UNIVERSITY OF MASSACHUSETTS AMHERST

May 4, 2020

Defense Technical Information Center
ATTN: DTIC-O
8725 John J. Kingman Road
Ft. Belvoir, VA 22060-6218

Dear DTIC Officer,

Enclosed, please find two copies of the Final Technical Report and the completed DD0882 form on the grant DARPA HR0011-16-1-0006 "Superior Artificial Intelligence".

We prepared the reports following the specifications. However, we experienced delay in the distribution in the past month, unfortunately, due to the virus constraints, for which I apologize.

Sincerely,

Robert Kozma, Principal Investigator
Co-Director, BINDS Lab
College of Information and Computer Sciences
University of Massachusetts Amherst
140 Governors Drive
Amherst, MA 01003
413-577-4282 office
rkozma@cs.umass.edu

Encl: 2 Final Technical Reports
2 Forms DD0882

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
4/21/2020

2. REPORT TYPE
Final Technical Report

3. DATES COVERED *(From - To)*
09/01/2016 - 12/31/2019

4. TITLE AND SUBTITLE
Superior Artificial Intelligence

5a. CONTRACT NUMBER
HR011-16-1-0006

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Robert Kozma
& members of the Biologically-Inspired Neural and Dynamical Systems Laboratory (BINDS)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Massachusetts Amherst
Office of Grants and Contracts Administration
Research Administration Building
70 Butterfield Terrace
Amherst, MA 01003-9242

8. PERFORMING ORGANIZATION REPORT NUMBER
GPID#: 24235

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Advanced Research Projects Agency (DARPA)
Contracts Management Office (CMO)
675 N. Randolph Street
Arlington, VA 22203-2114

10. SPONSOR/MONITOR'S ACRONYM(S)
DARPA/MTO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
HR0011623458

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
None.

14. ABSTRACT

The project goal is to substantially advance the state-of-art of artificial intelligence (AI). As a major approach to achieve this goal, we increase our understanding of the mechanisms that underlie biological intelligence. We avoid methods and systems that may appear intelligent, but are, in fact, hard coded and not intelligent. Many current systems give the appearance of intelligence; however, these systems cannot adapt to changing circumstances or truly interpret and make decisions based on input. Systems that can achieve flexible adaptation and decision-making represent the future of computer technology and are the target of this Superior AI project.

This work builds energy aware neurocomputers to solve problems that cannot be addressed by today's AI technology. Successful completion of this work allows going beyond the state-of-the-art AI, which is represented by Deep Learning (DL). In the overall problem setting of DL, resource constraints are often ignored, or have secondary role. DL typically requires huge amount of data/ time/ parameters/ energy/ computational power, which are not readily available in various scenarios. Target applications include rapid response to emergency situations based on incomplete and disparate information, supporting graceful degradation in the case of physical damage or resource constraints, and real time speech recognition in noisy and cluttered background. In spite of the drastic cuts in project budget by the Program Manager from Year 2, and his demand of eliminating some of the tasks, reorganize the others, and completely deleting the final year of this 4-year project, several breakthrough results have been accomplished during Y1-3, as planned.

15. SUBJECT TERMS
Artificial Intelligence, Deep Learning, Neuromorphic Computing, Power Consumption.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Carol Sprague, Director OGCA
Unclassified	Unclassified	Unclassified	None	31	**19b. TELEPHONE NUMBER** *(include area code)* 413-545-0698

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

REPORT OF INVENTIONS AND SUBCONTRACTS

(Pursuant to "Patent Rights" Contract Clause) (See Instructions on back)

Form Approved
OMB No. 9000-0095
Expires Jan 31, 2008

PLEASE DO NOT RETURN YOUR COMPLETED FORM TO THE ABOVE ORGANIZATION. RETURN COMPLETED FORM TO THE CONTRACTING OFFICER.

1.a. NAME OF CONTRACTOR/SUBCONTRACTOR	Robert Kozma, U. Mass, Amherst
c. CONTRACT NUMBER	HR0011-16-1-0006
b. ADDRESS (Include ZIP Code)	140 Governors Dr., Amherst, MA 01003
d. AWARD DATE (YYYYMMDD)	20160901
2.a. NAME OF GOVERNMENT PRIME CONTRACTOR	DARPA
c. CONTRACT NUMBER	HR0011-16-1-0006
b. ADDRESS (Include ZIP Code)	Defense Advanced Research Project Agency, Contracts Management Office, 675 N. Randolph St., Arlington, VA 22203
d. AWARD DATE (YYYYMMDD)	20160901
3. TYPE OF REPORT (X one)	a. INTERIM ☐ b. FINAL ☒
4. REPORTING PERIOD (YYYYMMDD)	a. FROM 20160901 b. TO 20191231

SECTION I - SUBJECT INVENTIONS

5. "SUBJECT INVENTIONS" REQUIRED TO BE REPORTED BY CONTRACTOR/SUBCONTRACTOR (If "None," so state)

NAME(S) OF INVENTOR(S) (Last, First, Middle Initial) a.	TITLE OF INVENTION(S) b.	DISCLOSURE NUMBER, PATENT APPLICATION SERIAL NUMBER OR PATENT NUMBER c.	ELECTION TO FILE PATENT APPLICATIONS (X) d. (1) UNITED STATES (a) YES	(1) UNITED STATES (b) NO	(2) FOREIGN (a) YES	(2) FOREIGN (b) NO	CONFIRMATORY INSTRUMENT OR ASSIGNMENT FORWARDED TO CONTRACTING OFFICER (X) e. (a) YES	(b) NO
None	None	None						

f. EMPLOYER OF INVENTOR(S) NOT EMPLOYED BY CONTRACTOR/SUBCONTRACTOR

(1) (a) NAME OF INVENTOR (Last, First, Middle Initial)	(2) (a) NAME OF INVENTOR (Last, First, Middle Initial)
(b) NAME OF EMPLOYER	(b) NAME OF EMPLOYER
(c) ADDRESS OF EMPLOYER (Include ZIP Code)	(c) ADDRESS OF EMPLOYER (Include ZIP Code)

g. ELECTED FOREIGN COUNTRIES IN WHICH A PATENT APPLICATION WILL BE FILED

(1) TITLE OF INVENTION	(2) FOREIGN COUNTRIES OF PATENT APPLICATION

SECTION II - SUBCONTRACTS *(Containing a "Patent Rights" clause)*

6. SUBCONTRACTS AWARDED BY CONTRACTOR/SUBCONTRACTOR (If "None," so state)

NAME OF SUBCONTRACTOR(S) a.	ADDRESS (Include ZIP Code) b.	SUBCONTRACT NUMBER(S) c.	FAR "PATENT RIGHTS" d. (1) CLAUSE NUMBER	(2) DATE (YYYYMM)	DESCRIPTION OF WORK TO BE PERFORMED UNDER SUBCONTRACT(S) e.	SUBCONTRACT DATES (YYYYMMDD) f. (1) AWARD	(2) ESTIMATED COMPLETION
None	None						

SECTION III - CERTIFICATION

7. CERTIFICATION OF REPORT BY CONTRACTOR/SUBCONTRACTOR (Not required if: (X as appropriate)) ☐ SMALL BUSINESS or ☒ NONPROFIT ORGANIZATION

I certify that the reporting party has procedures for prompt identification and timely disclosure of "Subject Inventions," that such procedures have been followed and that all "Subject Inventions" have been reported.

a. NAME OF AUTHORIZED CONTRACTOR/SUBCONTRACTOR OFFICIAL (Last, First, Middle Initial)	b. TITLE	c. SIGNATURE	d. DATE SIGNED
Robert Kozma	Visiting Professor	DocuSigned by: Robert Kozma 71E8D27DB6284B8...	4/21/2020

DD FORM 882 INSTRUCTIONS

GENERAL

This form is for use in submitting INTERIM and FINAL invention reports to the Contracting Officer and for use in reporting the award of subcontracts containing a "Patent Rights" clause. If the form does not afford sufficient space, multiple forms may be used or plain sheets of paper with proper identification of information by item number may be attached.

An INTERIM report is due at least every 12 months from the date of contract award and shall include (a) a listing of "Subject Inventions" during the reporting period, (b) a certification of compliance with required invention identification and disclosure procedures together with a certification of reporting of all "Subject Inventions," and (c) any required information not previously reported on subcontracts containing a "Patent Rights" clause.

A FINAL report is due within 6 months if contractor is a small business firm or domestic nonprofit organization and within 3 months for all others after completion of the contract work and shall include (a) a listing of all "Subject Inventions" required by the contract to be reported, and (b) any required information not previously reported on subcontracts awarded during the course of or under the contract and containing a "Patent Rights" clause.

While the form may be used for simultaneously reporting inventions and subcontracts, it may also be used for reporting, promptly after award, subcontracts containing a "Patent Rights" clause.

Dates shall be entered where indicated in certain items on this form and shall be entered in six or eight digit numbers in the order of year and month (YYYYMM) or year, month and day (YYYYMMDD). Example: April 2005 should be entered as 200504 and April 15, 2005 should be entered as 20050415.

1.a. Self-explanatory.

1.b. Self-explanatory.

1.c. If "same" as Item 2.c., so state.

1.d. Self-explanatory.

2.a. If "same" as Item 1.a., so state.

2.b. Self-explanatory.

2.c. Procurement Instrument Identification (PII) number of contract (DFARS 204.7003).

2.d. through 5.e. Self-explanatory.

5.f. The name and address of the employer of each inventor not employed by the contractor or subcontractor is needed because the Government's rights in a reported invention may not be determined solely by the terms of the "Patent Rights" clause in the contract.

Example 1: If an invention is made by a Government employee assigned to work with a contractor, the Government rights in such an invention will be determined under Executive Order 10096.

Example 2: If an invention is made under a contract by joint inventors and one of the inventors is a Government employee, the Government's rights in such an inventor's interest in the invention will also be determined under Executive Order 10096, except where the contractor is a small business or nonprofit organization, in which case the provisions of 35 U.S.C. 202(e) will apply.

5.g.(1) Self-explanatory.

5.g.(2) Self-explanatory with the exception that the contractor or subcontractor shall indicate, if known at the time of this report, whether applications will be filed under either the Patent Cooperation Treaty (PCT) or the European Patent Convention (EPC). If such is known, the letters PCT or EPC shall be entered after each listed country.

6.a. Self-explanatory.

6.b. Self-explanatory.

6.c. Self-explanatory.

6.d. Patent Rights Clauses are located in FAR 52.227.

6.e. Self-explanatory.

6.f. Self-explanatory.

7. Certification not required by small business firms and domestic nonprofit organizations.

7.a. through 7.d. Self-explanatory.

