AsiaPacificNeuralNetworkSociety **ICONIP 2019**

12th – 15th Dec 2019 - Sydney, Australia

One Analog Neuron Cannot Recognize Deterministic Context-Free Languages

Jiří Šíma & Martin Plátek

sima@cs.cas.cz

Institute of Computer Science
Czech Academy of Sciences, Prague

The Computational Power of Neural Networks (NNs)

(discrete-time recurrent NNs with the saturated-linear activation function)

depends on the information contents of weight parameters:

1. **integer** weights: **finite automaton** (Minsky, 1967)
2. **rational** weights: **Turing machine** (Siegelmann, Sontag, 1995)

   polynomial time $\equiv$ complexity class P

   polynomial time & increasing **Kolmogorov complexity** of real weights $\equiv$ a proper **hierarchy** of nonuniform complexity classes **between P and P/poly**

   (Balcázar, Gavaldà, Siegelmann, 1997)
3. arbitrary **real** weights: **"super-Turing"** computation (Siegelmann, Sontag, 1994)

   polynomial time $\equiv$ nonuniform complexity class P/poly

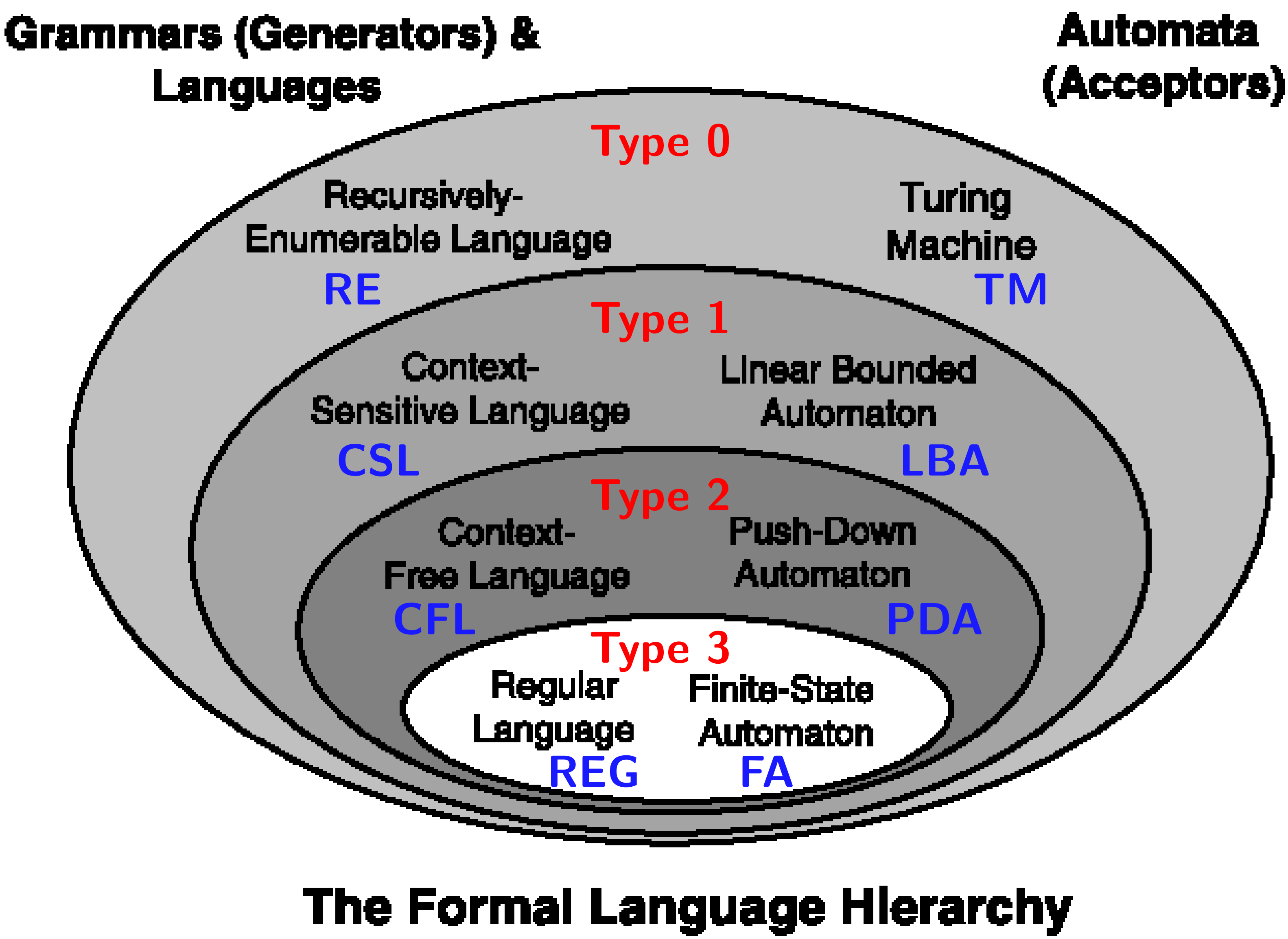
   exponential time $\equiv$ any I/O mapping

