# Combinatorial Capacity of modReLU Complex Networks: VC-Dimension Bounds and Lower Limits

**Anonymous authors**
**Paper under double-blind review**

## Abstract

Complex-valued neural networks (CVNNs) are increasingly used in settings where both magnitude and phase of the signal carry information. In particular, deep networks with the modReLU activation function have become standard in applications such as MRI reconstruction, radar, and complex-valued time-series modeling. While approximation properties of such networks have recently been analyzed in detail, their statistical capacity in the sense of VC-dimension has not, to the best of our knowledge, been studied.

In this paper we formalize a natural class of fully connected deep complex-valued networks with modReLU activation and real sign output, and view them as binary classifiers on $\mathbb{R}^{2d}$ via the usual realification. Using tools from real algebraic geometry and a VC-dimension bound for semi-algebraic concept classes due to Goldberg and Jerrum, together with quantitative bounds for quantifier elimination, we prove that for any architecture with $W$ real parameters and depth $L$, the VC-dimension of the corresponding hypothesis class is at most on the order of $W^2 \log W$, with a universal constant independent of the particular architecture.

On the other hand, by restricting to real inputs and parameters and exploiting results of Harvey, Liaw, and Mehrabian and of Bartlett et al. on deep networks with piecewise-linear activations, we obtain lower bounds of order $WL \log(W/L)$ for suitable depth-$L$ architectures within the modReLU class. Thus the VC-dimension of these networks grows at least linearly in both $W$ and $L$, and at most quadratically in $W$ up to a logarithmic factor. Closing this gap is an interesting open problem.

## 1 Introduction

In many signal-processing problems the natural data space is complex-valued. Examples include MRI, radar, wireless communication channels, and certain Fourier or wavelet representations of time series. For such problems, complex-valued neural networks (CVNNs) have been proposed as a more natural alternative to real networks; see, for instance, (10; 11; 12; 18; 13; 15) and the references therein. In these architectures all weights, biases, and activations are complex, and the network computes a map from $\mathbb{C}^d$ to $\mathbb{C}^k$.

Recently, there has been considerable progress on the approximation theory of CVNNs. In particular, Geuchen and Voigtlaender (6) established explicit quantitative error bounds for approximating smooth functions on compact subsets of $\mathbb{C}^d$ by complex-valued networks with the modReLU activation function, and showed that the rates are optimal up to logarithmic factors. Follow-up work has treated phenomena such as depth separation and universality for narrow complex networks (7; 12). A recent survey by Lee, Hasegawa, and Gao (13) gives a broad overview of complex activations, Wirtinger calculus, and optimization aspects.

In contrast, the statistical learning theory of CVNNs is still largely undeveloped. For real networks, a rich body of work has characterized the VC-dimension and pseudo-dimension of various architectures since the classical paper of Baum and Haussler (5) and the monograph of Anthony and Bartlett (1). For example, deep ReLU networks with $W$ real parameters and depth $L$ are now known to have VC-dimension of order $WL \log W$, and this dependence is essentially tight (9; 3). Related results characterize the pseudo-dimension of real networks in regression settings (16; 20).

For complex-valued networks there are, to the best of our knowledge, no analogous capacity results in terms of VC-dimension. The only global complexity bounds we are aware of are recent norm-based generalization bounds that control the generalization gap via a spectral-complexity measure for CVNNs (21), which is a different notion from combinatorial capacity. The main purpose of this work is to begin closing this gap by providing VC-dimension bounds for a widely used complex-valued architecture family.

