

Wireless Bayesian Neural Networks with Self-Assembly DNA Memory and Spin-Torque Oscillators

Ethan Chen, Jiachen Xu, Jian-Gang (Jimmy) Zhu, and Vanessa Chen
Department of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA, USA
Email: {ethanchen, jiachen, jzhu, vanessachen}@cmu.edu

Abstract—Wireless Bayesian neural networks (WBNNs) have been proposed to address the problem of energy efficiency and design complexity for training and classification in resource-constrained edge devices. By introducing thermally-activated DNA actuators and magnetically-tuned spin-torque oscillators (STOs), WBNNs are capable of learning from small data sets and resolving overfitting issues to achieve accurate classification results. To efficiently generate Gaussian variables, this work presents electromagnetically-coupled STOs that can inherently create programmable spectrum distributions for variational inference on Bayesian neural networks (BNNs). Specifically, a nanoscale heterostructure that monolithically integrates DNA origami with STOs is presented to perform multiplication-and-accumulation (MAC) with BNNs using a maximum amount of Gaussian variables, including: 1) STOs with weighted bias currents to set probability distributions and generate oscillation signals over a wide frequency range through interacting with an external magnetic field; (2) DNA origami that can selectively integrate wireless signals from various STOs to transform received energy into programmable magnetic fields. Simulation results demonstrate that the proposed WBNN can achieve an accuracy rate of higher than 96% while consuming 625 μ W.

Keywords—machine learning, DNA memory, spin-torque oscillators, Bayesian neural networks, brain-inspired computing

I. INTRODUCTION

Machine learning algorithms has been successfully developed to create intelligent machine for analysis of complex signals and images, such as support vector machine (SVM) [1][2], common spatial pattern (CSP) [3][4], linear discriminant analysis (LDA) [5][6], and artificial neural networks (ANNs). Among the neural network methods, convolutional neural networks (CNNs) are popular deep learning tools for image processing [7]. However, CNNs require huge amounts of data for regularization, so the algorithms quickly over-fit when processing small amounts of data. On the contrary, neural networks with Bayesian posterior inference are robust to overfitting, offering uncertainty estimates, and thus can efficiently learn from small datasets. Standard neural networks with Bayesian models offer a probabilistic interpretation of deep learning models by inferring distributions over the models'

weights. However, modeling a distribution over the kernels of a CNN usually requires tremendous amounts of computational power due to the vast number of parameters and large models, so inferring model posterior in a Bayesian neural network is a difficult task. Thus, posterior approximations are often employed with variational inference for model deployment to represent the posterior probability distributions on the weights with ensemble learning.

In this approach the posterior is modeled with a simple variational distribution and the likelihood distribution's parameters will closely resemble the true posterior distribution [8]-[10]. Nevertheless, the variational approach for approximating the posterior can be computationally expensive because the use of Gaussian approximating distributions may increase the number of model parameters significantly. To realize a computationally efficient platform that can quickly learn from small datasets with a small number of parameters, we propose wireless Bayesian neural networks (WBNNs) with the use of integrated DNA origami and STOs that are interacted with magnetic fields to generate required probability distributions for the approximate posteriors with the variational inference method.

II. BAYESIAN NEURAL NETWORKS

Real-time classification requires low prediction latency, sufficient accuracy, and low power overhead to be performed in edge devices. The paper presents wireless neural networks with Bayesian learning algorithms to overcome the fundamental trade-off between accuracy and computational complexity. The Bayesian neural network architecture is shown in Fig. 1, where each weight is assigned a probability distribution. Assume we have a training set $\mathcal{D}$, consisting of N input-output pairs: $\mathcal{D} = [(X^n, y^n) | n = 1, 2, \dots, N]$ where X is an input vector and y is the corresponding class label consisting of K classes. The objective is to use a neural network to model the input-output relation.

From the sum rule and the product rule of probability:

$$P(x) = \sum_y P(x, y) \quad (1)$$

This work was supported in part by the National Science Foundation CARRER program under Grant No. 1953801 and the Data Storage Systems Center (DSSC) at Carnegie Mellon University.

