\right)$. Fix a network N with α inhibitors and let X be the input for which, by Theorem 7, N requires at least DC rounds in expectation starting from initial configuration $\mathcal{C}_0$ with input X and $\mathrm{Y}^0 = \vec{0}$. In the following proof, we will actually exploit the fact that the lower bound in Theorem 7 applies to the time it takes to reach a WTA state with *constant probability* (a stronger time measure than expected time).

We work with the *execution tree* T which includes all possible DH round executions of N starting from $\mathcal{C}_0$. The tree T has depth DH where each layer corresponds to the configuration of the network in each round t. Each node u at level t is labeled by an $(n+\alpha)$-length binary vector $Q(u)$ describing the firing states of the outputs and inhibitors in round t, i.e., the firing states of the outputs in sub-round $(t,2)$ and the firing states of the inhibitors in sub-round $(t,3)$. Node u has $2^{n+\alpha}$ children, with the edge to each child labeled with the transition probability between the configuration in u to the child configuration. The root node r is labeled with $\mathcal{C}_0$. The mass of node u is given by the product of edge weights on its path to r. It is the probability of reaching u's configuration through that execution path. We call a node u a *reset* node (resp., *WTA* node), if in the configuration $Q(u)$ no output fires (resp., exactly one output with active input fires).

To lower bound $\mathcal{HT}(\mathrm{N})$ we will show that the probability to reach a non-WTA leaf node when starting from the root r is at least $1/n^2$, and thus the probability to reach a WTA leaf node is at most $1 - 1/n^2 < 1 - 1/n^c$, contradicting a w.h.p. runtime of $\leq DH$ rounds.

Our strategy is based on traversing the tree in an asynchronous manner from the root to (sufficiently many) non-WTA leaf nodes with sufficiently high total probability mass. For a given node u in layer t, we may move to a subset of its *non-WTA* children nodes in layer $t+1$. We call this move a *small jump*. Alternatively, we may make a *large jump*, moving DC steps from u and proceeding the traversal from a subset of *non-WTA* leaf nodes of $T_{DC}(u)$ (the height DC subtree rooted at u). With each jump starting at u, we loose some probability mass – the idea is to show that we do not loose it too quickly.

In more detail, in each step of our traversal, we maintain a collection of non-WTA nodes. When arriving a node u in the traversal, we consider its configuration $Q(u)$ and look at the probability that the next round is a *reset* round (with 0 firing outputs) given $Q(u)$. We show that if the probability of having at most 1 firing outputs in the next round is $\geq 1/\log\log n$, the probability of having a reset (no firing outputs) is large – i.e., $\geq 1/(\log\log n)^3$.

In this case we continue traversal only from the children of u that are reset nodes. For each of these children v, let $T_{DC}(v)$ be the execution tree of depth DC rooted at v. By the lower bound in Theorem 7, the probability to reach a non-WTA leaf node in $T_{DC}(v)$ starting from $Q(v)$ is at least a constant. So from each reset-node v, we make a large jump to the leaves of $T_{DC}(v)$. Overall, we maintain a $\Theta(1/(\log\log n)^3)$ fraction of the probability mass of u in making this large jump. Since such a jump can occur at most $DH/DC = \log n/\log\log\log n$ times, we maintain at least a $1/(\log\log n)^{3DH/DC} \geq 1/n^2$ fraction of the probability mass throughout the traversal.

On the other hand, when arriving a node u for which the probability of having at most 1 firing output in the next round is less than $1/\log\log n$, we make a small jump to the children of u in which the number of firing outputs is at least 2 (and hence which are non-WTA nodes). This jump maintains $1 - 1/\log\log n$ of the probability mass and since such a jump

can happen at most DH times, we again maintain $(1-1/\log\log n)^{DH} \geq 1/n^2$ of the original probability. Overall, through making both large and small jumps, at the end of the traversal, we reach a set of non-WTA nodes containing at least a $1/n^2$ fraction of the probability mass in the DH level execution tree. This gives us our high probability time lower bound. See Appendix B.6 for a complete analysis and Fig. 3 for an illustration of the execution tree. ◀

In Appendix C, we extend our lower bounds (for both expected and high probability time) to the case where the α auxiliary neurons can be both excitatory and inhibitory neurons. The more general bound holds under the restriction that outputs with no active input are not allowed to fire during the execution. Only competing outputs (that have a positive signal from their inputs) ever fire.

5 Discussion

We hope that this paper is a starting point for further investigation into stochastic spiking networks from an algorithmic perspective, which investigates fundamental tradeoffs between biological resources and identifies basic building blocks and principles for algorithm design in neural settings.

We focus on a restricted class of three layer networks, in which auxiliary neurons are not interconnected. This models the generally restricted connectivity structure that inhibitory neurons appear to have in biological networks and lets us give both very strong upper bounds and matching lower bounds. Still, it would be interesting to understand the effect of connections between auxiliary neurons. We have preliminary work showing that some speedups are possible in these more general networks, however obtaining any non-trivial lower bounds would be very interesting.

Studying other important primitives aside from the binary version of WTA that we focus on would also be interesting. We again have preliminary work on *non-binary WTA* in which the network must choose the input with the highest, or near highest firing rate as the winner. There are many other problems to consider.

Our model attempts to be biologically plausible enough to capture high level behavior, yet not be overly complex. However, many modeling assumptions are possible, and we hope that future work explores if changes to the model can lead to significant differences in computational power or algorithmic techniques. As an example, for simplicity we considered a synchronous model, however, asynchrony seems to be an important part of neural computation which would be valuable to study.

Finally, we note that significant theoretical work attempts to understand how neural networks can *learn* through the modification of synapse weights as their endpoints fire more or less frequently [46, 36]. The most common model for how synapse weights evolve is the *hebbian learning* rule, which is itself the focus of a vast literature. Merging the view of neural networks as executing algorithms given predetermined network parameters with understanding of learning would be very interesting. Can a WTA network 'evolve' naturally via simple learning rules? How do fixed network motifs such as WTA circuits interact with more flexible 'learning' networks?

Acknowledgments

We are grateful to Mohsen Ghaffari for noting the general upper bound network construction and for many helpful discussions on the lower bound proof. We would also like to thank Nir Shavit, Rati Gelashvili, and Sergio Rajsbaum for insightful discussions.

References

1 David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for boltzmann machines. *Cognitive science*, 9(1):147–169, 1985.

2 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.

3 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.

4 Sander M Bohte, Joost N Kok, and Han La Poutre. Error-backpropagation in temporally encoded networks of spiking neurons. *Neurocomputing*, 48(1):17–37, 2002.

5 Romain Brette, Michelle Rudolph, Ted Carnevale, Michael Hines, David Beeman, James M Bower, Markus Diesmann, Abigail Morrison, Philip H Goodman, Frederick C Harris Jr, et al. Simulation of networks of spiking neurons: a review of tools and strategies. *Journal of computational neuroscience*, 23(3):349–398, 2007.

6 Lars Buesing, Johannes Bill, Bernhard Nessler, and Wolfgang Maass. Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS Comput Biol*, 7(11):e1002211, 2011.

7 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.

8 Shlomi Dolev. *Self-stabilization*. MIT press, 2000.

9 Shlomi Dolev, Amos Israeli, and Shlomo Moran. Uniform dynamic self-stabilizing leader election. *IEEE Transactions on Parallel and Distributed Systems*, 8(4):424–440, 1997.

10 A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9(4):292–303, 2008.

11 Michael Fischer and Hong Jiang. Self-stabilizing leader election in networks of finite-state anonymous agents. In *International Conference On Principles Of Distributed Systems*, pages 395–409. Springer, 2006.

12 Wulfram Gerstner and Werner M Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002.

13 Sonia M Gómez-Urquijo, Concepción Reblet, José L Bueno-López, and Iñaki Gutiérrez-Ibarluzea. Gabaergic neurons in the rabbit visual cortex: percentage, layer distribution and cortical projections. *Brain research*, 862(1):171–179, 2000.

14 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.

15 Stefan Habenschuss, Zeno Jonke, and Wolfgang Maass. Stochastic computations in cortical microcircuit models. *PLoS Comput Biol*, 9(11):e1003311, 2013.

16 John J Hopfield, David W Tank, et al. Computing with neural circuits- a model. *Science*, 233(4764):625–633, 1986.

17 Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.

18 Eugene M Izhikevich. Which model to use for cortical spiking neurons? *IEEE transactions on neural networks*, 15(5):1063–1070, 2004.

19 Zeno Jonke, Stefan Habenschuss, and Wolfgang Maass. Solving constraint satisfaction problems with networks of spiking neurons. *Frontiers in neuroscience*, 10, 2016.

20 Samuel Kaski and Teuvo Kohonen. Winner-take-all networks for physiological models of competitive learning. *Neural Networks*, 7(6-7):973–984, 1994.

21 Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pages 115–141. Springer, 1987.

22 John Lazzaro, Sylvie Ryckebusch, Misha Anne Mahowald, and Caver A Mead. Winner-take-all networks of o (n) complexity. Technical report, DTIC Document, 1988.

23 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.

24 Nancy Lynch. A hundred impossibility proofs for distributed computing. In *Proceedings of the eighth annual ACM Symposium on Principles of distributed computing*, pages 1–28. ACM, 1989.

25 Nancy A Lynch. *Distributed algorithms*. Morgan Kaufmann, 1996.

26 Wolfgang Maass. On the computational power of noisy spiking neurons. *Advances in neural information processing systems*, pages 211–217, 1996.

27 Wolfgang Maass. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9):1659–1671, 1997.

28 Wolfgang Maass. Neural computation with winner-take-all as the only nonlinear operation. In *NIPS*, pages 293–299. Citeseer, 1999.

29 Wolfgang Maass. On the computational power of winner-take-all. *Neural computation*, 12(11):2519–2535, 2000.

30 Wolfgang Maass. Noise as a resource for computation and learning in networks of spiking neurons. *Proceedings of the IEEE*, 102(5):860–880, 2014.

31 Marvin Minsky and Seymour Papert. Perceptrons. 1969.

32 Steven J Nowlan. Maximum likelihood competitive learning. In *NIPS*, pages 574–582, 1989.

33 Matthias Oster, Rodney Douglas, and Shih-Chii Liu. Computation with spikes in a winner-take-all network. *Neural computation*, 21(9):2437–2465, 2009.

34 Matthias Oster and Shih-Chii Liu. Spiking inputs to a winner-take-all network. *Advances in Neural Information Processing Systems*, 18:1051, 2006.

35 Christos H Papadimitriou and Santosh S Vempala. Unsupervised learning through prediction in a model of cortex. *arXiv preprint arXiv:1412.7955*, 2014.

36 Christos Papadimitrou, Samantha Petti, and Santosh Vempala. Cortical computation via iterative constructions. *arXiv preprint arXiv:1602.08357*, 2016.

37 Josep L Rossello, Vincent Canals, Antoni Morro, and Antoni Oliver. Hardware implementation of stochastic spiking neural networks. *International journal of neural systems*, 22(04):1250014, 2012.

38 Lisa Roux and György Buzsáki. Tasks for inhibitory interneurons in intact brain circuits. *Neuropharmacology*, 88:10–23, 2015.

39 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.

40 BL Sabatini and WG Regehr. Timing of synaptic transmission. *Annual Review of Physiology*, 61(1):521–542, 1999.

41 H Sebastian Seung. Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron*, 40(6):1063–1073, 2003.

42 Michael N Shadlen and William T Newsome. Noise, neural codes and cortical organization. *Current opinion in neurobiology*, 4(4):569–579, 1994.

43 Simon J Thorpe. Spike arrival times: A highly efficient coding scheme for neural networks. *Parallel processing in neural systems*, pages 91–94, 1990.

44 Leslie G Valiant. *Circuits of the Mind*. Oxford University Press on Demand, 2000.

45 Leslie G Valiant. A neuroidal architecture for cognitive computation. *Journal of the ACM (JACM)*, 47(5):854–882, 2000.

46 Leslie G Valiant. Memorization and association on a realistic neural model. *Neural computation*, 17(3):527–555, 2005.

47 Wei Wang and Jean-Jacques E Slotine. K-winners-take-all computation with neural oscillators. *arXiv preprint q-bio/0401001*, 2003.

48 Alan L Yuille and Norberto M Grzywacz. A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Computation*, 1(3):334–347, 1989.

A Additional Discussion

A.1 Related Work

Spiking Neural Networks.

A vast literature studies computation in stochastic spiking neural networks. Work includes detailed models aimed at matching biological observations [12, 18], large scale simulation in hardware and software [5, 37], attempts to understand general properties of computation in these networks [6], the design of specific algorithms [4, 41], and theoretical investigation of computational power [26, 15]. For instance, it has been shown that deterministic spiking networks can simulate Turing machines and that stochastic spiking networks can implement MCMC sampling [6]. As is popular in the biologically-inspired algorithms literature, spiking networks have been used as heuristic 'stochastic search' solvers for NP-hard constraint satisfaction problems, such as Sudoku and TSP [19].

Our model can be seen as a discrete version of the continuous model discussed in by Maass in [30] or as a noisy version of the deterministic model in [27]. In addition to being stochastic, in comparison to the model of [27], our response latency Δ is constant for all connections in the network. Additionally, we have just a single round memory – each neuron's membrane potential is affected just by spikes of neighboring neurons in the same or immediately preceding round of computation. We note that if connections are allowed between auxiliary neurons, a longer memory can be easily be implemented within our general model.

Self-Stabilization in Distributed Computing.

The notion of self-stabilization goes back to Dijkstra in 1973. A self-stabilizing system can automatically recover following the occurrence of transient faults. The goal in this area is to design systems that converge to a desired behavior from any arbitrary starting point [8, 25]. Among the tremendously broad work, perhaps the most relevant to this work is self-stabilizing algorithms for leader election [9, 11].

In a stochastic neural network, self-stabilization is a necessity. Both changes to the given input as well as random deviations of the system from a converged state require the network to re-converge. Hence, we insure that all our networks converge to WTA from any initial network configuration and are self-stabilizing. This property does not hold in many previously studied WTA implementations for spiking networks [33].

Valiant's Neuroidal Model.

Valiant considers a model of neural computation in which abstract neurons (which he calls *neuroids*) are connected via a random network of synapses [44]. He discusses how these neurons can learn representations of real world objects whose perception stimulates the network in certain ways. As in our model, neurons fire in response to a membrane potential given by a weighted sum of firing neighbors. Differently, synapse weights evolve in response to increased firing of their end points, which allows *learning* to occur within the network. This learning ability is the primary focus of Valiant's work and of follow up work on the model. For example, recently, [35] extended understanding of how reasonably complex learning and pattern matching tasks can be performed in this model.

Our work deviates is somewhat more 'algorithmic' than the work of Valiant, focusing how basic takes can be computed using a set of neurons with a fixed set of synapses and bias

values. We do not consider how, for example, our WTA networks could form within a larger neural circuit through learning of appropriate synapse weights. Following previous work [28] we think of WTA networks as fundamental primitives of neural circuits on top of which high level algorithms, such as learning algorithms, can be built.

A.2 Biological Motivation for Network Dynamics

The timing of neural spikes is determined by two biological parameters, namely, the *refractory period* β and the *response latency*, Δ. The refectory period is the time during which stimulus given to the neuron would not cause a second action potential. The response latency is the delay between the time the action potential reaches the presynaptic terminal of the input neurons and the time the postsynaptic output neuron sends out an action potential (assuming it does). In our setting we consider the case where $\Delta < \beta$ since for connected neurons in close proximity to each other, and inhibitory neurons with primarily local connections, the response delay is a few hundred of micro-seconds whereas the refractory time is several milliseconds [40]. WTA networks are basic, local neural primitives that are not believed to involve long range connections, justifying our assumption.