**Contributions.** At a high level, the contributions of this paper are as follows.

- We formalize a natural class of fully connected deep complex-valued networks with modReLU activation and real sign output, and view them as binary classifiers on $\mathbb{R}^{2d}$ via realification.
- We show that the corresponding decision regions form a semi-algebraic family in $(\theta, x)$ whose description complexity—in terms of the number and degree of defining polynomials—can be bounded, in a way that depends only on the architecture and the number $W$ of real parameters.
- Using a VC-dimension bound for semi-algebraic concept classes due to Goldberg and Jerrum, combined with quantitative bounds on quantifier elimination, we derive an upper bound of order $W^2 \log W$ for the VC-dimension of this class. To the best of our knowledge, this is the first general VC-dimension upper bound for fully connected deep complex-valued networks.
- By specializing to real inputs and constraining parameters appropriately, we show that our class contains, as a subclass, real-valued networks with a fixed piecewise-linear activation. Results of Harvey, Liaw, and Mehrabian (9) and Bartlett et al. (3) then imply lower bounds of order $WL \log(W/L)$ for suitable architectures, showing that the VC-dimension must grow at least linearly in both $W$ and $L$.
- We discuss how the same approach could be extended to other complex activations and to pseudo-dimension, and outline several open problems, including whether the $W^2 \log W$ upper bound can be sharpened towards $WL \log W$ in the complex-valued setting.

## 2 Preliminaries

In this section we collect the definitions and auxiliary results that will be needed. We assume basic familiarity with VC-dimension; for background see, e.g., Chapter 3 of (19) or (1; 20).

**Semi-algebraic sets and VC-dimension.** A set $S \subset \mathbb{R}^n$ is called *semi-algebraic* if it can be expressed as a finite Boolean combination of polynomial equalities and inequalities in the coordinates $x_1, \ldots, x_n$. The *description complexity* of a semi-algebraic set refers to the number of such polynomial conditions and the maximum degree among them. A family of sets $\{S_\theta\}_{\theta \in \Theta}$ (where $\Theta$ is an index set) is called semi-algebraic in $\mathbb{R}^n$ if the set

$$\{(\theta, x) \in \Theta \times \mathbb{R}^n : x \in S_\theta\}$$

is a semi-algebraic subset of $\mathbb{R}^{\dim(\Theta)+n}$.

We will use the following general result on the VC-dimension of semi-algebraic set families due to Goldberg and Jerrum (see also (1, Theorem 8.3) and (4, Theorem 7.38)).

**Theorem 1 (Goldberg–Jerrum)** *Let $F = \{S_\theta\}_{\theta \in \Theta}$ be a family of subsets of $\mathbb{R}^n$ parameterized by $\Theta \subset \mathbb{R}^W$. Suppose there exist integers $s \geq 1$ and $d \geq 1$ and polynomials $p_1, \ldots, p_s \in \mathbb{R}[U_1, \ldots, U_W, X_1, \ldots, X_n]$ of degree at most $d$ such that for all $\theta \in \Theta$ and $x \in \mathbb{R}^n$,*

$$x \in S_\theta \iff \Phi\big(\mathrm{sign}(p_1(\theta, x)), \ldots, \mathrm{sign}(p_s(\theta, x))\big) = \text{true},$$

*for some fixed Boolean formula $\Phi$. Then there exists a universal constant $C > 0$ such that*

$$\mathrm{VCdim}(F) \;\leq\; C\,W\,\log(s\,d).$$

*In words, if each set $S_\theta$ is definable by a Boolean combination of a fixed number $s$ of polynomial inequalities of degree at most $d$, and the parameter space has dimension $W$, then the VC-dimension is at most on the order of $W \log(sd)$.*

**Complex-valued neural networks and modReLU.** Let $d \geq 1$ and $L \geq 1$ be integers. A fully connected feedforward complex-valued network (CVNN) of depth $L$ and input dimension $d$ consists of:

- layer widths $n_0 = d,\ n_1,\ \ldots,\ n_{L-1},\ n_L = 1$,
- complex weight matrices $W^{(\ell)} \in \mathbb{C}^{n_\ell \times n_{\ell-1}}$ and complex biases $b^{(\ell)} \in \mathbb{C}^{n_\ell}$ for $\ell = 1, \ldots, L-1$,
- output-layer parameters $W^{(L)} \in \mathbb{C}^{1 \times n_{L-1}}$ and $b^{(L)} \in \mathbb{C}$.

