Beyond PDP: the Frequency Modulation Neural Network Architecture

Hideto Tomabechi*and Hiroaki Kitano†
Center for Machine Translation
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213 U.S.A.
tomabech@cs.cmu.edu, hiroaki@cs.cmu.edu

Abstract

This paper proposes the Frequency Modulation Neural Network as an alternative to current neural-net models. This proposal is for an architecture of a heterogeneous neural-network in which information is propagated using frequency modulation of pulses oscillated by groups of neurons. The FMNN model enables operations including variable-binding, sequential recognitions and predictions. The use of FM signals for communication among neural clusters also enables the model to avoid communication bottlenecks arising in most massively parallel computer architectures.

1 Introduction

One weakness of the traditional neural network architecture is commonly known as the *binding problem*. Another weakness are difficulties in predicting and recognizing sequential patterns such as word sequences. Due to these problems in the traditional neural network models, application of the model to any serious natural language processing has not been conducted. In the Frequency Modulation Neural Network (FMNN)[1], one solution to the *binding problem* is attained through propagation of the activation source information through the frequency modulation of pulses generated by certain groups of neurons. Combined with local modular circuits which enable sequential prediction and activation, our model attains structured marker passing from the neural level. First, we briefly summarize the marker-passing and constraint propagation schemes which we assume as our basis of natural language and inferential processing and describe what needs to be attained by the neural network. Then we describe some features of the FMNN architecture relevant to such tasks.

*Also with the Laboratory for Computational Linguistics, Carnegie Mellon University.

†Also with NEC Corporation.

[1]Also see [Kitano and Tomabechi, ms.] for mathematical details and the descriptions of some of the circuits not included in this paper.

2 Marker-Passing and Constraint Propagation

2.1 Marker-Passing Models

The spreading-activation marker-passing based models of cognitive processing reflect the highly interactive and parallel nature of human cognitive activities[2]. One of the major differences between the marker-passing based models and the neural network models is that the marker-passing scheme allows some tokens (typically the sources of activation) to be passed with the activation propagations. The Direct Memory Access (DMA) model [Riesbeck and Martin, 1985] [Tomabechi, 1987] is a marker-passing scheme with case-based inference capability. In the DMA model, the spreading-activation of activation source tokens enables the models to instantiate generalized memory structures to capture the specific meaning of the input utterance. The tokenized activation source propagation can be seen as one way of solving the binding problem. Recently, researchers at Carnegie Mellon have adopted the DMA paradigm for various natural language and inference tasks. [Tomabechi and Tomita, 1988a] is a DMA natural language interface and inferences for development of knowledge-based systems and in [Tomabechi and Tomita, 1988b], DMA based contextual inferencing was integrated into a unification-based (LFG, [Kaplan and Bresnan, 1982]) phoneme-parser ([Saito and Tomita, 1988]) as a part of real-time speaker-independent speech-to-speech translation system ([Tomabechi et. al., 1989a]). Independently, a basic DMA algorithm has been implemented on VLSI chips [Kitano, 1988]. Also, the first DMA based speech-to-speech translation has been publicly demonstrated at Carnegie Mellon[3] ([Tomabechi et. al., 1989b], [Kitano et. al., 1989b]).

2.2 Constraint Propagation

The Massively-parallel Constraint Propagation (MCP, [Tomabechi and Tomita, ms.]) scheme extends the DMA philosophy and proposes the models to propagate constraints as well as other information such as the tokens for the sources

[2]Such as [Quillian, 1969], [Charniak, 1983], [Hirst, 1984], [Riesbeck and Martin, 1985], [Charniak, 1986], [Norvig, 1987], [Tomabechi, 1987], and [Hendler, 1988].

[3]Other efforts include the application of DMA to phoneme-based parsing [Tomabechi et. al., 1988], discourse-based processing [Kitano et. al., ms.], cost-based ambiguity resolution [Kitano et. al., 1989a] and concurrent parsing and generation [Kitano, 1989].

of activations. The Head-driven Massively-parallel Constraint Propagation (HMCP) [Tomabechi and Levin, ms.] proposes a scheme of propagating head-feature constraints to increase DMA's capacity for handling syntactic phenomena such as word order, agreement, case marking, inflectional and expletive morphology, control, and unbounded dependencies. In the MCP paradigm, words (or some smaller linguistic units) in the input string trigger the propagation of constraint information through the network. The propagated information includes the source of the activation, syntactic head-features, etc.. Concepts that represent heads of phrases contain bundles of syntactic features which constrain their complements. When concept activations of complements collide with activations of heads, the syntactic features of the complement and head are unified with each other.