Every round corresponds to an interval between two pulses of the inputs (hence a round lasts β milliseconds). At the beginning of every round, the input layer spikes (at sub-round $(t, 1)$ in the notation of our discrete model). The spikes generated by the inputs invoke an alternating dynamic between the three layers in the circuit. Specifically, with a delay of δ milliseconds after the input's spike, the outputs spike with probability proportional to their total synaptic strengths (in sub-round $(t, 2)$). As shown in equation (1), this potential incorporates any spikes which occurred within a β millisecond preceding window – the input spikes in sub-round $(t, 1)$ (Δ milliseconds before), the inhibitor spikes in sub-round $(t-1, 3)$ ($\beta - \Delta$ milliseconds before), and the neuron's own self-excitatory output spike in sub-round $(t-1, 2)$, β milliseconds before. Δ milliseconds after the outputs spike, the inhibitors spike in sub-round $(t, 3)$, again incorporating spikes that occurred with a β millisecond window, which due to their limited connectivity structure, just includes the spikes of Y in $(t, 2)$.

B Missing Proofs and Auxiliary Claims

Throughout, we make use of the following Corollary of the Chernoff bound.

▶ **Theorem 12** (Simple Corollary of Chernoff Bound). *Suppose $X_1, X_2, \ldots, X_\ell \in [0, 1]$ are independent random variables. Let $X = \sum_{i=1}^{\ell} X_i$ and $\mu = \mathbb{E}[X]$. If $\mu \geq 5 \log n$, then w.h.p. $X \in \mu \pm \sqrt{5\mu \log n}$, and if $\mu < 5 \log n$, then w.h.p. $X \leq \mu + 5 \log n$.*

B.1 WTA with Two Inhibitors

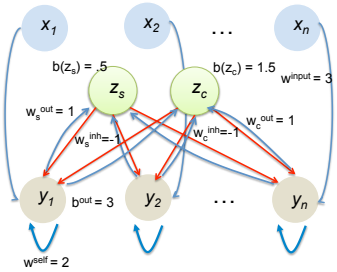

Figure 2 Two Inhibitor WTA Network

Proof of Theorem 3 (Two Inhibitor Upper Bound)

Formally the parameters of the network are set as follows: assume w.l.o.g. that $\lambda = 1/(c_1 \log n)$ for large constant c_1. For both inhibitors, set the excitatory output to inhibitor weights to ${\mathrm{w}^{\mathrm{out}}}_s = {\mathrm{w}^{\mathrm{out}}}_\ell = 1$ and $b(z_s) = .5$, $b(z_c) = 1.5$. Thus, by equation (2) z_s fires w.h.p. in sub-round $(t, 3)$ whenever at least one output fires in sub-round $(t, 2)$, and z_c fires w.h.p. whenever at least two outputs fire.

Set the inhibitor to output weights to ${\mathrm{w}^{\mathrm{inh}}}_s = {\mathrm{w}^{\mathrm{inh}}}_\ell = -1$, the excitatory input to output connection weight to $\mathrm{w}^{\mathrm{input}} = 3$, and the excitatory output to output self-loop to $\mathrm{w}^{\mathrm{self}} = 2$. Finally, set the output bias to $\mathrm{b}^{\mathrm{out}} = 3$.

The above parameters insure that only outputs corresponding to firing inputs ever fire w.h.p. Additionally, if we have not yet reached WTA and both z_s and z_c fire in sub-round $(t, 3)$, any output that fired in sub-round $(t, 2)$ will fire with probability $1/2$ in sub-round $(t + 1, 2)$. If we have reached WTA and just z_s fires, any output (the winner) that fired in round t will fire in round $t + 1$ w.h.p. In either case, any output that did not fire in round t will not fire w.h.p. in round $t + 1$.

We now give a formal proof of the theorem. First note that if the input $\mathrm{X} = \vec{0}$ then in every round, each output has potential $pot(y_j, t) \leq \mathrm{w}^{\mathrm{self}} - \mathrm{b}^{\mathrm{out}} = -1$ and so, recalling that $\lambda = 1/(c_1 \log n)$, fires with probability at most $\frac{1}{1+e^{c_1 \log n}} \leq 1/n^c$ for some large constant c in any round. So w.h.p. no outputs fire in each round, which is the valid output given $X = \vec{0}$ and so N trivially converges to WTA. So for the remainder of the section we focus on the case in which X has at least one firing input. We show that N satisfies the following conditions, which imply Theorem 3:

▶ **Claim 13** (Stability). *If* N *satisfies WTA in round* t *with* $y_j^t = 1$*, then* N *satisfies WTA in round* $t + 1$ *with* $y_j^{t+1} = 1$ *w.h.p.*

▶ **Claim 14** (Convergence). *Letting* $t = c_2 \log n$ *for constant* c_2*, for any input* X *with* $\|X\|_1 \geq 1$ *and any starting configuration* C^0*,* N *satisfies WTA in round* $C^{t'}$ *for some* $t' < t$*, with constant probability.*

Since Claim 14 holds for any starting configuration, we can simply apply it $\Theta(\log n)$ times to show that w.h.p. within $\Theta(\log^2 n)$ rounds, there will be a round in which WTA is satisfied, and hence N will converge to WTA by Claim 13. Additionally, it gives $\mathcal{ET}(\mathrm{N}) = O(\log n)$ as letting c_1 be the constant probability of reaching WTA in $O(\log n)$ rounds, we have:

$$\mathcal{ET}(\mathrm{N}) = O\left(\sum_{i=0}^{\infty}(1 - c_1)^i \cdot c_1 \log n\right) = O(\log n).$$

This gives us Theorem 3.

Proof of Claim 13. N satisfies WTA in round t with output y_j firing, so we have

$$pot(z_s, t) = 1 \cdot {\mathrm{w}^{\mathrm{out}}}_s - b(z_s) = .5 \text{ and } pot(z_c, t) = 1 \cdot {\mathrm{w}^{\mathrm{out}}}_s - b(z_c) = -.5.$$

Thus, recalling that $\lambda = 1/(c_1 \log n)$, in round t z_s fires with probability $\frac{1}{1+e^{-.5c_1 \log n}} \geq 1 - 1/n^c$ for large c and z_c fires with probability $\frac{1}{1+e^{.5c_1 \log n}} \leq 1/n^c$ for large c. So w.h.p. just z_s fires in round t. This gives that w.h.p.

$$pot(y_j, t + 1) = (1 \cdot {\mathrm{w}^{\mathrm{inh}}}_s) + (0 \cdot {\mathrm{w}^{\mathrm{inh}}}_\ell) + (1 \cdot \mathrm{w}^{\mathrm{self}}) + \mathrm{w}^{\mathrm{input}} - \mathrm{b}^{\mathrm{out}} = -1 + 2 + 3 - 3 = 1.$$

So y_j fires with probability $\frac{1}{1+e^{c_1 \log n}} \geq 1 - 1/n^c$ in round $t+1$. In contrast, for any $j' \neq j$, $y_{j'}$ does not fire in round t so we have w.h.p.

$$pot(y_{j'}, t+1) \leq (1 \cdot \mathrm{w^{inh}}_s) + (0 \cdot \mathrm{w^{inh}}_\ell) + (0 \cdot \mathrm{w^{self}}) + \mathrm{w^{input}} - \mathrm{b^{out}} = -1 + 3 - 3 = -1.$$

Therefore y'_j fires with probability $\leq 1/n^c$ in round $t+1$ so WTA is satisfied with output y_j firing in round $t+1$ w.h.p. ◀

Proof of Claim 14. Recall that we only consider $\|X\|_1 \geq 1$ as convergence to WTA is trivial when $X = \vec{0}$. We analyze three simple cases depending the initial configuration C^0:

Case 0: No output y_j with $x_j = 1$ fires in Y^0.

We first consider the subcase that no output (regardless of the value of x_j) fires in Y^0. In this case, $pot(z_s, 0) = -b(z_s) = -.5$ and $pot(z_c, 0) = -b(z_c) = -1$ so neither inhibitor fires w.h.p. in round 0. So w.h.p. all outputs with firing inputs have $pot(y_j, 1) \geq \mathrm{w^{input}} - \mathrm{b^{out}} = 0$ and so fire with probability $\geq 1/2$ in round 1. Since, $\mathrm{X} \neq \vec{0}$, with constant probability at least one of these outputs fires in round 1, in which case we appeal to Cases 1 and 2 below (where we re-label C^2 as the initial configuration C^0.).

Next consider the case when at least one output fires in Y^0, but all firing outputs correspond to non-firing inputs. In this case, we have $pot(z_s, 0) \geq 1 \cdot \mathrm{w^{out}} - b(z_s) \geq .5$ and so z_s fires w.h.p. in round 0. As noted, in any round, any output y_j with $x_j = 0$ has $pot(y_j, t) \leq \mathrm{w^{self}} - \mathrm{b^{out}} = -1$ and so does not fire w.h.p. Additionally, since every output with $x_j = 1$ has $y^0_j = 1$, these outputs have $pot(y_j, 1) = \mathrm{w^{inh}}_s + \mathrm{w^{input}} - \mathrm{b^{out}} = -1 + 3 - 3 = -1$ and so do not fire w.h.p. in round 1. So w.h.p. in round 1 no outputs fire and we are in the first case above.

Case 1: Exactly one output y_j with $x_j = 1$ fires in Y^0.

By Claim 13 and the fact that outputs with $x_j = 0$ do fire w.h.p. in any round, N satisfies WTA in round 1 and so immediately converges to WTA.

Case 2: More than one output y_j with $x_j = 1$ fires in Y^0.

Let k_t be the number of *active* outputs in round t – that is outputs corresponding to firing inputs that fire in round t. For any round with $k_t \geq 2$, we have $pot(z_s, t) \geq 2\,\mathrm{w^{out}} - b(z_s) = 1.5$ and $pot(z_c, t) \geq 2\,\mathrm{w^{out}} - b(z_c) = 1$. So both inhibitors fire in round t w.h.p. Conditioning on this event, all active outputs have:

$$pot(y_j, t+1) = (1 \cdot \mathrm{w^{inh}}_s) + (1 \cdot \mathrm{w^{inh}}_\ell) + (1 \cdot \mathrm{w^{self}}) + \mathrm{w^{input}} - \mathrm{b^{out}} = -1 - 1 + 2 + 3 - 3 = 0$$

and so fire with probability $1/2$ in round $t+1$. All inactive outputs, which did not fire in round t, do not have an active self loop and hence have $pot(y_j, t) = -2$ and don't fire in round $t+1$ w.h.p. (as discussed, all outputs with $x_j = 0$ also do not fire w.h.p.)

Conditioning on this event, with probability $1/2$, $k_{t+1} \leq k_t/2$. Further,

$$\Pr[k_{t+1} = 0] = 1/2^{k_t} \text{ and } \Pr[k_{t+1} = 1] = k_t \cdot (1/2^{k_t}) \geq \Pr[k_{t+1} = 0].$$

So the probability of reaching $k_{t+1} = 1$ and hence N converging to WTA is at least as high as the probability of overshooting WTA and having no outputs firing in round $t+1$.

Conditioning on the fact that z_s and z_c fire in every round in which $k_t \geq 2$ and that no output which was inactive in round t fires in round $t+1$, whenever $k_t \geq 2$ it decreases by a

factor of $1/2$ in round $t+1$ with good probability. So w.h.p. within $O(\log(k_0)) = O(\log n)$ rounds there is a round t with either $k_t = 1$ or $k_t = 0$. $k_t = 1$ is at least as likely as $k_t = 0$ so with constant probability, N converges to WTA within $O(\log n)$ rounds. ◀

Two Inhibitor Lower Bound

▶ **Theorem 15.** *For any basic WTA network* N *with* $\alpha = 2$ *inhibitors,* $\mathcal{ET}(\mathrm{N}) = \Omega(\log n/\log\log n)$ *and* $\mathcal{HT}(\mathrm{N}) = \Omega(\log^2 n/\log\log^2 n)$.

The key idea is that the use of a stability inhibitor z_s and a convergence inhibitor z_c in the algorithm is not just a design choice, but is required for *any* near-optimal two inhibitor WTA network.

▶ **Claim 16.** *For any basic WTA network* N *with* $\alpha = 2$ *inhibitors and* $\mathcal{ET}(\mathrm{N}) = O(\log^3 n)$, *one inhibitor* z_s *fires w.h.p. in sub-round* $(t,3)$ *if at least one output fires in sub-round* $(t,2)$. *The second inhibitor* z_c, *does not fire w.h.p. in* $(t,3)$ *if just a single output fires in* $(t,2)$.

Proof. Assume for contradiction that both inhibitors fire with probability $\omega(1/n^c)$ in sub-round $(t,3)$ after just a single output fires in sub-round $(t,2)$. Then, after a round t in which $z_s^t = z_c^t = 1$, any output y_j with $x_j = 1$ and $y_j^t = 1$ must fire w.h.p. in round $t+1$. This is because once N converges to WTA, when the single winning output fires in sub-round $(t,2)$, by our assumption, with relatively high $\omega(1/n^3)$ probability, both z_s and z_c fire in sub-round $(t,3)$. Even if this event occurs, the winning output must fire w.h.p. in round $t+1$ to maintain WTA w.h.p.

However, if we let $\mathrm{X} = \vec{1}$ and $\mathrm{Y}^0 = \vec{1}$, then for some constant c_1, all outputs will continue firing for $\omega(n^{c_1})$ rounds w.h.p. even if both z_s and z_c fire in every round. This contradicts our assumed $O(\log^3 n)$ runtime. Hence we have that at least one of the inhibitors, which we label z_c, fires with probability $O(1/n^c)$ in sub-round $(t,3)$ if just a single output fires in sub-round $(t,2)$.

Similarly, assume for contradiction that z_s *does not fire* with probability $\omega(1/n^c)$ in sub-round $(t,3)$ if a single output fires in sub-round $(t,2)$. Then, it must be that even if neither inhibitor fires in sub-round $(t,3)$, any output y_j that did not fire in sub-round $(t,2)$ (i.e. $y^t = 0$), must also not fire w.h.p. in sub-round $(t+1,2)$. This is because, by our assumption, after WTA is reached, with probability $(1 - O(n^c)) \cdot \omega(1/n^c) = \omega(1/n^c)$ neither inhibitor will fire in sub-round $(t,3)$ when just the single winning output fires in sub-round $(t,2)$. Still, all non-winning outputs must continue not firing in round $t+1$ to maintain WTA w.h.p.

However, if we let $\mathrm{X} = \vec{1}$ and $\mathrm{Y}^0 = \vec{0}$, since even when neither inhibitor fires in round t, each output does not fire in round $t+1$ w.h.p. if it did not fire in round t, it will take $\omega(n^{c_1})$ rounds (for some constant c_1) before even a single output fires w.h.p. contradicting our assumed $O(\log^3 n)$ runtime.

◀

The above claim allows us to strongly constrain the behavior of the network based on the action of the inhibitors z_s and z_c. Let p_0 be the probability that an output y_j fires in round $t+1$ given that $y^t = 0$, $x^t = 1$ and $z_s^t = z_c^t = 0$.