Motivation: filling the gap between **integer** and **rational** weights

w.r.t. **Chomsky hierarchy**:

regular (Type 3) $\times$ **recursively enumerable** (Type 0) languages

The Traditional Chomsky Hierarchy



Neural Networks with Increasing Analogicity

from **binary** $(\{0,1\})$ to **analog** $([0,1])$ neurons' states

α**ANN** = a **binary-state** NN with **integer** weights
+ α **extra analog-state** neurons with **rational** weights

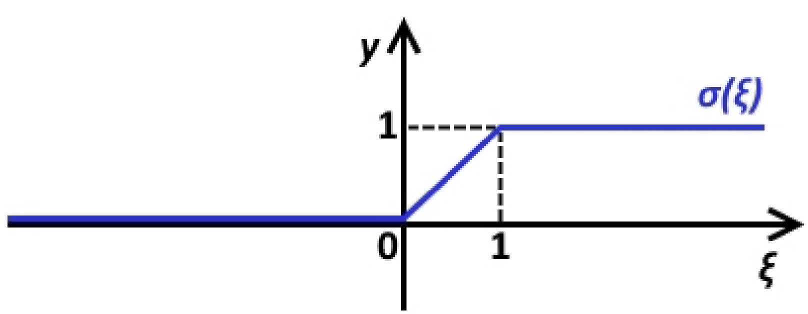
$$y_j^{(t+1)} = \sigma_j\left(\sum_{i=0}^{s} w_{ji} y_i^{(t)}\right) \qquad j = 1,\ldots,s$$

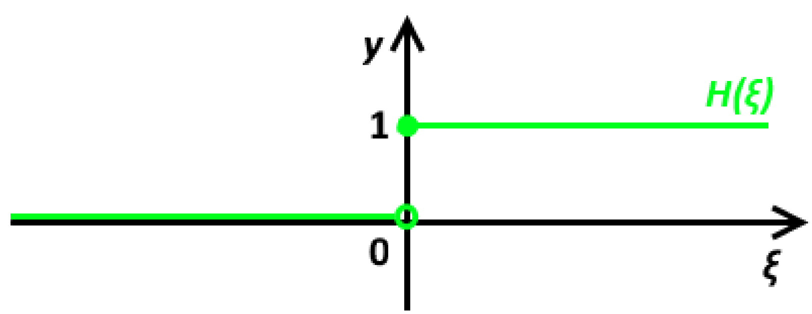
updating the neuron states

$$\sigma_j(\xi) = \begin{cases} \sigma(\xi) = \begin{cases} 1 & \text{for } \xi \geq 1 \\ \xi & \text{for } 0 < \xi < 1 \\ 0 & \text{for } \xi \leq 0 \end{cases} & j = 1,\ldots,\alpha \\[2ex] H(\xi) = \begin{cases} 1 & \text{for } \xi \geq 0 \\ 0 & \text{for } \xi < 0 \end{cases} & j = \alpha+1,\ldots,s \end{cases}$$