Below are the node entries taken from [Tomabechi and Levin, ms.] representing[4] the lexical concepts for the verbs *give* and *try*:

```
(lex-node *GIVE
 (is-a (*ACTION))
 (phonology </g/ /i/ /v/>)
 (syn-head-feature ((MAJ V) (VFORM FIN) (AUX
MINUS)))
 (giver (*PERSON 1))
 (receiver (*PERSON 2))
 (given (*OBJECT 3))
 (subcat <(NP[NOM] 1), (NP[ACC] 2), (NP[ACC]
3)>))

(lex-node *TRY
 (is-a (*ACTION))
 (phonology </t/ /r/ /a/ /i/>)
 (syn-head-feature ((MAJ V) (VFORM FIN) (AUX
MINUS)))
 (trier (*PERSON 1))
 (circumstance (*ACTION 2))
 (subcat <(NP[NOM] 1), ((((MAJ V)
                          (VFORM INF))
                         subcat<(NP 1)>)
                        2)>))
```

The list (NP[NOM] 1) in the subcat feature is a short-hand for (((MAJ N) (CASE NOM)) 1).

The list in *subcat* represents the postulation for the constraints that are to be satisfied by the nodes that will fill the conceptual roles (that are co-numbered). The contents of the *syn-head-feature* are the head-features that are propagated with the specific node activations along with other information. In the entry for *try* we can see that *subcat* specifies that it subcategorizes for an unsaturated verb phrase as an instance to fill the *circumstance* role (whose abstraction is the concept *ACTION). An input string such as *He* can trigger the propagation of information such as NP[NOM] as well as the semantic propagation through the abstraction and other links. The HMCP models demonstrate that constraint propagation schemes can handle sentential constructions such as control (as in *He tried to give Mary the book*) assigning the correct interpretations and attaining the correct grammaticality judgement.

[4]The lexical entries are originally taken from the representation under HPSG (Head-driven Phrase Structure Grammar, [Pollard and Sag, 1987]). HMCP attains the HPSG unification based analysis under the DMA framework.

3 MSP vs Conventional Architecture

Although the DMA models and the MCP (HMCP) models are appealing schemes for natural language processing and inference, because these models require propagation of information (i.e., sources of activation and constraint information), it has been difficult to support them in the conventional neural-net architecture. In this paper, we refer to the structured marker-passing algorithms that the DMA and MCP models assume in their implementation as 'Massively-parallel Structured-marker Passing' (MSP) algorithms. Because the MSP algorithms assume a massively-parallel machine architecture, the implementations (such as on MULTILISP, [Halstead, 1985]) have been generally slow. The conventional massively-parallel machines (such as *Connection Machine*, [Hillis, 1985]) are not desirable for MSP algorithms because they are incapable of propagating information without large overhead. In other words, the massively-parallel machine architecture, which assumes the propagation of small amounts of scalar values is not suitable for MSP algorithms.

4 Frequency Modulation Neural Network

We would like to introduce a model of neural network and a possible hardware architecture which directly support the MSP paradigm. We call this model the Frequency Modulation Neural Network (FMNN) and the major features which distinguish FMNN from other neural network are as follows:

- Simulates spikes instead of activation levels.
- Assumes neural clusters with certain circuit topologies: Heterogeneous Neural Network.
- Patterns of spikes, modulated as FM signals, have significance in propagating information.

4.1 Simulating Spikes

Unlike most neural networks which simulate the activation level of each node, the FMNN simulates the spikes which each neuron emits. Spikes are electric pulses emitted from neurons to propagate through axons. We are proposing to simulate the neural spikes because (1) we are interested in simulating a detailed biological process, and (2) patterns of pulses can carry information impossible to transmit in past neural network architecture. Biologically, we are not able to define the precise roles of spikes in the human brain with regard to cognitive activities. It is also probably true that mere simulation of spikes will not contribute much to our understanding of the physiological nature of the brain, because other factors, such as chemical substances, neurogenesis, and plasticity of the brain, are also involved. However, we expect that spike simulation may lead to the discovery of new computational features of the brain[5]. Apart from biological observations, the use of spikes in the neural network makes the flow of structured information attainable.

[5]In fact, [Loeb, 1985] has shown that differences between arrival times down to 10 microseconds can be detected by neurons leading to [Sejonwski, 1986]'s analysis that submillisecond timing information could be important in the cerebral cortex as well.

4.2 Heterogeneous Neural Network

The second assumption of the FMNN is the heterogeneous structure. We assume the existence of several types of groups of neurons and certain local and global structures in the network. Most neural networks assume homogeneity. Although these models employ certain global structures, such as layers composed of input, hidden, and output units [Rumelhart et. al., 1986] or a recurrent structure [Elman, 1988], each node is assumed to have homogeneous characteristics. In addition, the existence of a local structure is not assumed. Motivations for assuming heterogeneous structure are based on both biological and technical considerations. Biological observations on neurogenesis [Murphy et. al., 1983] [Nottebohm, 1989], the existence of local structure of neurons [Rosen et. al., 1989], and the existence of the specific global innate configurations for the specific tasks such as the Papez circuit involving the hippocampus [Papez, 1937] indicate that neural circuits in our brain have certain innate structures already encoded in the DNA. In addition, strong evidence of hypercolumn in the A17 field of visual cortex, models of the hippocampus [Zipser, 1985] and studies on the topological circuits of neurons [von der Malsburg, 1985] lead us to assume a structured functional module which we call *hypermodule*. We claim such local circuits provide technical benefits because they will allow us to predefine highly functional neural circuits without abandoning the basic features of the neural network.