Given an input $z^{(0)} = x \in \mathbb{C}^d$, the network computes recursively for $\ell = 1, \ldots, L-1$:

$$u^{(\ell)} \ = \ W^{(\ell)} z^{(\ell-1)} + b^{(\ell)} \ \in \ \mathbb{C}^{n_\ell},$$
$$z_j^{(\ell)} \ = \ \sigma(u_j^{(\ell)}), \qquad j = 1, \ldots, n_\ell,$$

and the output $F_\theta(x) \in \mathbb{C}$ is given by

$$F_\theta(x) \ = \ W^{(L)} z^{(L-1)} + b^{(L)},$$

where $\theta$ denotes the collection of all real parameters (the real and imaginary parts of all weights and biases). The total number of real parameters is

$$W \ = \ 2 \sum_{\ell=1}^{L} n_\ell n_{\ell-1} \ + \ 2 \sum_{\ell=1}^{L} n_\ell.$$

Here $\sigma : \mathbb{C} \to \mathbb{C}$ is a fixed complex activation function applied coordinate-wise. We work with the *modReLU* activation, introduced by Trabelsi et al. (18) and further analyzed by Geuchen and Voigtlaender (6). In polar form $z = re^{i\theta}$ with $r = |z| \geq 0$ and $\theta \in [0, 2\pi)$, modReLU is defined by

$$\sigma_{\mathrm{modR}}(z) \ = \ \begin{cases} 0, & r + \beta \leq 0, \\ (r+\beta)e^{i\theta}, & r + \beta > 0, \end{cases} \tag{1}$$

where $\beta \in \mathbb{R}$ is a (learnable) real bias parameter. In real coordinates $z = x + iy$ (so that $r = \sqrt{x^2 + y^2}$), the activation can be described piecewise in terms of $x$ and $y$.

The overall network defines a complex-valued function $F_\theta : \mathbb{C}^d \to \mathbb{C}$, and for binary classification we use the sign of the real part as output:

$$h_\theta(x) \ = \ \mathrm{sign}\big(\Re(F_\theta(x))\big) \ \in \ \{+1, -1\}. \tag{2}$$

We view $h_\theta$ as a classifier on $\mathbb{R}^{2d}$ via the standard identification of $\mathbb{C}^d$ with $\mathbb{R}^{2d}$. More formally, let $\Psi : \mathbb{R}^{2d} \to \mathbb{C}^d$ be the bijection that groups real coordinates into complex coordinates. For $\tilde{x} \in \mathbb{R}^{2d}$, we write $x = \Psi(\tilde{x}) \in \mathbb{C}^d$ and define

$$h_\theta(\tilde{x}) \ = \ \mathrm{sign}\big(\Re(F_\theta(x))\big). \tag{3}$$

Let $\mathcal{H}_{W,L}^{\mathrm{modR}}$ denote the set of all classifiers of the form (3) obtained by varying $\theta \in \mathbb{R}^W$ over all possible choices of weights and biases for a fixed architecture $(n_0, \ldots, n_L)$ as above. (Note that $W$ depends on the architecture; we keep this dependence implicit to avoid cumbersome notation.)

## 3 VC-dimension upper bound

**Theorem 2 (VC-dimension of modReLU networks)** *There exists a universal constant $C > 0$ such that for every choice of layer widths $(n_0, \ldots, n_L)$ with total number of real parameters $W \geq 2$, the VC-dimension of $\mathcal{H}_{W,L}^{\mathrm{modR}}$ satisfies*

$$\mathrm{VCdim}\big(\mathcal{H}_{W,L}^{\mathrm{modR}}\big) \ \leq \ C\, W^2 \, \log W.$$

*In particular, for any fixed depth $L$, $\mathrm{VCdim}(\mathcal{H}_{W,L}^{\mathrm{modR}}) = O(W^2 \log W)$ uniformly over architectures with $W$ parameters.*

The remainder of this section is devoted to the proof.

### 3.1 Semi-algebraic representation of the decision map

The key step is to show that the family

$$\{\tilde{x} \in \mathbb{R}^{2d} : h_\theta(\tilde{x}) = +1\}$$

is semi-algebraic in $(\theta, \tilde{x})$, with description complexity controlled (in an appropriate sense) by $W$ and the architecture.