Sponsor: Defense Advanced Research Project Agency
Grant No.: DARPA-HR0011-16-l-0006
Title: Superior Artificial Intelligence (AI)

FINAL TECHNICAL REPORT

Awardee
Organization: College of Information and Computer Sciences
Robert Kozma, Principal Investigator
University of Massachusetts Amherst
140 Governors Drive
Amherst, MA 01003
rkozma@cs.umass.edu

Grant Officer's
Representative: Dr. Thomas Renz, Program Manager
Emerging Unconventional Processing Systems
Air Force Research Laboratory
Information Directorate AFRL/RITB
525 Brooks Road
Rome, NY 13441-4505
Phone: 315-330-3423
Thomas.Renz@us.af.mil

Table of Contents

1. Executive Summary

1.1 Project Objectives

The project goal is to substantially advance the state-of-art of artificial intelligence (AI). As a major approach to achieve this goal, we increase our understanding of the mechanisms that underlie biological intelligence. We avoid methods and systems that may appear intelligent, but are, in fact, hard coded and not intelligent. Many current systems give the appearance of intelligence; however, these systems cannot adapt to changing circumstances or truly interpret and make decisions based on input. Systems that can achieve flexible adaptation and decision-making represent the future of computer technology and are the target of this Superior AI project.

This work builds energy aware neurocomputers to solve problems that cannot be addressed by today's AI technology. Successful completion of this work allows going beyond the state-of-the-art AI, which is represented by Deep Learning (DL). In the overall problem setting of DL, resource constraints are often ignored, or have secondary role. DL typically requires huge amount of data/ time/ parameters/ energy/ computational power, which are not readily available in various scenarios. Target applications include rapid response to emergency situations based on incomplete and disparate information, supporting graceful degradation in the case of physical damage or resource constraints, and real time speech recognition in noisy and cluttered background.

In spite of the drastic cuts in project budget by the Program Manager from Year 2, and his demand of eliminating some of the tasks, reorganize the others, and completely deleting the final year of this 4-year project, several breakthrough results have been accomplished during Y1-3, as planned. He we summarize the areas with breakthrough achievements:

1.2 Main Breakthrough Achievements

1.2.A Developed a novel energy-aware oscillatory memory array (CAN)

Developed an energy aware computing paradigm, which is based on coupled CAN (Capillary-Astrocyte-Neuron) arrays motivated by brain structure and operation. CAN arrays serve as basic building blocks of novel dynamic memory devices.

Built a coupled array of CAN units. Produced collective behaviors such as synchronous activity and coordinated firing in specific narrow-band oscillatory frequencies. Modulated the frequency of collective oscillations based on the connection strength between oscillators and implemented a correlational learning rule. Developed metrics to evaluate energy efficiency of computing. In the effort to construct artificially intelligent devices that can perform perceptual, motor, and cognitive tasks at the level that matches human performance, one important measure of the efficiency is the power consumption in comparison to brain. This property of CAN arrays is a key factor to alleviate the exponentially increasing computational demand of today's computers, faced by reaching the limits of Moore's law, when targeting to achieve human-like performance.

1.2.B. Established a software framework (BindsNET) for oscillatory neural memory.

Created the BindsNET software platform, which provides a unified environment to construct hierarchical computing solutions using energy aware computational and dynamic memory principles, using PyTorch framework.

BindsNet is based on spiking neural computing units, which are connected according a hierarchy of architectures of increasing complexity, and trained using unsupervised, reinforcement, and supervised learning. The BindsNET repository has been released as an open source at GitHub. In addition to the source code, we made it easy to install using pip package manager for a range of practical applications. BindsNET has been extended to solve advanced AI machine learning tasks using convolutional networks, pattern matching, reinforcement learning, Q learning, and Deep Learning in multilayer architectures. Beyond standard image recognition tasks, pipelines have been developed in BindsNET to address challenges in dynamically changing environments.

1.2.C. Applied BindsNET to solve practically important AI tasks with results matching or even exceeding the performance of the best AI in the world.

Implemented local and global learning in spiking NNs using BindsNET platform. Showed that our results are at the same level or in some case superior to existing approaches.

Incorporated various local (unsupervised) and global (supervised) learning rules on the software platform and combined them to achieve improved performance. Converted traditional Deep Learning (DL) network into Spiking Neural Networks (SNN) and demonstrated that SNN can be used in reinforcement learning paradigm. Implemented the developed learning approaches in several AI problems, such as classification and computer games. Showed that our implementations achieve similar performance as state of art, in spite of significant simplifications in the simulation of the network components. Importantly, our SNN-based approach produced improved robustness compared to top-of-the line Deep Learning solutions.

1.3 Organizational Issues and Pitfalls

Year 1&2:

Our project started with four Tasks, which continued during the first 2 years. Specifically, in Year 1&2, our project had the following 4 Tasks:

1. ***Local Learning:*** Developing biologically inspired learning algorithms with local learning rules to significantly improve learning efficiency;
2. ***Neuromorphic Architecture & Dynamics:*** Modeling the brain's layered architecture and dynamics to improve state-of-the-art deep learning neural networks;
3. ***Energy Efficiency:*** Drastic improvement of energy efficiency of AI computing through modeling energy utilization in brains sustaining higher cognitive functions;
4. ***Hardware for Oscillator Computing:*** Building cortical array architectures in analog hardware with periodic/quasi-periodic dynamics for neural computation using less energy.

Year 3 (Year 4 has been eliminated):

During Year 2, our DARPA PM requested eliminating the hardware Task 4. Accordingly, in Year 3 the project had only 3 Tasks, and these tasks have been reorganized following the PM requests. In Year 4, our funding was completely cut, in spite of the fact that our demonstrated excellent results with respect to state-of-art AI. We objected the PM decision, which we considered unjustified and not warranted by the results. In Year 3, we had the following 3 tasks:

1. ***Energy Efficiency (CAN):*** Expanded the CAN model (units) to CAN memory arrays. CAN arrays serve as memory devices encoding input data into sequences of oscillatory patterns.
2. ***Software Platform (BindsNET):*** Combined oscillatory/spiking computing and various learning approaches/architectures, including global gradient descent, local plasticity, reinforcement, and transfer learning.
3. ***Solving AI Problems:*** Implemented the developed oscillatory machine learning/AI approach to demonstrate superior performance in several practical machine learning tasks.

1.4 Overall Achievement Statement

In spite of the pitfalls due to the unsupportive acts of the PM, the Superior AI project achieved its stated goals, within the constraints of the reduced 3-year span with 3 Tasks, instead of the originally planned 4-project with 4 main tasks.

2. Summary of Major Achievements

2.1 Task1: Developing Local Distributed Learning and Integrate with Different Learning Modules (Task 1 in Y1&2, Task 2 in Y3)

2.1.1 Task Statement

Biologically realistic spiking neural networks are often referred to as the third generation of neural network models since they directly hold the capability for the processing of time-varying input. Currently no commonly accepted, effective learning algorithm exists for spiking neurons. To achieve successful AI implementations, it is crucial to build learning rules for spiking neural network. Our results achieved at the earlier phase of the project, local learning using spike timing dependent plasticity (STDP) has demonstrated advantages in reduced memory requirement, massive parallel processing, and online learning with very rapid convergence. Due to the unsupervised nature of STDP, its accuracy is lower than supervised DL, thus combination with alternative learning schemes can be beneficial. We also need an efficient learning algorithm for learning in CAN arrays that scales well with large populations of neurons.

2.1.2 Developed Solutions and Results

a) Developed a convolutional neural network with patch connectivity trained by STDP. We implemented various learning methods to solve classification tasks. We combined unsupervised learning (STDP, Hebbian) with reinforcement components for improved accuracy. In addition to STDP rule, we developed various local (unsupervised) learning algorithms, which can be used in combination with global (supervised) and reinforcement learning, to solve practical machine learning tasks.
b) We improved on previous work combining Reinforcement Learning (RL) and local rules, which resulted in network instability and poor convergence. We improved previous algorithms updating a fraction of the connections at any given time, with decreasing percentages plastic over time.
c) STDP learning performs computations using data that are available locally at each node; therefore, it does not need to communicate with an external memory unit. This eliminates the requirement for external memories during the computing process and the computational unit acts as memory, thus ideally suited for memristor hardware implementations.
d) Our approach has the crucial advantage that it allows massive parallelization due to the local nature of learning, while Deep Learning must cache the results in layer-wise computation and wait for the outcomes of the previous step from each consequent layer.
e) Our approach converges up to 10-times faster to the specific accuracy level than deep learning, which often indicates a reasonable compromise between somewhat decreased accuracy of unsupervised learning at the price of increased speed of computational

convergence.

f) In order to link our approach to existing feature extraction algorithms, including multilayer convolutional Deep Learning, we implemented a topological learning rule according to which nearby nodes are more sensitive to similar input patterns. We obtained an ordered map of feature detector neurons, which can be used to achieve improved accuracy as an intermediate layer in a classifier CNN.

g) Going beyond image classification task, we implemented software framework to work efficiently in dynamic machine learning problems, such as game playing with ATARI and geo-spatial anomaly detection with Numenta testbed. Our spiking NN shows improved performance w.r.t. Deep Q-Learning (DQL) in the ATARI breakout benchmark task, when transferring DQL weights to our proposed stochastic spiking NN.

h) We developed the software platform BindsNET, which provides a unified framework to construct hierarchical computing solutions using energy aware computing and dynamic oscillatory memory principles. We released on github the BindsNET repository as an open source package. We are adding more features to BindsNET that allow the implementation of CAN arrays and relevant learning algorithms.

2.2 Task 2: Designing a Hierarchy of Network Structure and Dynamics Motivated by Brains (Task 2 in Y1&2, Task 1 in Y3)

2.2.1 Problem Statement

Today's digital computers encode data in fixed strings of digits, in which memory capacity is proportional to the available memory units. Dynamic encoding produce explosion of memory capacity, with potential exponential memory with respect to memory units.