4.3 Frequency Modulation

We use Frequency Modulation (FM) signals as one of the basic communication mechanisms in our model. When we look into the physiological data of signal transmission between neurons, frequency of pulses varies depending on the strength of the input given to the neuron. [Treutlein and Schulten, 1985] analyzed the Fokker-Plank equation corresponding to the stochastic Bonhoeffer-Van der Pol (BvP) model, and concluded that the noise level can be employed to tune firing frequency of Hodgkin-Huxley type neurons. In their analysis, the noisy BvP model tends to limit cycle in which the mean period of pulses is depending upon the noise level. This means that **the more energy** the cell gets, **the higher the frequency** of impulses it emits[6]. This property has been known since [Adrian, 1946]; however, it has been neglected by the past neural network models. Moreover, there exist neurons whose activation levels vary dependent on the regular brain wave (θ-Cells associated with the hippocampal θ-wave.). Also, the importance of arrival time of impulses and relative timing of impulses for processing in the cortex has been noted (such as [Sejonwski, 1986] and [Sachs et. al., 1983]); however, no neural network model so far has captured this phenomenon[7]. This characteristic is well simulated when we consider FM as its approximation. FM is a kind of an angle modulation technique originally developed for telecommunication of signals [Marubayashi, 1981]. A signal is encoded into a carrier wave ($Acos(\omega_c t + \theta_c)$) by modulating its phase angle ($\theta = \omega_c t + \theta_c$). In the FM, the phase is modulated according to:

$$\theta_c = k \int f(t)dt$$

$\frac{d\theta}{dt}$ is called as 'instantaneous frequency' in which it follows:

$$\frac{d\theta}{dt} = \omega_c + kf(t)$$

Since we are going to deal with impulses instead of analog waves, instantaneous frequency is a probability density of the impulse. Apart from physiological aspects, the use of FM signals as representation and communication scheme of a massively parallel computer would provide significant advantages over traditional computer architecture.

5 Modular Neural Circuits in the FMNN

In this section, we describe some of the circuit topologies of modular neural circuits and their behavior. We will be discussing the circuits which are of particular relevance from the viewpoint of inferencing and natural language processing. Relevant work has been done extensively by [Amari, 1978]. A basis of our analysis also assumes [Amari, 1978], although we made some reformulation in order to simulate spikes[8].

5.1 Simple Random Circuit

A simple random circuit is a neural circuit in which each neuron is connected to the others at random [Amari, 1978]. The external behavior of this circuit is almost equivalent to the McCulloch-Pitt model assumed in most neural network theories [McCulloch and Pitts, 1943]. This circuit is also the simplest form of a bi-stable circuit. A simple random circuit stays at the low level equilibrium until input exceeding the threshold comes in. In this case, the circuit will transit to the high level equilibrium. As the signal goes off, the system will go back to the low level equilibrium. The other way to attain a simple bistable circuit is by connecting two excitory neural groups and one inhibitory group. An implication of such a circuit is that this neural complex can act as a node which is commonly assumed in connectionist (such as [Waltz and Pollack, 1984] and [Bookman, 1987]) literature. The circuit is a threshold device and it holds its activation for a certain period of time. A sequence of pulses emitted from such a circuit is random and has no significant meaning as a carrier of information.

5.2 Induced Excitory Oscillator

Induced Excitory Oscillators (IEOs) are types of circuits that generate pulses of a certain frequency whenever a certain external stimulus is given. This is a kind of bi-stable circuit in which one of the attractors is a periodic attractor[9]. An oscillator unit of neurons generates a group of pulses of a certain frequency. Each unit has its own frequency called its 'Characteristic Frequency' or 'Eigen Frequency'. Small

[6] [Sejonwski, 1981] introduces an idea of skeleton filter – the temporary network of neurons near threshold that can linearly transform correlations between spikes on input lines.

[7] One of the few exceptions may be [von der Malsburg and Schneider, 1986].

[8] The behavior of the neural circuits in this paper has been analyzed in part by using a neural spike simulator.

[9] Biologically, circuits with coherent oscillatory capability are interesting with respect to their relation to Stimulus-Evoked (SE) -resonance of 35-85Hz found in the primary visual cortex called *Coherent SE-resonance* [Eckhorn et. al., 1988].

circuits consist of a few neurons and with a large number of such circuits with a phased array configuration, output spikes form a simple sequence of pulses with a specific interval. Moreover, a circuit with a group of neurons that outputs a sequence of pulses with a specific modulation is also possible. Such circuits tend to limit a cycle. A spontaneous frequency or a probability density of the pulses(P_i) follows:

$$P_{t+1} = F(P_i) \quad t = 1, 2, \cdots, p$$
$$P_{p+1} = P_1$$

In the modular neural circuits, we only require IEOs to satisfy a weak form of the *Poincare-Bendixson-Theorem* [Beltrami, 1987]. The theorem indicates that for the system $\dot{x} = f(x), x$ in R^2, an orbit γ tends to a limit cycle Γ if $\gamma^+ = \Gamma$ and Γ is a non-constant closed orbit. γ^+ is the positive limiting set which is a set of points p in R^2 for which

$$||x(t_n) - p|| \rightarrow 0$$

Our neural circuit is stable if there is a cycle γ which is asymptotically stable. This would require the existence of an open neighborhood Ω_0 such that every orbit starting from Ω_0 has γ as its positive limiting set. This is known as a weak form of the *Poincare-Bendixson-Theorem*. The circuit that entails limit cycles that satisfy this theorem is useful since it continuously emits characteristic frequency, until the energy is dispersed or inhibition is imposed, which can identify the source of activation.