We first note that all operations used in the network computation can be described by polynomial equalities and inequalities in real coordinates. The following real functions and relations are semi-algebraic:

1. Addition and multiplication: $(u, v) \mapsto u + v$, $(u, v) \mapsto uv$ on $\mathbb{R}$.
2. Absolute value and maximum: $u \mapsto |u|$, $(u, 0) \mapsto \max\{u, 0\}$.
3. Square root (on $t \geq 0$): $t \mapsto \sqrt{t}$.
4. The complex modulus and the condition $|z| + \beta > 0$ for $z = x + iy \in \mathbb{R}^2$ and $\beta \in \mathbb{R}$.
5. The modReLU activation $\sigma_{\mathrm{modR}} : \mathbb{R}^3 \to \mathbb{R}^2$, mapping $(x, y, \beta)$ to $(\Re(\sigma_{\mathrm{modR}}(x + iy)), \Im(\sigma_{\mathrm{modR}}(x + iy)))$, as defined in (1).

Items (1) and (2) are standard: the graphs of addition and multiplication are algebraic sets in $\mathbb{R}^3$; the absolute value function satisfies $|u| = v$ if and only if $v \geq 0$ and $v^2 = u^2$, and $\max\{u, 0\}$ can be expressed via $u$ and $|u|$ as $(u+|u|)/2$. For (3), the graph of the square root function on $t \geq 0$ is given by the semi-algebraic set

$$\{(t, r) \in \mathbb{R}^2 : r \geq 0,\ r^2 = t\}.$$

For (4), write $|z| + \beta > 0$ as

$$\exists r \in \mathbb{R}\ \big(r \geq 0,\ r^2 = x^2 + y^2,\ r + \beta > 0\big).$$

The set of quadruples $(x, y, \beta, r)$ satisfying these polynomial (in)equalities is semi-algebraic, and by the Tarski–Seidenberg theorem its projection onto the $(x, y, \beta)$–coordinates, which is precisely

$$\{(x, y, \beta) : |x + iy| + \beta > 0\},$$

is also semi-algebraic.

For (5), recall that $\sigma_{\mathrm{modR}}(z)$ equals 0 when $r+\beta \leq 0$, and equals $(r+\beta)z/r$ when $r+\beta > 0$ (with $r = |z| > 0$), and that by definition $\sigma_{\mathrm{modR}}(0) = 0$. Introduce an auxiliary variable $r$ and consider the subset of $\mathbb{R}^6$ with coordinates $(x, y, \beta, u, v, r)$ given by the disjunction

$$\big(r \geq 0,\ r^2 = x^2 + y^2,\ r + \beta \leq 0,\ u = 0,\ v = 0\big)$$

(inactive region) and

$$\big(r > 0,\ r^2 = x^2 + y^2,\ r + \beta > 0,\ ur = (r + \beta)x,\ vr = (r + \beta)y\big)$$

(active region), together with the special case $x = y = 0,\ u = v = 0$ (which is compatible with both regimes and matches the convention $\sigma_{\mathrm{modR}}(0) = 0$). This is a finite Boolean combination of polynomial equalities and inequalities, hence defines a semi-algebraic subset of $\mathbb{R}^6$. Its projection onto the $(x, y, \beta, u, v)$–coordinates is exactly the graph of $\sigma_{\mathrm{modR}}$, and therefore, again by the Tarski–Seidenberg theorem, this graph is semi-algebraic.

The above argument is not constructive in the sense that we do not write the defining polynomials for the projection explicitly, but instead use the existence of an equivalent quantifier-free description guaranteed by quantifier elimination. This is sufficient for our purposes, since we only need to know that there is some finite collection of polynomials (with degree bounded in terms of the architecture) whose sign pattern determines the decision.