2.2.2 Produced Research Outcomes

Develop more complex architectures with feedback loops and recurrent connections, which allow memory arrays with dynamic encoding, with applications for neuromorphic hardware. We identify lattice architectures with feedback connections producing long oscillations in their activity patterns, which can be used as memory patterns. Properly tuned recurrent networks with long oscillations can be used to build powerful dynamical memories. We integrate the results on lattice dynamics with the energy aware computing principles with CAN units.

a) Percolation processes have been implemented over graphs with 2D layers of excitatory nodes coupled with inhibitory nodes. The observed spatio-temporal dynamics have been evaluated both theoretically (proofs) and computationally (simulations). The obtained results serve as mathematical basis of a novel AI approach using sequential pattern-based computing

b) The model has been simplified to allow rigorous mathematical analysis, while it maintained key properties of the spatio-temporal dynamics. We derived parametric approaches to control the dynamics to produce sequences of spatial patterns.

c) We demonstrated that starting from a background oscillatory state, various inputs lead to oscillatory patterns specific to that input, and the system returns to the background oscillatory state once the input is removed. These results serve as a conceptual framework for building spiking neural network-based dynamical pattern-based memories.

d) We provide theoretical proof of the existence of a range of initialization probabilities (seeds) that maintain sustained oscillations with limit cycle or chaotic behavior as the function of reset level (m) and the type of update rule (k). We showed that the oscillations are robust to changes in initialization for most parameter values, thus exhibiting strong mixing and effective memory-less dynamics within the specific range of parameters, with the exception of larger reset values (m).

e) Computer simulations provided evidence of chaotic behavior with fractal boundary regions for higher values of the reset parameter (m). Extensive simulations produced a diagram with bifurcation and tri-furcation properties, evidencing highly intermingled limit cycle and chaotic attractor basins.

f) These results provide key insights into dynamical memory properties of the percolation model, when input patterns can be encoded into attractors with limit cycles of specific lengths, or chaotic attractors (practically infinite cycle length). We studied how to store and recall specific inputs in the attractors with multiple wings. The breakthrough aspects of these results for AI rest in the exponential memory capacity attractors with very long cycles (possibly chaos), and the instant recall of the stored patterns without the need for lengthy search and related convergence process required in more traditional, convergence-based (fixed point) memory devices.

g) We described the properties of very long cycles (VLC), up to a duration of 10^5 time steps. A key result is that VLCs can lead to the emergence of metastable spatial activity patters, which are sustained for several 100, or 1000 time steps, they evolve slowly, and ultimately dissolve to random background activity. Crucially, these metastable patterns emerge spontaneously and predictably, and clearly distinct from the fluctuating background. In the context of the dynamic computing principles of our project, these metastable patterns are the candidates of intermittent 'symbols,' that are the basis of the computation in our dynamical memory device.

2.3 Task3: Design and Analysis of Energy Efficient Computing Models (Task 3 in Y1&2, Task 1 in Y3)

2.3.1 Problem Statement

Studying efficient energy consumption in brain's intelligence helps us to design AI that is energy efficient. It is crucial for new AI to develop energy aware model for higher energy efficiency in AI hardware. Energy efficiency is a fundamental requirement to achieve superior AI beyond the state-of-art. We develop superior AI in which energy constraint leads to intelligence. Moreover, developing energy coupling models for brain neural activity is a key component to interpret brain imaging data by fMRI that are based on BOLD (blood oxygen level dependent).

2.3.2 Produced Research Outcomes

Developed the CAN model (Capillary-Astrocyte-Neuron) by modeling the brain energy use, in which spiking neuron populations are combined with metabolic equations in a unique way.

a) One salient point of the CAN model is the coupling of physiological processes at a broad range of time scales with many orders of magnitude differences: (1) fast neuron spiking at milliseconds scale; (2) metabolic processes at a time scale of fractions of a second; (3) and vascular effects of blood flow (energy supply) at much smaller time scales of 10s or more. CAN allows to study coupling between computing (spiking neurons) and energy unit (astrocytes).
b) We demonstrated that the amount of available energy modulates the oscillation frequency of interacting computational units (spiking neurons); more energy inflow (via vascular input) produces increased frequency of oscillations (spiking). Moreover, for a given level of energy supply, we can produce transitions between highly synchronous oscillations (resonances) and desynchronization effects by changing the coupling coefficient from the computational units (neurons) to the metabolic components (astrocytes) acting as a control parameter.
c) We analyzed in CAN arrays the coupling between fast computing (spiking neurons) and slower energy supply (astrocytes), as well as the frequency modulation and synchronization effects due to energy constraints. The introduced synchronization control algorithm is crucial in our approach, as synchronization-desynchronization transitions create the sequential memory patterns as key to the brain-inspired encoding and recall.
d) The demonstrated synchronization control algorithm is a crucial advantage of our approach, as synchronization-desynchronization transitions allow rapid response to dynamically changing external conditions in the environment and thus they manifest critical components of an emergency response system.
e) In the case of a single CAN oscillatory units, we studied the feasible range of parameters maintaining oscillations of energy variables in a stable range, at the same time producing quantifiable variation in the oscillatory frequency (computational capability) as the function of available energy resources. It is also important that the feasible parameters produce high synchrony of oscillations in the single CAN computational unit.
f) We developed the blueprint of coupled CAN oscillator arrays. The input data are encoded in the spatially distributed oscillations in these arrays coupled with modifiable links (synapses), following the STDP learning rule.
g) We implemented a learning algorithm between CAN units of a CAN array based on Hebbian correlation principles, when the connection between two units is strengthened (weakened) when the activities of these nodes are positively (negatively) correlated. The learning algorithm is operational, and it serves as a basis for encoding and recalling input data for actual machine learning tasks.
h) We analyzed in CAN arrays the coupling between fast computing (spiking neurons) and slower energy supply (astrocytes), as well as the frequency modulation and synchronization effects due to energy constraints. The introduced synchronization control algorithm is crucial in our approach, as synchronization-desynchronization transitions create the sequential memory patterns as key to the brain-inspired encoding and recall.

2.4 Task 4: Super-Turing Analog Oscillatory Hardware Platform (Task 4 in Y1&2)

2.4.1 Problem Statement

This task supports other parts of the project by producing a fabricated chip that implements oscillatory computing. We build an electrical circuit implementing the energy aware computational model. This includes the construction and experimental evaluation of analog hardware and computer simulations to study their capabilities and limitations for special purpose computation. We fine-tune the dynamics, including frequency of the oscillations and the value of the amplitude synchronization measure. These circuits can be built from analog components and run at lower power.

2.4.2 Produced Research Outcomes

Designed a chip for neuromorphic computing. The implementation includes simulations of architectures with feed-forward and feedback connections; the connections are fixed (non-adaptable) at this stage. The system receives inputs from a computer interface and the oscillatory pattern readouts are evaluated using a multi-channel analyzer.

a) ASIC chip has been designed, which functions as a Quasi Periodic Oscillatory Machine (QPOM) for pattern recognition. The spiking neural network-based Q-POM performs real-time computations on continuous streams of data. The information is stored as spatio-temporal patterns inside the recurrent neural network during training phase. We have simulated the music application in software, and the specifications for neuromorphic chip were finalized.
b) The architecture of chip was then finalized after considering different design alternatives with minimum area and power in mind. The RTL design and behavioral simulation of the chip is completed except for the interface module. Design libraries for chip have been procured from MOSIS. The memory compilers were obtained from ARM.
c) To facilitate experiments, we have built a 16-neuron system with patch panel for quick rewiring and circuit evaluation with a multichannel digital signal analyzer; see photo in full report at the end of this document. Though this seems like a small system, it is important to remember that a ring of 16 neurons, of the type we have been experimenting with, is able to support 143 stable memory states. In previous reports we showed an exponential curve of the number of states vs. the number of neurons.
d) Aiming at FPGA platform, we implemented a single Izhikevich neuron and two interacting neurons in Verilog. We also implemented the multiply and accumulate model for creating a network. Behavioral simulation was done with 2 neurons. In the future we will add the metabolic system with two equations to produce a very simple CAN model.
e) We planned to scale up the design to a network of 48 CAN units with neurons and corresponding metabolics. We plan to synthesize design and put binary in ZedBoard and explore analog implementation. These plans were not realized as Task 4 on hardware work has been terminated in Year 2.

3. Publications/presentations produced in the project

Book/Edited Volume (1):

1. R. Kozma. C. Alippi, Y. Choe, C. Morabito, (eds.) *"Artificial Intelligence in the Age of Neural Networks and Brain Computing,"* Academic Press, Cambridge, MA, USA; pp. 1-315. (2018).

Book Chapter (1):

2. R. Kozma (2018) Computers versus Brains: Game is Over or More to Come? In: *"Artificial Intelligence in the Age of Neural Networks and Brain Computing,"* R. Kozma. C. Alippi, Y. Choe, C. Morabito. (Eds.) ISBN 9780128154809, Imprint, Academic Press, USA.

Journal Papers (13):

3. Heck, D. H., Kozma, R., & Kay, L. M. (2019). The rhythm of memory: how breathing shapes memory function. *Journal of Neurophysiology*, *122*(2), 563-571. https://www.physiology.org/doi/full/10.1152/jn.00200.2019
4. Hazan, H., Saunders, D. J., Sanghavi, D. T., Siegelmann, H., & Kozma, R. (2019). Lattice Map Spiking Neural Networks (LM-SNNs) for Clustering and Classifying Image Data, *Annals of Mathematics and Artificial Intelligence*, pp. 1-24. *https://arxiv.org/pdf/1906.11826.pdf*
5. Patel, D., Hazan, H., Saunders, D. J., Siegelmann, H., & Kozma, R. (2019). Improved robustness of reinforcement learning policies upon conversion to spiking neuronal network platforms applied to ATARI games. *Neural Networks*, 120, 108-115. *https://arxiv.org/abs/1903.11012*
6. Saunders, D. J., Patel, D., Hazan, H., Siegelmann, H. T., & Kozma, R. (2019). Locally Connected Spiking Neural Networks for Unsupervised Feature Learning. *Neural Networks,* 119, pp. 332-340. *https://arxiv.org/abs/1904.06269*
7. Kozma, R. (2019). The key role of coexisting opposite dynamic behaviors in brains and in cognition. Comment on" Chimera states in neuronal networks: A review" by M. Perc et al., *Physics of Life Reviews*, *28*, 140-141. https://doi.org/10.1016/j.plrev.2019.03.009
8. Janson, S., R. Kozma, R. Miklos, Y. Sokolov (2019) "Bootstrap percolation on a random graph coupled with a lattice," *Discrete Applied Math.,* 258, 152-165. https://www.sciencedirect.com/science/article/pii/S0166218X18306140
9. Hazan, H., D. Saunders, H. Khan, D.T. Sanghavi, H.T. Siegelmann, R. Kozma (2018) BindsNET: A machine learning-oriented spiking neural networks library in Python, *Frontiers in Neuroinformatics*. 12, 89. DOI: 10.3389/fninf.2018.00089 https://www.frontiersin.org/articles/10.3389/fninf.2018.00089/full
10. Kozma, R., J.J.J. Davis (2018) Why Do Phase Transitions Matter in Minds? *J. Consciousness Studies*, 25(1-2), 131-150.
11. Myers, M.H., R. Kozma (2018) Mesoscopic neuron population modeling of normal/epileptic brain dynamics, *Cognitive Neurodynamics,* 12 (2), 211-223.
12. Kozma, R., R. Noack (2017) "Freeman's Intentional Neurodynamics," *Chaos &*

Complexity Letters, 11(1), 94-103.