saturated-linear function

Heaviside function



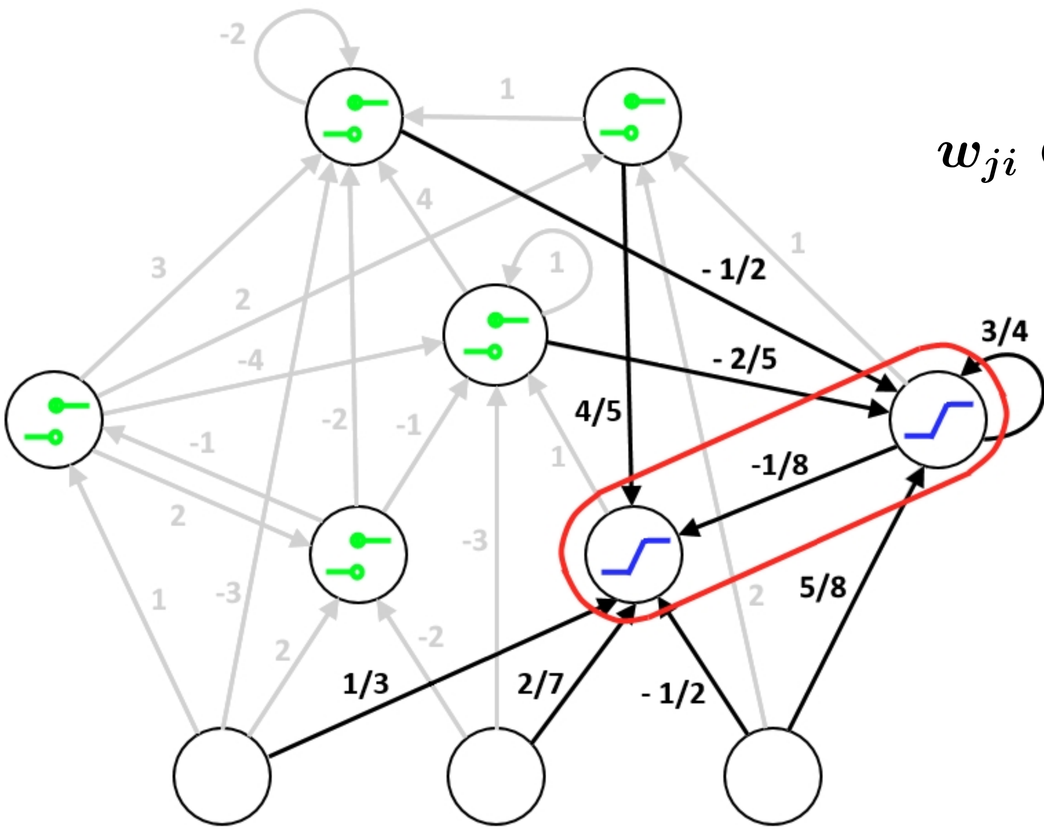


Neural Networks with Increasing Analogicity

equivalently from **integer** to **rational** weights

α**ANN** $=$ a **binary-state** NN with **integer** weights

$+$ α **extra** **analog-state** neurons with **rational** weights



$$w_{ji} \in \begin{cases} \mathbb{Q} & j = 1, \ldots, \alpha \\ \mathbb{Z} & j = \alpha + 1, \ldots, s \end{cases}$$

$$i \in \{0, \ldots, s\}$$

The Analog Neuron Hierarchy

the computational power of αANNs

increases with the number α of extra analog-state neurons:

integer weights $\downarrow$ ... **rational** weights $\downarrow$

$$\textbf{FAs} \equiv \text{0ANNs} \subseteq \text{1ANNs} \subseteq \text{2ANNs} \subseteq \text{3ANNs} \subseteq \ldots \equiv \textbf{TMs}$$

$\uparrow$ **Type 3** — $\times$ **Chomsky hierarchy** **Type 1, 2 ?** — $\uparrow$ **Type 0**

The Separation of 1ANNs: 0ANNs $\subsetneqq$ 1ANNs (Šíma, 2017):

- upper bound: **1ANNs** $\subset$ **LBAs** $\equiv$ **CSLs (Type 1)**
- lower bound: **1ANNs** $\not\subset$ **PDAs** $\equiv$ **CFLs (Type 2)**

$$L_1 = \left\{ x_1 \ldots x_n \in \{0,1\}^* \,\middle|\, \sum_{k=1}^{n} x_{n-k+1} \left(\tfrac{3}{2}\right)^{-k} < 1 \right\} \in \textbf{1ANNs} \setminus \textbf{CFLs}$$

Quasi-Periodic Numbers (Šíma, Savický, 2017):

for a fixed real **base (radix)** β ($|\beta| > 1$) and a finite set $A \neq \emptyset$ of real **digits**, we say that a real number x is **quasi-periodic** if every its **β-expansion**

$$x = (0\,.\,a_1\,a_2\,a_3\ldots)_\beta = \sum_{k=1}^{\infty} a_k\,\beta^{-k} \qquad \text{where} \quad a_k \in A$$

(i.e. non-standard positional numeral system) is **eventually quasi-periodic:**

$$\Big(0\,.\,\underbrace{a_1\ldots a_{m_1}}_{\text{preperiodic part}}\;\underbrace{a_{m_1+1}\ldots a_{m_2}}_{\text{quasi-repetend}}\;\underbrace{a_{m_2+1}\ldots a_{m_3}}_{\text{quasi-repetend}}\;\underbrace{a_{m_3+1}\ldots a_{m_4}}_{\text{quasi-repetend}}\;\ldots\Big)_\beta$$

such that

$$(0\,.\,\overline{a_{m_1+1}\ldots a_{m_2}})_\beta = (0\,.\,\overline{a_{m_2+1}\ldots a_{m_3}})_\beta = (0\,.\,\overline{a_{m_3+1}\ldots a_{m_4}})_\beta = \cdots$$

