

Genetic specification of recurrent neural networks

Draft - with errors and incomplete, not peer reviewed, unpublished

Bill Howell, Ottawa January 2006 - includes some far-fetched implications & "Things to ponder"

Abstract – Computational Neuro-Genetic Modeling (CNGM) is discussed from the perspective of building Artificial Neural Network architectures starting with substantially pre-defined modules and processes (DNA-ANNs). This is equivalent to assuming that DNA code in a neuron can ultimately specify function, process and some level of data abstraction beyond the immediate role of genes to produce proteins or to regulate processes, and using that basis as a metaphor for DNA-ANNs. A robust and diverse set of ensembles or modules of DNA-ANNs is sought that is sufficient for a given problem domain, and that generalizes well. The potential advantages that might be derived from highly evolved, fine-grained hybrid genetic/connectionist systems, and some of the implementation challenges that they could present are discussed.

I. INTRODUCTION

ANNs were originally inspired by the search for ways to explain the functioning of the brain and to replicate its great computational powers (Anderson J.A. Anderson, E. Rosenfeld, (editors) Talking nets: An oral history of neural networks. Cambridge, MA USA: Bradford Books, 1998.

This book provides an excellent context to early ANN research. This includes several references to the work of McCulloch&Pitts (eg W. McCulloch, W. Pitts, "A logical calculus of ideas immanent in nervous activity", 1943, reprinted in Neurocomputing.) and many other early researchers and how their thinking developed.). The brain will likely remain a source of inspiration and a far-off golden standard for very advanced computational capabilities for quite some time to come, so it is not surprising to see ANN research consider ways to incorporate new concepts from genetics, even if this is somewhat speculative at the current time.

Much of the focus of ANN research in the past has been on building connectionist architectures with various fixed activation functions for the nodes or neurons, using a variety of training approaches, and sometimes employing growth or pruning strategies to evolve a network structure. Often the objective has been to develop powerful general learning algorithms with somewhat specialized ANN architectures to tackle a broad class of problems, for example the ubiquitous Multi-Layer Perceptron (MLP) with backpropagation

Bill Howell is with the Mining and Mineral Sciences Laboratories of the Department of Natural Resources Canada in Ottawa. (e-mail: BillHowell@rogers.com) Note that this paper is a personal project and it does not reflect the interests, opinions and activities of the Government of Canada.

(Werbos P.J. Werbos, The roots of backpropagation: from ordered derivatives to neural networks and political forecasting, New York: John Wiley & Sons:1994. The book includes Werbos' Harvard PhD thesis from 1974., Parker D. Parker, "Learning logic", Invention report S81-64, File 1, Office of Technology Licensing, Stanford University, 1982., LeCunn Y. , Rummelhardt, Hinton, McLelland D.E. Rummelhart, G.E. Hinton, R.J. Williams, "Learning internal representations by error propagation", In J.L. McClelland, D.E. Rummelhart, & the PDP Research Group, Parallel distributed processing, explorations in the microstructure of cognition, Cambridge: MIT Press, 1986.), Adaptive Resonance Theory (Carpenter and Grossberg G.A. Carpenter, S. Grossberg, (editors) Pattern recognition by self-organizing neural networks, London: Bradford Book, MIT Press, 1991. This book provides an excellent history of the ART architecture, as well as work by other early leaders in the neural network field.), and Self Organizing Maps (Kohonen T. Kohonen, refer to J.A. Anderson, E. Rosenfeld, (editors) Talking nets: An oral history of neural networks.

Cambridge, MA USA: Bradford Books, 1998. This book provides an excellent context to early ANN research. This includes several references to the work of McCulloch&Pitts (eg W. McCulloch, W. Pitts, "A logical calculus of ideas immanent in nervous activity", 1943, reprinted in Neurocomputing.) and many other early researchers and how their thinking developed. and G.A. Carpenter, S. Grossberg, (editors) Pattern recognition by self-organizing neural networks, London: Bradford Book, MIT Press, 1991. This book provides an excellent history of the ART architecture, as well as work by other early leaders in the neural network field. for the context in which Kohonen developed Self Organizing Maps (SOMs).).

Evidence is rapidly accumulating that non-protein-coding DNA, sometimes in the past popularly referred to as "junk DNA", is not all simply waste coding interspersed between gene-coding regions (exons), but that it may have many very important roles in directing cellular processes beyond protein coding (Marcus G. Marcus, The birth of the mind: how a tiny number of genes creates the complexities of human thought, New York: Basic Books, 2004. This book (along with Pinker's "Blank slate") is an essential read, and its concepts and Marcus' current work are a basis for the current paper.) and even beyond well-known regulatory functions (Mattick J.S. Mattick, "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms"

BioEssays, vol 25 pp930-939, Oct. 2003). This is an exciting area of research for biology and neuroscience, and it will likely influence thinking in the humanities, cognitive sciences and Computational Intelligence. Several researchers are looking into how one might incorporate

DNA-like capabilities into Artificial Neural Networks (ANNs), or apply ANN methods to better understand and model the role of genetics in biological NNs. This work, currently described by Kasabov's phrase "Computational Neuro-Genetic Modeling" N. Kasabov, L. Benuskova, S.G. Wysoski, "Computational neurogenetic modelling: Gene networks within neural networks" Proceedings of the International Joint Conference on Neural Networks, Budapest, IEEE, is proceeding even though our knowledge of the biological processes is still in its infancy.

Work relating genetics and neuron or brain information processing is very broad, and is discussed in more detail in Section II.A. However, it is worthwhile to quickly list a few subjects:

- DNA as a memory molecule in early work (Crick and others B.F. Vanyushin, N.A. Tushmalova, L.V. Gus'kova, "Methylation of brain DNA as an index of participation of the genome in mechanisms of individual acquired memory", Doklady Akademii Nauk SSSR, vol219, pp742-744, November 1974. translated by Plenum Publishing, New York, 1975. to D.H. Adams, "Triplet code-independent programming of living systems organisation by DNA: the link with intelligence and memory", Medical Hypothesis, vol44, pp419-427, 1995.), although experiments haven't supported this;
- CNGMs as a means of modelling the the EEG signature of gene networks (Kasabov N. Kasabov, L. Benuskova, S.G. Wysoski, "A Computational Neurogenetic Model of a Spiking Neuron", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1768, pp 446-451, 31 July – 4 August 2005.);
- Ontogeny, or growth of the brain, and how genes can efficiently do this (Storjohann and Marcus R. Storjohann, G.F. Marcus, "NeuroGene: Integrated simulation of gene regulation, neural activity and neurodevelopment", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1646, pp428-433, 31 July – 4 August 2005.);
- Nature versus nurture – models of the relative importance of genetics and learning to retinotopic mappings (Thivierge J-P. Thivierge, E. Balaban, "Faithful Retinotopic Maps with Local Optimum Rules, Axonal Competition, and Hebbian Learning", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1491, pp2760-2765, 31 July – 4 August 2005.);

As a complement and contrast to the above work, this paper focuses on the genetic specification and operation of ANNs, simply referred to as DNA-ANNs, for which a core architecture and learning is pre-specified but the DNA-ANN may still learn and evolve over time in a variety of environments. The ultimate intent is to utilize a diverse combination of DNA-ANN architectures and learning methods such as mentioned in the second paragraph of this introduction. Additionally, there is some emphasis on

recurrent neural networks in this paper (RNNs, referred to as DNA-RNNs when incorporating genetics).

Section II of this paper summarizes three key sources of inspiration for DNA-RNNs: genetics; computational models of neuron and brain function; and trends with ANNs. The sub-section on genetics overviews some advances in interpreting the diverse roles of DNA, and more specifically overviews current work on the DNA specification of the brain and the way in which it is structured and functions. Section III lists the benefits that we hope to achieve by this pre-specification, and Section IV expands the discussion to issues raised by DNA-ANNs but which have a much broader significance, in particular with respect to our concepts of the mind.

II. INSPIRATION FOR DNA-ANNs

A. *Inspiration from Genetics: beyond protein coding and regulation*

1. *DNA beyond genes and immediate regulatory function*

Conventionally, "genes" might have been thought of as being "assembly language programming" for proteins, almost in a literal sense. This process involves finding the appropriate starting point on the DNA, culling out non-protein-coding sequences of mRNA (introns, part of npcDNA) from coding regions (exons) as indicated by spliceosomes. The mRNA is then used in ribosomes for the assembly of proteins. Occasionally, "decision points" occur in gene sequences that can result in different final proteins, depending on the regulatory function of a small component of npcDNA.

Additionally, a protein or enzyme may have different functions in different cell types, meaning that there is some degree of "overloading" (multi-functional roles) for some genes. This concept is revisited in Section III.

However, for a long time researchers have suspected that the role of DNA goes much further than merely coding for proteins, that life was not merely a "combinatorial soup" (Mattick J.S. Mattick, "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms" BioEssays, vol 25 pp930-939, Oct. 2003) that coincidentally gave rise to the tremendously detailed, complex and specific anatomy and physiology of the cells and of entire organisms (Mattick J.S. Mattick, "The hidden genetic program of complex organisms", Scientific American, pp60-67 Oct. 2004. See also <http://imbuq.edu.au/groups/mattick>, Ast G. Ast, "The alternative genome" Scientific American, vol 292 issue 4, Apr. 2005, pp58-65, Eddy S.R. Eddy, "Non-coding RNA genes and the modern RNA world" Nature Reviews Genetics, vol 2, pp 919-929, Dec. 2001, Gibbs W.W. Gibbs, "The unseen genome: Gems among the junk" Scientific American, vol 289 no 5, Nov. 2003 pp46-53).

Moreover, while it seemed that there should be a greater role for DNA than simply as "assembly language programming", it was also clear that the "genes" of eukaryotic organisms

(having cells with a nucleus as opposed to prokaryotic organisms with non-nuclear cells) account for only a small portion of the total DNA. For example, earlier estimates were that only ~1.5% of the 3.2 Giga base-pairs of amino acids in human DNA codes for proteins (less now, given the lower estimates of today's total gene counts!), and yet most DNA is transcribed to RNA! Furthermore, (Mattick J.S. Mattick, "The hidden genetic program of complex organisms", Scientific American, pp60-67 Oct. 2004. See also <http://imbuq.edu.au/groups/mattick>) points out that there is a poor relation between an organism's complexity and the number of protein-coding genes, but there is a more consistent relation between an organism's complexity and the amount of npcDNA. His description of the Cambrian period bio-complexity explosion (~1 Gy ago), and the rest of his analysis provide a substantial basis for going beyond the "central dogma of biology", that DNA only codes for genes and or their direct regulation. Highly complex architecture argues for the need or advantage and power of precise plans and "drawings" that are highly specific, either in the case of human buildings or biological systems.

The past expression "junk DNA" had perhaps conveyed the message that much of the DNA may simply be a wasteland of leftover code from past or modified needs, code resulting from replication errors, or code that served as a buffer to protect against certain types of error. While some of the npcDNA may have arisen by such mechanisms, the impression that much of it is "junk" could possibly be one of the great misnomers of scientific history. Furthermore, perhaps genes are merely the simplest and lowest level of coding on the DNA, just a starting point for much more powerful functionality and abstractions to be discovered.

Changes in DNA or its expression can occur through mutation, crossover of genes (diploid cells and sexual reproduction), and epigenetically (including cell differentiation). Epigenetic changes affect which sections of DNA occur through changes in the availability or in the "locking up" of sections of chromatin, the structures comprised of disk-like histones around which DNA sections are coiled. Methylation of sites, for example, either mask or reveals DNA code for use in the cell. More recently, researchers are focusing on non-protein-coding RNA and micro-RNA. The latter consists of segments of RNA that can perhaps be as short as 15 to 35 base pairs, and are thought to play a role in regulating gene expression. The work of Kim, Krichevsky et.al. J. Kim, A. Krichevsky, Y. Grad, G.D. Hayes, K. Kosik, G.M. Church, G. Ruvkun, "Identification of many microRNAs that copurify with polyribosomes in mammalian neurons" PNAS vol 101 no 1, 06Jan04, pp360-365. focuses on the role of micro-RNA on neuron-related genetics, and shows that micro-RNA can play a gene-regulatory role by influencing spliceosome decision points.

