A Low-Energy Implementation of Finite Automata by Optimal-Size Neural Nets

Jiří Šíma

Institute of Computer Science
Academy of Sciences of the Czech Republic

Energy Aspects of Neural Networks

- the activity of neurons in the brain is quite sparse, with only about 1% of neurons firing (Lennie, 2003)
- biological neurons require more energy to transmit a spike than not to fire
- in contrast, the design of artificial neural circuits does not usually take the energy aspects into account, e.g. on average, every second unit fires during a computation

Energy Complexity of Threshold Circuits

Uchizawa, Douglas, Maass (2006):
energy complexity of **feedforward** perceptron networks (threshold circuits)
= the maximum number of firing units taken over all the input values to the circuit

- related by tradeoff results to other complexity measures:
  - network size = the number of neurons
    (Uchizawa, Takimoto, 2008; Uchizawa, Takimoto, Nishizeki, 2011)
  - circuit depth = parallel computation time
    (Uchizawa, Nishizeki, Takimoto, 2010; Uchizawa, Takimoto, 2008)
  - fan-in = the maximum number of inputs to a single unit
    (Suzuki, Uchizawa, Zhou, 2011)

- a tool for proving the lower bounds in circuit complexity
  (Uchizawa, Takimoto, Nishizeki, 2011)

? Energy Complexity of Recurrent Networks ?

Model of Recurrent (Perceptron) Networks

- Architecture: s computational units (neurons, perceptrons, threshold gates) $V = \{1, \ldots, s\}$ connected into a directed graph

  where s is the size of the network

- each edge from neuron i to j is labeled with an integer weight $w(i, j)$

  ($w(i, j) = 0$ iff there is no edge (i, j))

- Computational Dynamics: the evolution of network state

$$\mathbf{y}^{(t)} = (y_1^{(t)}, \ldots, y_s^{(t)}) \in \{0, 1\}^s$$

  at discrete time instant $t = 0, 1, 2, \ldots$

Computational Dynamics

1. initial state $\mathbf{y}^{(0)}$
2. at discrete time instant $t \geq 0$ the excitation

   $$\xi_j^{(t)} = \sum_{i=1}^{s} w(i,j) y_j^{(t)} - h(j) \quad \text{for } j = 1, \ldots, s$$

   where $h(j)$ is an integer threshold of unit j
3. at the next time instant $t+1$, the neurons $j \in \alpha_{t+1}$ from a selected subset $\alpha_{t+1} \subseteq V$ updates their states (outputs)

   $$y_j^{(t+1)} = \begin{cases} 1 & \text{for } \xi_j^{(t)} \geq 0 \\ 0 & \text{for } \xi_j^{(t)} < 0 \end{cases}$$

   while $y_j^{(t+1)} = y_j^{(t)}$ for $j \notin \alpha_{t+1}$

Energy Complexity $=$ the maximum number of firing neurons $\sum_{j=1}^{s} y_j^{(t)}$ at any time instant $t \geq 0$, taken over all possible computation

Recurrent Neural Networks as Language Acceptors

(Horne, Hush, 1996; Indyk, 1995; Siegelmann, Sontag, 1995 etc.)

- language (problem) $L \subseteq \{0,1\}^*$ over binary alphabet

- input string $\mathbf{x} = x_1 \ldots x_n \in \{0,1\}^n$ of arbitrary length $n \geq 0$ is sequentially presented bit after bit via input neuron in $\in V$,

$$y_{\text{in}}^{(\tau(i-1))} = x_i \qquad \text{for } i = 1, \ldots, n$$

  where integer $\tau \geq 1$ is a time overhead (period) for processing a single bit

- output neuron out $\in V$ signals whether $x \stackrel{?}{\in} L$,

$$y_{\text{out}}^{(\tau n)} = \begin{cases} 1 & \text{for } x \in L \\ 0 & \text{for } x \notin L \end{cases}$$

Computational Power of Recurrent Perceptron Networks

- recurrent networks having at most 2^s different network states from $\{0,1\}^s$ correspond to finite automata recognizing regular languages

- a deterministic finite automaton A with m states can simply be implemented using $2m+1$ neurons, one for each 0 or 1 state transition of A (Minsky, 1967)

  this naive $O(m)$ implementation requires only a constant energy

- optimal-size implementations of a deterministic finite automaton with m states by neural nets with $\Theta(\sqrt{m})$ neurons (Horne, Hush, 1996; Indyk, 1995).