**Lemma 1 (Semi-algebraic decision map)** *For the network class defined above, there exist integers $s \geq 1$ and $d \geq 1$, depending only on the architecture $(n_0, \ldots, n_L)$, and polynomials $p_1, \ldots, p_s \in \mathbb{R}[U_1, \ldots, U_W, X_1, \ldots, X_{2d}]$ of degree at most $d$ such that for all $\theta \in \mathbb{R}^W$ and all $\tilde{x} \in \mathbb{R}^{2d}$,*

$$h_\theta(\tilde{x}) = +1 \iff \Phi\big(\mathrm{sign}(p_1(\theta, \tilde{x})), \ldots, \mathrm{sign}(p_s(\theta, \tilde{x}))\big) = \mathrm{true},$$

*for some fixed Boolean formula $\Phi$ independent of $\theta$ and $\tilde{x}$. Moreover, there exist absolute constants $c_0, c_1 > 0$ such that one can choose $s$ and $d$ so that*

$$s \leq (c_0 W)^{c_1 W} \qquad \text{and} \qquad d \leq 2^{c_1 W}.$$

*In particular, the family of sets*

$$S_\theta := \{\tilde{x} \in \mathbb{R}^{2d} : h_\theta(\tilde{x}) = +1\}$$

*is semi-algebraic in $(\theta, \tilde{x})$ with description complexity controlled by $s$ and $d$, and $\log(sd) = O(W \log W)$ uniformly over architectures with $W$ parameters.*

**Proof.** For each layer $\ell$ and neuron $j$ we introduce real variables to represent the real and imaginary parts of the pre-activation and post-activation outputs for that neuron, for a generic input $\tilde{x}$ and parameter $\theta$. The relations between these variables, $\tilde{x}$, and $\theta$ are:

- affine linear equalities expressing the pre-activations as real linear forms in the previous layer's outputs and the real and imaginary parts of the weights and biases;
- the semi-algebraic constraints describing the modReLU activation, as in item (5) above, which relate the pre-activation $(x, y)$, the bias parameter $\beta$, the auxiliary radius variable $r$, and the post-activation output $(u, v)$;
- at the output layer, linear relations expressing the real part of $F_\theta(x)$ as a real linear combination of the last hidden layer outputs and the output weights and bias, together with the inequality $\Re(F_\theta(x)) > 0$ implementing the sign decision.

Collecting all these constraints yields a first-order formula in the language of ordered fields, with free variables $(\theta, \tilde{x})$ and finitely many auxiliary variables representing intermediate quantities (pre- and post-activations, radii $r$, etc.). The number of such auxiliary variables is bounded by a constant multiple of the total number of neurons, and hence is $O(W)$; each pre-activation introduces two real coordinates and each activation introduces one radius variable. Similarly, the number of atomic polynomial equalities and inequalities appearing in the formula is bounded above by a constant multiple of the total number of neurons and connections, and hence by $c_2 W$ for some absolute constant $c_2 > 0$. Moreover, the degrees of the polynomials involved are uniformly bounded (at most 2 for the constraints involving magnitudes, and 1 for the affine linear constraints).

Now consider the set

$$G = \{(\theta, \tilde{x}) \in \mathbb{R}^{W+2d} : \Re(F_\theta(\Psi(\tilde{x}))) > 0\}.$$

By construction, $(\theta, \tilde{x}) \in G$ if and only if there exist auxiliary variables satisfying all of the aforementioned polynomial equalities and inequalities together with the additional inequality $\Re(F_\theta(x)) > 0$. Hence $G$ is definable by a first-order formula in the theory of real closed fields whose atomic predicates are polynomial (in)equalities of degree at most 2, and the number of atomic predicates is at most $c_2 W$.

By quantifier elimination for real closed fields (Tarski–Seidenberg; see, e.g., (17; 4)), there exists an equivalent quantifier-free formula in which $G$ is defined by Boolean combinations of polynomial equalities and inequalities in the free variables $(\theta, \tilde{x})$ alone. Quantitative bounds for quantifier elimination (see, for instance, (4, Theorem 14.16)) imply that there are constants $c_0, c_1 > 0$ such that one can choose a quantifier-free formula with at most

$$s \leq (c_0 W)^{c_1 W}$$

distinct polynomials, each of degree at most

$$d \;\le\; 2^{c_1 W},$$

in the variables $(\theta, \tilde{x})$. (The precise form of these bounds is not important; what matters is that $s$ and $d$ are at most exponential in a linear function of $W$.) Writing these polynomials as $p_1, \ldots, p_s$ and absorbing the structure of the quantifier-free formula into a Boolean function of their signs yields the desired representation

$$(\theta, \tilde{x}) \in G \iff \Phi\big(\operatorname{sign}(p_1(\theta, \tilde{x})), \ldots, \operatorname{sign}(p_s(\theta, \tilde{x}))\big) = \text{true}.$$