There have been a couple of strange anomalies in the application of genetic theory to ANNs:

- Duplication - Marcus G. Marcus, The birth of the mind: how a tiny number of genes creates the complexities of human thought, New York: Basic Books, 2004. This

book (along with Pinker's "Blank slate") is an essential read, and its concepts and Marcus' current work are a basis for the current paper. has described the great power of DNA duplication for biological systems, wherein a duplicate copy of DNA coding offers the opportunity of introducing errors, crossover, or mutations in the code while still retaining the functionality of the original copy (or visa versa). This provides a huge improvement in the ability to carry out "genetic experimentation" in natural systems. However, although one may interpret some earlier papers as being somewhat in this light (Fogel D.B. Fogel, (editor) Evolutionary computation: The fossil record: Selected readings on the history of evolutionary algorithms.

Piscataway, NJ: IEEE Press, 1998), this does not seem to have been widely adapted in the Computational Intelligence field prior to ~1998.

- "Nature versus Nurture" - Even more importantly, we should consider that the "nature versus nurture" debate has been largely anachronistic ever since a rough estimate of the the quantity of DNA base pairs and of the number of neurons and synapses have been available. As pointed out by Pinker S. Pinker, The blank slate, New York: Penguin Books, 2003. Pinker's book provides a much broader and deeper analysis of many of the concepts discussed in tgis paper. and Marcus G. Marcus, The birth of the mind: how a tiny number of genes creates the complexities of human thought, New York: Basic Books, 2004. This book (along with Pinker's "Blank slate") is an essential read, and its concepts and Marcus' current work are a basis for the current paper., DNA cannot fully specify the complete details of the brain, and it seems clear to this author that some randomness in addition to repetitiveness must result from ontogeny (growth) of the brain, and that this is NOT "environmentally driven". We have at the least a three-legged stool (nature, ontogeny, nurture), and most likely there are more stool legs than that. Additionally, we still cannot yet measure the complete DNA nor epigenetics for individuals in experiments, and don't know what most variations will mean. Therefore in a practical sense our measure of "nature" is far from complete. While dichotomies for a complex systems are useful for pedagogical reasons and to stimulate thinking when introducing a subject, if one is not careful they may become less a tool for enlightenment and more a tool for the marketing of beliefs and opinions.

2. DNA and the brain

[Minsky M. Minski and S. Papert, *Perceptrons*, MIT Press, Cambridge, MA, 1969.. Quoted by Emil M. Petriu, "Neural Networks: Basics" (tutorial)

http://ebrains.la.asu.edu/~jennie/tutorial/Emil_M_Petriu.pdf]: "The marvelous powers of the brain emerge not from any single, uniformly structured connectionist network but from highly evolved arrangements of smaller, specialized networks which are interconnected in very specific ways."

Clearly, brain growth or ontogeny is a very highly ordered and directed process, and while there is a high degree of variability in comparing brains and brain function between individuals (based on anatomy, fMRI etc), there is also an extremely high degree of commonality and regularity. Strikingly similar adult human behavior by identical twins separated at birth and the theory of Universal Generative Grammar for linguistics (Chomsky N. Chomsky, as quoted in Pinker 1994) are good examples of arguments for the pre-specification not just of "normal" anatomy and physiology, but of a great deal of "initial content" for the brain (data, functions, operating systems and likely far beyond that to use an admittedly poor metaphor from computing, but one that is easy to relate to). Steven Pinker's book "The blank slate" S. Pinker, The blank slate, New York: Penguin Books, 2003.

Pinker's book provides a much broader and deeper analysis of many of the concepts discussed in this paper. summarizes a wide range of results, many of which have been known for a long time, that dispel the "blank slate" concept of the brain learning everything from scratch. The expression "blank slate" will be used often in this paper to contrast DNA-ANNs, which have at least some pre-specified initial content, with ANNs that only start to learn or evolve when presented with data.

Researchers have long conjectured about a more direct information processing role for DNA in the brain. Several researchers over time have proposed that DNA would have excellent characteristics as a "memory molecule", providing enormous information storage capabilities (B.F. Vanyushin, N.A. Tushmalova, L.V. Gus'kova, "Methylation of brain DNA as an index of participation of the genome in mechanisms of individual acquired memory", Doklady Akademii Nauk SSSR, vol219, pp742-744, November 1974. translated by Plenum Publishing, New York, 1975.-D.H. Adams, "Triplet code-independent programming of living systems organisation by DNA: the link with intelligence and memory", Medical Hypothesis, vol44, pp419-427, 1995.), but this could not be substantiated by experiments. However it should be noted that conventional genomics, which relates genes to proteins and regulatory functions, is vastly easier than trying to relate DNA coding to brain structure, function and information content, neither of which can be directly measured in detail. That problem will get much worse as higher and higher levels of abstract brain function are addressed.

As mentioned in the Introduction, ANN models are being developed to describe the genetic specification of the growth and function of the brain and sensory organs (Storjohann and Marcus R. Storjohann, G.F. Marcus, "NeuroGene: Integrated simulation of gene regulation, neural activity and neurodevelopment", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1646, pp428-433, 31 July – 4 August 2005., Thivierge J-P. Thivierge, E. Balaban, "Faithful Retinotopic Maps with Local Optimum Rules, Axonal Competition, and Hebbian Learning", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1491, pp2760-2765, 31 July – 4 August 2005.), to model gene regulatory networks for spiking

neurons (Kasabov et al N. Kasabov, L. Benuskova, S.G. Woski, "A Computational Neurogenetic Model of a Spiking Neuron", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1768, pp 446-451, 31 July – 4 August 2005.), and to explore the possibility of a more direct "information processing" role for npeDNA in ANNs (Howell W.N. Howell, Presentation, Workshop on Computational Neuro-Genetic Modelling, International Joint Conference on Neural Networks 2006, Montreal,, presentation available from author.). Marcus convincingly argues that the mind is substantially specified by genes in a manner that maximizes the effect of their information content. This occurs because the genes don't specify all of the details. More importantly they specify how structures are built and learn. Learning through interaction with the environment fills in and tunes the brain.

More recently, work by Michael Meaney and associates in Montreal (I.C.G Weaver, N. Cervoni, F.A. Champagne, A.C. D'Alessio, S. Sharma, J.R. Seckl, S. Dymov, M. Szyf, M.J. Meaney, "Epigenetic programming by maternal behavior". Nature Neuroscience vol 7, no8, Aug. 2004, pp847-854, M.J. Meaney, M. Szyf, "Maternal care as a model for experience-dependent chromatin plasticity?". Trends in Neurosciences, Vol.28 No.9 September 2005.) have established strong evidence that epigenetic modifications may occur during the first week of a rat's life due to its mother nurturing behavior, and that these epigenetically-based behavioral modifications can persist throughout the life of the rat. Furthermore, there is a high likelihood that this will be passed on to future generations (albeit there will presumably be a fresh influence due to each new generation of mothers). This is extremely interesting – not only does it provide a concrete example of epigenetically-driven changes in behavior, it also provides a basis for "Lamarckian heredity" of the mind whereby traits set during the life of an individual may be passed on, in contrast to the conventional "Mendelian heredity" for which the die is set at conception, and changes during an individual's life are not passed on other than through mutation, crossover, duplication of DNA code etc. Before being confident about the existence of Lamarckian heredity, the results would have to be confirmed over several generations of rats, and of course we would need experimental results for other examples as well, both behavioral and knowledge-related.

As mentioned in the introduction, given these growing indications of extensive genetic and epigenetic "pre-programming" of the brain, and possibly more direct "live" interactions between DNA, individual neurons and brain function, attempts at integrating gene-like capabilities into ANNs have begun. "Computational Neuro-Genetic Modeling" (CNGM) is perhaps best defined by the question posed by Kasabov N. Kasabov, abstract for WCCI06 special session on Computational Neuro-Genetic Modelling (CNGM): http://www.wcci2006.org/WCCI-Web_Special_Session.html::

"Can they [CNGMs] facilitate the construction of [ANN] models that have flexible architectures, rapid trainability, adaptability to new environments, and the

capacity to facilitate knowledge representation? How can we use these models to improve our understanding of the brain and to find cures for brain diseases?"

This paper discusses the "biological metaphors" of neurons and genetics with respect to the specification and operation of ANNs. The paper also applies the "programming metaphor" when discussing the brain. For example, it is assumed that npcDNA ("junk DNA") hosts higher-level coding related to brain architecture, physiology, ontogeny (growth), function at different levels of abstraction, "operating systems" or higher level goal directed processes. Presumably the "junk DNA" also codes for several more layers of powerful concepts/ processes that we are not even aware of yet in computing science.

3) *Influence of brain function on DNA expression*

The whole "programming metaphor" for the DNA pre-specification and operation of the brain is hypothetical and only weakly indicated by experimental results. Much of the neurological experimental evidence is more suitable for establishing the "feedforward" influence of DNA on information content and processing. However, it is assumed in this paper for DNA-RNNs (artificial neural nets, not biological brains, but using biology as a metaphor and visa versa) that the information content of the processing activities of a neuron or region of the brain can of itself influence npcDNA expression (a information "feedback" influence on selecting DNA code). There doesn't seem to be much direct evidence for that biologically, but it will be interesting to see if this theme emerges from experiments in the future.

Clearly organisms respond to their environment, and this involves a huge cascade of gene expression and regulation at every level of the organism, and one might argue that these could be more or less "fixed" responses. But learning in the brain would require far more flexibility and rapidity of change than might exist or be allowable in most physiological responses. In a sense the immune system might provide an analogous example that must change very quickly to a rapidly changing external environment.

Meaney's results as discussed previously show behavioral changes in infant rats that are linked with epigenetic changes brought on by their mother's nurturing behavior, and that is at least a start. Extensive research has been carried out to model the role of hormones and more specifically neuromodulators (Doya, Dayan and Hasselmo K. Doya, P. Dayan, M. Hasselmo, (guest editors) "2002 Special Issue: Computational Models of Neuromodulation" Neural Networks, Vol 15 Nos 4-6, June-July 2002.). But while this work models the influence of neuromodulators on the functioning of neurons or regions of the brain, it does not show specific mechanisms for the influence of information processing on DNA expression other than for physiological requirements.

Perhaps future work by Krichevsky and her collaborators as mentioned in the previous subsection will establish more of a direct role for DNA code on information processing and, visa versa, a direct influence of neural information processing on npcDNA expression (perhaps including micro-RNA).

However, even though the biological parallels have not yet been established, the central assumption of this paper is that the "DNA programming code" of DNA-ANNs not only can pre-specify architecture, function and processes, but that it in turn can be directly invoked by the information-processing that is occurring in a neuron or "ensembles or modules" of neurons. Biologically and neurologically that remains speculation.