5.3 Resonance Circuit

Resonance circuits react only to a sequence of pulses whose frequency matches the characteristic frequency of the circuit. Such circuits can be built by combining a series of nodes forming a loop and some inhibitory connections to the input node. When a pulse comes in, the input node gets activated and emits an impulse to the next node in the loop. That node emits an impulse to the next node in turn. At the same time, the node emits an impulse to the inhibitory node in which the inhibitory link is connected to the input node. Thus, the input node can be activated until the inhibition is removed. By adjusting the size of loops and inhibitions, the characteristic frequency of the circuit can be adjusted.

5.4 Sequential Prediction and Activation Circuit

The capability to handle natural language requires the system to recognize sequential patterns. This type of nodes is assumed in the DMA type connectionist (or pseudo-connectionist) architecture (such as [Riesbeck and Martin, 1985] and [Tomabechi, 1987]), in which time-sequenced activation of nodes is assumed to model subcategorization of syntactic units, episodic knowledge invocations, etc. In our model, sequential pattern recognition and predictions are attained by assuming a series of interconnected multi-stable modular circuits. As one example of such circuits, we can assume a complex of modular nodes each having inactive, prediction and recognition states. When input signals are less than a certain level (L_2), the circuit stays inactive. Once the input strength exceeds that threshold, the excitation level goes up to the pre-excitory level. However, if the input strength is below the excitation threshold (L_3), the excitation level stays at prediction level for a while and goes back to an inactive state quickly. The circuit jumps up to a recognition state when the input signals are more than L_3. A recognition state is a stable state, but returns to an inactive state as a result of dispersion of the energy. Inter-module connections are created so as to transfer sufficient impulses to the next module so that the next module can be activated to the prediction state whenever the current module gets to the recognition state. Such network configuration has not been assumed in the traditional neural network. However, we assume the existence of such local circuit topologies in the neural network. The excitation level of the l-th multi-stable node can be described by:

$$E_l = \alpha_l(S_l + \sum_{i=1}^{2} \omega_{l-1,i}^{l,1} F(U_i^{l-1}) - h)$$

where E_l, S_l, U_i^{l-1} denotes the excitation level of the l-th node, external stimulus to the l-th node, and excitation level of the i-th element of the l-1-th node. $\alpha_l, \omega_{l-1,i}^{l,1}, h$ denotes a coefficient, weights, and threshold, respectively. The activation level of the node that packages the series of sequential nodes is:

$$E_{CSC} = \sum_{l=0}^{n} \alpha_l(S_l^{all} - h_l) - h_{CSC}$$

One interesting feature of this circuit is that it can handle the transposition of activation sequences by utilizing the cost of excitation propagation (instead of not recognizing the incorrectly ordered sequence at all). To be more specific, a correct input activation sequence will excite E_{CSC} node at a normal level, say 100. However, an incorrect ordering can exist at a lower level, say 60. This is because an incorrectly ordered sequence will consume more energy than the correctly ordered one due to the pre-excitation mechanism which carries the next module to the prediction state. One of the reasons that we use groups of neurons instead of asymmetrically connected single neurons is thay they can be combined with resonance circuits to handle context-free rules with some syntactic constraints which are implicitly encoded in form of the modulation frequency. For example, we can implicitly encode the obliqueness order of HPSG [Pollard and Sag, 1987].

6 Knowledge Representation in FMNN

A complex knowledge structure can be built by modulating carrier waves by a modulation signal. Suppose that a certain group of neurons is representing the cluster for some instance of an already known person such as JOHN, and this cluster of neurons has the eigen-frequency of 200 Hz. Now suppose some activations have triggered the syntactic realization that an input noun phrase that activated the clusters for JOHN concurrently activated another group of neurons that is representing the feature *case nominative* whose characteristic frequency was 5000 Hz. While propagation for the activations of these clusters are performed, the carrier frequency representing *case nominative* can be modulated by a signal representing JOHN. Such a scheme for modulation is nothing new to the current communication technology as has already been well accepted and performed by FM radios, TVs and other frequency modulated methods of communication. Intuitively, our concept of variable binding is similar to FM broadcasting in which carrier frequency of 80MHz

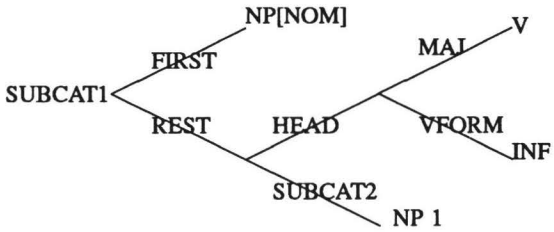

Figure 1: Example of a Simple Data Structure

(FM-Tokyo) carries the music of the Boston Symphony Orchestra (BSO) as a modulation signal (value to be bound). In this case, the music of BSO is bound to FM-Tokyo. A complex structure as seen in Figure-1 (which partially captures the *subcat* feature for *give* as seen previously) can be built by modulating signals.