Finally, note that $h_\theta(\tilde{x}) = +1$ is equivalent to $(\theta, \tilde{x}) \in G$, so the level sets $S_\theta$ are exactly the fibers of $G$ in the $\tilde{x}$–coordinates. The bound on $\log(sd)$ follows immediately from the bounds on $s$ and $d$:

$$\log(sd) \;\le\; \log s + \log d \;\le\; c_1 W \log(c_0 W) + c_1 W = O(W \log W).$$

This proves the lemma.

### 3.2 Proof of the VC-dimension bound

We now combine Lemma 1 with Theorem 1 to prove Theorem 2. Consider the family $\mathcal{F} = \{S_\theta\}_{\theta \in \mathbb{R}^W}$ defined by

$$S_\theta := \{\tilde{x} \in \mathbb{R}^{2d} : h_\theta(\tilde{x}) = +1\}.$$

By Lemma 1, there exist polynomials $p_1, \ldots, p_s$ in $\mathbb{R}[U_1, \ldots, U_W, X_1, \ldots, X_{2d}]$ of degree at most $d$ (with $s$ and $d$ depending only on the architecture $(n_0, \ldots, n_L)$ and the choice of activation) such that each $S_\theta$ is describable by a fixed Boolean formula in terms of the signs of $p_i(\theta, \tilde{x})$. Thus we may apply Theorem 1. Since $\dim(\Theta) = W$, we obtain

$$\mathrm{VCdim}(\mathcal{H}^{\mathrm{modR}}_{W,L}) = \mathrm{VCdim}(\mathcal{F}) \;\le\; C\, W\, \log(s\, d),$$

for some universal constant $C > 0$. By Lemma 1 we have $\log(sd) = O(W \log W)$ uniformly over architectures with $W$ parameters, so there exists a constant $C' > 0$ such that

$$\mathrm{VCdim}(\mathcal{H}^{\mathrm{modR}}_{W,L}) \;\le\; C\, W \cdot C' W \log W \;=\; (CC')\, W^2 \log W.$$

Renaming the product $CC'$ as $C$ yields the claimed bound in Theorem 2.

### 3.3 Lower bounds and comparison

We now record a lower bound, obtained by specializing known results for real-valued networks with piecewise-linear activations to an appropriate subclass of the modReLU architecture.

**Proposition 1 (Lower bound)** *Let $W, L$ be positive integers with $W$ sufficiently large compared to $L$. Then there exists a fully connected depth-$L$ complex-valued network with modReLU activation and at most $W$ real parameters whose associated classifier class (with sign output as in* (2)*) has VC-dimension at least*

$$c\, W\, L \log \frac{W}{L},$$

*for some universal constant $c > 0$. In particular, for suitable architectures, the VC-dimension of $\mathcal{H}^{\mathrm{modR}}_{W,L}$ grows at least on the order of $W\, L \log(W/L)$.*

**Proof sketch.** Harvey, Liaw, and Mehrabian (9) and Bartlett et al. (3) show that for any fixed piecewise-linear activation function $\psi : \mathbb{R} \to \mathbb{R}$ with a constant number of pieces, there exist depth-$L$ real-valued networks with activation $\psi$ and at most $W$ real parameters whose VC-dimension is at least $cW\, L \log(W/L)$ for some universal constant $c > 0$. Their results apply to arbitrary such $\psi$, not just ReLU.

Fix a real constant $\beta_0$ and consider the restriction of the modReLU activation to the real line with this fixed bias:

$$\phi_{\beta_0}(x) := \Re\big(\sigma_{\mathrm{modR}}(x + i0)\big).$$

For each $\beta_0$, $\phi_{\beta_0}$ is a piecewise-linear function of $x$ with at most three linear pieces (depending on the sign of $\beta_0$). Thus $\phi_{\beta_0}$ satisfies the assumptions of (9; 3), and hence there exist depth-$L$ real networks with activation $\phi_{\beta_0}$ and at most $W'$ real parameters whose VC-dimension is at least $cW'L\log(W'/L)$ for some $c > 0$ and all sufficiently large $W', L$.