B. Inspiration for DNA-ANNs from models of neurons, brain regions, and psychology

There has been a great deal of work over several decades on building ANN models of neurons, brain regions, sensory systems (olfactory by Freeman W.J. Freeman, "Stochastic Chaos Versus Deterministic Chaos: A Case for Analog Versus Digital Embodiment of Devices for Pattern Recognition", Proceedings of IJCNN 1999, International Joint Conference on Neural Networks. Washington, DC, paper#3012, 10-16 July 1999., Padgett G. Szekely, M.L. Padgett, G. Dozier, T.A. Roppel, "Odor detection using pulse coupled neural networks", Proceedings of IJCNN 1999, International Joint Conference on Neural Networks. Washington, DC, paper #3041, 10-16 July 1999. and their co-workers), motor systems (Grossberg J.W. Brown, D. Bulloch, S. Grossberg, "How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades", Neural Networks, vol 17, 2004, pp471-510.) and the effects of neuromodulators (Doya, Dayan and Hasselmo K. Doya, P. Dayan, M. Hasselmo, (guest editors) "2002 Special Issue: Computational Models of Neuromodulation" Neural Networks, Vol 15 Nos 4-6, June-July 2002. as cited above). Furthermore, many general computational learning methods have been based on or inspired from the field of psychology (Hebbian learning, ART, Backpropagation, K series chaotic ANNs). It has been natural to think in terms of specialized neuron arrangements, "modules", connections, specialized regions of the brain, and whole-brain processes, the latter exemplified by major efforts to computationally model the whole brain Editorial, "Modelling the brain: grey matter, blue matter", Economist, 11 June 2005, pp75-76. See also <http://bluebrainproject.epfl.ch/>.

There has also been a great deal of work on sensory and brain biomimetics and prosthesis (for example, retinal prosthetic work by groups led by Mark Humayun M. Humayan, D. Yanai, R.J. Grenberg, J. Little, B.V. Mech, M. Mahadevappa, J.D. Weiland, G.Y. Fujii, E. DeJuan, "Clinical results with the model 1 IRP implant", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, Plenary presentation, 31 July - 4 August 2005. and Eckhorn et al. R. Eckhorn, M. Eger, M. Wilms, T. Schanze, "Information transmission from

a retina implant to the cat visual cortex", Proceedings of IJCNN 2004, International Joint Conference on Neural Networks. Budapest, paper #1334, 2004., and a hippocampal prosthesis by Ted Berger's group T.W. Berger, J.J. Granacki, V.Z. Marmarelis, A.R. Tanguay, S.A. Deadwyler, G.A. Gerhardt, "Implantable Biomimetic Electronics as Neural Prostheses for Lost Cognitive Function", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper #1745 pp 3109-3114, 31 July – 4 August 2005.).

So what lessons or inspiration can we draw from neuron, brain, and psychology studies that can help us to develop DNA-ANNs? The short answer is that there will likely be a huge wealth of growth strategies, architectures, functions, and processes which should be directly relevant to DNA-ANNs, and the mind harbors capabilities that will continue to inspire research for a long time.

But there are at least some general observations that have often been made which are worth repeating here:

- Natural systems are rich, varied, dynamic and incredibly powerful at EVERY level – from the sub-neuron level neuron up to the whole brain. It goes without saying that this is a massively parallel architecture, commonly with recurrent connections.
- There is a huge diversity and redundancy of structures, and spiking. Non-stationary dynamics seems to be a key characteristic.
- The growth process of the brain involves making seemingly random connections between neurons, and the die-off of a significant fraction of the initial neurons. The response of an individual neuron or a brain region can vary dramatically from test to test in patch-clamp studies of neurons or EEG, PET, fMRI, and MEG experiments, such that averages over many repeat tests are used to extract the underlying signals from the "noise". In spite of this, there is a regularity to the architecture of the brain and its subsystems: the brain is capable of highly reproducible performance for challenging problems in very noisy environments, even where there may be significant damage to many of the neurons involved (graceful degradation of performance).
- We likely have a long ways to go before we can properly identify and understand processes at higher levels of abstraction, even though current models are showing promise.

Ideally, a more general discussion of neural and brain models (ANNs) would be provided here, but only a few examples are listed below:

- Frontal cortex and basal ganglia circuits (Brown, Bulloch & Grossberg J.W. Brown, D. Bulloch, S. Grossberg, "How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades", Neural Networks, vol 17, 2004, pp471-510.);

- Multi-modular brain (Taylor J. Taylor, "Neural Networks of the Brain: Their Analysis and Relation to Brain Images", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper #1410, pp1603-1608, 31 July – 4 August 2005.);
- Thalamo-cortical loops (Hecht-Nielsen R. Hecht-Nielsen, "A theory of thalamocortex", In Computational models for neuroscience: human cortical information processing, London: Springer-Verlag, pp 85-124, 2003.);
- Temporal sequences in frontal lobes (Taylor & Taylor N.R. Taylor, J.G. Taylor, "Learning to Generate Temporal Sequences by Models of Frontal Lobes", Proceedings of IJCNN 1999, International Joint Conference on Neural Networks. Washington, DC, paper #563, 10-16 July 1999.);
- Learning categories (Grossberg, Carpenter, Ersoy S. Grossberg, G.A. Carpenter, B. Ersoy, "Brain categorization: learning, attention, and consciousness", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper #1287, pp1609-1614, 31 July – 4 August 2005.); and
- Purkinje cell models W.L. Dunin-Barkowski, S.N. Markin, L.N. Podladchikova, D.C. Wunsch, "Climbing fibre Purkinje cell twins are found", Proceedings of IJCNN 1999, International Joint Conference on Neural Networks. Washington, DC, paper #444, 10-16 July 1999..

The work of Storjammen and Marcus R. Storjohann, G.F. Marcus, "NeuroGene: Integrated simulation of gene regulation, neural activity and neurodevelopment", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1646, pp428-433, 31 July – 4 August 2005. is providing a first computational framework for the genetic specification of growth and end structure of "modules" that produce the right kind of functionality. Kasabov et al N. Kasabov, L. Benuskova, S.G. Wysoski, "Computational neurogenetic modelling: Gene networks within neural networks" Proceedings of the International Joint Conference on Neural Networks, Budapest, IEEE, N. Kasabov, L. Benuskova, S.G. Wysoski, "A Computational Neurogenetic Model of a Spiking Neuron", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1768, pp 446-451, 31 July – 4 August 2005. have worked on modelling the EEG signal characteristics of brain regions based on the kinetics of gene networks.

This sub-section focuses on some key current trends of most interest to DNA-ANNs.

C. Inspirations for DNA-ANNs from trends with ANNs

The previous sub-sections have reviewed some of the trends and implications of research that links genetics with the partial pre-specification of neuron and brain architecture and

function, and the possibility (unproven) of active participation of DNA coding in learning and processing. This section gives a very brief overview of trends in ANNs that are also a source of inspiration for Computational Neuro-Genetic Modelling (CNGM).

Please note that biological and psychological terminology is used extensively in the rest of this paper even though most of the material refers to ANNs. Take this terminology in the metaphorical sense.

1) *Diversity of approaches*

As mentioned in the Introduction, there is a great diversity of ANN architectures and of approaches for their learning and evolution. The overall intent of DNA-ANNs as proposed in this paper is to make use of as many of the available ANN architectures and approaches as required or as possible. The overall assumption is that there will be a very diverse toolbox of approaches accommodated by DNA-ANNs, and that problem domain specific "systems" will arise from this toolbox. The various approaches would likely have a track record of reliable and robust performance either for very general problem classes, or for very specific functionalities or classes of problems. There is not an intent nor a requirement that only one approach or combination will "win out", so an ensemble solution can arise as a "diverse community" of effective approaches along with problem-specific evolved solutions.

2) *Local, incremental learning approaches*

Neural Gas Models T.M. Martinez, K.J. Schulten, "A neural gas model learns topologies", *Artificial Neural Networks*, pp397-402, 1991. P.A. Estevez, C.J. Figueroa, K. Saito, "Cross-entropy approach to data visualization based on the neural gas network", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper #1585, pp2724-2729, 31 July – 4 August 2005. and *Evolving Connectionist Systems (ECOS)*, Kasabov N. Kasabov, *Evolving connectionist systems: Methods and applications in bioinformatics, brain study, and intelligent machines*, London, UK: Springer-Verlag, 2003.) are examples of ANNs which have been designed to learn incrementally as new data is encountered. This may fit in well with DNA-ANNs, where a diverse population of substantially pre-specified DNA-ANNs could potentially be evolved in a more concise and effective manner.

3) *Multi-phase NN Architecture: "Crystalline to Gaseous"*

An interesting recent trend has been to generate substantially fixed weight neural networks for which only a small portion of the weights undergo changes during the learning phase, such as the Extreme Learning Machine (Huang G-B Huang, Q-Y Zhiu, C-K Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks" *Proceedings of IJCNN 2004, International Joint Conference on Neural Networks*. Budapest, paper# 1220, pp 985-990, 25-29 July 2004., D. Wang, G-B Huang, "Protein Sequence Classification Using Extreme Learning Machine"

Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1257 pp ???, 31 July – 4 August 2005.) and Echo State Networks (Jaeger H. Jaeger, "Adaptive nonlinear system identification with echo state networks," *NIPS 2002*., H. Jaeger, "Reservoir riddles: suggestions for Echo State Network research", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1767, pp 1460-1462, 31 July – 4 August 2005.). Although there are challenges and limitations (Prokhorov D.V. Prokhorov, G.V. Puskorius, L.A. Feldkamp, "Dynamical neural networks for control", in J. Kolen, S. Kremer, (eds) *A field guide to dynamical recurrent networks*, IEEE Press, 2001, pp.257-289.), this provides some insight into how one might do even better by rapidly integrating and training DNA-ANN "modules" by pre-specifying architectures that allow the modules to be very rapidly and efficiently trained and evolved.

4) *Networks, hierarchies of NNs:*

There has been strong interest in building ensembles and hierarchies of many varieties of ANNs (Carpenter and Martens G.A. Carpenter, S. Martens, "Self-Organizing Hierarchical Knowledge Discovery by an ARTMAP Information Fusion System", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1049, pp 452-456, 31 July – 4 August 2005., Cuadros-Vargas E. Cuadros-Vargas, R.A.F. Romero, "Special Session: Constructive/Hierarchical Self-Organizing Maps", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, , 31 July – 4 August 2005. This session included 6 papers and there were many other similar papers during the conference as well:). Clearly this is of direct relevance to DNA-ANNs, which must rapidly and effectively "restructure" combinations of pre-specified ANNs (both problem-class-specific and general). But while current efforts mainly focus on ensembles and hierarchies of ANNs of the same basic architecture by adjusting parameters to create some diversity, DNA-ANNs will face the additional challenge of finding general methods to make diverse ANN architectures (and input data type and number!) work together and learn effectively.

5) *RNNs and Approximate Dynamic Programming (ADP)*

Of particular relevance to this paper are Recurrent Neural Networks (RNNs), which are a very general and powerful class of ANNs that are challenging to train, but they can handle dynamical systems well and they are an important tool for tackling high-level decision making and adaptation challenges such as Approximate Dynamic Programming (ADP) and control (Werbos P.J. Werbos, "Neurocontrol and supervised learning: an overview and evaluation", and "Approximate dynamic programming for real-time control and neural modelling". in D.A. White, D.A. Sofge (editors), *Handbook of intelligent control*. New York: Van Nostrand Reinhold, 1992., Prokhorov, Puskorius, Feldkamp D.V. Prokhorov, G.V. Puskorius, L.A. Feldkamp, "Dynamical neural networks for control", in J. Kolen, S. Kremer, (eds) *A field guide to dynamical recurrent networks*, IEEE Press,

2001, pp.257-289., Venayagamoorthy G.K. Venayagamoorthy "Dynamic optimization of a multimachine power system with a FACTS device using identification and control ObjectNets", IAS2004, 0-7803-8487-3/04/\$20.00 © 2004 IEEE, Si et.al. J. Si, A.G. Barto, W.B. Powell, D. Wunsch, (eds) Handbook of learning and approximate dynamic programming. New Jersey, USA: IEEE Press and Wiley-Interscience, 2004.). RNNs are of particular relevance to the whole concept of DNA-ANNs, because many of the powerful capabilities of RNNs are also hoped-for advantages of DNA-ANNs (listed in the next section). A key question is whether DNA-ANNs will show any clear advantage over current RNN architectures and their training/ learning/ evolving strategies. This is also the reason that DNA-RNNs are the class of DNA-ANNs with some profile in this paper.