We assume that one group of neurons is oscillating at a certain frequency. A series of pulses are emitted toward a group of neurons which eventually emit frequency modulated pulses. A modulation signal is given from the group of neurons which oscillate in the frequency representing the value to be bound to the variable. The resulting pulse is modulated based on the frequency of the modulation signal. In case of figure-1, the data structure would be represented in the following equations:

$$
\begin{aligned}
A &= A\cos(\omega_{subcat1} t + \theta_{subcat1}) \\
\theta_{subcat1} &= \int f_{first}(t)dt + \int f_{rest}(t)dt \\
f_{first}(t) &= A_{first}\cos(\omega_{first} t + \int f_{NP[NOM]}(t)dt) \\
f_{rest}(t) &= A_{rest}\cos(\omega_{rest} t + \int f_{HEAD}(t)dt \\
&\quad + \int f_{SUBCAT2}(t)dt) \\
f_{HEAD}(t) &= A_{HEAD}\cos(\omega_{HEAD} t + \int f_{MAJ}(t)dt \\
&\quad + \int f_{VFORM}(t)dt) \\
f_{SUBCAT2}(t) &= A_{SUBCAT2}\cos(\omega_{SUBCAT2} t + \int f_{NP1}(t)dt) \\
f_{MAJ}(t) &= A_{MAJ}\cos(\omega_{MAJ} t + \int f_{V}(t)dt) \\
f_{VFORM}(t) &= A_{VFORM}\cos(\omega_{VFORM} t + \int f_{INF}(t)dt)
\end{aligned}
$$

7 FMNN Machine

Building an FMNN Machine should be the best way to take advantage of this architecture. It is analogous to the motivation for building the *Connection Machine* ([Hillis, 1985]) for implementing the connectionist network. In this section, we discuss some of the possible implementation strategies and outline relevant technologies.

7.1 Processors

A straightforward way of implementing the FMNN model is to build a VLSI neural network chip with heterogeneous connections and spike generation capabilities as described in this paper. One other approach is to build components that are equivalent to each local neural circuit in their functions. Such components would include energy-spike converters, oscillators, resonators, etc.. Several technologies are available to actually implement such functional modules. Analog circuit technology is a well-established technology that can attain functionalities of intended modules. Especially, recent studies on analog VLSI [Ryckebusch et. al., 1989] may provide a hardware basis from the VLSI level. Digital Signal Processing (DSP) technology is one other possibility which can be more compatible with current computer architecture. One advantage of using DSP technology is that the filtering and control of signals can be more precise than with analog technology. Also, it is be possible to implement FMNN using current computer technology. However, it would be more likely to be a simulation of FMNN rather than a direct hardware implementation. Figure-2 is a summary of differences of the FMNN model and other models of computation.

7.2 Communication

In the FMNN Machine, communication between processor units is performed by sending sequences of pulses. The conventional approach of communicating pulses between processors is using bus, N-cube, and a network which has problems such as load bottlenecks, routing, packet collision, etc.. A new approach which we are proposing is to send them using carrier waves. The pulses are carried on a specific carrier frequency possessed by each generator. Resonators in each node can tune to the desired carrier wave in order to establish appropriate connections. This communication method enables us to send many independent pulse sequences at a time; thus enabling us to simulate a very large-scale neural-network (VLNN) in which a massive amount of information is transmitted throughout the network. As briefly described earler, an FM radio is a good analogy to our approach, where the signal is sent by a certain carrier wave and an audio signal is modulated by frequency. To send a huge amount of impulses, communication media need to have a large capacity to communicate on many independent channels. Recent advancements in communication technologies including Optical Wavelength Division Multiplexing or Optical Frequency Division Multiplexing may provide such capabilities.

8 Discussion: Beyond PDP

8.1 Bridging the Gap between Neural Network and Cognitive Processing

The fundamental premise of connectionism has been that individual neurons *do not transmit large amounts of symbolic information* ([Feldman and Ballard, 1982]). We are extending this claim to add a phrase that, *however, a group of neurons is capable of oscillating pulses to carry information.* As a basis of this extension, we are assuming the existence of local functional circuits which we refer to as *hypermodules.* The impact of this addition leads to the significant enhancement of the neural network because; (1) we now have a

	Von Neumann	Conventional Neural-Net	FMNN
Representation	Bit pattern	Digital Activation	FM pulses
Data Structure	Pointer	Association of Nodes	Modulation and Superposition of pulses
Computation	Bit operation	Spreading Activation	Resonance and Spreading of Signals
Processor	CPU (One or few)	CPU (Massive, Homogeneous)	Non-linear Device (Massive, Heterogeneous)