13. Kozma, R., W.J. Freeman (2017) Cinematic operation of cerebral cortex interpreted via critical transitions in self-organized dynamical systems. *Frontiers in Syst. Neurosci.*11(10) https://www.frontiersin.org/articles/10.3389/fnsys.2017.00010/full
14. Kay, L., R. Kozma (2017) Walter Freeman – A Tribute, *Neuron,* Vol. 94 (4), 705-707. http://www.cell.com/neuron/fulltext/S0896-6273(17)30363-X
15. Heck, D., S. McAfee, Y. Liu, A. Babajani-Feremi, R. Rezaie, W.J. Freeman, J.W. Wheless, A.C. Papanicolaou, M. Ruszinko, Y. Sokolov, R. Kozma (2017) "Breathing as a fundamental rhythm of brain function," *Frontiers in Neural Circuits*, *10*, 115. https://www.frontiersin.org/articles/10.3389/fncir.2016.00115/full

Conference papers (14):

16. Aenugu, S., A. Sharma, S. Yelamarthy, H. Hazan, P. Thomas, R. Kozma, Reinforcement learning with a network of spiking agents, *Neural Information Processing Systems (NeurIPS2019), Neuro-AI Real Neurons-Hidden Units Workshop,* Vancouver, Canada, December, 2019.
17. Kozma, R., R. Noack, H.T. Siegelmann (2019) Models of Situated Intelligence Inspired by the Energy Management of Brains. *Proc. IEEE Inf. Conf. Systems, Man, and Cybernetics, SMC2019*, October 5-9, 2019, Bari, Italy, IEEE Press.
18. Davis, J.J.J., R. Kozma (2019) Interpretation of Mesoscopic Neurodynamics by Simulating Conversion Between Pulses and Waves, *Proc. 2019 IEEE/INNS Int. Joint Conf, Neural Networks (IJCNN2019),* Budapest, Hungary, July 14-19, 2019, IEEE Press.
19. Kozma, R., R. Noack (2018) Energy-Awareness in Brains and in Brain-Inspired Models of Computing, *Proc. IEEE Inf. Conf. Systems, Man, and Cybernetics, SMC2018*, October 7-10, 2018, Miyazaki, Japan.
20. Kozma, R., R. Ilin, H. T. Siegelmann (2018) Evolution of Abstraction Across Layers in Deep Learning Neural Networks, in Proc. 3rd INNS Conference on Big Data and Deep Learning 2018 (BDDL2018). April, 2018, Bali, Indonesia, *Procedia in Computer Science, Elsevier*, Vol. 144, pp. 203-213, *Best Paper Award.*
21. Davis, J.J.J., R. Kozma, Visualization of Human Cognitive States Monitored by High-density EEG. Arrays Proc. 3rd INNS Conference on Big Data and Deep Learning 2018 (BDDL2018), April, 2018, Bali, Indonesia, *Procedia in Computer Science,* Elsevier, Vol. 144, pp. 219-231.
22. Saunders, D.J., H. T. Siegelmann, R. Kozma (2018) STDP Learning of Image Patches with Convolutional Spiking Neural Networks, *IEEE/INNS Int. Joint Conf, Neural Networks (IJCNN2018), World Congress on Computational Intelligence,* July 8-13, 2018, Rio de Janeiro, Brazil, pp. 4906-4912, IEEE Press.
23. Andrade, G., M. Ruszinko, R. Kozma (2018) Graph Models of Neurodynamics to Support Oscillatory Associative Memories. *IEEE/INNS Int. Joint Conf, Neural Networks (IJCNN2018), World Congress on Computational Intelligence,* July 8-13, 2018, Rio de Janeiro, Brazil, pp. 5052-5059, IEEE Press.

24. Hazan, H., D. Saunders, D.T. Sanghavi, H.T. Siegelmann, R. Kozma (2018) Unsupervised Learning with Self-Organizing Spiking Neural Networks, *IEEE/INNS Int. Joint Conf. Neural Networks (IJCNN2018), World Congress on Computational Intelligence*, July 8-13, 2018, Rio de Janeiro, Brazil, pp. 493-498, IEEE Press.

25. Noack, R., J.J.J. Davis, C. Manjesh, R. Kozma (2017) Neuro-Energetic Aspects of Cognition - The Role of Pulse-Wave-Pulse Conversion in the Interpretation of Brain Imaging Data, *IEEE 2017 Symp. Series on Computational Intelligence (SSCI2017),* Nov. 29-Dec.1, 2017, Hawaii, US, pp. 1-8.

26. Davis, J.J.J., C. R. Kozma (2017) Amplitude-Phase Relationship of Brain Dynamics Viewed by ECoG using FIR-Based Hilbert Analysis, *IEEE 2017 Symp. Series on Computational Intelligence (SSCI2017),* Nov. 29-Dec.1, 2017, Hawaii, US, pp. 1-8.

27. Kozma, R. (2017) A Cognitively Motivated Algorithm for Rapid Response in Emergency Situations, *IEEE 2017 Cogn Situation Management Conf. CogSIMA2017,* March 27-31, 2017, Savannah, GA, DOI: 10.1109/COGSIMA.2017.7929580.

28. Ilin, R., T. Watson, R. Kozma (2017) Representations in Deep Neural Networks for Image Processing, *IEEE/INNS Int. Joint Conf. Neural Networks (IJCNN2017),* May 14-19, 2017, Anchorage, AK, USA, pp. 768-774.

29. Noack, R. C. Manjesh, H. Siegelmann, R. Kozma (2017) Resting State Neural Networks and Energy Metabolism. *IEEE/INNS Int. Joint Conf. Neural Networks (IJCNN2017),* May 14-19, 2017, Anchorage, AK, USA, pp. 228-235.

Presentations (12):

30. *Tutorial, R. Kozma and R. Ilin:* "Energy Constraints in Cognitive Processing – The Role of Constraint Satisfaction in Emergent Awareness," *IEEE Conf. Cognitive Situation Management (CogSIMA2019),* Las Vegas, April 8, 2019.

31. *Keynote Talk,* "Cognitive Phase Transitions in the Cerebral Cortex," *27th International Conference on Artificial Neural Networks (ICANN2018),* Rhodes, Greece, October 5-7, 2018.

32. Saunders, D.J., H. T. Siegelmann, R. Kozma (2018) STDP Learning of Image Patches with Convolutional Spiking Neural Networks, *IEEE/INNS Int. Joint Conf. Neural Networks (IJCNN2018), World Congress on Computational Intelligence,* July 8-13, 2018, Rio de Janeiro, Brazil.

33. Andrade, G., M. Ruszinko, R. Kozma (2018) Graph Models of Neurodynamics to Support Oscillatory Associative Memories, *IEEE/INNS Int. Joint Conf. Neural Networks (IJCNN2018), World Congress on Computational Intelligence,* July 8-13, 2018, Rio de Janeiro, Brazil.

34. Hazan, H., D. Saunders, D.T. Sanghavi, H.T. Siegelmann, R. Kozma (2018) Unsupervised Learning with Self-Organizing Spiking Neural Networks, *IEEE/INNS Int. Joint Conf. Neural Networks (IJCNN2018), World Congress on Computational Intelligence*, July 8-13, 2018, Rio de Janeiro, Brazil.

35. *Keynote Talk*, "Energy Constraints in Cognitive Processing – The Role of Constraint Satisfaction in Emergent Awareness," *IEEE Cognitive Situation Management Conference (CogSIMA2018)*, Boston, MA, USA, June 12-14, 2018.

36. Ilin, R., T. Watson, R. Kozma. "Abstraction Hierarchy in Deep Learning Neural Networks," *IEEE/INNS International Joint Conf. Neural Networks, IJCNN2017*, May 14-19, 2017, Anchorage, AK, USA.

37. Noack, R., C. Manjesh, M. Ruszinko, H. Siegelmann, R. Kozma. "Resting State Neural Networks and Energy Metabolism," *IEEE/INNS International Joint Conf. Neural Networks, IJCNN2017*, May 14-19, 2017, Anchorage, AK, USA.

38. *DARPA Site Visit - Superior AI Project*, May 11, 2017, 8:30am – 3pm, 10 Presentation, 1 hardware demo.

39. *Plenary Talk*, "Deep Learning in Brains and Machines - A Neuro-Energetics Perspective," Gulf of Mexico Spring School (GMSS): Deep Learning and Applications, Florida A&M University, Tallahassee, FL, April 16-18, 2017.

40. Kozma, R. "A Cognitively Motivated Algorithm for Rapid Response in Emergency Situations," *IEEE Conference on Cognitively Motivated Situation Management and Decision Support Conference (CogSIMA 2017)*, March 28, 2017, Savannah, GA.

41. Kozma, R., E. Rietman. Kick-off Meeting of DARPA Superior Intelligence Project, November 3, 2016, DARPA Office, Arlington, VA.

4. Comparison of Actual Accomplishments with the Goals and Objectives

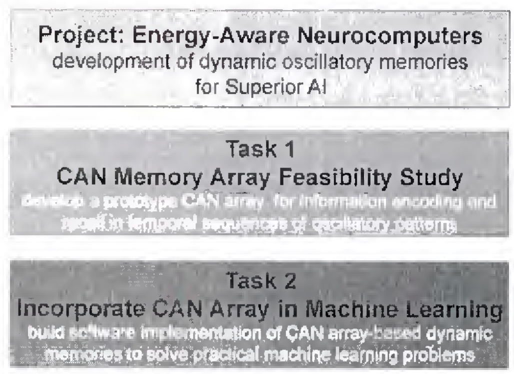
4.1 Objectives and Goals according to Research Description Document (RDD) Exhibit B - dated: August 30, 2016, and descoped version, dated February, 7, 2019.

The original project goals and objectives are summarized in Table 1. Note that Year 1&2 were performed according to this plan. However, at the end of Year 2, Task 4 (Super-Turing analog hardware) has been eliminated. Moreover, in Year 3, the remaining 3 tasks have been reorganized. Finally, Year 4 funding has been cut by the DAPRA PM in spite of the project progressing according to the stated tasks and goals, with achievements surpassing the stated goals and objectives.

Table 1. Summary of Superior AI Project Objectives and Goals
According to RDD, Exhibit B, August 30, 2016

tasks	year 1	year 2	year 3	year 4
1: Biological inspired efficient learning	Technical report on new learning concepts	Research paper on new learning concepts	Improved brain-inspired controller	Packaged trainable software, paper
2: Human brain architecture and AI	Network architecture: Bio-deep-recurrent	Learn abstractions and classification	Incorporate timeseries, model, software	Neural Compiler and concluding paper
3: Unique brain mechanisms	Energetic analysis of brain sections	Energy use from embryo to adult	Dynamics of cognitive tasks, multi-scales	Software, research papers
4: Super-Turing analog hardware	Construct array cortical architectures	Connect to sensory inputs, context activity	Associations and complex feedback	Phase-space study, research paper

Table 2. Superior AI Project task structure in Year 3, following the revisions implemented based on the PM request; Febr-7-2019.