A recent paper (Santiago and Lendaris R. Santiago, G.G. Lendaris, "Reinforcement Learning and the Frame Problem", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1659, pp 2971-2976, 31 July – 4 August 2005.) claims that RNNs can overcome a fundamental limitation of Artificial Intelligence (AI) – the "Frames Problem" whereby breaking down large, complex problems to facilitate training over smaller classes of the problem domain leads to an explosion in the number of classes, which in itself impedes overall solutions. While perhaps this issue needs further work using "conventional RNNs" (in particular those based on Adaptive Critics and/or BPTT), it may hold lessons to help guide the development of DNA-RNNs. This subject also leads into the theme of a recent workshop on achieving functional integration led by Hussain T.S. Hussain, "Workshop: Achieving Functional Integration of Diverse Neural Models", International Joint Conference on Neural Networks. Montreal, 31 July – 4 August 2005. As this was a workshop, presentations were not published, but the author/titles are listed below, and presentations are available on the website:.

6) Signal processing and information theoretics

Signal processing techniques have been successfully applied to the difficult problem of training recurrent neural networks (RNNs) (Puskorius & Feldkamp G.V. Puskorius, L.A. Feldkamp, "Decoupled extended Kalman filter training of feedforward layered networks", International Joint Conference on Neural Networks. Seattle, vol. 1, pp771-777, 1991., G.V. Puskorius, L.A. Feldkamp, "Extensions and enhancements of decoupled extended Kalman filter training", International Conference on Neural Networks, Houston, vol. 3, pp1879-1883, 1997., Wan and van der Merwe E.A. Wan and R. van der Merwe, "Chap 5: Dual Extended Kalman Filter Methods" and "Chap. 7: The Unscented Kalman Filter", in S. Haykin, (editor) Kalman Filtering and Neural Networks, Wiley Publishing., 2001) in combination with the calculation of error derivatives for each node using Back Propagation Through Time (BPTT). Signal processing techniques are even a component of some of the more advanced language translation techniques

(Isabelle P. Isabelle, National Research Council of Canada, Ottawa, comment made 12 Oct. 2005 at Ottawa IEEE Computer Society sponsored event: NRC-IIT Technology Showcase": leading edge IBM human language translation software uses signal processing technology.).

From the perspective of DNA-ANNs the successes of signal processing are interesting first because of their direct relevance to RNNs, which are a primary application for the DNA-ANNs. Furthermore, there is hope that advanced signal processing techniques will be developed to train RNNs WITHOUT the necessity of calculating derivatives (Wan and van der Merwe E.A. Wan and R. van der Merwe, "Chap 5: Dual Extended Kalman Filter Methods" and "Chap. 7: The Unscented Kalman Filter", in S. Haykin, (editor) Kalman Filtering and Neural Networks, Wiley Publishing., 2001).

Secondly, signal processing approaches sometimes resemble iterative-substitution solutions and vaguely resemble information-theoretic techniques (Principe D. Erdogmus, K.E. Hild, J.C. Principe, "Independent components analysis Renyi's mutual information and Legendre density estimation", Proceedings of IJCNN 2001, International Joint Conference on Neural Networks. Washington, DC, paper #595_3, pp 2762-2767, 15-19 July 2001., K-H. Jeong, J-W. Xu, J.C. Principe, "An Information Theoretic Approach to Adaptive System Training Using Unlabeled Data", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper #1719, pp191-196, 31 July – 4 August 2005.), yet contrast with matrix solutions for linear systems, gradient-descent approaches, and several approaches that incorporate searches that are at least partially random, such as evolutionary computations, particle swarms, and chaos.

Will there be more substantial concepts and approaches to learn from signal processing, information theoretics from ANNs and connecting fibers in neurology? It certainly seems to be a promising area.

7) Pods – ensembles or modules, with a pre-specified core

For the sake of clarity in following sections it will help to define the concept of "Pods":

- "Pod", or "pre-configured ensemble architectures" - refers to specific ANN architectures and their functions and "processes" that have been defined or have evolved for a specific set of ANN modules or ensembles (the terms "module" and "ensemble" are used interchangeably in this paper, and these may be a single neuron, groups of neurons, or groups or groups, etc). The pre-configuration could occur at all levels of a structure, the structure can be arbitrarily connected and nested (pods within pods, pods as members of several other pods) and could include:
 - physical structure and connections of the DNA-RNN;
 - functionalities, processes and communications;

- pre-set weights for "crystalline or fixed weight" phases of an ANN (see Section II.C.3 above). level of neurons, ensembles connections hierachies, configurations, and weight settings (the weights for); and
- methods for learning and evolving.
- Pod versus Ensemble - One might consider pods to merely be ensembles of ANNs (ANN modules), but the desire here is to put the emphasis on the alternate pre-configured architectures and not just general learning of a grouping of ensembles (typically with one class of problems in mind).
- Pre-configuration – doesn't mean that the "DNA" of a "DNA-RNN" is the only pre-configuration that can occur. Here the pre-configuration applies to any pod or its modification or evolved form, whether pre-specified right from the start, whether it is learned/ evolved during a single problem, or whether it evolves over many applications (projects, researchers, etc) of a pod over time.

In summary, this sub-section listed some of the trends in ANNs that may be of particular relevance to the development of DNA-ANNs, but the expectation is that much of what is being done with ANNs will be useful or applicable to DNA-ANNs, so the intent is not to be exclusive. Furthermore, in spite of a great deal of progress since McCulloch and Pitts initial work J.A. Anderson, E. Rosenfeld, (editors) Talking nets: An oral history of neural networks. Cambridge, MA USA: Bradford Books, 1998.

This book provides an excellent context to early ANN research. This includes several references to the work of McCulloch&Pitts (eg W. McCulloch, W. Pitts, "A logical calculus of ideas immanent in nervous activity", 1943, reprinted in Neurocomputing.) and many other early researchers and how their thinking developed., and in spite of powerful tools and concepts that we now have at our disposal, there is still an enormous gap between the capabilities of ANNs and the full power of the brain.

III. WHAT MIGHT WE HOPE TO ACHIEVE WITH DNA-ANNs?

How **might** the incorporation of DNA-like capabilities into ANNs benefit their design, learning, evolution and function? For the sake of brevity, potential benefits are listed below with a minimum of explanations, but keep in mind that any advantage of DNA-ANNs over other ANNs remains to be demonstrated. Note that many of the points raised in Section II provide background for the DNA-ANN wish lists below.

A. "Starting with the right answer"

Starting with the right answer, or approximately the right structure, function, processes and network pre-training, is the most obvious potential advantage of pre-specified DNA-ANNs. In general we of course don't know the answer

before we start, but a good guess can go a long way towards finding a good solution quickly.

- Evolution has a very long time to develop extremely powerful methods – and perhaps "true learning" requires evolutionary processes (Fogel D.B. Fogel, Blondie 24: Playing at the edge of AI. San Diego, USA: Academic Press, 2002. p301 for comments on learning and evolution). It may be too much to ask that "excellent, real-time" ANNs arise quickly from "blank slate" ANNs that have no training and an inappropriate structure, while DNA-ANNs will naturally possess a diverse "toolbox" of proven solutions to a wide variety of problems. "Appropriate" modules could be selected from those available, then they might be quickly adaptable to the problem at hand. Obviously a desirable characteristic of DNA-ANNs would be the automatic evolution of new capabilities over time, including novel strategies and principles in both general and specialized domains.
- In essence, we can take advantage of "societal/ species experience" across time and environments well beyond an individual's own experience.
- The evolution of stable systems of complex modules may greatly facilitate their application, interpretation and the extraction of meaning from their connections and behavior when modeling a system. This type of "interpretability" may help contribute to building robust, effective systems of ANNs and hybrid systems involving ANNs. There are also several related issues:
 - Small-world universal function approximation – what is the smallest set of ANNs, of various functional capabilities (general to specific), that is sufficient to solve most of the problems in a domain of interest? Are there general design principles for deriving a reduced set (something like design and analysis patterns for software development (E. Gamma, R. Helm, R. Johnson, J. Vlissides, Design patterns: Elements of reusable object-oriented software. Reading, MA: Addison-Wesley, 1995, M. Fowler, Analysis patterns. Menlo Park, CA: Addison-Wesley, 1997., B.P. Douglass, Real-time design patterns: Robust scalable architecture for real-time systems. Boston, MA: Addison-Wesley, 2003.))? If ANNs of sufficient size do become universal function approximators, that doesn't mean that they can be trained in an effective, timely and quality manner v- afterall, a three layer MLP is just that.
 - False confidence in good fits - This is the converse situation whereby models of sufficiently complex systems may tend to become "small-world universal function approximators", and while these may fit data extremely well, there may not be a strong assurance that the model is conceptually well founded. Perhaps Global Circulation Models for climate are an example. In any case, one hope is that DNA-ANNs will have the property of developing "sensible" models for problems, models that minimize the chances of arriving at "wild" or phenomenologically erroneous solutions. Ockham's

razor comes to mind – see point #B below, but perhaps there is more to this.

- Reiterating the point above in another way, we are looking for the [evolution, emergence] of [enduring, robust, powerful] representations of [risks, processes, situations, systems, data]. We are hoping to simultaneously achieve goals that may appear to be conflicting when "learning from the blank slate":
 - excellent AND fast;
 - richness of representation and understanding AND with only a minimum of input data and environmental situations. Contrast the hidden attributes and incomplete understanding when starting from a "blank slate" versus the depth and power of representations and functionality when using DNA-ANNs that have evolved over time and over a very diverse universe of environmental challenges.
- We can avoid an excessive reliance on general, powerful learning methods, and base learning more on the capabilities and power of a long evolutionary heritage that is pre-programmed (including very general & powerful learning methods).
- The brain and modern computers have huge memory space and processing capacity. Take advantage of this! This means that we are not restricted to a small set of tools, nor to a single learning theory/method. That doesn't mean that very general powerful learning methods aren't important or that they won't be commonly used, it's just that there is no reason to be constrained to a small subset of approaches where highly specialised solutions work much better.

B. Higher levels of abstraction

Following on with the concept of rich representations as discussed above, what we would also like to achieve, and expect to be present with DNA-ANNs, are ever-evolving higher and higher level abstractions of the environments we encounter and the toolsets that we are using.

- Abductive reasoning (reasoning by simili and metaphor versus inductive, deductive and transductive logic) - Reasoning by simili or metaphor (pattern matching and beyond) could be one of the more important learning methods at higher levels of abstraction. This kind of reasoning would facilitate problem solving and innovation through restructuring existing solution for different problems. Furthermore, a rich DNA-RNN environment with "lessons from the past" is essential for this. Is abductive reasoning actually more common and powerful than inductive and deductive logic, especially after an ANN or individual gains experience and wisdom?
- Meaning - Perhaps by defining and identifying relatively-common pods with well known characteristics, it will become easier to describe the functioning of ANNs to predict their behaviors for different problem domains, and therefore to visualize how to build completely new pods and systems. This is related to the concept of "logic is an emergent property"

for complex systems (see Section IV.A below).