Such a real network can be viewed as a special case of our complex-valued modReLU architecture by:

- restricting the input to lie in $\mathbb{R}^d \subset \mathbb{C}^d$ (identified with $x + i0$),
- constraining all weights and biases in the complex network to be real (i.e., imaginary parts fixed to zero),
- fixing the modReLU bias parameter of each unit to be the constant $\beta_0$.

Under these restrictions, the complex-valued network computes exactly the same real-valued function on $\mathbb{R}^d$ as the corresponding real network with activation $\phi_{\beta_0}$. The number of free real parameters in this subclass of complex networks is $W'$, whereas the total number of real parameters (if one counts the fixed bias parameters as well) is at most $W \leq c_1 W'$ for some architecture-dependent constant $c_1$ (since the number of units is at most linear in the number of weights and biases).

It follows that for any sufficiently large $W, L$ one can choose $W' \geq W/c_1$ and obtain a depth-$L$ complex modReLU network with at most $W$ real parameters whose classifier class has VC-dimension at least

$$c\,W'L\log\frac{W'}{L} \;\geq\; \frac{c}{c_1}\,WL\log\frac{W}{c_1 L} \;\geq\; c_2\,WL\log\frac{W}{L}$$

for some universal constant $c_2 > 0$ and all sufficiently large $W, L$ with $W$ larger than a fixed multiple of $L$. This proves the proposition.

Combining Proposition 1 with Theorem 2 we obtain, for suitable depth-$L$ architectures and large $W$,

$$c_2\,WL\log\frac{W}{L} \;\lesssim\; \mathrm{VCdim}(\mathcal{H}^{\mathrm{modR}}_{W,L}) \;\lesssim\; W^2\log W.$$

Thus the VC-dimension grows at least linearly in both $W$ and $L$, and at most quadratically in $W$ up to a logarithmic factor. Narrowing this gap—for instance by sharpening the upper bound to something closer to $WL\log W$ in the complex-valued setting—is left as an open problem.

## 4 Discussion and related work

We have shown that the VC-dimension of deep complex-valued networks with modReLU activation and $W$ real parameters is at most on the order of $W^2\log W$, uniformly over fully connected architectures of a given size. The proof relied on two main ingredients: (i) the modReLU network computation can be encoded as a semi-algebraic family with description complexity controlled by the architecture and the number of parameters, and (ii) semi-algebraic concept classes of this type have VC-dimension bounded by $O(W\log W)$ in terms of the number of parameters and the logarithm of the number and degree of defining polynomials, via the Goldberg–Jerrum theorem. Quantitative bounds on quantifier elimination then imply that $\log(sd) = O(W\log W)$, yielding the $O(W^2\log W)$ bound.

On the lower-bound side, by restricting to real inputs and parameters and fixing the modReLU bias parameter across units, we obtain a subclass of our complex-valued architecture that falls within the scope of the nearly tight VC-dimension bounds of (9; 3) for networks with piecewise-linear activations. This yields lower bounds of order $WL\log(W/L)$ for suitable depth-$L$ architectures within the modReLU class. Taken together, these results show that the VC-dimension of deep modReLU networks grows at least linearly in both the number of parameters and the depth, and at most quadratically in the number of parameters up to a logarithmic factor.

Our work suggests a number of directions for further research. First, it would be interesting to explore whether the $W^2\log W$ factor can be improved, perhaps by exploiting finer properties of the modReLU activation and the specific structure of complex multiplication, in analogy with the sharpened analyses for ReLU networks in (9; 3). Second, one may ask whether similar VC-dimension bounds hold for other complex activations beyond modReLU, or for convolutional and recurrent complex-valued networks. These extensions are not automatic, as they would likely require new semi-algebraic encodings or bounding techniques beyond the direct analogues of known analyses of real ReLU networks. Finally, one could study pseudo-dimension and other complexity measures for complex-valued architectures, and compare the resulting generalization guarantees to norm-based bounds such as those of (21).