▶ **Claim 17.** *For any basic WTA network* N *with* $\alpha = 2$ *inhibitors and* $\mathcal{ET}(\mathrm{N}) = o(\log^2 n)$, $p_0 = \omega(1/\log^2 n)$.

Proof. Consider X with just two firing inputs $x_1 = 1$ and $x_2 = 1$. For any round t in which $y_1^t = y_2^t = 0$, the probability that y_1 or y_2 fires in round $t+1$ is *at most* p_0 – since the

firing of z_s or z_c can only decrease the probability of the outputs firing. Assuming by way of contradiction that a $p_0 \leq c_1/\log^2 n$ for some constant c_1, starting from $\mathrm{Y}^0 = \vec{0}$, with constant probability, neither output will fire for $\Omega(\log^2 n)$ consecutive rounds, and so N cannot converge to WTA in expected $o(\log^2 n)$ rounds. ◀

Let p_{out} be the probability that output y_j fires in round $t+1$ given $y^t = 1$, $x^t = 1$ and $z_s^t = z_c^t = 1$.

▶ **Claim 18.** *For any basic WTA network* N *with* $\alpha = 2$ *inhibitors and* $\mathcal{ET}(\mathrm{N}) = o(\log^2 n)$, $p_{out} = \omega(1/\log^6 n)$.

Proof. Consider X with $\Theta(\log^4 n)$ firing inputs and initial configuration Y^0 where $y_j = 1$ for all j with $x_j = 1$. Consider some round t in which at least two outputs (corresponding to firing inputs) have fired in all rounds $t' \leq t$. If either (or both) of z_s or z_c do not fire in round t, then since they face at most as much inhibition as when the network has converged to WTA, all outputs with firing inputs that fired in round t fire w.h.p. in round $t+1$. However, if both z_s and z_c do fire in round t, if $p_{out} = O(1/\log^6 n)$ then with probability $\leq (1-p_{out})^{\Theta(\log^4 n)} = 1 - \Theta(1/\log^2 n)$ *no* output corresponding to a firing input fires in round $t+1$. Since by Claim 16 a single inhibitor firing is enough to maintain convergence to WTA, once these outputs do not fire in some round t, they do not fire again w.h.p. until a round in which neither z_s or z_c fire. Then by Claim 17 and a Chernoff bound (Theorem 12) $\omega(\log^2 n)$ of them fire w.h.p.

So overall, we alternate between having many (between $\omega(\log^2 n)$ and $\Theta(\log^4 n)$) outputs corresponding to firing inputs and 0 outputs with firing inputs. Each time we have many firing outputs, with probability at least $1 - \Theta(\log^2 n)$ we have no firing outputs in the next round. So it takes at least $\Omega(\log^2 n)$ rounds before we have a round with exactly one valid firing output with constant probability, contradicting our assumed runtime of $\mathcal{ET}(\mathrm{N}) = o(\log^2 n)$. ◀

With the above claims in place, we are ready to prove Theorem 15. Consider $\mathrm{X} = \vec{1}$ and initial configuration with $\mathrm{Y}^0 = \vec{1}$. Let $k_t = \|Y^t\|_1$ be the number of outputs that fire in round t. Now, if y_j fires in round t, then it fires with probability *at least* p_{out} in round $t+1$, since p_{out} is the firing probability with maximum inhibition. Let $d = c_1 \log n/p_{out}$ for some constant c_1. By Claim 18, $d = O(\log^7 n)$ and since $p_{out} \leq 1$, trivially $d = \Omega(\log n)$. Starting from Y^0 with all outputs firing, for $t = c_2 \frac{\log(d/n)}{\log p_{out}}$ for sufficiently small c_2 we have that any output fires in all rounds up to t with probability $\theta\left(p_{out}^t\right) = \omega\left(\frac{d}{n}\right)$. So by a Chernoff bound (Theorem 12) w.h.p. $\omega(d)$ outputs fire in all rounds $t' \leq t$.

Let t_f represent the first round in which $\leq d$ outputs fire. By our argument above, w.h.p.

$$t_f = \Theta(\log(n/d)/\log(1/p_{out})) = \Theta(\log n/\log(1/p_{out})) = \Omega(\log n/\log\log n) \tag{3}$$

by Claim 18. This gives us $\mathcal{ET}(\mathrm{N}) = \Omega(\log n/\log\log n)$. So it just remains to show our lower bound on $\mathcal{HT}(\mathrm{N})$.

Since $> d$ outputs fire in round $t_f - 1$, again by a Chernoff bound, w.h.p. $k_{t_f} \geq d \cdot p_{out} = \Omega(\log n)$. Consider any round $t > t_f$ in which $k_{t'} > 1$ for all $t' \leq t$. If either of z_s of z_c do not fire in round t, then $k_{t+1} = k_t > 1$ w.h.p. Otherwise, $\Pr[k_{t+1} = 1] = k_t \cdot p_{out}(1-p_{out})^{k_t - 1}$ and:

$$\Pr[k_{t+1} = 0] = (1-p_{out})^{k_t} = \Pr[k_{t+1} = 1]] \cdot \frac{1-p_{out}}{k_t p_{out}} \geq \Pr[k_{t+1} = 1]] \cdot \frac{1}{\log^8 n}$$

where we use the fact that $k_t \leq d = O(\log^7 n)$ and $1 - p_{out} \geq \log n$ or else by (3) we would already not reach WTA w.h.p. in $O(\log^2 n)$ rounds.

So, the probability that $k_{t+1} = 0$ is high (within a polylog n) factor of the probability that $k_{t+1} = 1$. So, with probability at least $\Omega(1/\log^8 n)$, t_f is followed by a reset round in 0 outputs fire before a round in which a single output fires. Further, once such a reset round occurs, then no output will fire until z_s and z_c don't fire in a round (and hence inhibition is lower than it is after convergence to WTA) in which case by Claim 17 $\omega(n/\log^2 n)$ outputs will fire. So w.h.p. there will be $\Omega(\log n/\log\log n)$ rounds before another round in which ≤ 1 outputs fire.

Overall, in order to have a round in which exactly 1 output fires w.h.p. requires $\Omega(\log n/\log(\log^8 n)) = \Omega(\log n/\log\log n)$ resets, each taking $\Omega(\log n/\log\log n)$ rounds, and giving our final lower bound of $\Omega(\log^2 n/\log\log^2 n)$.

B.2 WTA with One Inhibitor

One Inhibitor Lower Bound

▶ **Theorem 19.** *For any basic WTA network* N *with* $\alpha = 1$ *inhibitors,* $\mathcal{ET}(\mathrm{N}) = \Omega(n^c)$.

We fix any constant c and assume by way of contradiction that there is a network N which converges to WTA in $O(n^c)$ rounds in expectation. Let z denote the single inhibitor in N. We first argue that N must be at least somewhat active – given no firing activity from the outputs Y and the inhibitor z, each output connected to an active input should fire with reasonably high probability.

▶ **Claim 20** (Sufficiently Active Network). *If* $z^t = 0$ *then each output* y_j *with* $x_j = 1$ *and* $y_j^t = 0$ *fires in round* $t+1$ *with probability* $\Omega(1/n^c)$.

Proof. Let X be an input in which exactly one input x_j fires and let $Y^0 = \vec{0}$. The time for N to converge to WTA is lower bounded by the time required for y_j to fire at least once.

Let p_0 be the probability that y_j fires in round $t+1$ if $y_j^t = 0$ and $z^t = 0$ and let p_1 be the probability that y_j fires in round $t+1$ if $y_j^t = 0$ and $z^t = 1$. $p_1 \leq p_0$, so as long as y_j does not fire in round t, it fires with probability at most p_0 in round $t+1$. If $p_0 \leq c_1/n^c$ for some constant c_1 then starting from C_0, with constant probability, y_j will not fire for $\Omega(n^c)$ consecutive rounds. By our assumption that N converges to WTA in $O(n^c)$ rounds in expectation, we have $p_0 = \Omega(1/n^c)$. ◀

We next show that the inhibitor z must fire in round t w.h.p. whenever at least one output fires, in order to maintain stability once WTA has been reached.

▶ **Claim 21** (Stability). *For any configuration* $\mathcal{C}^t$ *of* N*, if at least one output neuron fires in round* t *(i.e.* $\|\mathrm{Y}^t\|_1 \geq 1$*),* z *fires in round* t *w.h.p.*

Proof. Consider input $\mathrm{X} = \vec{1}$. Let t be a round in which WTA is satisfied (exactly one output y_j fires while no other outputs fire). Using the notation of Claim 20, the probability that a non-firing output fires in round $t+1$ is:

$$\Pr[z^t = 1|\,\mathrm{Y}^t] \cdot p_1 + \Pr[z^t = 0|\,\mathrm{Y}^t] \cdot p_0.$$

By Claim 20 we have $p_0 \geq c_1/n^c$ for some constant c_1. Since N converges to WTA it must be that w.h.p. in round $t+1$, y_j continues firing and no other output fires. So we have, for some large constant c_2:

$$\begin{aligned}\Pr[z^t = 1|\,\mathrm{Y}^t] \cdot p_1 + \Pr[z^t = 0|\,\mathrm{Y}^t] \cdot p_0 &\leq 1/n^{c_2}\\ \Pr[z^t = 0|\,\mathrm{Y}^t] \cdot c_1/n^c &\leq 1/n^{c_2}\\ \Pr[z^t = 0|\,\mathrm{Y}^t] &= O(1/n^{c_2-c})\end{aligned}$$

which gives the claim as long as $c < c_2$ since exactly one output fires in Y^t. The probability that z fires when > 1 output fires is at least as large due to the excitatory nature of the outputs. ◀

Finally, by way of contradiction, we show that when z fires, any output must stop firing with reasonably high probability. Otherwise, starting with multiple firing outputs, it will take too long to converge to WTA. As we will see this convergence requirement conflicts with the stability requirement of Claim 21 since it means that the winning output will stop firing with reasonably high probability after convergence to WTA.

▶ **Claim 22** (Convergence). *If $z^t = 1$ then y_j with $y_j^t = 1$ and $x_j = 1$ does not fire in round $t+1$ with probability $\Omega(1/n^c)$.*

Proof. Let p denote the probability that an output which corresponds to a firing input and which fires in round t does not fire in round $t+1$ given that $z^t = 1$. We want to show that $p = \Omega(1/n^c)$.

Let $\mathrm{X} = \vec{1}$ and let t be any round in which at least two outputs fire. By Claim 21, $z^t = 1$ w.h.p. and at least two outputs fire in round 1 with probability $(1-p)^2 \geq 1 - 2p$. If we start from $Y^0 = \vec{1}$, then w.h.p. at least two outputs will fire in $\Theta\left(\frac{1}{p}\right)$ consecutive rounds. By assumption N converges to WTA within $O(n^c)$ rounds in expectation so we must have $p = \Omega(1/n^c)$. ◀

Putting it all together, consider an execution that satisfies WTA in round t with exactly one output y_j firing. Then, by Claim 21, z fires in round t w.h.p. Thus, by Claim 22, y_j stops firing in round $t+1$ with probability $\Omega(1/n^c)$, in contradiction to the fact that the network must eventually converge to WTA and have y_j fire for n^{c_1} consecutive rounds for some large constant c_1. We briefly note that the above lower bound can be matched with a trivial single inhibitor algorithm.

▶ **Observation 23.** *There is basic network* N *with $\alpha = 1$ inhibitors with $\mathcal{ET}(\mathrm{N}) = O(n^c)$.*

Proof. The single inhibitor z simply fires w.h.p. in round t whenever ≥ 1 outputs fire in round t. The weights are set such that when $z^t = 1$ and $y_j^t = 1$, y_j fires in round $t+1$ with probability $1/n^{c+1}$. If z does not fire, any y_j with $x_j = 1$ fires w.h.p.

It is not hard to see that starting with any input, we will reach a round satisfying WTA within $O(n^c)$ rounds in expectation and after this round is reached, WTA will be maintained for $O(n^{c-1})$ additional rounds in expectation (and so $O(n^{c-2})$ w.h.p.). ◀

B.3 WTA with $O(\log n)$ Inhibitors

Proof of Theorem 4 ($O(\log n)$ Inhibitor Upper Bound). Recall that we assume w.l.o.g. $1/\lambda = c_1 \log n$ for some constant c_1. We set $\mathrm{w}^{\mathrm{input}} = 3$, $\mathrm{w}^{\mathrm{self}} = 2$, and $\mathrm{b}^{\mathrm{out}} = 3$. In this way, exactly as in the two inhibitor network analyzed in Section B.1, any output y_j with $x_j = 0$ will have $pot(y_j, t) \leq \mathrm{w}^{\mathrm{self}} - \mathrm{b}^{\mathrm{out}} = -1$ in every round t and so will not fire w.h.p. in any round.

Our network has $\alpha = \lceil \log n \rceil$ inhibitors. The first is a stability inhibitor z_s, which behaves exactly as the stability inhibitor in the two inhibitor network analyzed in Section B.1. ${\mathrm{w}^{\mathrm{out}}}_s = 1$, $b(z_s) = 0.5$ and ${\mathrm{w}^{\mathrm{inh}}}_s = -1$. z_s fires w.h.p. in sub-round $(t, 3)$ if ≥ 1 output fires in sub-round $(t, 2)$ and does not fire w.h.p. if no output fires. We also have $\alpha - 1$ convergence inhibitors $z_1, ..., z_{\alpha-1}$. For each z_i, $b(z_i) = 2^i - .5$ and ${\mathrm{w}^{\mathrm{out}}}_i = 1$. Therefore, z_i fires w.h.p. in round t whenever $\geq 2^i$ outputs fire in the round. It does not fire w.h.p. if

$< 2^i$ outputs fire. We set the inhibitor weight from z_1 to each output to be $\mathrm{w^{inh}}_1 = -1$. For each $i \in 2, ..., \alpha - 1$ we set $\mathrm{w^{inh}}_i = -\lambda \cdot \log_2(e)$.

We can see that the stability Claim 13 holds just as it does in the two inhibitor network analyzed in Section B.1. Specifically, if just a single output y_j with $x_j = 1$ fires in some round t, w.h.p. z_s will fire while the convergence inhibitors will all not fire. So we will have:

$$pot(y_j, t+1) = \mathrm{w^{inh}}_s + \mathrm{w^{input}} + 1 \cdot \mathrm{w^{self}} - \mathrm{b^{out}} = -1 + 3 + 2 - 3 = 1$$

so y_j fires w.h.p. in round t+1. At the same time for $j' \neq j$, since $y_{j'}$ does not fire in round t:

$$pot(y_{j'}, t+1) = \mathrm{w^{inh}}_s + \mathrm{w^{input}} + 0 \cdot \mathrm{w^{self}} - \mathrm{b^{out}} = -1 + 3 + 0 - 3 = -1$$

so $y_{j'}$ will not fire in round $t+1$. So, once a single y_j with $x_j = 1$ fires in some round t, N will converge to WTA w.h.p. We now show that N reaches such a round in $O(1)$ expected time.