- Symbolic logic – immediately it comes to mind as being extended from abductive reasoning. When does symbolic (rule or belief based) reasoning take precedence, and is this related to the degree of abstraction?
- Coherence - Point #A above ("Starting with the right answer") also infers, for DNA-ANN modules which "fit together" with other modules (DNA-ANN pods), that they must "co-evolve". That doesn't mean that all new modules have to fit with many others – a "stand-alone" module might legitimately lack coherence with other modules yet still serve a purpose well.
- Systems-level tools - While concepts and toolsets move to higher levels of abstraction and complexity, we might expect the need for systems-level approaches to building and operating the complexity. Which of the conventional IT tools will apply, and how will these have to evolve? For example:
 - design/ analysis patterns – as mentioned in point #1 above ;
 - global brain model projects may illustrate these challenges Editorial, "Modelling the brain: grey matter, blue matter", Economist, 11 June 2005, pp75-76. See also <http://bluebrainproject.epfl.ch/>.
- Problem decomposition and modularisation seems to be a "natural" way to analyze complex systems. DNA-RNNs should help this (see "logic is an emergent property" in Section IV.A and "abductive reasoning" immediately above in II.B, which relate directly to this).
- Ockham's razor T.H. Leahey, A history of psychology: Main currents in psychological thought. 6th edition, New Jersey, USA: Pearson Education, 2004. reference to William of Ockham on page 95 (Ockham's razor) (all other things being equal, take the simplest explanation) is a natural for ensembles of ANNs, and applies very well to the DNA-ANN concept:
 - use algorithms/ theories where appropriate (constant, linear, gentle non-linear, strong non-linear, chaotic, discontinuous);
 - extend to the next level of complexity when necessary; and
 - step up into complexity for the purposes of generalization to cover a greater portion of the problem domain.

C. Rapid training, learning and evolution

Many researchers are looking for orders of magnitude faster training while still generating accurate/ robust solutions. Local learning methods have contributed to this (Kasabov ECOS), and recent examples include "Extreme Learning Machines" (Guang-Bin Huang et.al. G-B Huang, Q-Y Zhiu, C-K Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks" Proceedings of IJCNN 2004, International Joint Conference on Neural Networks. Budapest, paper# 1220, pp 985-990, 25-29 July 2004.D. Wang, G-B Huang, "Protein Sequence Classification Using Extreme Learning Machine" Proceedings of IJCNN 2005, International Joint Conference

on Neural Networks. Montreal, paper#1257 pp ???, 31 July – 4 August 2005.) and "Echo State Networks" (Jaeger H. Jaeger, "Reservoir riddles: suggestions for Echo State Network research", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1767, pp 1460-1462, 31 July – 4 August 2005. and Prokhorov D. Prokhorov, "Echo State Networks: Appeal and Challenges", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1002, pp 1463-1466, 31 July – 4 August 2005.) as described in Section II.C.3 above. Of course fast specialized modules can always be combined with powerful general learning techniques for unusual/ difficult problems, where more time will be required to learn/ evolve.

- The clear preference is for training time to almost disappear if quality solutions can still be generated – in other works we should be targeting one-shot learning where that is possible!! And how do we know when this should be possible?
- Rapid reconfiguration of pods (ensembles of ANN modules) – a capability to rapidly rearrange DNA-ANN pods along "high likelihood solutions" arrangements would be desirable. In the limit, dynamic structures would allow pods to switch and evolve in real time! If that "easy approach" doesn't work then longer term more exhaustive evolution may be required.
- Data (delivery mechanism) – DNA-ANNs offer a unique opportunity – that code segments can identify data (DNA or RNA code keys physically bring data and destination together!), and "data" can drive code and architecture. A mental picture is that of RNA transcription, where chunks of micro-RNA and introns are shed off, and can potentially work in parallel as interacting code and data. Perhaps there is something that can be done with a "computational soup" (This is reminiscent of Stuart Kauffman's work S. Kauffman, *At home in the universe*. New York: Oxford University Press, 1995, however, whereas Kauffman spoke of the benefits of being at the edge of chaos, current research is showing the benefits of being fully chaotic).
- Recurrent Neural Networks (RNNs) are especially difficult to train, and benefits here may be of greatest value.

D. Quality of solutions/ models

Hopefully, DNA-ANNs will provide the robustness, reliability, and accuracy of solutions that have evolved over a broad range of real-world situations encountered over evolutionary times ("tried and true"). This bears some relationship to point B above as it leads into "bigger picture", better solutions that simultaneously address other needs, threats and opportunities that may be associated with the immediate objective in mind. However, achieving this with pods will likely be harder than for most of the specialized ANNs currently used.

- Plasticity versus stability - This classical challenge for ANNs is to retain what is learned, while training on new data. Many ANNs have addressed this problem, but in a sense DNA-ANNs may be the penultimate solution, if

the learning is closely related to pre-specified "content"!

- Resistance to the effects of damage and disease - DNA-ANNs should have regenerative capabilities, at least for their pre-specified components. And while ANNs have always been recognized for their amazing robustness, DNA-ANNs could take this to quite a different level. Lamarckian heredity makes this even more interesting! If cortical neurons can be "reborn", that clearly would provide an AMAZING plasticity of the brain! However epigenetic changes won't be sufficient unless they can be communicated from "experienced" to new neurons. See the next Section III for further discussion of Lamarckian heredity.

E. Resource utilization

Because of their "hybrid symbolic/ connectionist" nature, DNA-ANNs may offer some unique opportunities for the efficient and effective utilization of "brain resources".

- Perhaps biological neurons cannot fire too long due to energetic, ion, neurotransmitter and other physiological demands, with the consequences that:
 - they must rotate tasks to have competitive system performance, and to minimize downtime; and
 - the "spacially wandering" processing of tasks would lead to qualitatively different computational capabilities.
- Share/ swap/ rotate tasks – Taking the previous insight further, it's easy to imagine the advantage of "migrating computations" in a biological NN. But the capability of doing "wandering computations" will itself give rise to new capabilities:
 - This will help to balance workload across the brain, allowing a restoration of ions, energy, and neurotransmitters, and a period of rest for neurons.
 - Individual neurons don't have to operate at maximum power for long periods of time even though the problem or its solution may actually require that a constant high output be provided.
 - There could result a tremendous robustness of the brain's capabilities as these will be relatively insensitive to local neuronal damage, disease and dysfunction.
- Functional overloading – having multiple functionalities share the same NN module will clearly yield an efficiency of resources. But as hinted in sub-sections above it may possibly lead to some kind of "morphing" evolutionary process towards more powerful learning, architectures and abstraction.
 - Functional overloading – reminds one of the effects of neuro-modulators (Kenji Doya K. Doya, P. Dayan, M. Hasselmo, (guest editors) "2002 Special Issue: Computational Models of Neuromodulation" *Neural Networks*, Vol 15 Nos 4-6, June-July 2002. K. Doja, "Metalearning and neuromodulation" in Doja, Dayan, Hasselmo, *Neural Networks*, 2002, pp495-506 The author provides a simple model of interactions for dopamine, serotonin, acetylcholine and noradrenaline.) and gene networks (Kasabov N. Kasabov, L. Benuskova, S.G. Wysoski,

- "Computational neurogenetic modelling: Gene networks within neural networks" Proceedings of the International Joint Conference on Neural Networks, Budapest, IEEE, N. Kasabov, L. Benuskova, S.G. Wysoski, "A Computational Neurogenetic Model of a Spiking Neuron", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1768, pp 446-451, 31 July – 4 August 2005.), including a capacity for meta learning.
- Beyond merely "switching" between functionalities, the same NN module could simultaneously be part of many different "models" and processes.
 - The ability to handle variable object inputs and problem types should be a characteristic that arises from functional overloading.
 - DNA code overloading - This is not the same as functional overloading for an ANN module as discussed in the previous point, but instead refers to the potential re-use of DNA coding for different purposes in different neurons, or for different computational processes in the same neuron.
 - Given that there are 3.2 Gbp (base pairs) of amino acids in the human DNA – for any "program" of length x bp, what is the probability of finding the code on an individual's genome (assuming that there is some way of accessing all code on the genome). Clearly this relates to the length of x , given that the probability is 1.0 for one "legitimate" bp, and pretty well zero for $x =$ length of the individual's entire genome.
 - For code of "moderate length" (say 100,000 bps), there is probably a reasonable chance of finding code very close to that being sought, but that may have a few errors. So could "errors" result simply because of the convenience of using an approximate sequence of code, rather than from any change in the code or its expression?
 - DNA code overloading provides for a way of "re-interpreting DNA" for a completely different application.

F. Non-linear dynamical systems modelling and control

This author feels that RNNs, and especially those for control and Approximate Dynamic Programming, are a particularly important class of ANNs for which real advantages of DNA-RNNs must eventually be shown. Merely using DNA-ANNs as a pre-specification for ANNs is fine, but it doesn't give much more than simply what has already happened – many researchers over time creating highly specific and effective ANNs through a variety of approaches, and that activity will certainly continue. But the power of RNNs may be currently limited by the difficulties with training, evolving and applying them, and if DNA-RNNs, together with many other approaches to this challenge, could make substantive contributions in this area, then perhaps not only will it help with the implementation of current RNNs, but it may also contribute to the development of much more powerful

RNNs. This is where the term "pod" is helpful, as defined in point II.C.7 above.

- Dynamic transitions and performance in non-stationary environments – Given that RNNs and their ensembles are characterised by recurrent feedback, how can one rapidly evolve systems that are robust with respect to non-stationary environments where changes may cascade in an unstable fashion through the RNN? DNA-RNN pods offer an opportunity to implement "self-muting" or "self-stabilizing" mechanisms to reduce these problems, or conversely, to promote instability when that is required. Examples of issues in this area include:
 - Dynamic transitions during learning/evolution and control actions - Current RNNs already learn how to do this quite well (as previously cited Puskorius and Feldkamp G.V. Puskorius, L.A. Feldkamp, "Decoupled extended Kalman filter training of feedforward layered networks", International Joint Conference on Neural Networks. Seattle, vol. 1, pp771-777, 1991., G.V. Puskorius, L.A. Feldkamp, "Extensions and enhancements of decoupled extended Kalman filter training", International Conference on Neural Networks, Houston, vol. 3, pp1879-1883, 1997., Wan and van der Merwe E.A. Wan and R. van der Merwe, "Chap 5: Dual Extended Kalman Filter Methods" and "Chap. 7: The Unscented Kalman Filter", in S. Haykin, (editor) Kalman Filtering and Neural Networks, Wiley Publishing., 2001), plus Kozma and Meyers R. Kozma, M. Myers, "Analysis of Phase Transitions in KIV with Amygdala During Simulated Navigation Control", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper #1675, pp126-130, 31 July – 4 August 2005.). Naturally, it would be preferable that cycling through "pods" would be possible in real time.
 - Searching problem/solution state space – It may be desirable to "destabilize" ANNs to cycle through possible approaches to solving problems. An example might be chaotic searches through weight-space as a means of "parameter adjustment, or cycling through many "pods" of a very diverse set of ANNs, wherein each of these ensemble-architectures is effective for different classes of problems. And if there isn't an easy solution on hand, then "strategically" building new "pre-configured ensemble-architectures" should be a capability of DNA-RNNs. This provides a "strategic" capability for DNA-RNNs, in that completely new approaches/ techniques/ concepts would arise, as might be appropriate for Approximate Dynamic Programming (ADP). Evolutionary theory, particle swarms, and chaos theory are examples of techniques for implementing searches that are at least partially randomized, and these can be used in combination with each other or in conjunction with gradient-based methods like BPTT.
 - Variable object inputs/ problem types - as with

functional overloading, it would be a useful capability if DNA-ANNs were designed to be able to use a different number of inputs, a variety of input "types" and missing data. Actually, that capability may be essential in many real-world situations, and its the sort of thing that is difficult to do in a general sense starting from a "blank slate".

- Instability-breaking – Is it possible to "break through" the Lyapunov and Slotine J.-J.E. Slotine, W. Lohmiller, "Modularity, evolution, and the binding problem: a view from stability theory", Neural Networks, vol 14, 2001, pp137-145. criteria for stability constraints simply because the "behavior" of a DNA-RNN isn't "fixed"? Can DNA-RNNs allow non-convex S matrices to provide stable short-term dynamics that enhance learning speeds and response times while ensuring a stable, robust, response?
- Ultimately, to use software programming analogies, perhaps what we are looking for are "strategies, design patterns, evolutionary principles, system engineering principles" to copy, create, and evolve DNA-RNN pods, and that these strategies would be provided as part of the "DNA code". What are the principles and means of more effectively evolving infrastructure and processes for meta-or-abstract level data / functions/ and processes?