Figure 2: Comparison of Traditional Architecture with FMNN

heterogeneous network with highly functional local circuits of neurons that are oscillating, receiving and modulating signals, (2) the FM signals propagated in the network can contain information such as which oscillating node initiated the propagation and what constraints to inhibit/delay/impose and slow/accelerate certain propagation. Especially the capacity to propagate information results in the capacity to perform operations such as variable-binding (by specifying where the oscillation was initiated) and structure-building (modulating the oscillations on top of receiving frequencies and utilizing the variable-binding capacity). This is significant for cognitive research which requires propagation of information and constraints. Indeed the network is capable of supporting the *massively-parallel structured-marker passing* (MSP) from the hardware level. As discussed in detail in [Tomabechi and Tomita, ms.], the MSP network is a model that is capable of performing inferencing and natural language processing that is not currently possible with the conventional connectionist model due to its lack of capacity to propagate information (which is done in MSP as a passing of *structured-markers*). Also as reported in [Kitano and Tomabechi, ms.], we can build circuits of neurons within the FMNN architecture to capture nodes that respond to certain *features* of sensory input such as sound-wave length and light-wave length[10]. Representation of knowledge in our model is very different from other existing models. Especially, our model does not assume a static representation of knowledge as it has been the case in other models. We are assuming constant *information flow* in the brain (as opposed to the specific portion of the network statically representing some knowledge) and the knowledge is implicitly represented as time-varying signals.

8.2 Physiological Relevance

From the physiological point of view, recent studies of neurophysiology including the one by Eckhorn's group [Eckhorn et. al., 1988] provide some biological relevance to our model. They report stimulus dependencies of oscillatory responses[11]. Moreover, coherence of SE-resonance was found within a vertical cortex column, between neighbouring hypercolumns, and between two different cortical areas. They claim that "SE-resonance can be phase-locked within half an oscillation cycle up to a distance of 10mm at 50Hz." and conclude that coherent SE-resonances are likely to constitute a second higher stage of sensory coding. They further claim that "We are convinced that SE-resonances are a general phenomenon, forming the basis of a correlation code which is used within and between different sensory systems and perhaps even throughout the entire brain." These Eckhorn's findings are consistent with our speculations on the basis of the FMNN model which assumes the significance of resonance frequency created by a group of neurons.

[10]Such a property of neurons are known to exist ([Crick and Asanuma, 1986], [Zeki, 1978], [Michael, 1978]), but has not been captured in the past neural network models.

[11]"The type of visual stimulation was found to influence the frequency of the dominant spectral peak of oscillatory responses." and "The mean frequency of the spectral peak increased by 5Hz if the stimulus drift velocity was doubled."

9 Conclusion

We have proposed an FMNN architecture which is a neural network with a heterogeneous composition and with a capacity to attain variable binding and structure building using the frequency modulation of pulses by groups of neurons. Assumption of *hypermodule* allows us to assume modular circuits with sequence prediction and recognition capability. Cognitive processing using the MSP models can be attained by combining *hypermodules*. In our model, cognitive activities are performed through the modulation and propagation of pulses in a neural network, whereas the conventional neural network models have been based on the propagation of scalar values. Although the formulation of the details of the formal character of the FMNN architecture is yet to be completed, the introduction of the hypothesis for frequency modulation of neural activation pulses is a significant enhancement to the conventional neural network architecture. Three assumptions we introduced in our model (i.e., spike simulation, a heterogeneous neural network, and frequency modulation) provide our model with an *information flow* and local functional capability which conventional neural networks have not attained. Linkage of distant neural circuit through oscillatory signals is one of the interesting features postulated in our model. This is also interesting from the viewpoint of neuro-physiology supported by the discovery of coherent oscillations [Eckhorn et. al., 1988]. We have also proposed that it is possible to build an FMNN machine and that an FMNN machine would be able to perform tasks such as inferencing and natural language processing that are run under a simulated massive-parallelism in the massively-parallel structured-marker passing (MSP) models at the hardware level.

Acknowledgements

The authors wish to thank Jaime Carbonell, Masaru Tomita, Hitoshi Iida, James McClelland and David Touretzky. We would also like to thank other members of the Carnegie Mellon community for fruitful discussions. Margalit Zabludowski was especially helpful in preparing the final version of this paper.

References

[Adrian, 1946] Adrian, E.D. *The Physical Background of Perception,* Clarendon Press, Oxford, 1946.

[Amari, 1978] Amari, S. *Shinkei Kairomou no Suuri (The Mathematics of Neural Circuits).* Sangyou Toshyo, 1978

[Beltrami, 1987] Beltrami, E., *Mathematics for Dynamic Modeling*, Academic Press, 1987.

[Bookman, 1987] Bookman, L.A. 'A Microfeature Based Scheme for Modelling Semantics'. In *Proceedings of the IJCAI-87*, 1987.

[Charniak, 1983] Charniak, E. 'Passing Markers: A theory of Contextual Influence in Language Comprehension'. Cognitive Science 7. 1983.

[Charniak, 1986] Charniak, E. 'A neat theory of marker passing'. In *Proceedings of the AAAI-86*, 1986.

[Crick and Asanuma, 1986] Crick, F. and Asanuma, C. 'Certain Aspects of the Anatomy and Physiology of the Cerebral Cortex'. In J. McClelland and D. Rumelhart (Eds.) *Parallel Distributed Processing*. MIT Press, 1986.