Consider any round $t > 0$ in which $k_t \geq 2$ outputs fire. We can assume that all these outputs corresponding to firing inputs since as discussed, outputs corresponding to non-firing inputs do not fire w.h.p. in any round. For some i we have $k_t \in [2^i, 2^{i+1})$ and so w.h.p. in round t, $z_s, z_1, ..., z_i$ fire while all other inhibitors do not fire (note that $\alpha - 1 = \lceil \log n \rceil - 1$ and so even if n outputs fire, all inhibitors fire). We thus have, w.h.p. for any active output y_j with $y_j^t = 1$ and $x_j = 1$:

$$\begin{aligned} pot(y_j, t+1) &= \mathrm{w^{self}} + \mathrm{w^{input}} - \mathrm{b^{out}} + \mathrm{w^{inh}}_s + \mathrm{w^{inh}}_1 + \sum_{j=2}^{i} \mathrm{w^{inh}}_j \\ &= 2 + 3 - 3 - 1 - 1 - (i-1)\lambda = (i-1)\lambda \cdot \log_2(e). \end{aligned}$$

So y_j fires in round $t+1$ with probability:

$$p(y_j, t+1) = \frac{1}{1 + e^{(i-1)\lambda \log_2(e)/\lambda}} = \frac{1}{1 + 2^{i-1}}$$

Since $k_t \in [2^i, 2^{i+1})$, we have $1 \leq \frac{k_t}{1+2^{i-1}} \leq 4$ and so can bound the probability that exactly one output that was active in round t fires in round $t+1$ as:

$$\begin{aligned} k_t \cdot \frac{1}{1+2^{i-1}} \cdot \left(1 - \frac{1}{1+2^{i-1}}\right)^{k_t - 1} &\geq \left(1 - \frac{1}{1+2^{i-1}}\right)^{k_t - 1} \\ &\geq \left(1 - \frac{1}{1+2^{i-1}}\right)^{4(1+2^{i-1})} \\ &\geq \frac{1}{4^4}. \end{aligned}$$

So, with constant probability exactly one output that fired in round t also fires in round $t+1$. Any output that did not fire in round t has potential $\leq \mathrm{w^{input}} - \mathrm{b^{out}} + \mathrm{w^{inh}}_s + \mathrm{w^{inh}}_\ell = -2$ and so does not fire with high probability. So, with constant probability, exactly one output y_j with $x_j = 1$ fires, and so N converges to WTA.

We conclude by noting that, by the arguments of Claim 14 for our two inhibitor network, with constant probability, starting with any $\mathrm{Y^0}$ we in fact have a round with $k_t \geq 1$ firing outputs all with active inputs within constant rounds. So from any starting configuration, we converge to WTA with constant probability in $O(1)$ rounds. Repeating this constant probability argument gives both $\mathcal{ET}(\mathrm{N}) = O(1)$ and $\mathcal{HT}(\mathrm{N}) = O(\log n)$. ◀

$\Omega(\log n)$ High Probability Runtime Lower Bound

▶ **Theorem 24.** *Any basic WTA network* N*, with any number of inhibitors, has* $\mathcal{HT}(\mathrm{N}) = \Omega(\log n)$.

Proof. We show that any network N requires $\Omega(\log n)$ rounds before a round t in which WTA is satisfied w.h.p. This immediately gives our lower bound on convergence time.

Consider input $X = \vec{1}$ (so any output is a valid winner) and any round t such that WTA has not been satisfied for any $t' < t$. That is, in no round t' does exactly one output y_j fire. Let W_t be the event that in round t exactly one output fires and hence WTA is satisfied. We claim that $\Pr[W_t = 1 \mid C^{t-1}] \leq c$ for any configuration C^{t-1} of N in round $t-1$ and some universal constant c. That is, no matter the network configuration in round $t-1$, WTA will only be achieved with constant probability in the next round. Hence, as long as the initial output configuration Y^0 is one in which WTA is not satisfied, for $t = O(\log n)$, with probability at least $(1-c)^t = \Omega(1/n^{c'})$, for some constant c', WTA will not be satisfied in any even round up to t. This gives that $\mathcal{HT}(\mathrm{N}) = \Omega(\log n)$. There are two cases:

Network Reset.

$Y^{t-1} = \vec{0}$. In this case, no output fired in round $t-1$. Since all outputs are identical w.r.t their edge weights and bias values, conditioned on the behavior Z^{t-1} of the inhibitors in round $t-1$, all outputs will fire independently with some fixed probability p in round t. For any p and any $n \geq 2$, the probability that exactly 1 will fire in round t is:

$$\Pr[W_t = 1 \mid C^{t-1}] = n \cdot p(1-p)^{n-1} \leq \frac{1}{2}.$$

No Reset.

$\|Y^{t-1}\|_1 \geq 2$ – i.e. there are at least 2 firing outputs in round $t-1$. Let O_1 be the set of firing outputs in round $t-1$ and O_0 be the set of non-firing outputs. Conditioned on Z^{t-1}, any output in O_1 fires independently with some probability p_1 in round t and any output in O_0 fires with some probability p_0. Further, $p_0 \leq p_1$ since the only difference in membrane potential between the neurons in O_0 and O_1 will be whether their excitatory self loop is active.

For $a \in \{0,1\}$ let V_a be the event that exactly 1 output from O_a fires in round t. Clearly, $W_t \subseteq V_1 \cup V_0$. For any p_1, $\Pr[V_1 \mid C^{t-1}] = |O_1| \cdot p_1(1-p_1)^{|O_1|-1} \leq 1/2$ since we have not reached WTA and so $|O_1| \geq 2$. If $|O_0| = 0$, then vacuously, $\Pr[V_0 \mid C^{t-1}] = 0$ and hence $\Pr[W_t \mid C^{t-1}] \leq 1/2$. Alternatively, If $|O_0| \geq 2$ then we also have $\Pr[V_0 \mid C^{t-1}] \leq 1/2$ and, since all outputs fire independently conditioned on C^{t-1},

$$\Pr[W_t \mid C^{t-1}] \leq 1 - \Pr[\neg(V_1 \cup V_0)] \leq 1 - (1-1/2)^2 = 3/4.$$

Finally, if $|O_0| = 1$ either $p_0 \leq 1/2$, in which case $\Pr[V_0 \mid C^{t-1}] \leq 1/2$ and we again have $\Pr[W_t \mid C^{t-1}] \leq 3/4$ or $p_0 \geq 1/2$ in which case $p_1 \geq p_0 \geq 1/2$, and the probability that at least two outputs from O_1 fire is at least $1/4$ and hence WTA is achieved with probability at most $3/4$. ◀

B.4 WTA with $\alpha \geq 2$ Inhibitors

Proof of Theorem 5. We first describe the network construction in detail. As in our previous networks, we have a stability inhibitor z_s that fires w.h.p. whenever ≥ 1 outputs fire in

round t. This inhibitor ensures that in round $t+1$ w.h.p. only outputs that fired in round t (and hence have an active self loop) will fire in round $t+1$.

We set the excitatory input to output connection weight to $\mathrm{w}^{\mathrm{input}} = 3$, the excitatory output self-loop to $\mathrm{w}^{\mathrm{self}} = 2$, and the output bias to $\mathrm{b}^{\mathrm{out}} = 3$. For the stability inhibitor we set the excitatory output to inhibitor weight ${\mathrm{w}^{\mathrm{out}}}_s\, 1$, $b(z_s) = .5$, and ${\mathrm{w}^{\mathrm{inh}}}_s = -1$ just as we did in the two inhibitor algorithm.

We have θ groups each containing $\lceil (\log n)^{1/\theta} \rceil$ convergence inhibitors, $Z_1, Z_2, ..., Z_\theta$ where we denote $Z_i = \{z_{i,1}, z_{i,2}, ..., z_{i,\lceil (\log n)^{1/\theta} \rceil}\}$. We set ${\mathrm{w}^{\mathrm{out}}}_i = 1$ for all $i \in Z_1, Z_2, ..., Z_\theta$ and $b(z_{i,j}) = 2^{jd_i} - .5$. In this way, when $k_t \in \left[2^{jd_i}, 2^{(j+1)d_i}\right)$ w.h.p. $z_s, Z_1, ..., Z_{i-1}, z_{i,1}, ..., z_{i,j}$ all fire while the remaining inhibitors do not. We set ${\mathrm{w}^{\mathrm{inh}}}_{i,j}$ such that

$$pot_{i,j} = \mathrm{w}^{\mathrm{input}} + \mathrm{w}^{\mathrm{self}} + {\mathrm{w}^{\mathrm{inh}}}_s - \mathrm{b}^{\mathrm{out}} + \sum_{\{(k,l)|k<i \text{ or } l \leq j\}} {\mathrm{w}^{\mathrm{inh}}}_{k,l}$$

satisfies:

$$p_{i,j} = \frac{1}{1 + e^{-pot_{i,j}/\lambda}} = \frac{c_1}{2^{jd_i}}$$

for some small constant c_1. For simplicity of presentation, we do not explicitly calculate out these weights. However, it is clear that choosing correct weights $p_{i,j}$ decreases as most inhibitors fire and the sigmoid function is continuous and decreases monotonically as pot_i decreases. We are now ready to analyze the network behavior in detail.

No Firing Inputs.

As in the two inhibitor network, any y_j with $x_j = 0$, has maximum potential is $\mathrm{w}^{\mathrm{self}} - \mathrm{b}^{\mathrm{out}} = -1$ (even when no inhibitors fire) so and will not fire w.h.p. outside of the initial configuration Y^0. ($p(y_j, t) \leq \frac{1}{1+e^{1/\lambda}} \leq 1/n^c$ for any t since $\lambda = 1/c_1 \log n$). If $\mathrm{X} = \vec{0}$, this implies that a valid WTA state in which no outputs fire will be converged to w.h.p. trivially. We now focus on the case when $\| \mathrm{X} \|_1 \geq 1$.

Maintaining WTA (Stability).

If just a single output y_j corresponding to an active input ($x_j = 1$) fires in round t then w.h.p. by Claim 13 in Appendix B.1, N converges to WTA. This is because w.h.p. just z_s will fire in round t and y_j has potential

$$pot(y_j, t+1) = (1 \cdot {\mathrm{w}^{\mathrm{inh}}}_s) + (0 \cdot {\mathrm{w}^{\mathrm{inh}}}_\ell) + (1 \cdot \mathrm{w}^{\mathrm{self}}) + \mathrm{w}^{\mathrm{input}} - \mathrm{b}^{\mathrm{out}} = -1 + 2 + 3 - 3 = 1.$$

So y_j fires with probability $\frac{1}{1+e^{c_1 \log n}} \geq 1 - 1/n^c$ in round $t+1$. In contrast, for any $j' \neq j$, $y_{j'}$ does not fire in round t so has

$$pot(y_{j'}, t+1) \leq (1 \cdot {\mathrm{w}^{\mathrm{inh}}}_s) + (0 \cdot {\mathrm{w}^{\mathrm{inh}}}_\ell) + (0 \cdot \mathrm{w}^{\mathrm{self}}) + \mathrm{w}^{\mathrm{input}} - \mathrm{b}^{\mathrm{out}} = -1 + 3 - 3 = -1.$$

Therefore y'_j fires with probability $\leq 1/n^c$ in round $t+1$ so WTA is satisfied with output y_j firing in round $t+1$ w.h.p.

Converging to WTA.

It now just remains to show that with constant probability, within $O(\theta)$ rounds, there is at least one round in which exactly one output y_j with $x_j^t = 1$ fires. By the stability argument above once such a round occurs, N will converge to WTA w.h.p.

By the arguments of the convergence Claim 14 for the two inhibitor network, with constant probability, starting with any Y^0 we in fact have a round with $k_t \geq 1$ firing outputs all with active inputs within constant rounds. If $k_t = 1$ then N converges to WTA and we are done. So it suffices to consider the case when $k_t \geq 2$.

If $k_t \in \left[2^{jd_i}, 2^{(j+1)d_i}\right)$ then w.h.p. $z_s, Z_1, ..., Z_{i-1}, z_{i,1}, ..., z_{i,j}$ fire while the other inhibitors do not and so in round $t+1$ any active output that fired in round t fires with probability $p_{i,j}$. So we have $E[k_{t+1}] \in [1, c_1 2^{d_i})$, and, so with at least constant probability by a Markov bound $k_{t+1} < 2^{d_i}$ if we set c_1 to a small constant.

Additionally, in any round with $k_t \geq 2$ conditioning on the high probability event that the correct inhibitors fire,

$$\Pr[k_{t+1} = 1] = k_t \cdot p_{i,j}(1 - p_{i,j})^{k_t - 1}$$

and:

$$\begin{aligned}\Pr[k_{t+1} = 0] = (1 - p_{i,j})^{k_t} &= \Pr[k_{t+1} = 1] \cdot \frac{(1 - p_{i,j})}{k_t p_{i,j}} \\ &\leq \Pr[k_{t+1} = 1] \cdot \frac{1}{2^{jd_i} \cdot c1/2^{jd_i}} \\ &\leq \frac{1}{c_1} \Pr[k_{t+1} = 1].\end{aligned}$$

So, the probability of having exactly one output fire and hence converging to WTA is within a constant factor of the probability or having 0 outputs fire and 'reseting' the network. So overall with constant probability, we reach such a round with $k_t = 1$ within just $O(\theta)$ rounds. Iterating this argument gives the expected and high probability runtime bounds of Theorem 5. ◀

Proof of Theorem 6. Again we have a stability inhibitor z_s that fires w.h.p. in sub-round $(t, 3)$ whenever ≥ 1 outputs fire in sub-round $(t, 2)$. We also have a 'base level' convergence inhibitor that fires w.h.p. whenever ≥ 2 outputs fire. When just z_s and z_ℓ fire in round t, any output (with an active input) that fired in round t fires with probability $1/2$ in round $t+1$.

We then employ $\alpha - 2$ additional convergence inhibitors $z_1, ...z_{\alpha-2}$. For $i \in 1, ..., \alpha - 2$ let

$$d_i = (\log n)^{i/(\alpha-1)}.$$

Letting k_t be the number of outputs that fire in round t, z_i fires w.h.p. in round t whenever $k_t \geq 2^{d_i}$. The synapse weights from the inhibitors to the outputs are chosen such that, when $k_t \in \left[2^{d_i}, 2^{d_{i+1}}\right)$, and hence $z_1, ..., z_i$ each active output (i.e. each y_j with $y_j^t = 1$ and $x_j = 1$) fires with probability:

$$p_i = \frac{c \log n}{d_i} = \frac{c \log n}{(\log n)^{i/(\alpha-1)}}$$

in round $t+1$. This probability is enough to ensure that within few rounds, we will have $< 2^{d_i}$ active outputs. Specifically, since $k_t \in \left[2^{d_i}, 2^{d_{i+1}}\right)$, for

$$r = \frac{\log k_t}{\log 1/p_i} \leq \frac{(\log n)^{(i+1)/(\alpha-1)}}{(\log n)^{i/(\alpha-1)} - \log(c \log n)} = O\left((\log n)^{1/(\alpha-1)}\right)$$

with high probability, there will be a round $r' = O(r)$ with $k_{t+r'} \leq 2^{d_i}$. At the same time, p_i is large enough that w.h.p. we will not overshoot WTA and have 0 firing outputs in round

$t + r'$. Even if $k_t = 2^{d_i}$ then we have $k_t \cdot p_i = c \log n$ and so, for large enough c, with high probability, by a Chernoff bound (Theorem 12) at least $O(\log n)$ outputs fire in round $t + 1$.