The project focus has been descoped towards creating energy-aware neurocomputers for superior AI, based on oscillatory memory arrays. Task 1 has the goal to conduct feasibility studies of CAN-based oscillatory memories. Task 2 incorporates the oscillatory memories to solution algorithms for Machine Learning problems.

Overall Project Achievement Statement

In spite of the pitfalls due to the unsupportive act of the PM, the Superior AI project achieved its stated goals, within the constraints of the reduced 3-year span with 3 Tasks, instead of the originally planned 4-project with 4 main tasks.

The goals and objectives specified for each task at a given phase of the project are given in details below, together with the delivered results.

Objectives and goals of the Superior AI project

- as specified in RDD Exhibit 2, see verbatim exerts from RDD below -

Task 1: Biologically-inspired efficient learning
The potential benefits include a learning rule, which is not only more robust computationally, but also can transfer the classification success of deep network to the temporal domain and include control as well. Since this research is challenging, we will start with simple architectures: multi-layer, small cortical like circuits, and then advance toward the rich structure taken from the human connectome architecture. Information regarding how the brain controls adaptation will be studied via neuroinformatics from existing databases, demonstrating changes among twins or different concentrations of chemicals correlated with plasticity. The software experiments will include: implementing changes in plasticity parameters, analysis of the systems by statistical and graphic studies, and testing on real-world data sets.

Task 2: Human Brain Architecture and AI
The risks are high. The challenge is in translating the neuroinormatics findings to artificial networks. Can be done in multiple ways but it will involve significant computational and mathematical modeling. Developing a compiler will be challenging, but personnel in the BINDS Lab has experience that we can draw on. As in Task 1, the systems will be studied by statistical and graphics techniques using real-world data sets.

Task 3: Unique Brain mechanisms (energy management)
Because our approach is two-fold with each approach orthogonal to the other, the risk is moderate. The main initial work will be data cleaning and parsing, all solvable through exploratory and graphical data analysis. The more challenging part will be to integrate these disparate approaches. It would be nice to combine the two orthogonal approaches. How to do that is not obvious and will require exploratory computational studies.
The first approach involves using transcriptome data overlaid on protein-protein interaction networks. From this we can compute the Gibbs free energy. Transcriptome data are available for human brain development, brain diseases, and brains that did not develop properly (e.g. too few synapses).
The second approach involves using ECoG, fMRI and MEG data from Allen Brain Atlas. Such data were collected during specific cognitive tasks. Using these data, we will be able to incorporate into neuronal modeling energy changes and thus build more biologically realistic AI systems.

Task 4: Super-Turing analog hardware
This is a high-risk task. The primary risk is that the phase space is too complex to de-convolute into a useful "algebra." That would be unfortunate. If that is the case, we can fall back on using statistics of the dynamics and build statistical models instead of group- theory models. Our approach is to build actual hardware systems, collect a statistical amount of input/output data representative of the overall dynamics, and use group theory (first choice) to develop an "algebra of design" to allow us to compute new architectures for specific tasks.

4.2 Deliverables and Results Produced in Phase 1&2 (Y1&2)

4.2.1 Task 1: Biologically inspired efficient learning algorithms (STDP)

The main goal of the task was to develop local learning rules to significantly improve the efficiency of learning in AI models using global learning rules. Develop and evaluate biologically inspired spike-time dependent plasticity (STDP) learning rules, which can be incorporated as learning algorithms in large-scale spiking neural networks. Implement the developed STDP rule in spiking neural networks inferred from brains networks, determined based on the massive data available from the brain connectome project. Expected advantages include drastically improved learning speed and the ability to provide robust learning based on small number of examples manifested in brains.

4.2.1.1 Deliverables in Task 1 (Phase 1)

1. *Compare existing STDP methods and determine winner STDP to be used in future algorithms.*
2. *Suggest possible improvements on the selected STDP learning.*
3. *Extract biologically realistic spiking NN from open source brain imaging data and extend detailed connectivity matrix to include around up to thousand regions.*

Completion criteria and performance metrics: *Candidate STDP methods are implemented and compared using MNIST/CIFAR datasets. Comparison is based on performance metrics: correct classification rate, learning speed, computational cost and complexity. A winner STDP is selected based on the specified performance criteria. In addition, we develop a large-scale neural network extracted from brain imaging data.*

4.2.1.2 Delivered outcomes in Task 1 (Phase 1), corresponding to the previously specified list of deliverables.

1. We implemented and compared 4 different spike time-dependent (STDP) rules. We selected for our purposed the "online" version of the rule, where we only keep track of a "trace", or a memory, of the most recent spike (*temporal nearest neighbor*), which has arrived at post- and presynaptic neurons. This is fast algorithm and produces comparable accuracy as other STDP rules. We re-implemented the spiking neural network (SNN) model from ETH (Diehl & Cook, 2015) *Diehl, P. & Cook, M. (2015). Unsupervised learning of digit recognition using spike-timing-dependent plasticity. Frontiers in Comp. Neurosci, 9. doi:10.3389/fncom. 2015.00099.*
2. We proposed several modifications to the spiking neuron learning rules and spiking neural network (SNN) architecture. (i) A key result is that using a simplified *leaky integrate-and-fire (LIF) spiking neuron model*, we get similar accuracy as the more detailed ETH model with synaptic dynamics, and we can achieve that with significantly reduced computational efforts. (ii) We also modified the SNN architecture by implementing a *layer of convolution "features" or "patches."* This layer follows a convolutional neural network approach. We use convolution windows of size $k \times k$, with a stride of size s. Our results show that for large k we obtain improved performance over the original ETH network. As k decreases, the performance of C-SNN degrades as expected, due to the fragmentation of the image in the various convolution patches. (iii) We modified our C-SNN model by including *Lattice Connectivity between pairs of neighboring convolutional patches (CLC-SNN).* The weights connectivity pairs of convolutional patches are learned via the same STDP rule, which is used to learn the weights from input to the patches. Our convolutional spiking neural network models with lateral patch connectivity show great potential because of their ability to scale to large problem instances

with modest increase in computational demand and memory use. (iv) We implemented a *topological learning rule* according to which nearby nodes are more sensitive to similar input patterns. As a result, we obtained an ordered map of feature detector neurons, to be used for improved accuracy as an intermediate layer in a classifier CNN.

3. We conducted detailed performance evaluation of the adopted STDP methods and SNN architectures. We did not conduct "*biologically realistic spiking NN from open source brain imaging data,*" contrary to what was planned originally. Instead, we used layered architectures, following the ETH literature and other resources. This was a more conservative approach, as this way we studied the consequences of the changes due to the introduced patch connectivity and topological mapping and had control over the effects in the performance. It is feasible that at later stage of the project we may conduct data mining of brain imaging data for SNN structures, however, at this stage we felt that using more regular structures helped to pinpoint the main connection between the SNN structure and classification performance. To test the performance of the SNN, we *experimented extensively with the MNIST digit recognition database.* This is widely used to evaluate classifier approaches. We have also experimented by CIFAR10 based on related projects. We compared the performance of various SNNs using metrics, such as correct classification rate, learning speed, and computational cost. These results are documented in detail and indicate the edge of the STDP learning in concerning memory use and learning speed, with reasonable compromise on the accuracy decrease due to the unsupervised nature of the methods. Future work will be aimed at using unsupervised and supervised/reinforced methods to leverage their advantages in a combined learning approach.

4.2.1.3 Deliverables in Task 1 during Y2:

1. *Operational CAN unit demonstrating the desired frequency modulation characteristics and changing synchronization in CAN due to energy availability/constraint.*
2. *Implemented an array of CAN units (100) with STDP rule on the connections between units.*
3. *Identified at least one task to be implemented in details, e.g., rapid response to emergency scenarios, graceful degradation, which is unsolvable by traditional tools. For this task, demonstrate classification by the CAN array and show the impact of energy constraint on performance.*

Completion Criteria/Metrics: *The synchronization and frequency modulation properties of the CAN module (with up to 10,000 units) are thoroughly tested and demonstrate coexistence of multiple frequencies (gamma carrier frequency 40-60Hz, and alpha gating frequency 6-10Hz). For a given level of energy supply, we must produce transitions between highly synchronous oscillations (resonances) and desynchronization effects by changing the coupling coefficient from the computational units (neurons) to the metabolic components (astrocytes) acting as a control parameter. Operational arrays of interacting CAN units, starting with a 10x10 array modeling a mesoscopic cortical region (of area up to 1cm^2). Implemented learning rules (STDP) between the array units and train the system to perform a classification task. We explore specific task domains, in which the novel energy-modulated CAN units provide solutions to problems unsolvable by existing methods.*

4.2.1.4 Summary of the delivered outcomes in Task 1 (Phase 2), corresponding to the previously specified list of deliverables.

1. Developed CAN array for oscillatory sequential memory, which is a coupled square lattice of 5x5 CAN units, each with 1,000 neurons. This leads to a network of 25,000 neurons, which is somewhat less than the target 10x10 array, but it is in line with the target deliverable.
2. We were successful in producing collective behaviors such as synchronous spiking activity and coordinated firing activity in the gamma oscillatory band (40–60 Hz). We modulated the collective oscillations based on the connection strength between oscillators.
3. Implemented a correlation-based Hebbian learning rule over the CAN array. The microscopic connections inside each CAN unit (1,000 neurons) were not changed during the learning. In addition to the microscopic STDP learning, we implemented Hebbian learning over the macroscopic connections between the CAN units, producing a mean-filed effect. The implemented hierarchy of microscopic-macroscopic neural populations facilitates effective scaling using the Hebbian learning, as motivated by principles of brain dynamics.
4. In the effort to construct artificially intelligent devices that can perform perceptual, motor, and cognitive tasks similar to those of humans and non-human mammals, one important measure of the efficiency of those AI designs is how their power consumption compares to that of the biological brain. In order to assess that efficiency, it is first necessary to form and outline a standard metric to compare the two. We evaluated suitable metrics to quantify energy efficiency in brains and computational models.
5. We described a metric based on spiking activity to identify the computational capacity of several representative mammalian brains as well as the energy cost in various neuromorphic computer hardware. We confirmed that the mammalian cortex the power consumption per neuron is roughly conserved across species, roughly 0.5 to 1*10^-9 W/neuron. In this metrics, present CPUs provide about 10 million times worse performance than mammalian brains. The metric improves with GPUs and FPGAs, ARMs. Still, the best available neuromorphic chip at present, TrueNorth lags cortical neurons by about 3 orders of magnitude.
6. The identified highly efficient energy utilization of biological spiking neurons, together with the sequential oscillating memory feature of CAN arrays, has the potential of alleviating the exponentially increasing computational demand of today's digital memories and mainstream AI solutions. Advanced GPU's used for AlphaGo require power in the order of several MW. With this trajectory, we will run out of computational and energy resources within a decade.
7. CAN array-based encoding in oscillatory sequential patterns can lead to drastic increase in memory (potentially exponential memory capacity). Thus, our approach with dynamic memory can provide solution to acute energy constraints reflected in Moore law and Landauer limit.