[Eckhorn et. al., 1988] Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. and Reitboeck, H. J. 'Coherent Oscillations: A Mechanism of Feature Linking in the Visual Cortex?' Biol. Cybern. 60, 121-130, 1988.

[Elman, 1988] Elman, J.L. 'Finding Structure in Time,' CRL technical report 8801, Center for Research in Language, University of California, San Diego, CA, 1988.

[Feldman and Ballard, 1982] Feldman, J.A. and Ballard, D.H. 'Connectionist models and their properties'. *Cognitive Science,* 1982.

[Halstead, 1985] Halstead, R. 'Multilisp: A language for Concurrent Symbolic Computation'. In *ACM Trans. on Prog. Languages and Systems.* 1985

[Hendler, 1988] Hendler, J. *Integrating Marker-Passing and Problem-Solving.* Lawrence Erlbaum Associates, 1988.

[Hillis, 1985] Hillis, D.W., *The Connection Machine.* MIT Press, 1985.

[Hirst, 1984] Hirst, G. 'A Semantic Process for Syntactic Disambiguation'. In *Proceedings of AAAI-84*. 1984.

[Kaplan and Bresnan, 1982] Kaplan, R. and Bresnan, J. Lexical Functional Grammar: A formal System for Grammatical Representation, In *The Mental Representation of Grammatical Relations*, ed J. Bresnan. MIT Press, 1982.

[Kitano, 1988] Kitano, H. 'Multilingual Information Retrieval Mechanism using VLSI,' In *Proceedings of RIAO-88,* 1988.

[Kitano, 1989] Kitano, H., 'A Massively Parallel Model of Natural Language Generation for Interpreting Telephony: Almost Concurrent Parsing and Generation,' In *Proceedings of the Second European Workshop on Natural Language Generation,* 1989.

[Kitano et. al., 1989a] Kitano, H., Tomabechi, H. and Levin, L. 'Ambiguity Resolution in DmTrans Plus,' In *Proceedings of European Chapter of the Association for Computational Linguistics,* 1989.

[Kitano et. al., 1989b] Kitano, H., Tomabechi, H., Mitamura, T. and Iida, H. 'A Massively Parallel Model of Speech-to-Speech Dialog Translation: A Step Toward an Interpreting Telephony,' Proceedings of European Conference on Speech Communication and Technology (EuroSpeech89), 1989.

[Kitano and Tomabechi, ms.] Kitano, H. and Tomabechi, H. Manuscript. 'Frequency Modulation Neural Network Theory: Physical, Biological and Cognitive Foundations for Very Large Heterogeneous Neural Network'. Carnegie Mellon University.

[Kitano et. al., ms.] Kitano, H., Iida, H., Mitamura, T. and Tomabechi, H. Manuscript, 'An Integrated Discourse Understanding Model for an Interpreting Telephony under the Direct Memory Access Paradigm'. Carnegie Mellon University.

[Kuffler et. al., 1984] Kuffler, S., Nicholls, J. and Martin, R. *From Neuron to Brain,* Sinauer Associates, Inc., 1984.

[Langton, 1988] Langton, C., (Ed.) *Artificial Life*, Addison-Wesley, 1988.

[Loeb, 1985] Loeb, G. E., 'The functional replacement of the ear'. Scientific American, Feb., 1985.

[Marubayashi, 1981] Marubayashi, G., *Tuushin Densou Kougaku (Transmission Theory and Its Applications),* Corona Publishing, 1981 (in Japanese).

[McCulloch and Pitts, 1943] McCulloch, W.S. and Pitts, W.H. *A Logical Calculus of the Ideas Immanent in Neural Nets*. Bull. Math. Biophys., 5, 115-133, 1943.

[Michael, 1978] Michael, C.R. 'Color vision mechanisms in monkey striate cortex: Dual-opponent cells with concentric receptive fields'. Journal of Neurophysiology, 41, 1978.

[Murphy et. al., 1983] Murphy, R.K., Johnson, S.E. and Sakaguchi, D.S. 'Anatomy and Physiology of Supernumerary Cercal Afferents in Crikets: Implications for Patter Formation," J. Neurosci. 3:312-325, 1983.

[Norvig, 1987] Norvig, P. 'Inference in Text Understanding'. In *Proceedings of the AAAI-87*, 1987.

[Nottebohm, 1989] Nottebohm, F. 'From Bird Song to Neurogenesis'. Scientific American, Feb., 1989.

[Papez, 1937] Papez, J. W. 'A proposed mechanism of emotion,' Archives Neurol. Psychiat., 38, pp. 725-744, 1937.

[Pollard and Sag, 1987] Pollard, C. and Sag, I. *Information-based Syntax and Semantics,* Chicago University Press, 1987.

[Quillian, 1969] Quillian, M.R. (1969) 'The teachable language comprehender'. BBN Scientific Report 10. 1969.

[Riesbeck and Martin, 1985] Riesbeck, C. and Martin, C. *Direct Memory Access Parsing*, YALEU/DCS/354, 1985.