Overall, within $O\left((\alpha - 2)(\log n)^{1/(\alpha-1)}\right)$ rounds, the number of active outputs falls within $[2, 2^{d_1}]$ w.h.p. Once k_t is in this range, just z_s and z_1 fire w.h.p. so our network is essentially identical to the two inhibitor network described in the previous section and analyzed in detail in Appendix B.1. We thus reach WTA with constant probability in $\Theta(\log 2^{d_1}) = \Theta\left((\log n)^{1/(\alpha-1)}\right)$ additional rounds, giving our final runtime bound of $O\left(\alpha(\log n)^{1/(\alpha-1)}\right)$.

We now formalize the above arguments. Following our earlier constructions, we set the excitatory input to output connection weight to $\mathrm{w}^{\mathrm{input}} = 3$, the excitatory output self-loop to $\mathrm{w}^{\mathrm{self}} = 2$, and the output bias to $\mathrm{b}^{\mathrm{out}} = 3$. Set the excitatory output to inhibitor weights ${\mathrm{w}^{\mathrm{out}}}_s = {\mathrm{w}^{\mathrm{out}}}_\ell = 1$, $b(z_s) = .5$, $b(z_\ell) = 1.5$, and ${\mathrm{w}^{\mathrm{inh}}}_\ell = {\mathrm{w}^{\mathrm{inh}}}_s = -1$ just as we did in the two inhibitor algorithm.

For the additional convergence inhibitors, set ${\mathrm{w}^{\mathrm{out}}}_i = 1$ for all $i \in 1, ..., \alpha - 2$ and $b(z_i) = 2^{d_i} - .5$. In this way, when $k_t < 2^{d_1}$, w.h.p. just z_s and z_1 fire, and each active output in round t has potential

$$pot(y_j, t + 1) = \mathrm{w}^{\mathrm{input}} + \mathrm{w}^{\mathrm{self}} + {\mathrm{w}^{\mathrm{inh}}}_s + {\mathrm{w}^{\mathrm{inh}}}_\ell - \mathrm{b}^{\mathrm{out}} = 3 + 2 - 1 - 1 - 3 = 0$$

and so fires with probability $p_1 = 1/2$ in round $t + 1$. We set ${\mathrm{w}^{\mathrm{inh}}}_i$ such that

$$pot_i = \mathrm{w}^{\mathrm{input}} + \mathrm{w}^{\mathrm{self}} + {\mathrm{w}^{\mathrm{inh}}}_s + {\mathrm{w}^{\mathrm{inh}}}_\ell - \mathrm{b}^{\mathrm{out}} + \sum_{j=1}^{i} {\mathrm{w}^{\mathrm{inh}}}_j$$

satisfies:

$$p_i = \frac{1}{1 + e^{-pot_i/\lambda}} = \frac{c \log n}{2^{d_i}}.$$

As in the proof of Theorem 5, we do not explicitly calculate out these weights. Roughly, ${\mathrm{w}^{\mathrm{inh}}}_i \approx \Theta(\frac{\lambda \log\log n}{\alpha - 1})$ such that when i inhibitors fire $p_i \approx \frac{1}{e^{-\Theta(\frac{i\lambda \log\log n}{\alpha-1})}} \approx \frac{c \log n}{2^{d_i}}$. It is clear that choosing correct weights is possible as $1/2 > p_1 > ... > p_{\alpha-1}$ and the sigmoid function is continuous and decreases monotonically as pot_i decreases.

By identical arguments to those in the proof of Theorem 5, we converge to WTA in constant rounds w.h.p. if there are no firing inputs or if a single output with a firing input fires in a round. Hence it just remains to show that with constant probability, within $O(\alpha(\log n)^{1/(\alpha-1)})$ rounds, there is at least one round in which exactly one output y_j with $x_j^t = 1$ fires.

Again, by the arguments of the convergence Claim 14 for the two inhibitor network, with constant probability, starting with any Y^0 we in fact have a round with $k_t \geq 1$ firing outputs all with active inputs within constant rounds. If $k_t = 1$ then N converges to WTA and we are done. So it suffices to consider the case when $k_t \geq 2$. In this case, as discussed if $k_t \in \left[2, 2^{d_1}\right)$ then w.h.p. just z_s and z_ℓ fire, and so each active output has potential

$$pot(y_j, t + 1) = (1 \cdot {\mathrm{w}^{\mathrm{inh}}}_s) + (1 \cdot {\mathrm{w}^{\mathrm{inh}}}_\ell) + (1 \cdot \mathrm{w}^{\mathrm{self}}) + \mathrm{w}^{\mathrm{input}} - \mathrm{b}^{\mathrm{out}} = -1 - 1 + 2 + 3 - 3 = 0$$

and fires with probability $1/2$ in round $t + 1$. All inactive outputs, which did not fire in round t, do not have an active self loop and hence have $pot(y_j, t) = -2$ and don't fire in round $t + 1$ w.h.p. (as discussed, all outputs with $x_j = 0$ also do not fire w.h.p.)

Conditioning on this event, with probability $1/2$, $k_{t+1} \leq k_t/2$ and by the arguments in Claim 14, we converge to WTA with constant probability within $O(k_t) = O(d_1) = O\left((\log n)^{1/(\alpha-1)}\right)$ rounds.

If $k_t \in \left[2^{d_i}, 2^{d_{i+1}}\right)$ for some $i \in 1, ..., \alpha - 2$ then as discussed, w.h.p. $z_s, z_\ell, z_1, ..., z_i$ all fire in round t while all other inhibitors do not fire. We thus have

$$\mathrm{E}[k_{t+1}] \geq 2^{d_i} \cdot p_i = \frac{2^{(\log n)^{i/(\alpha-1)}} \cdot c \log n}{2^{(\log n)^{i/(\alpha-1)}}} = c \log n$$

By a Chernoff bound (Theorem 12), w.h.p. k_{t+1} falls within a constant multiplicative factor of its expection. Thus, w.h.p. we still have $k_{t+1} \geq 2$. At the same time, w.h.p. $k_{t+1} \leq c_1 k_t \cdot p_i$ for some constant c_1. So overall, within $r = \frac{\log k_t}{\log 1/p_i} = O\left((\log n)^{1/(\alpha-1)}\right)$ rounds, w.h.p. $k_{t+r} < 2^{d_i}$. Within $\alpha - 2$ epochs of $O\left((\log n)^{1/(\alpha-1)}\right)$ rounds we thus have $k_t \in \left[2, 2^{d_1}\right)$ w.h.p. and then reach WTA withing $O\left((\log n)^{1/(\alpha-1)}\right)$ additional rounds with constant probability.

Iterating this constant proability argument gives the expected and high probability runtime bounds of Theorem 6. ◀

B.5 Missing Proofs for Main Lower Bound (Theorem 7)

B.5.1 Inhibitors are Nearly Deterministic for Most Density Classes

Proof of Lemma 8. By the definition of the set S, for $z \in S$ it holds that z fires in sub-round $(t, 3)$ with probability $1/(1 + e^{-pot_1(z)}) \geq 1/\log^{3c} n$ and hence $\mathrm{w^{out}}_z - b(z) \geq -3c \log\log n$. By our no-background noise assumption that neurons do not fire w.h.p. with no external input, we can assume $b(z) \geq 3 \log n$ and hence have $pot_2(z) = 2\,\mathrm{w^{out}}_z - b(z) \geq 2 \log n$. Thus, z fires with probability at least $1 - 1/n^2$ in sub-round $(t, 3)$. Overall, all the $|S| \leq O(\log n)$ inhibitors fire in sub-round $(t, 3)$, with probability at least $1 - 1/n$ as required. ◀

Proof of Lemma 9. In any round t, even if all n outputs fire in sub-round $(t, 2)$, the firing probability of each inhibitor in R in sub-round $(t, 3)$ is at most $1/\log^c n$ (or else the inhibitor would fall in C). Union bounding over the first $O(\log\log n)$ rounds of execution and the at most $O(\log n)$ inhibitors in R, we get that with probability at least $1 - 1/\log^{c-3} n$, none of these inhibitors fires in these rounds. ◀

Proof of Lemma 10. Let $k(z)$ be the smallest integer in $[1, n]$ such that z fires in sub-round $(t, 3)$ with probability at least $1/\log^c n$ when $k(z)$ outputs fire in sub-round $(t, 2)$. By the definition of C, when n outputs fire, z fires in the next sub-round with probability at least $1/\log^c n$, and hence $k(z)$ is well defined. In addition, since $z \notin S$, $k(z) \geq 2$.

Part (1) of the claim follows immediately by the definition of $k(z)$. To prove part (b), the key idea is to exploit the following gap in the behavior of $z \in C$: since z is not in S, the firing probability of z in steady state (with exactly one firing output) is *at most* $1/\log^{3c} n$. On the other hand, when there are at least $k(z) \geq 2$ active outputs, the firing probability of z is *at least* $1/\log^c n$. This implies that the sigmoid function which converts the number of firing inputs to z's firing probability must be steep enough such that z fires with good probability when $\geq 2k(z)$ outputs fire. By the fact that $z \notin S$, $pot_1(z) = \mathrm{w^{out}}_z - b(z) \leq -3c \cdot \log\log n$ and so $\mathrm{w^{out}}_z \leq b(z) - 3c \log\log n$. On the other hand, by the definition of $k(z)$, $\mathrm{w^{out}}_z$ cannot be too small since $pot_{k(z)}(z) = k(z) \cdot \mathrm{w^{out}} - b(z) \geq -c \cdot \log\log n$ so

$$k(z) \cdot \mathrm{w^{out}}_z \geq b(z) - c \log\log n. \tag{4}$$

Combining this we get: $k(z)b(z) - 3k(z) \cdot c \log\log n \geq b(z) - c \log\log n$ and so $b(z) \geq 3c \log\log n$. Using that and Eq. (4), we get: $pot_{2k(z)}(z) = 2k(z) \cdot \mathrm{w^{out}} - b(z) \geq 2b(z) - 2c \log\log n - b(z) = b(z) - 2c \log\log n \geq c \log\log n$. Hence, $1/(1 + e^{-pot_{2k(z)}(z)}) \geq 1 - 1/(\log^c n)$ as required. ◀

B.5.2 Detailed Description of the Prediction Process

In this section we describe the prediction process in more detail.

Inductive Assumptions.

For each round t, in showing that we are able to predict the behavior of N for a large number of inputs in round t, we make several inductive assumptions:

For two ranges of positive numbers $R_1 = [r_1, r_2]$ and $R_2 = [r_3, r_4]$ such that $r_1 \leq r_2 \leq r_3 \leq r_4$, and a positive number a, the ranges are called a-separated if $r_3/r_2 \geq a$. The *value* of the range $R_1 = [r_1, r_2]$ is taken to be r_1. We assume that for $\mathrm{X} \in \mathcal{X}_{t-1} \subset \mathcal{X}$ the ranges $R_{t-1}(\mathrm{X})$ are all a separated for some constant a and have minimum value $\Theta(\log^7 n)$. We also assume that our earlier predictions are accurate: for each $X \in \mathcal{X}_{t-1}$, $\widehat{R}_{t-1}(\mathrm{X}) \in R_{t-1}(\mathrm{X})$ and $\widehat{F}_{t-1}(\mathrm{X}) = F_{t-1}(\mathrm{X})$ with probability at least $1 - \Theta(1/\log n)$. We first show that these assumptions hold for round one:

Predicting the number of firing outputs in sub-round $(1, 2)$.

Since we consider the initial reset configuration $Y^0 = \vec{0}$ we have $\widehat{R}_0(\mathrm{X}_i) = 0$ for all X_i. Trivially we can set $\mathcal{X}_0 = \mathcal{X}$ – we deterministically know the behavior of all outputs in round 0. By our no-background noise assumption, for every $z \in \mathrm{Z}$, $b(z) = c \log n$, and so w.h.p. $\widehat{F}_0(\mathrm{X}_i) = 0$ for all X_i (no inhibitor fires in the initialization round). Let $\mathcal{X}_1^{large} = \{\mathrm{X}_i \mid 2^i \geq \log^9 n\}$ (note that $|\mathrm{X}_1^{large}| = \Theta(\log n)$). Let p_0 be the probability that an output fires in sub-round $(t+1, 2)$ given that no inhibitor and no output fires in round t (i.e, no output has an active self-loop). Since there are 2^i active *input neurons* in X_i, conditioned on the high probability event that $\widehat{R}_0(\mathrm{X}_i) = 0$ and $\widehat{F}_0(\mathrm{X}_i) = \vec{0}$, the expected number of firing outputs in sub-round $(1, 2)$ is $p_0 \cdot \mathrm{X}_i$. It is not hard to show that $p_0 = \Omega(1/\log^2 n)$ and by combining this fact with a Chernoff bound we have:

▶ **Claim 25.** *For every $\mathrm{X}_i \in \mathcal{X}_1^{large}$, w.h.p. the number of firing outputs in sub-round $(1, 2)$, $\widehat{R}_1(\mathrm{X}_i)$ is in the range $R_1(\mathrm{X}_i) = [(1 - 1/\log^3 n) \cdot p_0 2^i, (1 + 1/\log^3 n) \cdot p_0 2^i]$. Hence, the predicted output ranges for the inputs in $\mathcal{X}_1^{large}$ are $2(1 - 1/\log n)$ separated. Additionally each has minimum value $\Omega(\log^7 n)$.*

Proof. Let X_1 be a vector with exactly one firing input and let y_i be its corresponding output. Starting from $Y^0 = \vec{0}$, w.h.p., no inhibitor fires in round 0. If $p_0 < 1/\log^2 n$ then since p_0 rate is the maximum firing probability for y_j in sub-round $(t+1, 2)$ given that it didn't fire in sub-round $(t, 2)$, the network requires $\Omega(\log^2 n)$ rounds until y_j fires with constant probability and so at least that long to converge to WTA. So we can work in the case where $p_0 \geq 1/\log^2 n$.

For $\mathrm{X}_i \in \mathcal{X}^{large}$ we thus have the expected number of firing outputs in sub-round $(1, 1)$ is $p_0 \cdot 2^i \geq 1/\log^2 n \cdot \log^9 n = \log^7 n$. Since the random firings of the outputs are independent given the firing behavior of the inhibitors and since no inhibitors fire in sub-round $(0, 3)$ w.h.p. by a Chernoff bound (Theorem 12), we have that w.h.p. the number of firing outputs $\widehat{R}_1(\mathrm{X}_i)$ is in the range $(1 \pm 1/\log^3 n) \cdot p_0 \cdot 2^i$ for all $\mathrm{X}_i \in \mathcal{X}^{large}$. ◀

The above shows that the predicted ranges for all $\mathrm{X} \in \mathcal{X}_1^{large}$ are well separated, accurate, and have high value. We can now set $\mathcal{X}_1$ to include any $\mathrm{X} \in \mathcal{X}_1^{large}$ except possibly $|C| \leq \alpha$ inputs where $R_1(\mathrm{X})$ overlaps a critical region $K(z)$ for some $z \in C$. Since the remaining ranges do not overlap any critical regions, by Lemmas 8, 9, and 10 we are able to predict $\widehat{F}_1(\mathrm{X})$ with good probability, and so have all our inductive assumptions in round 1.

Predicting the number of firing outputs for $t \geq 2$.

We first define a subset of inputs $\mathcal{X}_t^{large} \subseteq \mathcal{X}_{t-1}$ for which we can predict the behavior of the outputs in N in sub-round $(t,2)$. Let $\mathcal{X}_t^{same} \subseteq \mathcal{X}_{t-1}$ be the largest subset of inputs whose predicted firing vector $F_{t-1}(\mathrm{X})$ for the inhibitors in sub-round $(t-1,3)$ is the *same*, and denote this common firing vector by F_{t-1}^*. Let $\mathcal{X}_t^{large}$ be the set of inputs in $\mathcal{X}_t^{same}$ after omitting $\Theta(\log\log n)$ inputs with the smallest range value in sub-round $(t-1,2)$.