4.2.2 Task 2: Modeling Brain Networks Architectures using Graph Theory Tools

The goal was to resolve the design bottleneck of the state-of-art of deep hierarchical neural networks by learning from the layered architecture in human brain during performing cognitive functions. Hierarchical deep neural networks are very popular and successful in many application areas. However, there are a number of unresolved questions why these networks are successful. In particular, it is not clear how knowledge is represented in deep layers and how the abstraction level changes from low abstraction at data level to high abstraction at decision/output layers. Understanding the way information is transformed and represented across layers, how features and produced and utilized in intermediate layers, will help to design more optimal networks for specific tasks.

4.2.2.1 Deliverables of Task 2 in Phase 1

1. *Extend the compression/granulation from the 188 regions used earlier to around 1,000 and extract multi-layer deep neural networks from the data. Report on the evolution of the abstraction level across the hierarchy.*
2. *Develop a model of brains processing sensory data (input) through many intermediate layers, to high-level symbolic knowledge and robust decisions.*
3. *Study the nature of the abstraction gradient; identify possible sudden changes in the abstraction gradient with potential functional/behavioral significance in the corresponding brain regions.*

Completion criteria and performance metrics: *Completed analysis of fMRI data with high spatial resolution up to 1000 units. Documented changes in the brain networks by evaluating network properties, such as number of hops, hub structure, and clustering of nodes to rich club/peripheral/feeder types. Algorithm developed to trace changes in abstraction gradient. Quantify the emerging symbolic abstraction level from sensory layer to deep layers.*

4.2.2.2 Summary of the delivered outcomes in Task 2 (Phase 1), corresponding to the previously specified list of deliverables.

1. To study various layered neural network structures, we started with (1) the all-to-all connected Hopfield network layer, using Hebbian learning. Next, we considered more complex and biologically plausible structures, as the (2) *combination of regular lattice layers and additional edges connecting remote nodes on the lattice*. In order to model the spiking process of neural populations, we designed a (3) *model of multi-layer structures with excitatory and inhibitory (reset) layers.* Typical model is with 10,000 excitatory units (100x100 lattice), while inhibitory layers have 2,500 (50x50) nodes. We studied the propagation of initial activation over the lattice using the mathematical tools of percolation theory. Various dynamical regimes are documented in these networks, such as *convergence to zero (all activations die out), convergence to unity (all sites become active), convergence to non-zero mean activity level (part of the sites are in non-zero steady-state), limit cycle oscillation (with specific period length), and quasi-periodic oscillations* (no period detected in the given observation window, typically 10,000 time steps). The current implementation of bootstrap percolation is done in python and was designed to offer as much flexibility as possible so as to allow for the implementation of a wide range of possible models. We also designed this so that we can painlessly incorporate different lattice structures and varied connections between inhibitory and excitatory vertices as needed.

2. The obtained results serve as theoretical basis of our new *superior AI approach using pattern-based computing.* We provided theoretical proof of the existence of a range of initialization probabilities that allow sustained oscillations with limit cycle or chaotic behavior as the function of reset level (m) and the type of update rule (k). We showed that the oscillations have a property of memory-less behavior for most parameter values, with the exception of larger reset values (m). We described the emergence of metastable spatial activity patters, which are sustained for several 100s or more time steps, they evolve slowly, and ultimately dissolve to random background activity. Crucially, these metastable patterns emerge spontaneously and predictably, and clearly distinct from the fluctuating background. *In the context of the dynamic computing principles of our project, these metastable patterns are the candidates of intermittent 'symbols,' that are the basis of the computation in our dynamical memory device.* These results provide key insights into dynamical *memory properties of the model, when input patterns can be encoded into attractors with limit cycles of specific length.* We studied how to store and recall specific inputs in the attractors with multiple wings. *The breakthrough aspects of these results for AI rest in the exponential memory capacity attractors with very long cycles (chaos), and the instant recall of the stored patterns without the need for lengthy search and related convergence process required in more traditional, convergence-based (fixed point) memory devices.*
3. *We developed a tool to assess knowledge representation in deep networks by progressing from the output towards the input.* The basic idea is that we use the activation at intermediate convolutional layers to classify the input data generating the given activation. Clearly, no classification occurs based on the raw input (zero abstraction level), while good classification (high level of abstraction) is possible based on the output activations, when the learning process converged. The important quest is how the abstraction level evolves across the layers? We implemented the extracted deep NNs in computational models and test how the different architectures influence the network performance in solving specific tasks. Using the CIFAR10 image dataset, the classification accuracy is evaluated. Our observations show: (1) There is a tendency of overall increase of the abstraction in the DL layers when moving deeper in the network from Input towards the Output Layer. (2) The general tendency of incremental change in the classification, however, has been interrupted by several jumps in the layer-by-layer classification accuracy. This conclusion is especially important when analyzing brain-imaging data (Taylor et al, 2015) displaying such layer-by-layer feature evolution. *Taylor, P., J.N. Hobbs, J. Burroni, H.T. Siegelmann, "The global landscape of cognition: hierarchical aggregation as an organizational principle of human cortical networks and functions, Scientific Reports (Nature Publ.), 5:18112, DOI: 10.1038/srep18112, 2015.*

<u>4.2.2.3 Deliverables in Task 2, during Y2</u>

1. *Complete implementation of reinforcement learning with STDP (ER-STDP) in spiking neural network.*
2. *Integrate ER-STDP with arrays of CAN units. Test the performance of the integrated system using MNIST and CIFAR datasets. Comparison is based on performance metrics: learning speed, classification rate, memory requirement, and computational cost.*
3. *Research paper summarizing the new concepts and tested results of a "teachable" brain-like AI.*

<u>Completion criteria and metrics</u>: *Complete an operating program and model of ER-STDP. Integrate STDP, RL-STDP over networks of energy-aware CAN units. Test the performance of the*

new model as compared to other unsupervised learning approaches, as well as to current spiking neural network control systems, particularly in regard to transfer learning and prediction.

4.2.2.4 Summary of the delivered outcomes in Task 2 (Phase 2), corresponding to the previously specified list of deliverables.

1. We developed a topological spiking neural network (SNN) with STDP learning rule. We showed that it significantly outperforms the state-of-art ETH approach, in terms employed learning patterns required to achieve a given accuracy. This is closely linked to online learning. In these results an n-gram-based evaluation approach has been very useful. The SNNs with the n-gram classification scheme can achieve a given performance level with 4-5 times less nodes and learning examples.
2. We implemented a conversion from BRIAN software to BindsNET, our Python-based library, in order to provide a software platform that allows scaling up our SNN model with CAN arrays in completing various machine learning tasks. BindsNET provides a unified framework to construct hierarchical computing solutions using energy aware computing and dynamic memory principles, building on core spiking NN computing units, which are connected according a specific network architecture and trained using unsupervised, reinforcement, and supervised learning.
3. The BindsNET repository has been released as an open source at GitHub. In addition to the source code, we made it easy to install using pip package manager and it been installed by about 1,000 users. We also have some discussions with users and requests for future features.
4. We tested the accuracy of BindsNET versus our BRIAN implementation of the ETH classification model. We showed that that BindsNET achieves the same performance as the ETH under BRIAN framework, in spite of significant simplifications in the simulation of the network components. The benefits using the simpler model is that we can achieve the same performance with LIF model that is notably less complex compare to the BRIAN model, and with simpler synapse rather the complex synapse used by BRIAN.
5. Beyond standard image recognition tasks, we developed pipelines in BindsNET to address dynamically changing environments and reinforcement learning in computer gaming, such as AI Gym Space Invaders and ATARI. We implemented various versions of reinforcement learning. For that end we use the Atari "break-out" game to train spiking neuronal network. We choose two approaches: (1) training spiking neuronal network using deep Q learning; (2) training regular network with deep Q learning algorithm, then using the trained network by replacing its regular neurons with spiking neurons. By choosing these two approaches, we show that spiking neurons can be used in reinforcement learning paradigm. The challenge is therefore to train the network with spiking neuronal network to achieve weights that provide good performance.
6. We currently extend BindsNET by new features, including convolutional networks, pattern matching, reinforcement learning, Q learning, and Deep Learning in multilayer architectures. We plan to use it for anomaly detection for prediction in time series, including Numenta benchmarks. We identify anomalies by showing that the BindsNET prediction significantly deviates from the observed input data stream.

4.2.3 Task 3: Brain energy utilization study for improved AI

The goal is to develop models of energy metabolism in brains during cognition, which is a requirement to achieve superior AI, at the same time leads to superior energy efficiency in hardware applications. Note that in Y1 we concentrated efforts on Task 3.1 (Energy mechanism and utilization at micro-meso-macro levels), while we did not work on Task 3.2 (Brain energy studies based on gene expression data), to better focus on our project. We have achieved the deliverables in Task 3.1 as detailed below, thus successfully complete the goals of Task 3.

4.2.3.1 Deliverables in Task 3 (Phase 1)

Task 3.1: Energy mechanism and utilization at micro-meso-macro levels

1. *Develop metabolism models that include cellular mechanisms of neuron-glia.*
2. *Incorporate the cellular models into a network of integrate and fire spiking neurons with excitatory-inhibitory (E-I) populations.*
3. *Include a connectivity pattern of neurons based on available brain connectome data.*
   Completion criteria and performance metrics: Multi-scale model is developed for neuron-glia cellular units with energy metabolism, which form a network of spiking neurons motivated by brain structures. The operation of the integrated model is tested and its performance is evaluated using entropy, pragmatic information, and free energy metrics.

Task 3.2: Brain energy studies based on gene expression data

1. *As with the other tasks, we are recruiting students. Realistically, at the end of the first year we should have the data sources collected, the software written for analyzing them, and because of the complexity of the analysis involved we should be the preliminary data analysis using the iris data set.*
2. *We should also have outlined a neural architecture and started computational experiments to address energy related issues (e.g. stochastic resonance).*

Completion criteria and performance metrics: Our energy-efficient neural network will be compared with a feedforward network on the iris dataset.

Note: To better focus the resources of our project, we concentrated efforts on Task 3.1, while we did not work on Task 3.2 in the reporting period. Therefore, no results are reported on Task 3.2.

4.2.3.2 Summary of the delivered outcomes in Task 3(3.1) in Phase 1 corresponding to the previously specified list of deliverables.