IV. THINGS TO PONDER, AND NOT JUST FOR DNA-ANNS

A. Symbolism

What about an "aptitude" for symbolism by DNA-RNNs? Section III.F on non-linear dynamics and control and the description of "pods" in Section II.C.7 push this question further than its introduction in Section III.B "Higher levels of abstraction". Already, one might interpret "Context Discerning" properties of some RNNs (Santiago R.A. Santiago, "Context discerning multifunction networks: reformulating fixed weight neural networks", Proceedings of IJCNN 2004, International Joint Conference on Neural Networks. Budapest, paper# 1743, pp 189-194, 25-29 July 2004.) as being potentially a more powerful form of symbolism than some of the results of neuro-fuzzy or fuzzy-neural systems. The latter sometimes end up with great numbers of rules, but not enough abstract (symbolic) meaning. Perhaps it is best not to have an exact or fuzzy description, but a rough, incomplete, and approximate symbolic answer that lends itself well to projecting or extrapolating to different situations, and to the "abductive reasoning" that was described earlier.

In the simple case, pods could simply be descriptors for the collection of "contexts" expressed by some RNNs. As one becomes more familiar with the capabilities and applications of pods, logic or meaning arises from one's experience in applying the pods.

"Logic is an emergent property" - This expression has been adapted from a saying for linguistics: "semantics is an emergent property". In keeping with the earlier points about symbolism, an "emergent logic" would help with explain-ability at a high level and the extrapolation of the capability of pods to new challenges, while there is still "computational truth" at the lower connectionist level. The hope is that the emergent logic/ symbolism would give rise to:

- clean abstractions of emergent, approximate logic versus the incomprehensible "reality in excessive detail" of connectionist systems that does not give people an intuitive understanding of a system;
- predictability and robustness of solutions that (like fuzzy systems in some respects) "surf over bumps and details" that might distract the training of very general, amorphous ANNs;
- greater functional/ mapping specificity – perhaps the "emergent logic" of DNA-ANN pods will give them an advantage over general, "blank slate" ANNs for new types of problems that have some resemblance to those that the pods evolved with. This hope for "emergent logic" is similar to the general hope that DNA-ANNS will have an advantage over normal learning/ training approaches starting from "blank slate" ANNs, at least for the classes of problems that the DNA-ANNS have evolved for;
- ease of combining pods; and
- significant advances in the conceptual power of ANNs (as discussed earlier for pods in Section III.B).

Category theory is slowly showing some results for Adaptive Resonance Theory (ART) networks (Healy, Caudell et.al. M.J. Healy, R.D. Olinger, R.J. Young, T.P. Caudell, K.W. Larson, "Modification of the ART-1 Architecture Based on Category Theoretic Design Principles", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1767, pp 1460-1462, 31 July – 4 August 2005.), and although it may be too early to apply it to DNA-ANNS, it may be another, very distinct, approach to attaining powerful symbolism through connectionist architectures.

B. "Multiple personalities" (DNA-ANNS, not real people!)

Consider a pod and its modular construction and multi-functional capabilities as discussed in Section III.A. If these capabilities are "organized" according to classes of problems or challenges, then one would want to "flip" from one context to another according to the current problem-solving environment and in order to explore different combinations to solve a problem. This would involve changes at every level on the pod(s), involving many changes in weights, connections, neuron behavior (transfer function) etc., and these changes might best be accomplished by a wholesale "flip" of the pod state. Borrowing terms from psychology, let's call these different states "behaviors" if they are relatively modest changes, and "personalities" if the states are radically different (for example, massive reconfiguration between different pods).

What we hope, as described in the last Section III.B, is to develop ways to effectively decompose and modularize problems – and ways to flip quickly between behaviors and, if necessary, personalities, but how does one reconstruct novel, complex solutions? Can this be done quickly? Perhaps this is another question that might lead into Classical AI or Design Patterns and other conventional software methodologies, but my feeling is that connectionist approaches will have much to offer before passing purely to the symbolic level, and work on ANN hierarchies may help develop tools for this (refer back to Section II.A.6 for a similar issue). Perhaps a very compact set of tools will evolve from that, as with the general power of back-propagation and Hebbian learning for ANN training.

However they arise, once "collectivities of behavior" have been established which are appropriate for a wide range of environmental circumstances, it may be relatively easy to "collapse" into the appropriate behavior or personality. This process brings to mind Walter Freeman's hypothesis of stochastic chaos in the brain W.J. Freeman, "Stochastic Chaos Versus Deterministic Chaos: A Case for Analog Versus Digital Embodiment of Devices for Pattern Recognition", Proceedings of IJCNN 1999, International Joint Conference on Neural Networks. Washington, DC, paper#3012, 10-16 July 1999., and jumping rapidly from one attractor basin to another.

But in thinking back to ANNs in general, but especially the comments of Santiago and Lendaris in Section II.C.5, there is certainly no reason to believe that the behaviors and personalities have to be discrete states, although some may be so. And looking back to Section III.A one should expect that an apparently infinite set of behaviors and personalities would arise from "small-world universal function approximation" capability of a surprisingly small set of "pods". Pinker also made comments along these lines (S. Pinker, *The blank slate*, New York: Penguin Books, 2003).

Pinker's book provides a much broader and deeper analysis of many of the concepts discussed in this paper. (p36,39,40). Furthermore, this phenomenal flexibility and plasticity doesn't prevent the core genetic set of behaviors and personalities from re-emerging, nor any of the learned/evolved intermediate (possibly Lamarckian) forms.

C. Lamarckian heredity and multiple inheritance

Lamarckian heredity is understood here to mean changes in DNA code or its "express-ability" which occur during an individual's life through learning or adaptation, and which may or may not be passed on to the next generation. This contrasts with the normal biological process of Mendelian heredity, where DNA coding or its expression are assumed to change only at conception (and from mutations, or errors during normal cell division). Both are still components of Darwin's general theory evolution – as with punctuated evolution etc.. Further thoughts on this topic include:

- In non-biological domains Lamarckian heredity has always been there, but goes by different names (eg

education, organizational change/ development and management theories and fashions). In each of these areas, lessons learned ARE passed to the next generation of employee, spin-off companies, other countries looking for a constitutional framework etc. Lamarckian-style co-evolution is particularly evident in competitive marketplaces. ;

- Does Lamarckian heredity apply biologically? Epigenetic changes for the functioning of the mind during an individual's life are discussed in several references (Marcus G. Marcus, *The birth of the mind: how a tiny number of genes creates the complexities of human thought*, New York: Basic Books, 2004. This book (along with Pinker's "Blank slate") is an essential read, and its concepts and Marcus' current work are a basis for the current paper., Meaney I.C.G Weaver, N. Cervoni, F.A. Champagne, A.C. D'Alessio, S. Sharma, J.R. Seckl, S. Dymov, M. Szyf, M.J. Meaney, "Epigenetic programming by maternal behavior". *Nature Neuroscience* vol 7, no8, Aug. 2004, pp847-854M.J. Meaney, M. Szyf, "Maternal care as a model for experience-dependent chromatin plasticity?". *Trends in Neurosciences*, Vol.28 No.9 September 2005.), but there still does not appear to be adequate direct evidence for multi-generational changes. It's still very early though, and the toolsets for studying this subject are improving rapidly.
- However, it's interesting to consider constraints with the genome. Mitochondrial DNA is completely separate from nuclear DNA, and it doesn't change through sexual mixing of coding. Perhaps here the life-threatening constraints (and requirement for optimal or competitive use of energy) are too severe to allow for much diversity. Somewhat less severe, but still high constraints on diversity might apply to critical functions (organs, the heart etc). But many features aren't particularly critical within certain bounds (height, weight), and cosmetic features are only loosely constrained, and indeed might be purposefully variable as indicated by much higher mutation rates for some of these features in some organisms. Pushing this to the ultimate extreme, it would seem that abstract thinking would be extremely free to mutate, crossover, and to undergo Lamarckian hereditary changes, as redundancy could handle dysfunctional "pods", and great diversity would have a good chance of still being useful in some manner.
- Lamarckian advantage - Consider for a moment the impact of Lamarckian heredity for the brain and thought. Passing on even small doses of learned or evolved pods (whether epigenetic or DNA code, whether data, structure or process), could be of incredible advantage to a population. This applies especially to where the new mental capability is highly dependent on pod structure (architecture, weights etc) that enable certain types of learning or mental capabilities much more than if an individual didn't possess those capabilities.
- Explosion of mental complexity - Think back to John Mattick's comments regarding an explosion of complexity in the physiology of life, perhaps due to non-

protein-coding DNA (npcDNA). Did humans go through something similar 40 to 100 kyear ago with language and other intellectual and social capabilities? Has this happened several times in history (over the last 7 to 8,000 years) in different regions?

- Perhaps by taking a Lamarckian perspective, we can push the frontiers of evolutionary theory (and not just Lamarckian heredity concepts).
- A pod may become a "member" of many systems/classes at the same time, sort of like multiple inheritance, but this is more like a multiple allegiance, or floating allegiance.
- Lamarckian heredity allows great flexibility, but in the end may not differ much from learning/ evolving ANNs unless the inheritance can be easily passed to other DNA-ANNs. Migrating pods from one DNA-ANN to another might be very difficult in general.

D. Co-expressed "DNA" – data, function, and process

Section III.E introduced the potential capability of "DNA code overloading" for DNA-ANNs, wherein the same "DNA" code could have different functionality for different neurons, functions, processes (and by extension for different behaviors and personalities). We can take this idea further using by "visually" considering the opportunities presented when "RNA" and "micro-RNA" segments are present for short times in the nucleus in normal biology. The central theme of this paper is to metaphorically assume that "junk DNA" can have a more direct "programming" role than merely being "assembly language programming" for proteins or providing the immediate regulation of protein production. Moreover, that theme when extended to DNA-ANNs also assumes that the information-processing activities can also more or less directly influence "DNA" expression – in other words it may launch other "DNA" code or access data of sorts.

So what are the opportunities presented when "DNA code" (RNA in the case of biology) is floating around during the expression of a code segment, and different pieces of that "floating code" could simultaneously be related to data, function, or process, and it is clearly correlated with the current DNA being expressed? In a biological sense one would assume that the bits are broken down to be re-used for regular cell biology. But if they are present long enough to be useful, the opportunities are intriguing.

Massive parallelism and recurrency are obvious potential properties of such a "combinatorial soup", but there is more to it. In a sense, any "bits" could potentially be interpreted EITHER as data, function or processes, or some programming concepts that is a blend of them all (and beyond Object Oriented Programming (OOPS)), and where all bits have the potential for interacting with one another.

There could be, for example, a data/code delivery mechanism wherein code segments, which are tags or code keys in and of themselves (by whole or by part), can "diffusively" seek other tags, creating self-assembly code.

E. Mindcode

This subsection is pure speculation and fantasy, but I think that it is a useful fantasy to drive lines of investigation and to maintain an awareness of the types of results that we should be looking for with modern genetics and the brain, as it is easier to find something when you are looking for it, or for something like it.

"Given that computer code is used to program computers, then mindcode..."

The perspective here isn't to "program" a child/adult brain by some external means, but rather to be able to interpret "junk DNA" coding (and other sources of coding such as epigenetics) that may define the basis of our brains from conception. What might such code tell us about ourselves and our history that is different from current psychology, sociology, anthropology, management theory, economics?

How might one read such mindcode? Will it clarify concepts that we do not yet understand? Will it lead us to new concepts that we don't yet even know exist? Or will it be impenetrable to a long time into the future?