Eventually we will show that $\mathcal{X}_t^{large}$ is a reasonably large set of inputs compared to $\mathcal{X}_{t-1}$, and hence we can continue predicting behavior for at least some inputs for a large number of rounds. But first we show how to predict $R_t(\mathrm{X})$ for every input $\mathrm{X} \in \mathcal{X}_t^{large}$.

Let p be the probability that an active output (one with $y_j^{(t-1,2)} = 1$) fires in sub-round $(t,2)$ given that the inhibitors fired in sub-round $(t-1,3)$ according to F_{t-1}^*. Since all inputs in $\mathcal{X}_t^{same}$ have the same predicted firing vector F_{t-1}^*, in each of them, an active output fires in sub-round $(t,2)$ with probability p. In addition, by induction for every $\mathrm{X} \in \mathcal{X}_t^{same} \subseteq \mathcal{X}_{t-1}$, $R_{t-1}(\mathrm{X})$ has a minimum of $\Theta(\log^7 n)$ predicted firing outputs. So inhibition in sub-round $(t-1,3)$ w.h.p. must be at least as high as it is once we have converged to WTA and just a single output is firing. Thus, any output that did not fire in sub-round $(t-1,2)$ must not fire w.h.p. in sub-round $(t,2)$, since non-firing outputs continue not to fire once WTA is converged to.

So just focusing on active outputs that fire in sub-round $(t,2)$, for every $\mathrm{X}_i \in \mathcal{X}_t^{same}$, let $R_{t-1}(\mathrm{X}_i) = [\ell_i, m_i]$ be the predicated range of firing outputs in sub-round $(t-1,2)$. Then the expected number of firing outputs in sub-round $(t,2)$ is in the range $[p \cdot \ell_i \ , \ p \cdot m_i]$. For every $\mathrm{X}_i \in \mathcal{X}_t^{same}$, let $R_t(\mathrm{X}_i) = [(1 - 1/\log^3 n) \cdot p\ell_i \ , \ (1 + 1/\log^3 n) \cdot pm_i]$.

We now observe that if the expected number of firing outputs is too small for even one of the inputs in $\mathcal{X}_t^{same}$, then it implies a lower bound of $\Omega(\log n)$ for $\mathcal{ET}(\mathrm{N})$. Essentially this is because if this is the case, with good probability, 0 outputs will fire in round t, and a reset configuration identical to Y^0 will occur. This will keep occurring, causing the network to have large runtime.

▶ **Observation 26.** *For every $t \geq 1$, if there exists $\mathrm{X} \in \mathcal{X}_t^{same}$, such that the smallest value of $R_t(\mathrm{X}_i)$ is less then $1/\log^4 n$, then $\mathcal{ET}(\mathrm{N}) = \Omega(\log n)$.*

Proof. Let $\mathrm{X} \in \mathcal{X}_t^{same}$ be such that $R_t(\mathrm{X})$ is less then $1/\log^4 n$. Then, given that the inhibitors fire according to the prediction F_{t-1}^* in sub-round $(t-1,3)$, by Markov inequality, the probability that the number of firing outputs in sub-round $(t,2)$ is at least 1 is less then $1/\log^4 n$. In other words, the conditional probability (where we condition on the prediction for round $t-1$) that a reset where 0 outputs fire happens in sub-round $(t,2)$ is at least $1 - 1/\log^4 n$. However, by our inductive assumption $\widehat{F}(\mathrm{X}) = F_{t-1}^*$ must be correct with probability at least $1 - 1/\log n$. Hence, with probability at least $1 - \Theta(\log n)$ $Y^t = \vec{0}$ and a reset round occurs. With constant probability this occurs $\Omega(\log n)$ times before WTA is ever reached. The observation follows. ◀

Hence, from now on, we assume the complementary case that the number of predicted firing outputs in sub-round $(t,2)$ is at least $1/\log^4 n$ for every $\mathrm{X} \in \mathcal{X}_t^{same}$. This allows us to show:

▶ **Claim 27.** *For every $\mathrm{X} \in \mathcal{X}_t^{large}$*

(1) Given that the inhibitors fire according to F_{t-1}^* in sub-round $(t-1,3)$, then with probability $1 - 1/n$, the number of firing outputs in sub-round $(t,2)$ is in the range $R_t(\mathrm{X})$.

(2) The set of ranges $R_t(\mathrm{X})$ for $\mathrm{X} \in \mathcal{X}_t^{large}$ are all a-separated for some constant a.

(3) $R_t(\mathrm{X})$ has value at least $\Omega(\log^7 n)$ for every $\mathrm{X} \in \mathcal{X}_t^{large}$.

Proof. Since for any $\mathrm{X} \in \mathcal{X}_t^{same}$ the predicted number of firing outputs is $\Omega(1/\log^4 n)$, and since the ranges are constant separated by our inductive assumption that the ranges $R_{t-1}(\mathrm{X})$ for $\mathrm{X} \in \mathcal{X}_{t-1}$ are separated, by omitting $\Theta(\log\log n)$ inputs from $\mathcal{X}_t^{same}$, the minimum number of firing outputs in the predicted ranges for the remaining set of inputs, namely, $\mathcal{X}_t^{large}$ is $\Omega(\log^7 n)$. Hence the true number of firing outputs is well concentrated around this expectation and so we have (1) by a Chernoff bound (Theorem 12).

Further, since we increase the width of the predicted range $R_t(\mathrm{X})$ by factor of at most $(1+1/\log^3 n)$ compared to the range $R_{t-1}(\mathrm{X})$, over all $O(\log\log n)$ rounds of prediction, the range is increased by at most a factor of $(1+1/\log^3 n)^{O(\log\log n)} \leq 1+O(1/\log^2 n)$. Since the ranges have separation 2 in the initialization round, they remain constant separated in round t, giving (2). ◀

Predicting $\widehat{F}_t(\mathrm{X})$ given the predicted range $R_t(\mathrm{X})$.

We first define the final subset $\mathcal{X}_t \subseteq \mathcal{X}_t^{large}$ of inputs for which round t is fully predicted (i.e., both the number of firing outputs in sub-round $(t,2)$ and the states of the inhibitors in sub-round $(t,3)$). The set $\mathcal{X}_t$ contains any $\mathrm{X} \in \mathcal{X}_t^{large}$ unless $R_t(\mathrm{X})$ intersects the critical range $K(z)$ for some convergence inhibitor $z \in C$. By Lemma 10, the firing state of each inhibitor $z \in C$ can be predicted with good probability as long as the number of firing outputs in previous sub-round is not in the critical range $K(z) = [k(z)/2, 2k(z)]$. In particular, if the range $R_t(\mathrm{X})$ falls below $k(z)/2$, then we predict that z does not fire in sub-round $(t,3)$. On the other hand, if the range $R_t(\mathrm{X})$ falls above $2k(z)$, then we predict that z fires in sub-round $(t,3)$. Regardless of the exact number of firing outputs in sub-round $(t,2)$, since $R_t(\mathrm{X})$ does not intersect the critical ranges of the inhibitors of C, we can predict with good probability the firing states of C in sub-round $(t,3)$ by Lemma 10. By Lemma 8, with probability at least $1-1/n$, all the stability inhibitors S fire in sub-round $(t,3)$ and by Lemma 9, with good probability, no inhibitor in R fires. So overall we can predict all inhibitor behavior with good probability. With the above in place we are finally have that our inductive assumptions hold in round t. We summarize:

▶ **Lemma 28.** *For every $t \geq 1$ it holds that:*

(Q1) For every $\mathrm{X} \in \mathcal{X}_t$, the predicted range of firing outputs $R_t(\mathrm{X})$ satisfies:

$$\Pr[\widehat{R}_t(\mathrm{X}) \in R_t(\mathrm{X}) \;\mid\; \widehat{F}_{t-1}(\mathrm{X}) = F_{t-1}(\mathrm{X})] \geq 1 - 1/n \;. \tag{5}$$

(Q2) The collection of predicted ranges $R_t(\mathrm{X})$ for $\mathrm{X} \in \mathcal{X}_t$ are all a-separated for some constant a and all have value at least $\Omega(\log^7 n)$.

(Q3) For every $\mathrm{X} \in \mathcal{X}_t$, the predicted firing pattern for the inhibitors satisfies

$$\Pr[\widehat{F}_t(\mathrm{X}) = F_t(\mathrm{X}) \;\mid\; \widehat{R}_t(\mathrm{X}) \in R_t(\mathrm{X})] \geq 1 - 1/\log^3 n \;. \tag{6}$$

The final step before giving our expected time lower bound is to show that $\mathcal{X}_t$ is reasonably large, so we are able to keep predicting the behavior of N for a number of outputs round after round. This follows from a few simple observations:

▶ **Observation 29.** $|\mathcal{X}_t^{same}| \geq |\mathcal{X}_{t-1}|/\alpha$.

Recall that $\mathcal{X}_t^{same}$ consists of the largest subset of $\mathcal{X}_{t-1}$ with the *same* predicted inhibitor behavior F_{t-1}^* in round $t-1$. Naively, there are 2^α possible predictions for F_{t-1}^* which gives that $|\mathcal{X}_t^{same}| \geq |\mathcal{X}_{t-1}|/2^\alpha$. In order to obtain the much stronger bound above, we again use Lemma 10 which shows that, as long as $\widehat{R}_{t-1}(\mathrm{X})$ does not intersect the critical region of any $z \in C$, the inhibitors behave with good probability as linear threshold circuits and so there are only α possible predictions $F_{t-1}(\mathrm{X})$.

Proof. Since by Lemma 10 each inhibitor $z \in C$ behaves with probability $1 - \log^c n$ as a threshold network in sub-round $(t, 3)$ (so long that the number of firing outputs in sub-round $(t, 2)$ is not in the critical range $K(z)$), the total number of different inhibitor firing state configurations (different $F_{t-1}(\mathrm{X})$ vectors predicted in the previous step) is bounded by $|C|$. To see this, since conditioning on the prediction $R_t(\mathrm{X})$ being correct, there is at least one firing output in sub-round $(t-1, 2)$, the inhibitors of S will fire w.h.p. Further the inhibitors R never fire with good probability, so the only varying part in $F_{j-1}(\mathrm{X})$ is the prediction for C and as discussed there are only $|C| \leq \alpha$ such possible predictions. ◀

▶ **Observation 30.** $|\mathcal{X}_t^{large}| \geq |\mathcal{X}_t^{same}| - O(\log\log n)$.

This is immediate as $\mathcal{X}_t^{large}$ was derived by removing $\Theta(\log\log n)$ of the inputs with the smallest predicted range values from $\mathcal{X}_t^{same}$.

▶ **Observation 31.** $|\mathcal{X}_t| \geq |\mathcal{X}_t^{large}| - O(\alpha)$.

This follows as $\mathcal{X}_t$ is derived by removing all inputs from $\mathcal{X}_t^{large}$ where $R_t(\mathrm{X})$ overlaps the critical region of some $z \in C$. By (Q2) the $R_t(\mathrm{X})$ are all constant separated so there can be at most $|C| = O(\alpha)$ which overlap critical regions. We are now ready to show:

▶ **Lemma 32.** $\mathcal{ET}(\mathrm{N}) = \Omega(\log\log n / \log\alpha)$.

Proof. We can continue predicting the behavior of N up to round t until we have $|\mathcal{X}_t| = \Theta(\log\log n)$ (at which point $\mathcal{X}_t^{large}$ may be empty and so we will have to stop simulation). Further, as long as we can predict for t rounds, by Lemma 28 we will know with good probability that at least $\Omega(\log^7 n)$ outputs are still firing for all $\mathrm{X} \in \mathcal{X}_t$. So with good probability WTA is not reached for those inputs, giving a lower bound of $\Omega(t)$ rounds in expectation to solve WTA.

Set $t = c_1 \log\log n / \log\alpha$ for small enough constant c_1 and recall that we can assume $\alpha = O(\log^{c_2} n)$ for small constant c_2 since otherwise our runtime bound is $\Omega(1)$ and so holds vacuously. By Observations 29, 30, and 31 after t rounds we have:

$$\begin{aligned} |\mathcal{X}_t| &\geq \frac{|\mathcal{X}_0|}{\alpha^t} - t \cdot \alpha - t \cdot O(\log\log n) \\ &\geq \frac{\log n}{\log^{c_1} n} - \log\log n \cdot \log^{c_2} n - (\log\log n)^2 = \Omega(\log^{1-c_1} n) \end{aligned}$$

and hence can predict for at least t rounds. This completes the proof. ◀

Monotonicity property of basic WTA networks.

We show that the WTA dynamic is monotone so long as there is at least one firing output. Intuitively, we show that all basic WTA networks pick a single winner by monotonically decreasing the number of firing outputs until just a single output is firing. The number of firing outputs only ever increases if the network 'overshoots' the WTA state and has a round in which no outputs fire.

▶ **Lemma 33.** *For any basic WTA network* N*, as long as the number of firing outputs is more than one, their number is monotone non-increasing. In particular, if at least one output fires in round t, w.h.p. , an output that did not fire in that round, will not fire again in round $t+1$.*

Proof. Given input X with at least one firing input neuron, the network N must eventually converge so that in every round exactly 1 output fires w.h.p. Consider a round t in this *steady state period*. Since all outputs have the same parameters (e.g., edge weights and bias values) and since the weight of the self-loop is positive, if output y_i fires in round t, it is at least as likely to fire in round $t+1$ as output y_j for any $j \neq i$. Additionally, conditioned on the configuration of the inhibitors in time t, the probability that each output fires in round $t+1$ is independent. Hence, it must be that w.h.p., if y_i fired in round t, it continues to fire in round $t+1$ and each y_j, which did not fire in round t does not fire in round $t+1$ with high probability.

Further, consider any round t with at least one firing output. Since all connections from the output layer are excitatory, the probability that any inhibitor in Z fires at the end of round t is at least as large as it is in the steady state of the network, and hence any output that does not fire in round t does not fire in round $t+1$ w.h.p. ◀

B.6 Complete Proof for High Probability Lower Bound (Lemma 11)

Let $Q_Y \subseteq \{0,1\}^n, Q_Z \subseteq \{0,1\}^\alpha$ be the vectors describing the firing states of the outputs and inhibitors in a given round. Let $Q = Q_Y \circ Q_Z \subseteq \{0,1\}^{n+\alpha}$ be a vector describing the firing states of the inhibitors and outputs. Let $P_{1,j}(Q)$ be the probability to achieve the WTA state in round j given Q, that is the probability that exactly one output fires in sub-round $(j,2)$ given that the firing states of the outputs (resp., inhibitors) in sub-round $(j-1,2)$ (resp., $(j-1,3)$) is Q_Y (resp., Q_Z). Similarly, let $P_{0,j}(Q)$ be the probability that *no output* fires in sub-round $(j,2)$ given Q, that is the probability that a reset event happens. Finally, let $P_{01,j}(Q)$ be the probability that a reset event or a WTA event happens in round j given that configuration in round $j-1$ is Q, hence $P_{01,j}(Q) = P_{1,j}(Q) + P_{0,j}(Q)$. We begin by claiming the following.