1. *We have developed a prototype model to simulate the neuroenergetics of glia-neuron assemblages in the mammalian brain.* The core model contains five differential equations, which simulate the behavior of these assemblages at three hierarchical scales. The innermost component is that of the individual *spiking neuron (computing component), which operates on a millisecond time scale (0.001s).* The mid-level component is our modeling of the *glial processes operating on a 100 milliseconds time scale (0.1s).* The most global component of the hierarchy is that of the fluctuations of *the cerebral blood flow to provide nutrient exchange to the glia-neuron assemblages, operating at the rate of roughly 0.1 Hz (10s)*, which is in line with cerebral blood flow fluctuation periods identified in blood oxygen-level dependent (BOLD) signals that inform the clinical fMRI. Timing is everything in this complex orchestration of exchange and recycling, and we have found that a model using the mentioned three times scales is able to describe the metabolic process effectively.
2. We have produced crucial progress in the energy-constrained spiking neural network model with 1000 neuron units. Our approach demonstrates that *the amount of available energy*

modulates the oscillation frequency of interacting computational units (spiking neurons); more energy inflow (via vascular input) produces increased frequency of oscillations (spiking). Moreover, for a given level of energy supply, we can produce *transitions between highly synchronous oscillations (resonances) and desynchronization effects by changing the coupling coefficient from the computational units (neurons) to the metabolic components (astrocytes) acting as a control parameter.* The demonstrated synchronization control algorithm is a crucial advantage of our approach, as synchronization-desynchronization transitions allow rapid response to dynamically changing external conditions in the environment and thus they manifest critical components of and emergency response system. *We have developed an interactive GUI interface* whereby the model parameters can be changed much more conveniently, and the results witnessed in real time; see Figure 1.

3. Going beyond the single oscillatory CAN units, we coupled together CAN oscillators using 2D lattice architecture. This *2D lattice will serve as a basis of the memory unit of the dynamic associative memory device.* For the component CAN units, we used parameters maintaining oscillations of energy variables in a stable range, at the same time producing quantifiable variation in the oscillatory frequency (computational capability) as the function of available energy resources. *We developed the blueprint of coupled CAN oscillator arrays,* for the time being a 3x3 array. The input data are encoded in the spatially distributed oscillations in these arrays coupled with modifiable links (synapses). In future studies, the connections between the oscillators are adapted following the STDP learning rule developed in Task 1 (Learning). The suitable architecture of larger arrays with lattice links and some shortcuts (between non-local lattice points) will be based on the results in Task 2 (Architecture).

4.2.3.3 Deliverables in Task 3, Y2:

1. *Implementation of new architectures for spiking neural networks with CAN units.*
2. *A code combining the CAN network architecture with STDP learning, to solve classification tasks.*
3. *Demonstration of the modulatory role of energy constraints on performance.*

Completion criteria: *Successful implementation of the recurrent network in combination with learning algorithms, such as STDP and use them for categorization tasks in temporal and spatial domains.*

4.2.3.4 Summary of the delivered outcomes in Task 3 (Phase 2), corresponding to the previously specified list of deliverables.

- We have introduced a lattice-based neural network with discrete time and space dynamics; the corresponding square-lattice topology supports CAN array architectures. The network consists of excitatory and inhibitory neurons and it is able to exhibit oscillatory dynamics. Various parametrizations influence the nature of the oscillators, producing phase transitions from fixed point, limit cycle, and non-periodic dynamics.
- We identified several model parameters which can control the oscillatory dynamics and that the oscillations are robust to input perturbations. Our model has advantages in digital computer implementations, as the discrete nature of the iterative dynamics makes it less susceptible to numerical errors while unfolding its dynamics.
- We studied various dynamic regimes in the spiking neural network architecture. An important achievement has been the analysis of the modulatory effect of input patterns introduce on dynamics of the quasi-chaotic oscillatory regime with very long oscillatory cycles. We

concluded that the inputs produced frequency modulation in the dynamics and visualized these effects using suitable graphics.

- We formalized the description of activity/inactivity clusters, which provide the mathematical tool of the effects that input stimuli have on the network. We conducted simulations with variable stimuli to see their effect on the network dynamics.
- Our results did not include explicit implementation of a specific learning rule but combined with learning rules developed in other tasks (Task 2), the outcomes have the potential of drastically increased memory capacity of CAN arrays as associative memories employed in specific machine-learning tasks.

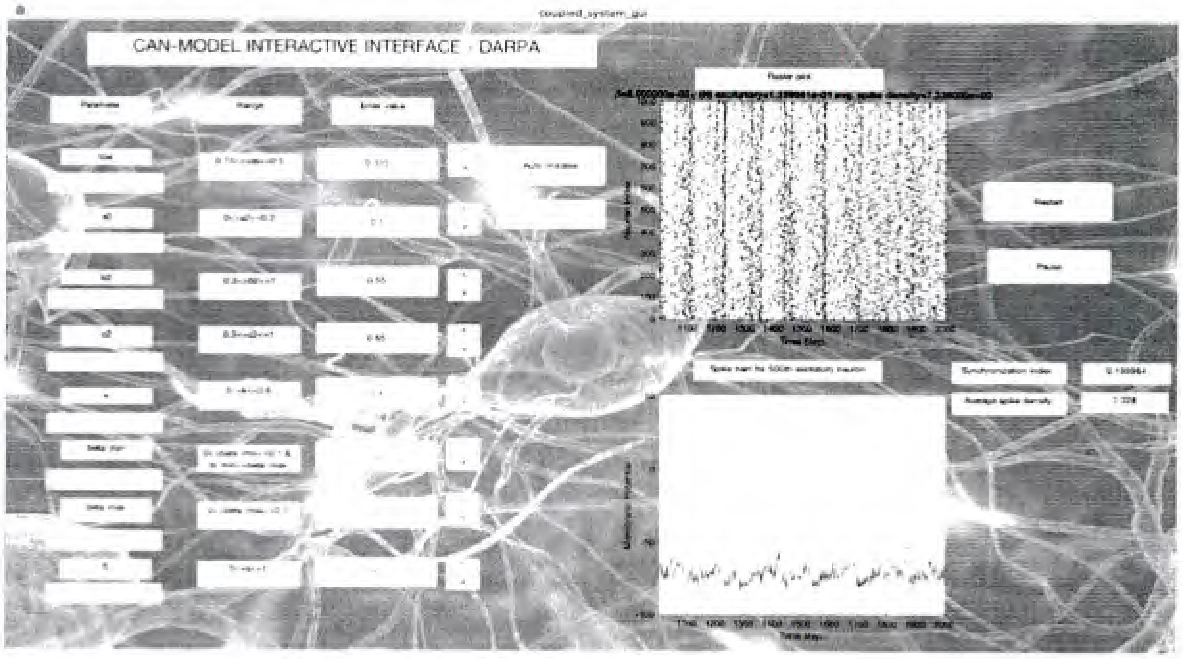


Figure 1. Snapshot of the CAN model GUI. We have developed an interactive GUI interface whereby the model parameters can be changed much more conveniently, and the results are displayed in real time.

4.2.4 Task 4: Periodic and Quasiperiodic Analog Hardware for Neural Computation

Objective is to build a new type of analog computational system that uses less energy. The system is modeled after the brain and not simply a layer-after-layer architecture. The basic computational elements in this new architecture are multi-state and quasiperiodic oscillators. In a real brain the oscillatory elements are assembled into complex networks. This suggests two approaches to our research; (1) Construction and investigation of analog multi-state and quasiperiodic oscillator hardware with the ultimate goal of collecting state diagrams so a compiler can be written to exploit this hardware. These circuits can be built from analog components and run at lower power; (2) Assemble software simulations comprised of power-law distribution of nodes and edges into computational systems for demonstration of conventional and novel applications. We will compare the ability of our oscillator circuits to store memory states with similar size Hopfield networks. That is, we will measure the number of stable oscillatory states as a function of the number of "neurons" and compare this with the number of fixed-point states in Hopfield networks for same numbers of "neurons."

4.2.4.1 Deliverables of Task 4 in Phase 1

1. *Simulations: At end of Phase 1, we should have made significant progress toward simulation of a robot in an artificial environment with real physics. The "brain" will be a simulation of a network of oscillatory circuits.*
2. *Hardware: At end of Phase 1, we should have circuits constructed and data collected on oscillator circuits ranging in size from 4 to 24 nodes, and some (4-16 nodes) of the individual oscillator circuit attractor diagrams completed.*
3. *Algebra of Design: Preliminary group theory analysis on the 4- to 16-node oscillator circuits will be completed. This preliminary work is required for compiler design.*

4.2.4.2 Summary of the delivered outcomes in Task 4 (Phase 1), corresponding to the previously specified list of deliverables.

1. *We developed Quasi-Periodic Oscillator Neural Networks (QPONN) to be used as computational engines.* Quasiperiodic oscillators form the basis of autonomic nervous systems in many animals. QPONN control processes dependent on adaptive behavior, such as walking and running. QPONNs achieve a large repertoire of interrelated, dynamic behaviors that are readily interchangeable for rapid adaptation in unfamiliar situations. However, they avoid the computationally intensive nature of maintaining a model of the world. And, they do not attempt to enumerate the intractable size of anticipating all possible scenarios. QPOs are ideal for encoding adaptive behaviors, which are responsive and can be complex. Furthermore, QPOs are utilized across the autonomic nervous system amongst animals in vastly diverse biomes. They generalize to different problem domains and through further evolutionary computation may become more specialized.
2. Robotic application domain of QPONN. Robots depend on adaptive behaviors for processes such as walking, flying, and swimming. And, they are subject to constraints on their design space, limiting their time and space complexity. The incorporation of QPOs is motivated by the hypothesis that the benefits observed in biology may confer analogous improvements in the adaptive behavior of robotic systems. QPOs will enable robots to have complex responses to unfamiliar situations without intensive computations or bulky storage requirements. To that end, we construct QPOs to form the basis of artificial autonomic nervous systems in robots. We worked on developing multiple environments that have varying subgoals and require

cumulative learning to traverse. The environments are: (1) mazes in a 2D environment. These are mostly simple environments that require much less computational power for quick iterative testing of the QPONNs during development; (2) mazes in a 3D environment that include pushing objects to create paths and avoiding moving obstacles.