It is NOT assumed that we will find a common "language" for MindCode, but rather that it may vary radically according to neuron types, connections, functionality and brain regions. Possibly, the active language might change from one instant in time to the next. Perhaps in some respects some of the MindCode will possess characteristics of human language (obviously coding for language will!) in its manner of operation more than present-day computer coding. In other words, as with semantics the meaning of segments of "MindCode" may not be unique or stable. While there isn't a basis for believing that MindCode should be this "fluid", given the continuous surprises we discover in the brain, it is not safe to be too comfortable that our computer programming metaphor will be anything close to the capabilities and power of what is really going on. More likely, the metaphor will be a temporary crutch, and the crutch will be quickly replaced based on what we learn from the brain, and so on iteratively.

Given that a great deal has already been done with linguistics, perhaps this is a starting point for beginning to understanding "Mindcode", if it exists, and if it can be found.

F. The dangers of metaphorical or abductive reasoning

There is a heavy application of abductive (or metaphorical) reasoning in this paper in going from biology to DNA-ANNs and visa versa. It is important to keep in mind that abductive (or metaphorical) reasoning can be misleading – saying that something is "like" something else can greatly help initial understanding and the selection of toolsets, but even subtle differences between a subject and its metaphor can mean that we shouldn't have too much confidence in the metaphor. That is certainly the case in this paper, as metaphors for programming, neurology, and genetics are liberally used.

Moreover, as stated immediately above it is likely that what is really going on in the brain is far ahead of computer science concepts and other or new areas of research, and that radical changes in thinking will occur more or less regularly for a long time to come. This process will be an inspiration for ongoing revolution in the humanities. That certainly implies that the concepts in this paper are also at best only a stopover in conceptual terms. We have barely even started yet.

V. CONCLUSION

The chief potential advantage of DNA-ANNs may be the ease and rapidity of evolving and training robust "pods" (genetically pre-specified core of DNA-ANN ensembles, and ensembles of ensembles, with a capability for further learning and evolution) for very diverse and non-stationary environments. There may be an especially important role for DNA-RNNs in controls and Approximate Dynamic Programming. However, many of these capabilities could easily be part of an RNN design package for an "RNN ensemble design engineer", so it's not clear if the incorporation of "DNA" into ANNs will really add much to the "hand-crafted evolution" that has always occurred with human scientific and engineering endeavors.

Clearly, approaches to specifying DNA-ANN "pods" are the second priority, drawing from a rich population of existing ANNs (general and highly specialized), and creating/evolving new compact pods and means for their integration. To be really useful, means are required of evolving large-scale DNA-ANN pods over a very large and diverse population of benchmark problems.

There is also a critical need to better understand the meaning of, and to find appropriate tools for, evolving and training pods at higher and higher levels of abstraction (meta-levels) – one can't help thinking that at some point this transitions into the symbolics of Artificial Intelligence (AI).

Finally, the overall vision of this paper is on the retention of a vast set of DNA-ANN tools and methods because reliable and robust approaches to highly varied problems will likely require a rich environment of evolved, proven tools, very much as implied by Minsky's quote in Section II.A.2. The issue becomes not one of finding a "grand unified field theory" of problem solving at this very early stage of the development of ANNs, but to evolve effective ways of coordinating a variety of these DNA-ANN methodologies to work together at different levels of abstraction.

ACKNOWLEDGEMENTS

This paper does not present any radically new concepts, but instead presents the established concepts of others in a specific, stylistic manner. The author would like to thank Nik Kasabov, Director of KEDRI at the Otago College of the Auckland University of Technology, who initiated discussions and work in this area, and Gary Marcus of New

York University for several references of particular importance to this topic, including his book "The birth of the mind". John Mattick, Steven Pinker, Danil Prokhorov, Lee Feldkamp, Guang-Bin Huang, and Herbert de Jaeger, Steven Grossberg, and John Taylor were also important influences, and the references list many others. Thanks to John Mielke for the Michael Meaney references.

The views expressed in this paper are personal and speculative. They are in no way related to the research, policies, viewpoints, and programs of my current employer, the Canadian federal government department "Natural Resources Canada". The author is unaware of any work underway or planned in this area within the Canadian federal government at this time.

REFERENCES

- [1] J.A. Anderson, E. Rosenfeld, (editors) *Talking nets: An oral history of neural networks*. Cambridge, MA USA: Bradford Books, 1998. This book provides an excellent context to early ANN research. This includes several references to the work of McCulloch&Pitts (eg W. McCulloch, W. Pitts, "A logical calculus of ideas immanent in nervous activity", 1943, reprinted in Neurocomputing.) and many other early researchers and how their thinking developed.
- [2] P.J. Werbos, *The roots of backpropagation: from ordered derivatives to neural networks and political forecasting*, New York: John Wiley & Sons:1994. The book includes Werbos' Harvard PhD thesis from 1974.
- [3] D. Parker, "Learning logic", *Invention report S81-64, File 1*, Office of Technology Licensing, Stanford University, 1982.
- [4] Y. LeCun, "Une procedure d'apprentissage pour reseau a seuil asymmetrique (a Learning Scheme for Asymmetric Threshold Networks)," in *Proceedings of Cognitiva 85*, Paris, pp. 599-604, 1985.
- [5] D.E. Rummelhart, G.E. Hinton, R.J. Williams, "Learning internal representations by error propagation", In J.L. McClelland, D.E. Rummelhart, & the PDP Research Group, *Parallel distributed processing, explorations in the microstructure of cognition*, Cambridge: MIT Press, 1986.
- [6] G.A. Carpenter, S. Grossberg, (editors) *Pattern recognition by self-organizing neural networks*, London: Bradford Book, MIT Press, 1991. This book provides an excellent history of the ART architecture, as well as work by other early leaders in the neural network field.
- [7] T. Kohonen, refer to J.A. Anderson, E. Rosenfeld, (editors) *Talking nets: An oral history of neural networks*. Cambridge, MA USA: Bradford Books, 1998. This book provides an excellent context to early ANN research. This includes several references to the work of McCulloch&Pitts (eg W. McCulloch, W. Pitts, "A logical calculus of ideas immanent in nervous activity", 1943, reprinted in Neurocomputing.) and many other early researchers and how their thinking developed. and G.A. Carpenter, S. Grossberg, (editors) *Pattern recognition by self-organizing neural networks*, London: Bradford Book, MIT Press, 1991. This book provides an excellent history of the ART architecture, as well as work by other early leaders in the neural network field. for the context in which Kohonen developed Self Organizing Maps (SOMs).
- [8] G. Marcus, *The birth of the mind: how a tiny number of genes creates the complexities of human thought*, New York: Basic Books, 2004. This book (along with Pinker's "Blank slate") is an essential read, and its concepts and Marcus' current work are a basis for the current paper.
- [9] J.S. Mattick, "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms" *BioEssays*, vol 25 pp930-939, Oct. 2003
- [10] N. Kasabov, L. Benuskova, S.G. Wysoski, "Computational neurogenetic modelling: Gene networks within neural networks" *Proceedings of the International Joint Conference on Neural Networks*, Budapest, IEEE, paper #1384, 2004

- [11] N. Kasabov, L. Benuskova, S.G. Wysocki, "A Computational Neurogenetic Model of a Spiking Neuron", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1768, pp 446-451, 31 July – 4 August 2005.
- [12] R. Xu, D.C. Wunsch II, "Gene Regulatory Networks Inference with Recurrent Neural Network Models" *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1728, pp286-291, 31 July – 4 August 2005.
- [13] R. Storjohann, G.F. Marcus, "NeuroGene: Integrated simulation of gene regulation, neural activity and neurodevelopment", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1646, pp428-433, 31 July – 4 August 2005.
- [14] W.N. Howell, Presentation, Workshop on Computational Neuro-Genetic Modelling, *International Joint Conference on Neural Networks 2006*, Montreal., presentation available from author.
- [15] B.F. Vanyushin, N.A. Tushmalova, L.V. Gus'kova, "Methylation of brain DNA as an index of participation of the genome in mechanisms of individual acquired memory", *Doklady Akademii Nauk SSSR*, vol219, pp742-744, November 1974. translated by Plenum Publishing, New York, 1975.
- [16] F. Crick, "Memory and molecular turnover", *Nature*, vol312, p101, 1984.
- [17] R. Bergin, "DNA as a basis for acquired long-term memory in neurons", *Medical Hypothesis*, vol 50, pp451-455, 1998.
- [18] S. Pena de Ortiz, Y.I. Arshavsky, "DNA recombination as a possible mechanism in declarative memory: a hypothesis", *Journal of Neuroscience Research*, vol63, pp72-81, 2001.
- [19] R.J. Greenspan, "RNA and memory: from feeding to localisation", *Current Biology*, vol13, ppR126-127, 18Feb. 2003.
- [20] A.H.F.M. Peters, D. Schubeler, "Methylation of histones: playing memory with DNA", *Current Opinion in Cell Biology*, vol17, pp230-238, 2005.
- [21] D.H. Adams, "Triplet code-independent programming of living systems organisation by DNA: the link with intelligence and memory", *Medical Hypothesis*, vol44, pp419-427, 1995.
- [22] J.-P. Thivierge, E. Balaban, "Faithful Retinotopic Maps with Local Optimum Rules, Axonal Competition, and Hebbian Learning", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1491, pp2760-2765, 31 July – 4 August 2005.
- [23] J.S. Mattick, "The hidden genetic program of complex organisms", *Scientific American*, pp60-67 Oct. 2004. See also <http://imbuq.edu.au/groups/mattick>
- [24] G. Ast, "The alternative genome" *Scientific American*, vol 292 issue 4, Apr. 2005, pp58-65
- [25] S.R. Eddy, "Non-coding RNA genes and the modern RNA world" *Nature Reviews Genetics*, vol 2, pp 919-929, Dec. 2001
- [26] W.W. Gibbs, "The unseen genome: Gems among the junk" *Scientific American*, vol 289 no 5, Nov. 2003 pp46-53
- [27] J. Kim, A. Krichevsky, Y. Grad, G.D. Hayes, K. Kosik, G.M. Church, G. Ruvkun, "Identification of many microRNAs that copurify with polyribosomes in mammalian neurons" *PNAS* vol 101 no 1, 06Jan04, pp360-365.
- [28] G. Storz, "An expanding universe of noncoding RNAs" *Science*, vol 296 pp 1260-1263, 17May02
- [29] D.B. Fogel, (editor) *Evolutionary computation: The fossil record: Selected readings on the history of evolutionary algorithms*. Piscataway, NJ: IEEE Press, 1998
- [30] M. Minsky and S. Papert, *Perceptrons*, MIT Press, Cambridge, MA, 1969.. Quoted by Emil M. Petriu, "Neural Networks: Basics" (tutorial) http://ebrains.la.asu.edu/~jennie/tutorial/Emil_M_Petriu.pdf
- [31] S. Pinker, *The language instinct: how the mind creates language*, New York: William Morrow & Company, 1994, Perennial Classics edition, 2000.
- [32] N. Chomsky, as quoted in Pinker 1994
- [33] N. Kasabov, abstract for WCCI06 special session on Computational Neuro-Genetic Modelling (CNGM): http://www.wcci2006.org/WCCI-Web_Special_Session.html]
- [34] S. Pinker, *The blank slate*, New York: Penguin Books, 2003. Pinker's book provides a much broader and deeper analysis of many of the concepts discussed in this paper.
- [35] I.C.G Weaver, N. Cervoni, F.A. Champagne, A.C. D'Alessio, S. Sharma, J.R. Seckl, S. Dymov, M. Szyf, M.J. Meaney, "Epigenetic programming by maternal behavior". *Nature Neuroscience* vol 7, no8, Aug. 2004, pp847-854
- [36] M.J. Meaney, M. Szyf, "Maternal care as a model for experience-dependent chromatin plasticity?". *Trends in Neurosciences*, Vol.28 No.9 September 2005.
- [37] N. Kasabov, Abstract for Special Session on Computational Neuro-Genetic Modelling, *International Joint Conference on Neural Networks 2005*, http://faculty.uwb.edu/ijcnn05/Special-Sessions/02_Kasabov_-_neuro-genetic_modelling.htm
- [38] K. Doya, P. Dayan, M. Hasselmo, (guest editors) "2002 Special Issue: Computational Models of Neuromodulation" *Neural Networks*, Vol 15 Nos 4-6, June-July 2002.
- [39] K. Doja, "Metalearning and neuromodulation" in Doja, Dayan, Hasselmo, *Neural Networks*, 2002, pp495-506 The author provides a simple model of interactions for dopamine, serotonin, acetylcholine and noradrenaline.
- [40] W.J. Freeman, "Stochastic Chaos Versus Deterministic Chaos: A Case for Analog Versus Digital Embodiment of Devices for Pattern Recognition", *Proceedings of IJCNN 1999, International Joint Conference on Neural Networks*. Washington, DC, paper#3012, 10-16 July 1999.
- [41] G. Szekely, M.L. Padgett, G. Dozier, T.A. Roppel, "Odor detection using pulse coupled neural networks", *Proceedings of IJCNN 1999, International Joint Conference on Neural Networks*. Washington, DC, paper #3041, 10-16 July 1999.
- [42] J.W. Brown, D. Bulloch, S. Grossberg, "How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades", *Neural Networks*, vol 17, 2004, pp471-510.
- [43] Editorial, "Modelling the brain: grey matter, blue matter", *Economist*, 11 June 2005, pp75-76. See also <http://bluebrainproject.epfl.ch/>
- [44] T.W. Berger, J.J. Granacki, V.Z. Marmarelis, A.R. Tanguay, S.A. Deadwyler, G.A. Gerhardt, "Implantable Biomimetic Electronics as Neural Prosthesis for Lost Cognitive Function", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper #1745 pp 3109-3114, 31 July – 4 August 2005.
- [45] M. Humayan, D. Yanai, R.J. Grenberg, J. Little, B.V. Mech, M. Mahadevappa, J.D. Weiland, G.Y. Fujii, E. DeJuan, "Clinical results with the model 1 IRP implant", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, Plenary presentation, 31 July – 4 August 2005.
- [46] R. Eckhorn, M. Eger, M. Wilms, T. Schanze, "Information transmission from a retina implant to the cat visual cortex", *Proceedings of IJCNN 2004, International Joint Conference on Neural Networks*. Budapest, paper #1334, 2004.
- [47] N.R. Taylor, J.G. Taylor, "Learning to Generate Temporal Sequences by Models of Frontal Lobes", *Proceedings of IJCNN 1999, International Joint Conference on Neural Networks*. Washington, DC, paper #563, 10-16 July 1999.
- [48] J. Taylor, "Neural Networks of the Brain: Their Analysis and Relation to Brain Images", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper #1410, pp1603-1608, 31 July – 4 August 2005.
- [49] R. Hecht-Nielsen, "A theory of thalamocortex", In *Computational models for neuroscience: human cortical information processing*, London: Springer-Verlag, pp 85-124, 2003.
- [50] S. Grossberg, G.A. Carpenter, B. Ersoy, "Brain categorization: learning, attention, and consciousness", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper #1287, pp1609-1614, 31 July – 4 August 2005.
- [51] W.L. Dunin-Barkowski, S.N. Markin, L.N. Podladchikova, D.C. Wunsch, "Climbing fibre Purkinje cell twins are found", *Proceedings of IJCNN 1999, International Joint Conference on Neural Networks*. Washington, DC, paper #444, 10-16 July 1999.
- [52] T.M. Martinez, K.J. Schulten, "A neural gas model learns neural topologies", *Artificial Neural Networks*, pp397-402, 1991.
- [53] P.A. Estevez, C.J. Figueroa, K. Saito, "Cross-entropy approach to data visualization based on the neural gas network", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper #1585, pp2724-2729, 31 July – 4 August 2005.
- [54] N. Kasabov, *Evolving connectionist systems: Methods and applications in bioinformatics, brain study, and intelligent machines*, London, UK: Springer-Verlag, 2003.
- [55] G-B Huang, Q-Y Zhiu, C-K Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks" *Proceedings of IJCNN 2004, International Joint Conference on Neural Networks*. Budapest, paper# 1220, pp 985-990, 25-29 July 2004.