▶ **Claim 34.** *For every round j and for every vector $Q \in \{0,1\}^{n+\alpha}$ in which there are at least two firing outputs (i.e., Q is neither a WTA state nor a reset state), and such that $P_{01,j}(Q) \geq \Theta(1/\log\log n)$, it holds that $P_{0,j}(Q) \geq \Theta(1/(\log\log n)^3)$.*

Proof. Since $P_{01,j}(Q) = P_{0,j}(Q) + P_{1,j}(Q)$, if $P_{0,j}(Q) \geq P_{01,j}(Q)/2$, then we are done. Hence, we can assume from now on that $P_{1,j}(Q) = \Theta(1/\log\log n)$. We will show that $P_{0,j}(Q) \geq P_{1,j}(Q)/(\log\log n)^2$, which will establish our claim.

Let p be the firing probability of an active output[8] in sub-round $(j,2)$ given Q and let $k \geq 2$ be the number of outputs that fire in round $j-1$ as specified by Q. Since Q has at least two firing outputs, w.h.p., only active outputs (those that fire in the previous round) can fire in the next round. The probability that the WTA state is achieved in round j is $P_{1,j}(Q) = k \cdot p \cdot (1-p)^{k-1}$ and the probability that a reset is achieved in round j is $P_{0,j}(Q) = (1-p)^k$.

We consider two cases depending whether the firing probability p is large or small. First, assume that $p \geq 0.1$ and set $r = c/\log\log n$. Since $P_{1,j}(Q) \geq r$, we have that $1-p \geq r/k$. We also have:

$$k(9/10)^{k-1} \geq k \cdot p \cdot (1-p)^{k-1} \geq r,$$

and hence $k \leq \Theta(\log\log n)$. Overall, $P_{0,j}(Q)/P_{1,j}(Q) = (1-p)/(kp) \geq (1-p)/k \geq r/k^2 \geq c/(\log\log n)^2$. Next, consider the complementary case where $p < 0.1$. Letting $y = kp/2$,

$$y \cdot e^{-y} \geq (kp/2)(1-p)^{k/2} \geq (k/2)p(1-p)^{k-1} \geq r/2,$$

[8] Recall that an output is active in round j if it fires in sub-round $(j-1,2)$.

hence $y \leq 2 \log 1/r = \Theta(\log \log \log n)$. Overall,

$$P_{0,j}(Q)/P_{1,j}(Q) = (1-p)/kp \geq \Theta(1/\log \log \log n).$$

◀

The Execution Tree.

A key tool used in this section is the notion *execution tree* that captures all possible transcripts that can evolve in a window of DH rounds when starting with the initial configuration C_0. The execution tree T is a tree of depth DH where each layer j corresponds to round j when running the network on the initial configuration C_0. Each node in T is labeled by an $(n+\alpha)$-length binary vector describing the firing configurations (or states) of the outputs and the inhibitors in a given round, and the edges are labelled by the transition probabilities. Hence, this tree describes all the possible firing states in a span of DH rounds when starting from the initial configuration C_0 (for which the time it takes to achieve WTA with constant probability is at least DC). The root r is labeled by the zero vector (since in round 0, no output fires and hence w.h.p also no inhibitor fires). For every $j \geq 2$, every node u in layer j is labeled by a vector $Q(u) = Q_Y(u) + Q_Z(u) \in \{0,1\}^{n+\alpha}$ describing the firing status of the outputs and the inhibitors in round j. Hence, each node has $2^{n+\alpha}$ children in the configuration tree. Every edge $e = (\pi(u), u)$ connecting u to its parent $\pi(u)$ in T is labeled by a probability $p(e)$ that the firing configuration in round j is $Q(u)$ given that the configuration in round $j-1$ is $Q(\pi(u))$.

Let $T_d(u)$ be the subtree of depth d rooted at u. When d is omitted $T(u)$ is simply the entire subtree of u in T.

For a leaf node $\ell \in T$, let $\mathcal{P}(\ell) = [r = u_0, u_1, \ldots, u_{DH}]$ be the path connecting ℓ to the root r in T. Let $p_{leaf}(u)$ be the probability that starting from r the firing configuration in each round $j \in \{0, \ldots, DH\}$ is $Q(u_j)$. Since there is an independence between the coin flips in every round j given the configuration in round $j-1$, we get that

$$\begin{aligned} &p_{leaf}(u) = \\ &\prod_{j=0}^{DH} \Pr[Q_Y(u_j) \text{ in round } (j,2) \;\mid\; Q_Y(u_{j-1}), Q_Z(u_{j-1}) \text{ in rounds } (j-1,2),(j-1,3)] \\ &\cdot \Pr[Q_Z(u_j) \text{ in sub-round } (j,3) \;\mid\; Q_Y(u_j) \text{ in sub-round } (j,2)] \\ &= \prod_{j=1}^{DH} p(e_j) \;\; \text{where} \;\; e_j = (u_j, u_{j+1}). \end{aligned}$$

For a node $u \in T$, let $Leaf(u)$ be the set of leaves in $T(u)$ and define

$$p_{node}(u) = \sum_{\ell \in Leaf(u)} p_{leaf}(u) \ ,$$

and for a subset of nodes U, let $p_{node}(U) = \sum_{u \in U} p_{node}(u)$. It is convenient to view $p_{node}(u)$ as the *weight* of tree $T(u)$. Hence, the weight of T is 1. In the same spirit, for a given subset of nodes U_i whose subtrees in T are vertex disjoint, we view $\sum_{u \in U_i} p_{node}(u)$ as the weight of the forest $\bigcup_{u \in U_i} T(u)$. We would like to show that:

$$\sum_{u \in Leaf(r)} \{p_{leaf}(u) \mid u \;\text{ is a WTA node }\} < 1 - 1/n^c \ , \tag{7}$$

In the next paragraphs, we will find a collection of non-WTA leaf nodes of large weight, i.e. of weight at least $1/n^2$ which will establish Eq. (7) for $c > 2$. To do that, we iteratively traverse the tree T from root to leaves, omitting undesired subtrees (and hence also leaf nodes) through the journey. This traversal is done in an asynchronous manner in the following sense: there are times that for a given node u in layer j, we move to a subset of its children in layer $j+1$, we call this move a *small jump* in the tree. In contrast, there are cases in which from a given node u in layer t, we jump DC layers in the subtree $T(u)$ and proceed the traversal from a subset of leaf nodes in the tree $T_{DC}(u)$ of depth DC, we call such a move a *large jump*. In the analysis part we will claim that by eliminating nodes in the tree T, we do not loose much weight, to deal with the fact that there are two types of jumps: small and large, we will employ an amortization claim that will enable us to bound the loss of weight layer by layer. See Fig. 3, for an illustration of the Execution Tree.

In each iteration $j \in \{1, \ldots, DH\}$, we maintain a collection of *non-WTA* nodes U_j whose subtrees in T are vertex disjoint. The final set U_{DH} will be a set of non-WTA leaf nodes for which we will show that their weight is at least $1/n^2$. Starting with $U_0 = \{r\}$, in every iteration $j \in \{1, \ldots, DH\}$, we have a set of nodes U_j that satisfy the following:

(A1) The subtrees $T(u)$, $u \in U_j$, are vertex-disjoint.
(A2) The distance of each node $u \in U_j$ from r is at least j.
(A3) No node in U_j is a WTA node.

In the high level, the nodes U_{j+1} are the leaf nodes of subtrees rooted at the nodes $u \in U_j$. Particularly, from each node $u \in U_j$, when constructing U_{j+1}, we omit part of the subtree $T(u) \subseteq T$ and replace u by a subset of nodes $V(u)$ in the subtree of u in T. The nodes $V(u)$ are subset of the leaf nodes of the subtree $T_{d(u)}(u)$ of depth $d(u)$ rooted at u. The value of the depth $d(u)$ is set to be either 1 or DC [9] depending on the configuration stored at node u. That is, either the nodes $V(u)$ are a subset of the children of u or that they are subset of the leaf nodes of the DC-depth tree rooted at u.

In the first case where $d(u) = 1$, we will show that we loose only $\Theta(1/\log\log n)$ of the weight of the tree $T(u)$, hence we keep $1 - \Theta(1/\log\log n)$ fraction of the weight. In the second case, we will show that we keep $\Theta(1/\log\log n)$ fraction of the weight of $T(u)$. The key observation here is to note that this cannot happen more than DH/DC times in a given branch, since the depth of the sub-tree of u is DC. In other words, on average, we maintain $\Theta(1/\log\log n)^{1/DC}$ of the weight per layer of the subtree $T_{DC}(u)$, and hence overall, after DH iterations, we maintain $1/n^2$ fraction of the total weight.

We first eliminate from the tree T all nodes u such that $Q_Y(u) = \vec{0}$ but $Q_Z(u) \neq \vec{0}$. Since the bias value of the inhibitors in $\Omega(\log n)$, we know that if no output fires in round j, then w.h.p. no inhibitor fires in that round. Let T' be the resulting tree. We first observe that by that step, we eliminate only $1/n^c$ of the total weight of the tree T.

▶ **Observation 35.** *The total weight of r in T' is at least $1 - 1/n^c$.*

From now on, we consider the tree T' and describe the iterative construction of the set U_j in details. Let $U_0 = \{r\}$. For $j \geq 1$ given U_j, the set U_{j+1} is obtained by defining for each node $u \in U_j$, a subset of non-WTA nodes $V(u)$ as described next.

Case 1: u is a reset node. Set the depth of the subtree to be $d(u) = \min\{DH - dist(u, r, T), DC\}$ and let $V(u)$ be the non-WTA nodes in the leaf nodes of $T_{d(u)}(u)$.

Since u is not a WTA node, it remains to consider the case where the number of active outputs in $Q(u)$ is at least 2. Recall that $P_{01,j}(Q(u))$ be the probability of achieving WTA

[9] To be more precise it is either 1 or $\min\{DC, DH - dist(r, u, T)\}$.

or reset in round $t+1$ given the configuration in round t is $Q(u)$. We distinguish between two cases depending on the value of $P_{01,j}(Q(u))$.

Case 2.1: $P_{01,j}(Q(u)) \geq \Theta(1/\log\log n)$**.** Let $V'(u)$ be the children of u in T that are reset-nodes. For each reset-node $w \in V'(u)$, let $V(w)$ be the non-WTA nodes in the leaf nodes of $T_{d(u)-1}(w)$ and let $V(u) = \bigcup V(w)$.

Case 2.2: $P_{01,j}(Q(u)) < \Theta(1/\log\log n)$**.** Let $V(u)$ be the children of u that have at least 2 active outputs in $Q(v)$ (hence $d(u) = 1$). This completes the definition of U_{j+1}.

To bound the weight of U_{DH}, we make use of the following claims that show that we do not loose too much weight in this traversal. Consider a node u and let $NW_{DC}(u)$ be the set of non-WTA leaves of the tree $T_{DC}(u)$.

▶ **Claim 36.** *If u is a reset node, then $p_{node}(NW_{DC}(u)) \geq c' \cdot p_{node}(u)$, for some constant c'.*

Proof. Let j be the layer of node u. Then by the selection of the initial configuration C_0, we know that the time it takes to achieve WTA with constant probability c when starting from C_0 is strictly larger than DC. Since a reset node is labelled with this same initial configuration, we get that $p_{node}(NW_{DC}(u)) \geq c' \cdot p_{node}(u)$ for $c' = 1 - c$. ◀

▶ **Claim 37.** *Let u be a node in layer j that satisfies Case (1) or Case (2.1), then $p_{node}(V(u)) \geq \Theta((1/\log\log n)^3) \cdot p_{node}(u)$.*

Proof. If u satisfies Case (1), the claim follows immediately by Cl. 36. We now consider the case where u satisfies Case (2.1). Recall that in this case the number of active outputs in $Q(u)$ is at least 2. Let A_0, A_1 be the set of children of u that are reset nodes, WTA nodes respectively. Let $A_{0,1} = A_0 \cup A_1$.

Then, since u satisfies Case (2.1), $p_{node}(A_{0,1}) \geq \Theta(1/\log\log n) \cdot p_{node}(u)$. In addition, since in $Q(u)$ there are at least two firing outputs, we can safely apply Cl. 34, to have that $p_{node}(A_0) \geq \Theta(1/(\log\log n)^2) \cdot p_{node}(A_1)$. Combining these two inequalities, we get that

$$p_{node}(A_0) \geq \Theta(1/(\log\log n)^3) \cdot p_{node}(u).$$

Next, by using Cl. 36, for every node $v \in A_0$ (which is a reset node), we have that $p_{node}(NW_{DC-1}(v)) \geq c' \cdot p_{node}(v)$. All together, we get that

$$\begin{aligned} p_{node}(NW_{DC}(u)) &\geq \sum_{v \in A_0} p_{node}(NW_{DC-1}(v)) \\ &\geq c' \cdot p_{node}(A_0) \geq \Theta(1/(\log\log n)^3) \cdot p_{node}(u) \ . \end{aligned}$$

Since $V(u) = A_0$, the claim follows. ◀

▶ **Claim 38.** *Let u be a node that satisfies Case (2.2), then $\sum_{w \in V(u)} p_{node}(w) \geq 1 - \Theta(1/\log\log n) \cdot p_{node}(u)$.*

Proof. By the definition of u, $P_{01,j}(Q(u)) < \Theta(1/\log\log n)$. Hence, letting $V(u)$ be the children of u that have at least 2 active outputs in $Q(v)$ (hence $d(u) = 1$), we have that $\sum_{w \in V(u)} \geq 1 - \Theta(1/\log\log n) \cdot p_{node}(u)$. ◀

Starting from a tree of weight 1, we would like to show that at the end of the process after at most DH iterations, the total weight of the leaf nodes U_{DH} is at least $1/n^2$. We now use Cl. 37 and 38 to prove the lower bound. By Cl. 37, when we consider $u \in U_j$ that satisfies either case (1) or case (2.1), we keep $\Theta(1/(\log\log n)^3)$ fraction of the weight but enjoy a large jump of DC layers in the sub-tree $T(u)$. Hence, on average, we keep

$\Theta(1/(\log\log n)^3)^{1/DC}$ fraction of the weight of $T(u)$ per layer. By Cl. 38, in type (2), we keep at least $1-\Theta(1/\log\log n)$ fraction of the weight of $T(u)$ when moving from a node u in layer i to a subset of its children $V(u)$ in layer $i+1$. Hence, on average in every iteration, we keep at least

$$\max\{(c/(\log\log n)^3)^{1/DC}, 1-c/\log\log n\}$$

fraction of the weight of the current forest. Hence, after $DH = DC\cdot\Theta(\log n/\log\log\log n)$ iterations, our total weight of the leaf set U_{DH} is at least

$$(\max\{(c/(\log\log n)^3)^{1/DC}, 1-c/\log\log n\})^{DH} \geq 1/n^2.$$

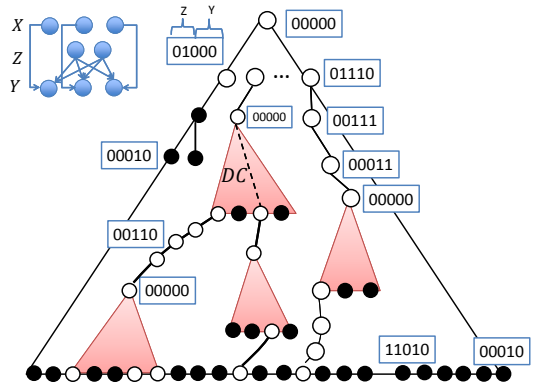

Figure 3 The Execution Tree. Shown is a schematic illustration of the Execution Tree for a small network with two inhibitors and three outputs. Every node u in layer j is labeled by a vector of length 5 describing an optional firing state for the inhibitors and outputs in round j. Each node has 2^5 children – covering all possible firing behaviors in round $j+1$. The weight nodes are non WTA nodes and the black nodes are the WTA nodes. When arriving a reset node u, a large jump is made by considering the leaf nodes of $T(u)$. When arriving a non-WR node, a small jump is made by considering subset of its children.