3. *Designed a chip for neuromorphic computing.* Quasi Periodic Oscillatory Neural Network (QPONN) is being designed. (i) Simulated the targeted applications in software to define the specifications of the QPONN. (ii) Based on simulation results, we have defined the requirements for the ASIC chip. (iii) The system perspective of how to drive the inputs to and from the chip have also been finalized. (iv) After exploring various design alternatives, the specifications of ASIC chip were decided, keeping in mind minimum area and power. (v) The design libraries for fabrication have been procured from MOSIS. (vi) The RTL design of the chip is in progress. Continuing the work on design libraries for chip have been procured from MOSIS. The memory compilers were obtained from ARM. The chip is scheduled for manufacturing in September and should be ready for evaluation early in `18.
4. To facilitate experiments, we have built a 16-neuron system with patch panel for quick rewiring and circuit evaluation with a multichannel digital signal analyzer; see photo in full report at the end of this document. Though this seems like a small system, it is important to remember that a ring of 16 neurons, of the type we have been experimenting with, is able to support 143 stable memory states. In previous reports we showed an exponential curve of the number of states vs. the number of neurons.
5. *NeuroComputing Hardware Lab:* This new lab has begun being setup since January 2017. The components that have come in include: (1) 2 Digital Oscilloscopes, (2) 3 Function Generators, (3) 3 Power supplies, (4) Toolkits, (5) Logic Analyzer, (6) Chips, (7) Arduino Uno, (8) Flux Pens, (9) Shelves for circuit components, (10) Tables and chairs. We got a laptop computer from the department (Windows 7, 8 GB, Intel i5, 64bit), which has been used to interface with the oscilloscopes. We installed Python and Java IDEs on the machine for our computational experiments.
6. In the field of *algebra of design*, we have found that group theory is well suited to describe oscillatory dynamics; we expect to extend it from a description of individual oscillators to coupled oscillators. An example of the four stable states for a 6-neuron machine described in in the full report. We are working on extending this to larger rings and coupled rings. We anticipate that the theory will also be applicable to describe sensor signal inputs to stable oscillatory state and describe the resulting dynamical changes.

4.2.4.3 Deliverables of Task 4 in Phase 2

1. *Electrical schematics for the neuromorphic hardware.*
2. *Detailed phase diagrams for frequency modulation and synchronization.*
3. *Comparison of the results obtained with analog hardware versus digital simulations.*

<u>*Completion criteria*</u>*: Experimental evaluation of the role of control parameters in the coupled model in producing synchronization-desynchronization phase transitions and frequency modulation effects due to energy constraints.*

4.2.4.4 Summary of the delivered outcomes in Task 4 (Phase 2), corresponding to the previously specified list of deliverables.

During the performance period (Year 2, Phase 2), in response to the requests of the PM, there have been changes in the task goals. Accordingly, no analog hardware implementation of CAN circuitry has been completed.

1. Efforts have been directed to the development of a neuromorphic chip, which is an ASIC chip designed to have multiple digitally coupled oscillators and to perform Hopfield network-based computation to perform pattern recognition.
2. Worked on a generalized Hopfield NN model to describe energy dynamics, in which plasticity to be controlled by energy dynamics. Another model was also studied which uses the simple Schmitt trigger neuron model to create ring oscillators, which rings are chained together into a lattice network structure.
3. Aiming at FPGA platform, we implemented a single Izhikevich neuron and two interacting neurons in Verilog. We also implemented the multiply and accumulate model for creating a network. Behavioral simulation was done with 2 neurons. In the future we will add the metabolic system with two equations to produce a very simple CAN model.
4. We plan to scale up the design to a network of 48 CAN units with neurons and corresponding metabolics. We plan to synthesize design and put binary in ZedBoard and explore analog implementation.
5. We implemented the neuron model proposed by Izhikevich. There are N copies of the module, which feeds current to N neurons. Each neuron has modules to perform the 2 equations of Izhikevich original neuron model. Fixed-point arithmetic has been utilized in the design. We implemented the Izhikevich neuron model in Verilog; behavioral simulation has been performed with 2 neurons. The Multiply and Accumulate module was also completed for creating a network.
6. Not completed: scaling up the design to a network of 48 neurons and add capability for 2 new equations to neuron for the energy component. Synthesize design and put binary in ZedBoard.

4.3 Deliverables and Results Produced in Phase 3 (Y3)

4.3.1 Task 1: Developing a prototype CAN array for energy aware computing

Our goal is to develop energy efficient dynamical memories and compare their performance with mainstream DL AI approaches. Brains need merely 20W of power for solving even the most complicated tasks required from human intelligence. This efficiency is about a million times higher than today's cutting-edge Deep Learning (DL) solutions developed to specific Machine Learning (ML) benchmark tasks, requiring power supply with many MWs capacity, which is unsustainable in the foreseeable future. CAN arrays are motivated by the extremely energy efficiency of our brains. Neuromorphic hardware platforms (TrueNorth/IBM, Loihi/Intel, etc.) imitating spiking communication between biological brain cells (neurons), have much better energy efficiency than leading DL computing. CAN arrays provide highly innovative and efficient computational algorithms to drastically expand the impact of neuromorphic chips.

4.3.1.1 Deliverables in Task 1 (Phase 3)

Demonstrating the feasibility of computing by CAN oscillatory arrays as memory devices.

1. *Develop tuning methods using some control parameters inside the CAN units, as well as by adjusting the coupling parameters between the CAN units (learning) to achieve sequential transitions between oscillating regimes.*
2. *Repeated transitions between synchronous and non-synchronous operating regimes produce the sequential patterns as basic elements of CAN dynamical memories.*
3. *Testing the feasibility of such encoding using simple test patters for classification task.*

Completion criteria: Demonstrated encoding and recall of input data using oscillatory CAN arrays.

4.3.1.2 Summary of the delivered outcomes in Task 1 (Phase 3), corresponding to the previously specified list of deliverables.

1. Implemented arrays of CAN units with modifiable connections between the units. We identified conditions in CAN arrays leading to the sequence of synchronization patterns.
2. A breakthrough result has been achieved when analyzing the dynamics of CAN units by showing the presence of hysteresis effect due to cusp bifurcation in the coupled neural and metabolic system. Namely, we observed that the space defined by the forward gain from neural to metabolic subsystems, and the feedback gain from metabolic to neural system has a bifurcation point leading to the split of a stable equilibrium to two stable and one unstable equilibrium. The parameters corresponding to the bifurcated states create the conditions of self-sustained oscillations, which provide the basis and feasibility of CAN arrays serving as dynamical memories.
3. Implemented Hebbian learning on a CAN array with 25,000 neuron components, which means a 5x5 array of CAN units, each with 1,000 neurons. In our architecture, input patterns directly project to the CAN columns. Based on biological motivation 40-60Hz band of oscillations have been used. We demonstrated the feasibility of CAN arrays as the memory devices. CAN arrays can be trained using Hebbian learning to store and recall input data using sequences of oscillatory patters. Simulation results show that the encoded information can be read out from these patterns, and it is statistically significant at level $p=0.05$, even 0.01, for properly tuned system.
4. Evaluated the benefits of the groundbreaking results in literature on the effect is called ephapsis

(Chiang et al., 2019)[1]. Accordingly, ephapsis denotes the influence of the firing rate of a neuron by the extracellular ionic loop currents of neighboring neurons, which on entering the neuron tend to hyperpolarize it and exiting tend to depolarize it. It has been shown experimentally that weak electric fields can entrain action potentials of neurons. Ephaptic coupling has been suggested as a mechanism involved in modulating neural activity from different regions of the nervous system. Our studies support that the generation of the slow periodic activity is within the dendritic areas. Further studies are needed to integrate these new results to our CAN model.

4.3.2 Task 2: Building software implementations of dynamical memories to solve practically relevant AI/ML tasks

This task focuses on the conversion of the project results with oscillatory neural memories to real life problems. As a unified software platform, we employ BindsNET environment developed at the previous phases of the project. BindsNET provides a framework to construct hierarchical computing solutions using energy aware computing and dynamic oscillatory memory principles, building on core spiking NN computing units, which are connected according a specific network architecture and trained using unsupervised, reinforcement, and supervised learning.

4.3.2.1 Deliverables in Task 2 (Phase 3)

1. *Demonstrate the capability to scale up the preliminary results with BindsNET to practical, highly-competitive ML tasks. In selecting suitable ML tasks, employ the previous implementations of the Atari computer game.*
2. *Show that the proposed approach provides competitive results w.r.t. leading ML solutions, such as Q-learning.*
3. *Demonstrate the effectiveness and robustness of the developed oscillatory neural network model, in comparison with other existing tools in the field.*

Completion criteria and metrics: Demonstrated feasibility of using spiking networks to competitive ML tasks. Evaluation metrics include gaming performance, robustness to random or non-random distortions in the input data and the decision making.

4.3.2.2 Summary of the delivered outcomes in Task 2 (Phase 3), corresponding to the specified list of deliverables.

1. The software environment BindsNET, which has been created earlier in this DARPA project, has been successfully released as open source software on github[2]. It has received many followers and it is a highly competitive software environment in the field of neuromorphic computing simulations.
2. Implemented the interface with the Atari 2600 breakout ML task and established a framework to solve this game task using our spiking neural network. At this point, we test the performance of spiking neuron architectures on the game of Atari 2600 Breakout.
3. Trained artificial neural networks with architecture corresponding to Deep Q-Learning algorithm and transferred the learned weights to the spiking neural network. Our key achievement shows that spiking neural networks are capable of outperforming top ML

[1] Chiang, C. C., Shivacharan, R. S., Wei, X., Gonzalez-Reyes, L. E., & Durand, D. M. (2019). Slow periodic activity in the longitudinal hippocampal slice can self-propagate non-synaptically by a mechanism consistent with ephaptic coupling. *The Journal of physiology, 597*(1), 249-269.
[2] https://github.com/BindsNET/bindsnet

approaches on Atari task achieved by Google and other leaders of the field. We are conducting extensive studies on various popular ML games to show the generalizability of our approach and achievement. The results achieved using our transfer learning approach, have the potential to be implemented on hardware platforms, such as neuromorphic chips.

4. We demonstrated that the robustness of the spiking neural network is significantly improved as compared to the widely used Deep Learning, even though part of the screen is occluded from the player. Our SNN produces better performance for occlusions, and it avoids the catastrophic drop in performance at some sensitive locations. We are preparing papers to report these breakthrough results at top conferences.
5. An additional key point is that our approach showed high robustness to perturbations, being natural (noise) or adversarial attack. This addresses a weak point of Deep Learning and our spiking method has a clear edge. It relates to adversarial AI, and we are involved in research related to this topic.

5. Closing Statement

The DARPA Superior Artificial Intelligence Project has been performed at the Biologically-Inspired Neural and Dynamical Systems (BINDS) Laboratory, University of Massachusetts - Amherst, College of Information and Computer Sciences (CICS). It has been the collective effort of a team of:

- 3 faculty members and senior personnel.
- 1 postdoc,
- 18 graduate students,
- 5 undergraduate students.

The project produced the following outcomes:

- 1 book volume,
- 13 journal papers,
- 14 conference proceedings,
- 12 talks and presentations.

The project achieved its stated goals, within the constraints of the reduced 3-year span with 3 tasks, instead of the originally planned 4-project with 4 main tasks.
At this time further support is actively solicited to complete the efforts which remained unfinished due to the early termination of the project.