- [56] D. Wang, G-B Huang, "Protein Sequence Classification Using Extreme Learning Machine" *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1257 pp ???, 31 July – 4 August 2005.
- [57] H. Jaeger, "Adaptive nonlinear system identification with echo state networks," *NIPS* 2002.
- [58] H. Jaeger, "Reservoir riddles: suggestions for Echo State Network research", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1767, pp 1460-1462, 31 July – 4 August 2005.
- [59] D. Prokhorov, "Echo State Networks: Appeal and Challenges", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1002, pp 1463-1466, 31 July – 4 August 2005.
- [60] G.A. Carpenter, S. Martens, "Self-Organizing Hierarchical Knowledge Discovery by an ARTMAP Information Fusion System", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1049, pp 452-456, 31 July – 4 August 2005.
- [61] E. Cuadros-Vargas, R.A.F. Romero, "Special Session: Constructive/Hierarchical Self-Organizing Maps", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, , 31 July – 4 August 2005. This session included 6 papers and there were many other similar papers during the conference as well:
- C. Hung, "A Constructive Self-Organising Model for A Non-Stationary Environment"
 - M. Dittenbach, A.s Rauber, "Constructive/Hierarchical Self-Organizing Maps"
 - J. Pakkanen, E. Oja, "Analyzing large data sets efficiently with The Evolving Tree"
 - E. Cuadros-Vargas, R.A.F. Romero, "The SAM-SOM and MAM-SOM Families"
 - A.T-W. Chan, J S-Y. Lee, "Anomaly Detection with the Growing Hierarchical Self Organising Map"
 - V. Baier, "Motion Processing and Prediction With a Hierarchical RSOM based Model"
- [62] P.J. Werbos, "Neurocontrol and supervised learning: an overview and evaluation", and "Approximate dynamic programming for real-time control and neural modelling". in D.A. White, D.A. Sofge (editors), *Handbook of intelligent control*. New York: Van Nostrand Reinhold, 1992.
- [63] D.V. Prokhorov, G.V. Puskorius, L.A. Feldkamp, "Dynamical neural networks for control", in J. Kolen, S. Kremer, (eds) *A field guide to dynamical recurrent networks*, IEEE Press, 2001, pp.257-289.
- [64] G.K. Venayagamoorthy "Dynamic optimization of a multimachine power system with a FACTS device using identification and control ObjectNets", *IAS2004*, 0-7803-8487-3/04/\$20.00 © 2004 IEEE
- [65] J. Si, A.G. Barto, W.B. Powell, D. Wunsch, (eds) *Handbook of learning and approximate dynamic programming*. New Jersey, USA: IEEE Press and Wiley-Interscience, 2004.
- [66] R. Santiago, G.G. Lendaris, "Reinforcement Learning and the Frame Problem", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1659, pp 2971-2976, 31 July – 4 August 2005.
- [67] T.S. Hussain, "Workshop: Achieving Functional Integration of Diverse Neural Models", *International Joint Conference on Neural Networks*. Montreal, 31 July – 4 August 2005. As this was a workshop, presentations were not published, but the author/titles are listed below, and presentations are available on the website: <http://openmap.bbn.com/~thussain/NeuralIntegration/index.html>
- M. de Kamps, "Large scale brain simulations are not a technical problem"
 - Iversen, N.K. Taylor, K. Brown, "Integrating Neural Network Strategies for Discrimination, Recognition and Clustering"
 - V. Eliashberg, "The ULC Project"
 - X. Gu, "Integrate Different Neural Models Using Double Channels of Pulse Coupled Neural Network"
 - L.A. Coward, "Practical Architecture Limits on Complex Learning Systems"
 - W.J. Puma-Villanueva, G.B. Bezerra, C.A. M. Lima, F.J. Von Zuben, "Improving Support Vector Clustering with Ensembles"
 - T.S. Hussain, "Explicit Learning Relationships within Neural Systems"
- [68] G.V. Puskorius, L.A. Feldkamp, "Decoupled extended Kalman filter training of feedforward layered networks", *International Joint Conference on Neural Networks*. Seattle, vol. 1, pp771-777, 1991.
- [69] G.V. Puskorius, L.A. Feldkamp, "Extensions and enhancements of decoupled extended Kalman filter training", *International Conference on Neural Networks*, Houston, vol. 3, pp1879-1883, 1997.
- [70] E.A. Wan and R. van der Merwe, "Chap 5: Dual Extended Kalman Filter Methods" and "Chap. 7: The Unscented Kalman Filter", in S. Haykin, (editor) *Kalman Filtering and Neural Networks*, Wiley Publishing., 2001
- [71] P. Isabelle, National Research Council of Canada, Ottawa, comment made 12 Oct. 2005 at Ottawa IEEE Computer Society sponsored event: NRC-IIT Technology Showcase": leading edge IBM human language translation software uses signal processing technology.
- [72] D. Erdogmus, K.E. Hild, J.C. Principe, "Independent components analysis Renyi's mutual information and Legendre density estimation", *Proceedings of IJCNN 2001, International Joint Conference on Neural Networks*. Washington, DC, paper #595_3, pp 2762-2767, 15-19 July 2001.
- [73] K-H. Jeong, J-W. Xu, J.C. Principe, "An Information Theoretic Approach to Adaptive System Training Using Unlabeled Data", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper #1719, pp191-196, 31 July – 4 August 2005.
- [74] D.B. Fogel, *Blondie 24: Playing at the edge of AI*. San Diego, USA: Academic Press, 2002. p301 for comments on learning and evolution
- [75] T.H. Leahey, *A history of psychology: Main currents in psychological thought*. 6th edition, New Jersey, USA: Pearson Education, 2004. reference to William of Ockham on page 95 (Ockham's razor)
- [76] E. Gamma, R. Helm, R. Johnson, J. Vlissides, *Design patterns: Elements of reusable object-oriented software*. Reading, MA: Addison-Wesley, 1995
- [77] M. Fowler, *Analysis patterns*. Menlo Park, CA: Addison-Wesley, 1997.
- [78] B.P. Douglass, *Real-time design patterns: Robust scalable architecture for real-time systems*. Boston, MA: Addison-Wesley, 2003.
- [79] S. Kauffman, *At home in the universe*. New York: Oxford University Press, 1995
- [80] L.A. Feldkamp, G.V. Puskorius, "Fixed-weight controller for multiple systems", *Proceedings of the 1997 International Conference on Neural Networks (ICNN'97)*. Houston, pp 773-778, 9-12 June 1997.
- [81] R.A. Santiago, "Context discerning multifunction networks: reformulating fixed weight neural networks", *Proceedings of IJCNN 2004, International Joint Conference on Neural Networks*. Budapest, paper# 1743, pp 189-194, 25-29 July 2004.
- [82] R. Kozma, M. Myers, "Analysis of Phase Transitions in KIV with Amygdala During Simulated Navigation Control", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper paper #1675, pp126-130, 31 July – 4 August 2005.
- [83] J.-J.E. Slotine, W. Lohmiller, "Modularity, evolution, and the binding problem: a view from stability theory", *Neural Networks*, vol 14, 2001, pp137-145.
- [84] M.J. Healy, R.D. Olinger, R.J. Young, T.P. Caudell, K.W. Larson, "Modification of the ART-1 Architecture Based on Category Theoretic Design Principles", *Proceedings of IJCNN 2005, International Joint Conference on Neural Networks*. Montreal, paper#1767, pp 1460-1462, 31 July – 4 August 2005.