Example: the plastic $\beta \approx 1.324718$ ($\beta^3 - \beta - 1 = 0$), $A = \{0, 1\}$

$$1 = (0\,.\,0\;\underbrace{100}\;\underbrace{0\,011\,011\,1}\;\underbrace{0\,011\,1}\;\underbrace{100}\ldots)_\beta$$

with quasi-repetends: $(0\,.\,\overline{100})_\beta = (0\,.\,\overline{0(011)^i1})_\beta = \beta$ for every $i \geq 1$

1ANNs with Quasi-Periodic "Weights" (QP-1ANNs):

w_{11} is the **self-loop weight** of the one analog-state neuron ($0 < |w_{11}| < 1$)

$\beta = 1/w_{11}$ is the **base**

$A = \left\{ \sum_{i=0\,;\,i\neq 1}^{s} \frac{w_{1i}}{w_{11}} y_i \,\middle|\, y_2, \ldots, y_s \in \{0,1\} \right\} \cup \{0, \beta\}$ are the **digits**

$X = \left\{ \sum_{i=0\,;\,i\neq 1}^{s} \frac{w_{ji}}{w_{j1}} y_i \,\middle|\, j \neq 1\,,\; w_{j1} \neq 0\,,\; y_2, \ldots, y_s \in \{0,1\} \right\} \cup \{0,1\}$

definition of a **QP-1ANN:** every $x \in X$ is **quasi-periodic**

(e.g. 1ANNs with Pisot β + other weights from $\mathbb{Q}(\beta)$ are QP-1ANNs)

Regular 1ANNs (even with real weights) (Šíma, 2017):

QP-1ANNs $\equiv$ 0ANNs $\equiv$ FAs $\equiv$ REG (Type 3)

Example: **1ANNs** with **rational** weights + the **self-loop weight** $w_{11} = 1/\beta$

where e.g. β is an integer **or** the plastic constant (≈ 1.324718)

or the golden ratio (≈ 1.618034)

The Collapse of the Analog Neuron Hierarchy (Šíma, 2018)

3ANNs $=$ **4ANNs** $=$ **5ANNs** $=$ $\ldots$ $\equiv$ **TMs** $\equiv$ **RE** **(Type 0)**

three analog-state neurons can simulate any **TMs**

The Separation of 2ANNs (Šíma, 2019)

1ANNs $\subsetneqq$ **2ANNs**

the "counting" language $L_\# = \left\{0^n 1^n \,\middle|\, n \geq 1\right\} \in$ **2ANNs** $\setminus$ **1ANNs**

$L_\#$ is a (non-regular) deterministic context-free language **(DCFL)**

accepted by a deterministic push-down automaton **(DPDA)**

- $L_\# \in$ **DCFLs** $\equiv$ **DPDAs** $\subset$ **2ANNs**

  two analog-state neurons can simulate any **DPDA**

- $L_\# \notin$ **1ANNs**

  one analog-state neuron cannot count up to n (even with real weights)

$\longrightarrow$ **DCFLs** $\equiv$ **DPDAs** $\not\subset$ **1ANNs**

The Main Result: The Stronger Separation of 2ANNs

$$(\textbf{DCFLs} \setminus \textbf{REG}) \subseteq (\textbf{2ANNs} \setminus \textbf{1ANNs})$$

or equivalently $(\textbf{DCFLs} \setminus \textbf{REG}) \cap \textbf{1ANNs} = \emptyset$

$$\textbf{1ANNs} \cap \textbf{DCFLs} = \textbf{0ANNs} \equiv \textbf{REG}$$

Theorem. *Any non-regular deterministic context-free language $\boldsymbol{L}$ cannot be recognized by any 1ANN with one extra analog unit having real weights.*

Idea of Proof:

by contradiction: suppose $\mathcal{N} \in$ **1ANNs** recognizes $\boldsymbol{L} \in$ **DCFLs** $\setminus$ **REG**

a construction of a bigger $\mathcal{N}_{\#} \in$ **1ANNs** which exploits $\mathcal{N}$ as its subnetwork (subroutine) for recognizing the counting language $\boldsymbol{L}_{\#}$

which implies $\boldsymbol{L}_{\#} \in$ **1ANNs** – a contradiction

The Simplest Non-Regular Deterministic CFLs

the counting language $L_\# = \{0^n 1^n \mid n \geq 1\}$ can be reduced through a Turing-like reduction to every language in the class **DCFLs** $\setminus$ **REG**:

Theorem. *For every non-regular deterministic context-free language* $L \subset \Sigma^*$ *over a finite alphabet* $\Sigma \neq \emptyset$*, there exist words* $u, w, z \in \Sigma^*$*, nonempty strings* $x, y \in \Sigma^+$*, an integer* $\kappa \geq 0$*, and languages* $L_k \in \{L, \overline{L}\}$ *for* $k \in K = \{-\kappa, \ldots, -1, 0, 1, \ldots, \kappa\}$*, such that for every pair of integers,* $m \geq 0$ *and* $n \geq \kappa$*,*

$$\left(ux^m wy^{n+k}z \in L_k \text{ for all } k \in K\right) \quad \textit{iff} \quad m = n\,.$$

Example: $L \subseteq \{0,1\}^*$ is composed of words that contain more 0s than 1s

$\longrightarrow$ u, w, z empty, $x = 0$, $y = 1$, $\kappa = 1$, $L_{-1} = L$, $L_0 = L_1 = \overline{L}$

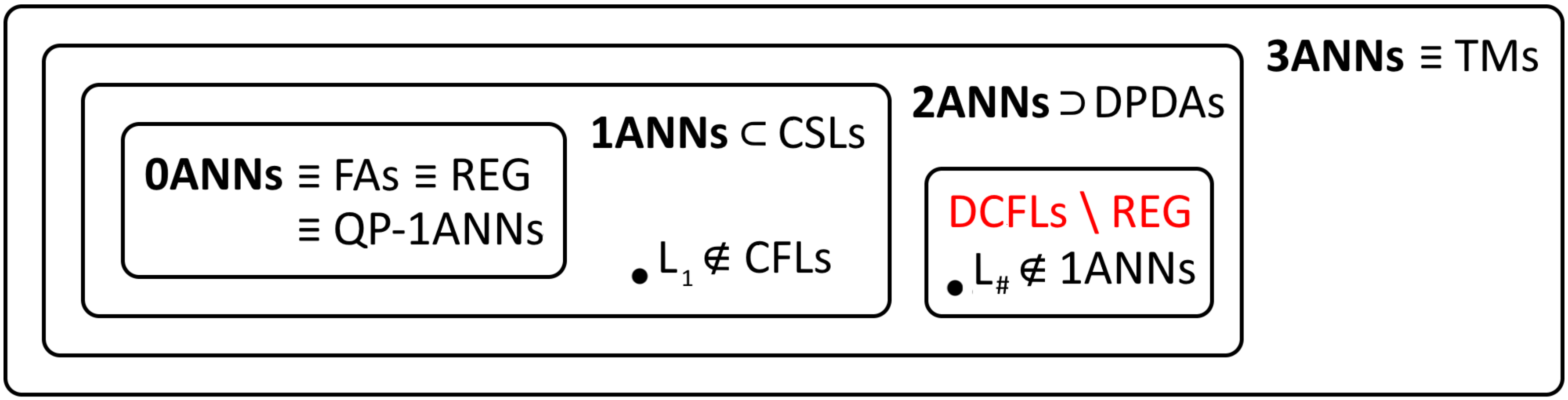
$$\left(0^m 1^{n-1} \in L_{-1} = L \;\&\; 0^m 1^n \in L_0 = \overline{L} \;\&\; 0^m 1^{n+1} \in L_1 = \overline{L}\right)$$

$$\text{iff } (m > n-1 \;\&\; m \leq n \;\&\; m \leq n+1) \text{ iff } m = n\,.$$

contribution to complexity theory: a counterpart to the hardest problem in a complexity class to which every problem is reduced (e.g. NP-completeness)

A Summary of the Analog Neuron Hierarchy

$$\text{FAs} \equiv \text{0ANNs} \subsetneq \text{1ANNs} \subsetneq \text{2ANNs} \subseteq \text{3ANNs} \equiv \text{TMs}$$



Open Problems:

- the separation of the 3rd level: **2ANNs** $\subsetneq$ **3ANNs** **?**
- strengthening the 2nd level separation to the **nondeterministic CFLs**:

$$(\textbf{CFLs} \setminus \textbf{REG}) \cap \textbf{1ANNs} = \emptyset \ \textbf{?}$$

- a proper "natural" hierarchy of NNs between integer and rational weights which can be mapped to known infinite hierarchies of REG/CFLs **?**