[Rosen et. al., 1989] Rosen, S.C. *et. al.* 'Selective Modulation of Spike Duration by Serotonin and the Neuropeptides, FMRFamide, SCP_B, Buccalin and Myomodulin in Different Classes of Mechanoafferent Neurons in the Cerebral Ganglion of *Aplysia*'. Journal of Neuroscience, Feb., 1989 9(2):390-402.

[Rumelhart et. al., 1986] Rumelhart, D.E., Hinton, G.E. and McCleland, J.L. 'A General Framework for Parallel Distributed Processing,' In McClelland J. and Rumelhart, D. (Eds.), *Parallel Distributed Processing*. MIT Press 1986.

[Ryckebusch et. al., 1989] Ryckebusch, S., Bower, J.M. and Mead, C. 'Modeling Small Oscillating Biological Network in Analog VLSI', In Touretzky, D. (Ed.) *Neural Information Processing System I.*, Morgan Kaufmann, 1989.

[Sachs et. al., 1983] Sachs, M.B., Voigt, H.F. and Young, E.D. 'Auditory nerve representation of vowels in background noise'. Journal of Neurophysiology, 50, 1983.

[Saito and Tomita, 1988] Saito, H. and Tomita, M. (1988) 'Parsing Noisy Sentences'. In *Proceedings of the COLING'88*.

[Scott and Muller, 1980] Scott, S.A. and Muller, K.J. *Devel. Biol.* 80: 345-363, 1980.

[Sejonwski, 1981] Sejonwski, T.J., 'Skelton filters in the brain'. In Hinton, G.E. and Anderson, J.A. (Eds.) *Parallel models of associative memory*. Erlbaum, 1981.

[Sejonwski, 1986] Sejonwski, T.J., 'Open Questions About Computation in Cerebral Cortex'. In McClelland J. and Rumelhart, D. (Eds.), *Parallel Distributed Processing*. MIT Press 1986.

[Tomabechi, 1987] Tomabechi, H., 'Direct Memory Access Translation'. In *Proceedings of IJCAI-87*, 1987.

[Tomabechi and Tomita, 1988a] Tomabechi, H. and Tomita, M., 'Application of the Direct Memory Access paradigm to natural language interfaces to knowledge-based systems'. In *the Proceedings of COLING'88*, 1988.

[Tomabechi and Tomita, 1988b] Tomabechi, H. and Tomita, M., 'The Integration of Unification-based Syntax/Semantics and Memory-based Pragmatics for Real-Time Understanding of Noisy Continuous Speech Input'. In *Proceedings of the AAAI-88*, 1988.

[Tomabechi et. al., 1988] Tomabechi, H. Mitamura, T. and Tomita, M. 'Direct Memory Access Translation for Speech Input:' A Massively Parallel Network of Episodic/Thematic and Phonological Memory. In *the Proceedings of the International Conference on Fifth Generation Computer Systems 1988 (FGCS'88)*. 1988.

[Tomabechi et. al., 1989a] Tomabechi, H., Saito, H. and Tomita, M. 'SpeechTrans: An Experimental Real-Time Speech-to-Speech Translation System'. In *Proceedings of the Spring Symposium of the AAAI* 1989.

[Tomabechi et. al., 1989b] Tomabechi, H., Kitano, H., Mitamura, T., Levin, L. and Tomita, M. *Direct Memory Access Speech-to-Speech Translation: A theory of Simultaneous Interpretation.* CMU-CMT-89-111. Carnegie Mellon University. 1989.

[Tomabechi and Tomita, ms.] Tomabechi, H. and Tomita, M. Manuscript. 'Massively Parallel Constraint Propagation: Parsing with Unification-based Grammar without Unification'. Carnegie Mellon University.

[Tomabechi and Levin, ms.] Tomabechi, H. and Levin, L. Manuscript. 'Head-features and subcategorization as interacting constraints in associative memory'. Carnegie Mellon University.

[Touretzky, 1988] Touretzky, D. 'Beyond Associative Memory: Connectionists Must Search for Other Cognitive Primitives'. In *the Proceedings of the 1988 AAAI Spring Symposium Series. Parallel Models of Intelligence*, 1988.

[Treutlein and Schulten, 1985] Treutlein, H. and Schulten, K. 'Noise Induced Limit cycles of the Bonhoeffer-Van der Pol Model of Neural Pulses,' Ber. Bunsenges. Phys. chem. 89, 710-718, 1985.

[von der Malsburg, 1985] von der Malsburg, C. 'Nervous Structures with Dynamical Links,' Ber. Bunsenges. Phys. Chem. 89, 703-710, 1985.

[von der Malsburg and Schneider, 1986] von der Malsburg, C. and Schneider W. "A Neural Cocktail-Party Processor," Biol. Cybern. 54, 29-40, 1986.

[Waltz and Pollack, 1984] Waltz, D. and Pollack, J. 'Phenomenologically plausible parsing'. In *Proceedings of the AAAI-84*, 1984.

[Zeki, 1978] Zeki, S.M. 'The cortical projections of foveal straite cortex in the rhesus monkey'. Journal of Physiology, 277, 1978.

[Zipser, 1985] Zipser D. 'A Computational Model of Hippocampal Place Fields,' Behavioral Neuroscience, Vol. 99, No. 5, 1006-1018, 1985.