? Energy complexity of an optimal-size neural network simulating a given deterministic finite automaton ?

Main Result: Tradeoff Between Energy and Time Overhead

Theorem 1 *For any function e satisfying $e = \Omega(\log s)$ and $e = O(s)$, a given deterministic finite automaton A with m states can be simulated by a neural network N of optimal size $s = \Theta(\sqrt{m})$ neurons with time overhead $\tau = O(s/e)$ per one input bit, using the energy $O(e)$.*

Simple Idea of Proof:

- each of the m states of A can be encoded using $p = \lceil \log \rceil + 1$ bits including a one-bit indicator of final states
- transition function $\delta : Q \times \{0, 1\} \longrightarrow Q$ of A (producing the new state from the old one and the current input bit) can be viewed as a vector Boolean function $\mathbf{f} : \{0, 1\}^{p+1} \longrightarrow \{0, 1\}^p$

Idea of Proof (Continuation)

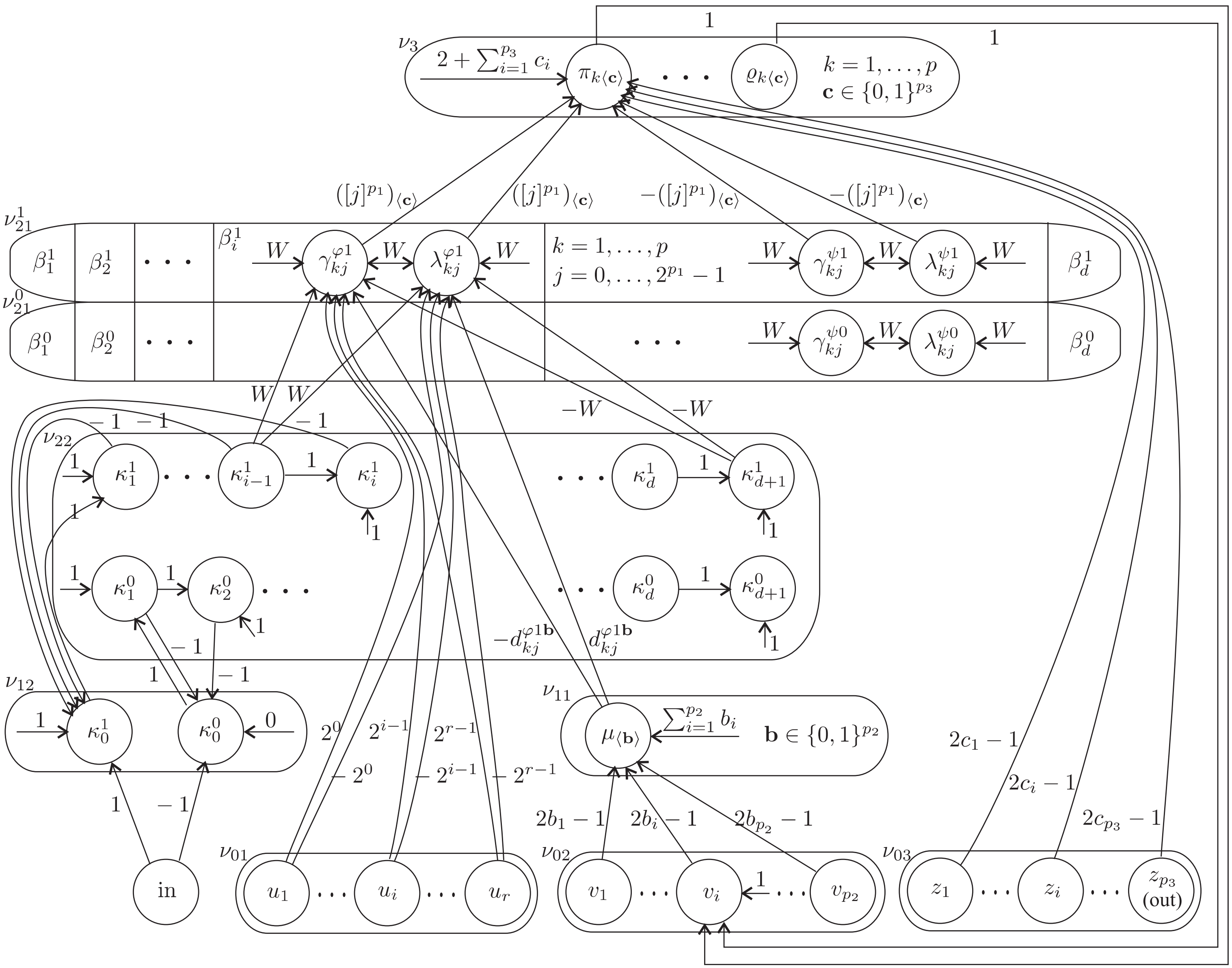
- function $\mathbf{f}$ is implemented by four-layer perceptron network of optimal size $s = O(\sqrt{2^p}) = O(\sqrt{m})$ using the method of threshold circuit synthesis due to Lupanov (1973)