C Extension to Excitatory Auxiliary Neurons

In this section, we consider the more general case where the auxiliary neurons can be either excitatory or inhibitory. Let α denote their number. We assume that outputs with no active input are not allowed to fire. Hence, in a given sub-round $(t,2)$, we consider two types of outputs that might fire: *active* outputs – those that fire in the previous round and hence have a positive feedback via the self-loop; and *inactive* outputs – those that did not fire in the previous round. Whereas in the inhibitory case, we could show that the dynamic is monotonic – hence incative outputs do not fire with high probability, here it is not the case. Specifically, it might be the case that the level of inhibition during the process to achieve the WTA state is *lower* than that in steady-state and hence inactive outputs (outputs that did not fire in the previous rounds) join the game in later rounds. In our lower bound proofs, we heavily used the monotonicity property as it allowed us focus only on the active outputs (those that fired in the previous rounds) and totally neglect the inactive ones. In this section, we revise the claims that are based on the monotonicity lemma and adapt the proof to the general case of excitatory and inhibitory neurons.

C.1 Extensions for the Lower Bound for Expected Time

We classify the auxiliary neurons as before into three classes S, C and R. Note that all the proofs that concern the predictability of the inhibitors, i.e., Lemmas 8,9,10 depend only on the potential functions of the inhibitors and not on their effect on the outputs. Since the excitatory auxiliary neurons have exactly the same potential functions, the proofs follow immediately.

The main adaptation is in the second part where we use the predictability of the auxiliary neurons to predict the network for at least one input configuration. We proceed by bounding the gap in potentials between active outputs and inactive outputs by showing that the weight of the self-loop is large.

▶ **Observation 39.** $\mathrm{w}^{\mathrm{self}} \geq 2c \cdot \log n$.

Proof. In the steady state situation, there exists one leader u that fires in each round w.h.p. $1 - 1/n^c$ for polynomially many rounds. On the other hand, all other outputs v that do not have the positive feedback from the self-loop fire with probability $1/n^c$. Hence for such a round t in steady state, we have: $pot_t(u) \geq c \log n$ and $pot_t(v) \leq -c \log n$. We get that $\mathrm{w}^{\mathrm{self}} = pot_t(u) - pot(v) \geq 2c \log n$. The observation follows. ◀

An immediate corollary of that is the following:

▶ **Corollary 40.** *Consider a sub-round $(t, 2)$ and let F_{t-1} be the firing configuration of the auxiliary neurons in sub-round $(t-1, 3)$. If the firing probability of an inactive output v (output that did not fire in the previous sub-round $(t-1, 2)$) in sub-round $(t, 2)$ is at least $1/n^c$, then the firing probability of an active output u in sub-round $(t, 2)$ is $\geq 1 - 1/n^c$.*

Proof. Since all outputs have the same connections to the auxiliary neurons, only difference in the potential of an inactive output and an active output is the weight of the self-loop. Hence, $pot_t(u) = pot_t(v) + \mathrm{w}^{\mathrm{self}} \geq -c \log n + 2c \log n \geq c \log n$, where the first inequality follows by plugging Obs. 39 and using the fact that the firing probability of v is $1/(1+e^{-pot_t(v)}) \geq 1/n^c$. Thus, u fires with probability $1/(1 + e^{-c \log n}) = 1 - 1/n^c$. ◀

We now consider the second part of the lower bound where we predict $\Omega(\log \log n / \log \alpha)$ rounds of the network for at least one density input class. Since in the zero round no-output fires and w.h.p. also no auxiliary neuron is firing (since their bias value is $\omega(\log n)$), predicting the number of firing outputs in round 1 is exactly the same as in the only-inhibitor case.

C.1.0.1 Predicting the number of firing outputs in round $t \geq 2$.

We first define a subset of inputs $\mathcal{X}_t^{large} \subseteq \mathcal{X}_{t-1}$ for which we can predict the behavior of the outputs in the network N in round t. Let $\mathcal{X}_t^{same} \subseteq \mathcal{X}_{t-1}$ be the largest subset of inputs whose predicted firing vector $F_{t-1}(\mathrm{X})$ for the auxiliary neurons in round $t-1$ is the *same*, and denote this common firing vector by F_{t-1}^*. Let $\mathcal{X}_t^{large}$ be the set of inputs in $\mathcal{X}_t^{same}$ after omitting $\Theta(\log \log n)$ inputs with the smallest range value in round $t-1$. Eventually we will show that $\mathcal{X}_t^{large}$ is a reasonably large set of inputs compared to $\mathcal{X}_{t-1}$, and hence we can continue predicting behavior for at least some inputs for a large number of rounds. But first we show how to predict $R_t(\mathrm{X})$ for every input $\mathrm{X} \in \mathcal{X}_t^{large}$.

Let p' be the firing probability that an inactive output (one with $y_j^{t-1} = 0$) fires in sub-round $(t, 2)$ given that the inhibitors fired in sub-round $(t-1, 3)$ according to F_{t-1}^*. Since all inputs in $\mathcal{X}_t^{same}$ have the same predicted firing vector F_{t-1}^*, in each of them, an inactive output fires in sub-round $(t, 2)$ with probability p'. Let p be the corresponding firing

probability of an active output. We now consider two cases depending on the value of p'. If $p' < 1/n^c$, we predict that no inactive output fires in that round. Note that this prediction holds with probability $\geq 1 - 1/n^{c-1}$. In such a case we only predict the range for the active outputs in the exact same manner as before. Note that when we predicted the range of firing active outputs in the previous section, we did not use the fact that the auxiliary neurons are inhibitory, only that all competing outputs whose cardinality is to be estimated fire with the same probability in that round.

Next, we consider the more interesting case where $p' \geq 1/n^c$, that is the inactive outputs have a fair chance of firing in sub-round $(t, 2)$. Here, we make use of Lemma 40 that says that with probability at least $1 - 1/n^{c-1}$, all active outputs (i.e., that fired in round $t-1$) fire in sub-round $(t, 2)$ as well. Let $k = 2^i$ be the number of active inputs in the vector X. Let $E_{t-1} = E(\widehat{R}_{t-1}(\mathrm{X}) \mid F_{t-2}(X))$ be the expected number of firing outputs in sub-round $(t-1, 2)$ given the predicted firing vector $F_{t-2}(X)$. Then, the expected number of firing outputs in sub-round $(t, 2)$ is

$$E(\widehat{R}_t(\mathrm{X}) \mid F_{t-1}(X)) = E_{t-1} + p' \cdot (k - E_{t-1}) = (1 - p') \cdot E_{t-1} + p' \cdot k.$$

▶ **Claim 41.** *Let* $\mathrm{X}_1, \mathrm{X}_2 \in \mathcal{X}_t$ *be such that* $||\,\mathrm{X}_1\,||_1 \geq 2||\,\mathrm{X}_2\,||_1$. *Then*

$$E(\widehat{R}_t(\mathrm{X}_1) \mid F_{t-1}(\mathrm{X}_1)) \geq 2E(\widehat{R}_t(\mathrm{X}_2) \mid F_{t-1}(\mathrm{X}_2)).$$

Proof. We will prove by induction on the number of rounds t. Let $k_j = ||\,\mathrm{X}_j\,||_1$ and $E_{j,t} = E(\widehat{R}_t(\mathrm{X}_j) \mid F_{t-1}(\mathrm{X}_j))$ for $j \in \{1, 2\}$.

Since $\mathrm{X}_1, \mathrm{X}_2 \in \mathcal{X}_t$, it holds that $\mathrm{X}_1, \mathrm{X}_2 \in \mathcal{X}_\ell$ for every $\ell \in \{1, \ldots, t\}$ hence $F_\ell(\mathrm{X}_1) = F_\ell(\mathrm{X}_2)$ for every $\ell \in \{1, \ldots, t\}$. For the base of the induction of round $t = 1$, this clearly holds since $E_{0,t} = p_0 \cdot k_j$, $j \in \{1, 2\}$, where p_0 is the firing probability of an output where in the previous round no one fired. Assume the claim holds up to round $t-1$. We have that $E_{j,t} = E_{j,t-1} + p' \cdot (k_j - E_{j,t-1}) = (1 - p')E_{j,t-1} + p'k_j$, for $j \in \{1, 2\}$. By the induction assumption for $t-1$, we get $E_{1,t-1} \geq 2 \cdot E_{2,t-1}$ and by definition $k_1 \geq 2 \cdot k_2$, overall $E_{1,t} \geq 2E_{2,t}$ as required. ◀

We get that the expected number of firing outputs (conditioned on the predictions) are 2-separated. Now, we can claim exactly as before that all these expected values should be $\Omega(1/\log^4 n)$ as otherwise there is at least one input configuration for which there is a reset (i.e., in the next round no output fires) for $\Omega(\log n)$ times (see Obs. 26).

Since all expected predictions for the number of firing outputs are $\Omega(1/\log^4 n)$, by removing the $\Theta(\log\log n)$ inputs from $\mathcal{X}^{same}$ (i.e., as given by set $\mathcal{X}^{large}$), we get that all expected numbers of firing outputs are $\Omega(\log^7 n)$ and hence the random variables $\widehat{R}_t(\mathrm{X})$ are well concentrated around their expectation. The remaining proof goes exactly the same as in the inhibitory-case.

C.2 Extensions for the Lower Bound for High Probability Time

We define the *weak* WTA state to be state in which exactly one active output is firing (but possibly many inactive firing outputs). Whenever we use the notion of WTA nodes in the proof of Lemma 11, we now use the notion of weak WTA nodes instead. The definition of a reset node remains as is, i.e., a node u such that in its configuration $Q(u)$ no output (of any type) fires.

Note that the lower bound proof for the expected time implies that there is an input X_0 such that with a good probability after $t = \Omega(\log\log n/\log\alpha)$ rounds there are still $\Omega(\log n)$

competing outputs. After $t+1$ rounds, either we can assume w.h.p. that no inactive output fires or that all the $\Omega(\log n)$ active outputs fire. Hence, the lower bound implies that after $t+1$ rounds, with good probability, the number of firing active outputs is $\Omega(\log n)$, implying that the network is in a *weak* WTA state. Let $P_{1,j}(Q)$ be the probability that exactly one *active* output fires in sub-round $(j,2)$ given that the auxiliary neurons fire in round $j-1$ according to Q. Similarly, let $P_{0,j}(Q)$ be the probability that *no active output* fires in sub-round $(j,2)$ given Q. Finally, let $P_{01,j}(Q)$ be the probability that at most one active output fires in round $(j,2)$ given that configuration in round $j-1$ is Q, hence, $P_{01,j}(Q) = P_{1,j}(Q) + P_{0,j}(Q)$. Since we consider only the active outputs, Cl. 34 follows as is. We now claim the following.

▶ **Corollary 42.** *For every round j and for every vector $Q \in \{0,1\}^{n+\alpha}$ in which there are at least two active firing outputs and such that $P_{01,j}(Q) \geq \Theta(1/\log\log n)$, it holds that there is a (total) reset in round j (i.e., no output fires) with probability at least $\Theta(1/(\log\log n)^3)$.*

Proof. Since in Q there are at least two firing *active* outputs, by Cl. 34, $P_{0,j}(Q) \geq \Theta(1/(\log\log n)^3)$. Hence the probability that no active output fires is at least $\Theta(1/(\log\log n)^3)$. We now claim that the probability that also no inactive output fires is at least $1-1/n^{c-1}$. Hence, by the independence between the output decisions (given the firing states of the inhibitors), we get that the probability that no output fires is at least $\Theta(1/(\log\log n)^3)$ as required.

Assume towards contradiction that inactive output fires with probability $\geq 1/n^c$. By Cor. 40, we get that an active output fires with probability at least $1-1/n^c$. Since in the previous round there are at least two firing *active* outputs, we get that with probability $\geq 1-1/n^c$ there are at least two firing outputs in sub-round $(j,2)$, contradiction to the assumption that $P_{01,j}(Q) \geq \Theta(1/\log\log n)$.

Thus we get that each inactive output fires with probability $< 1/n^c$, and with probability $\geq 1-1/n^c$ no inactive output fires. The claim follows. ◀

Equipped with Cor. 42 and the lower bound for expected time, we can now use the execution tree to show that the weight of non weak-WTA nodes is at least $1/n^2$. The same idea generally holds up to few adaptations. Recall that in our execution tree traversal, at step j we obtain a collection of non weak WTA nodes. That is nodes u with configuration $Q(u)$ which either there are at least two active outputs that are firing. For $j \geq 1$ given U_j, the set U_{j+1} is obtained by defining for each node $u \in U_j$, a subset of non weak WTA nodes $V(u)$ as described next.

Case 1: u is a reset node. Set the subtree depth $d(u) = \min\{DH - dist(u,r,T), DC\}$ and let $V(u)$ be the non weak WTA nodes in the leaf nodes of $T_{d(u)}(u)$. By the lower bound proof, the set $V(u)$ captures $1-1/\log n$ of the probability mass in $T(u)$.

Since u is a non weak WTA node, it remains to consider the case where the number of active firing outputs in $Q(u)$ is at least 2. Recall that $P_{0,1}(Q(u))$ is the probability that in sub-round $(j,2)$ at most one active output fires given that the configuration in round $j-1$ is $Q(u)$.

Case 2.1: $P_{0,1}(Q(u)) \geq \Theta(1/\log\log n)$. Let $V'(u)$ be the children of u in T that are reset-nodes. For each reset-node $w \in V'(u)$, let $V(w)$ be the non-WTA nodes in the leaf nodes of $T_{d(u)-1}(w)$ and let $V(u) = \bigcup V(w)$.

By Cl. 42, since the number of firing active outputs in $Q(u)$ is at least 2 and since $P_{0,1}(Q(u)) \geq \Theta(1/\log\log n)$, the probability for a (total) reset in the next round is at least $\Theta(1/(\log\log n)^3)$ and hence $V'(u)$ captures $\Theta(1/(\log\log n)^3)$ of the probability mass in $T(u)$. This will allow us to follow the same argument as before when following the case 2.1.

Case 2.2: $P_{0,1}(Q(u)) < \Theta(1/\log\log n)$. Let $V(u)$ be the children of u that have at least

2 *active* outputs in $Q(v)$ (hence $d(u) = 1$). Since $P_{0,1}(u) \leq \Theta(1/\log\log n)$, we capture $1 - \Theta(1/\log\log n)$ of the weight of the tree $T(u)$. This completes the definition of U_{j+1}. The argument that uses this case follows now the exact same line. In sum, either we capture only $\Theta(1/\log\log n)$ of the probability mass in such a case we have a large jump in the tree or that we capture $1 - \Theta(1/\log\log n)$ of the probability mass. As before using the amortization argument, overall the number of non weak WTA nodes can be bounded by $\geq 1/n^2$. The completes the extension to excitatory auxiliary neurons.