**Limitations.** Our analysis is restricted to fully connected feedforward architectures with modReLU activation, and we do not track constants or lower-order terms in the VC-dimension bounds. The semi-algebraic encoding relies on quantifier elimination, which is known to be computationally heavy; we use it only as a theoretical tool, not as a practical algorithm. The resulting $O(W^2 \log W)$ upper bound is unlikely to be optimal, as suggested by the nearly tight $O(WL \log W)$ bounds available for real-valued piecewise-linear networks (9; 3). We also do not address data-dependent or norm-based complexity measures for complex-valued networks, nor do we consider convolutional, recurrent, or attention-based architectures.

### Reproducibility Statement

This is a theoretical paper. All claims are stated as formal theorems, lemmas, or propositions, with proofs provided in the main text or as proof sketches that rely on standard results from real algebraic geometry and VC theory. No datasets, code, or hyperparameters were used. The results can be checked by following the derivations and cited references.

**Broader Impact Statement.** This work is purely theoretical and studies the combinatorial capacity of a class of complex-valued neural networks. We do not anticipate direct negative societal impacts beyond those already associated with general advances in machine learning theory. Any downstream impact will arise only through applications that use complex-valued networks; in those settings, standard considerations around fairness, robustness, and potential misuse of models still apply.

## References

[1] M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations.* Cambridge University Press, 1999.

[2] P. L. Bartlett, D. J. Foster, and M. Telgarsky. Spectrally-normalized margin bounds for neural networks. In *Advances in Neural Information Processing Systems 30*, pages 6240–6249. Curran Associates, Inc., 2017.

[3] P. L. Bartlett, N. Harvey, C. Liaw, and A. Mehrabian. Nearly-tight VC-dimension and pseudodimension bounds for deep neural networks. *Journal of Machine Learning Research*, 20(63):1–17, 2019.

[4] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in Real Algebraic Geometry.* Springer, 2nd edition, 2006.

[5] E. B. Baum and D. Haussler. What size net gives valid generalization? *Neural Computation*, 1(1):151–160, 1989.

[6] P. Geuchen and F. Voigtlaender. Optimal approximation using complex-valued neural networks. In *Advances in Neural Information Processing Systems 36*, 2023.

[7] P. Geuchen. Universal approximation with complex-valued deep narrow networks. *Constructive Approximation*, 2025.

[8] P. W. Goldberg and M. R. Jerrum. Bounding the VC-dimension of concept classes parameterized by real numbers. *Machine Learning*, 18(2–3):131–148, 1995.

[9] N. Harvey, C. Liaw, and A. Mehrabian. Nearly-tight VC-dimension bounds for piecewise linear neural networks. In *Proceedings of the Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 1064–1068. PMLR, 2017.

[10] A. Hirose. Complex-valued neural networks for signal processing. In S. Haykin and B. Kosko, editors, *Intelligent Signal Processing*, pages 471–509. Wiley–IEEE Press, 2003.

[11] A. Hirose. *Complex-Valued Neural Networks*. Springer, 2nd edition, 2012.

[12] Y. J. Kim and S. Adali. Approximation by fully complex multilayer perceptrons. *Neural Computation*, 15(7):1641–1666, 2003.

[13] C. Y. Lee, H. Hasegawa, and S. Gao. Complex-valued neural networks: A comprehensive survey. *IEEE/CAA Journal of Automatica Sinica*, 9(8):1406–1426, 2022.

[14] B. Neyshabur. Towards learning convolutions from scratch. arXiv preprint arXiv:1809.01062, 2018.

[15] J. W. Smith, P. J. Wolfe, and V. Tarokh. Complex-valued deep learning for signal processing. arXiv preprint arXiv:2309.07948, 2023.

[16] E. D. Sontag. VC dimension of neural networks. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 1188–1192. MIT Press, 2nd edition, 2003.

[17] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. University of California Press, 1951.

[18] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal. Deep complex networks. In *International Conference on Learning Representations*, 2018.

[19] V. N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

[20] M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019.

[21] H. Chen, F. He, S. Lei, and D. Tao. Spectral complexity-scaled generalisation bound of complex-valued neural networks. *Artificial Intelligence*, 322:103951, 2023.