- the recurrent connections leading from the fourth to the first layer replace the current state by the new one

- the dominant-size layer of $\Theta(s)$ neurons is properly partitioned into $O(s/e)$ blocks of $O(e)$ units each

- control units ensure that these blocks are updated successively one by one so that the energy consumption $O(e)$ is guaranteed while the time overhead for processing a single input bit is $O(s/e)$

Technical Schema of Low-Energy Neural Automata



The Lower Bounds

Theorem 2 *Let* $\tau \log \tau = o(\log s)$. *There exists a neural network of size* s *neurons simulating a finite automaton with time overhead* τ *per one input bit which needs energy* e *such that* $\log e = \Omega_\infty\left(\frac{1}{\tau}\log s\right)$.

Idea of Proof: the technique due to `Uchizawa, Takimoto (2008)` based on communication complexity argument

Corollary 1

1. If $\tau = O(1)$, *then* $e \geq s^\delta$ *for some* δ *such that* $0 < \delta < 1$ *and for infinitely many* s. $\times$ *our construction* $e = O(s)$

2. If $\tau = O(\log \log s)$, *then* $e = \Omega_\infty\left(s^{\frac{1}{\log^\delta s}}\right) = \Omega_\infty\left(2^{\log^{1-\delta} s}\right)$ *for any* δ *such that* $0 < \delta < 1$.

3. If $\tau = O(\log^\alpha s)$ *for some* $0 < \alpha < 1$, *then* $e = \Omega_\infty\left(s^{\frac{\log \log s}{\log^\delta s}}\right) = \Omega_\infty\left((\log s)^{\log^{1-\delta} s}\right)$ *for any* δ *such that* $\delta > \alpha$. $\times$ *our construction* $e = O\left(\frac{s}{\log^\alpha s}\right)$

Conclusions and Open Problem

- We have, for the first time, applied the energy complexity measure to recurrent neural nets which has recently been introduced and studied for feedforward perceptron networks.

- We have presented a low-energy implementation of finite automata by optimal-size neural nets with the tradeoff between the time overhead for processing one input bit and the energy varying from the logarithm to the full network size.

- We have also achieved lower bounds for the energy consumption of neural finite automata which are valid for at most sublogarithmic time overheads and are still not tight.

- An open problem remains for further research whether these bounds can be improved.