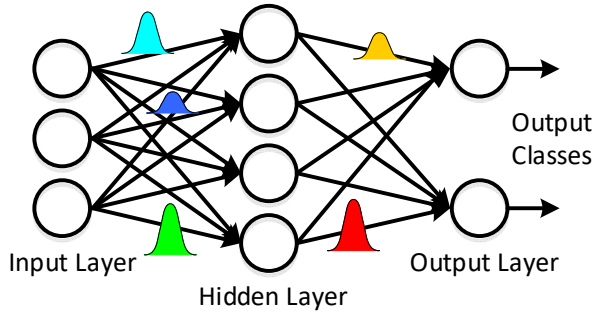


Fig. 1 Illustration of model structure of Bayesian neural networks in which each weight is assigned a probability distribution.

and

$$P(x, y) = P(x)P(y|x), \quad (2)$$

the Bayesian learning algorithm can be written as

$$P(\theta|\mathcal{D}, m) = \frac{P(\mathcal{D}|\theta, m)P(\theta|m)}{P(\mathcal{D}|m)} \quad (3)$$

where $P(\mathcal{D}|\theta, m)$ is the likelihood of parameters θ in model m , $P(\theta|m)$ is the prior probability of θ , $P(\theta|\mathcal{D}, m)$ is the posterior probability of θ given data $\mathcal{D}$. To predict the results, the intractable integral needs to calculate

$$P(x|\mathcal{D}, m) = \int P(x|\theta, \mathcal{D}, m)P(\theta|\mathcal{D}, m)d\theta. \quad (4)$$

Markov Chain Monte Carlo (MCMC) sampling, which is a general and powerful approximate inference method, is still computationally intense. The variational method is another commonly used approach to approximate the posterior probability.

To train neural networks, the conventional approach is to use Maximum Likelihood Estimation (MLE) that maximizes the likelihood, $P(\mathcal{D}|\theta)$. Given a model weighted by a prior probability, the other way to obtain a point estimate of an unknown quantity on the basis of the observed data is to find the value which maximizes the posterior probability, $P(\theta|\mathcal{D})$, with Maximum a Posteriori (MAP) learning, which is equivalent to the MLE with an additional regularization term. To efficiently compute neural networks with uncertainty on the weights, the algorithm Bayes by Backprop was proposed to optimize a well-defined objective function [11].

While the variational posterior distribution is assumed to be a Gaussian distribution, a sample of the weights $\mathbf{w}$ can be obtained by shifting and scaling the unit Gaussian variables with a mean μ and a standard deviation σ , where the non-negative vector of standard deviations is derived from the variational posterior parameters $\theta = (\mu, \rho)$ by:

$$\sigma = \ln(1 + \exp(\rho)) \quad (5)$$

Hence, a sample of the weights can be expressed as

$$\mathbf{w} = \mu + \epsilon \circ \ln(1 + \exp(\rho)) \quad (6)$$

where $\epsilon \sim \mathcal{N}(0, 1)$ is a vector of independent unit Gaussian variables, and $\circ$ denotes the element-wise multiplication operation. In the Bayes by backprop algorithm, the cost function is approximated by using sampled weights and a scale mixture of two Gaussians from the prior distribution. To calculate the samples of the weights of the Bayesian neural networks, unit Gaussian variables need to be generated efficiently. In this work a wireless machine learning architecture based on a Bayesian method is exploited to estimate the class probability for a given input from the MNIST dataset. As shown in Fig. 1, each weight requires to sample a unit Gaussian random variable that is generated by the electromagnetically coupled STOs and integrated by DNA origami to performance inference with the proposed wireless Bayesian neural network.

III. SPIN-TORQUE OSCILLATORS FOR GENERATING PROBABILITY DISTRIBUTIONS

Spin-torque oscillators (STOs) exhibit resonant oscillations with non-linear transformation functions, which are utilized as inherent activation functions in the WBNN. The STOs can be integrated at high density because of their nanoscale size and low-power consumption. Thus, oscillators are exploited as processing units in this work. Fig 2(a) illustrates the schematic of the STO design: an elliptical shaped magnetic element is placed on top of a heavy-metal wire. The current flow in the heavy metal wire generates a perpendicular pure spin current injected into the magnetic layer via spin Hall effect, as illustrated

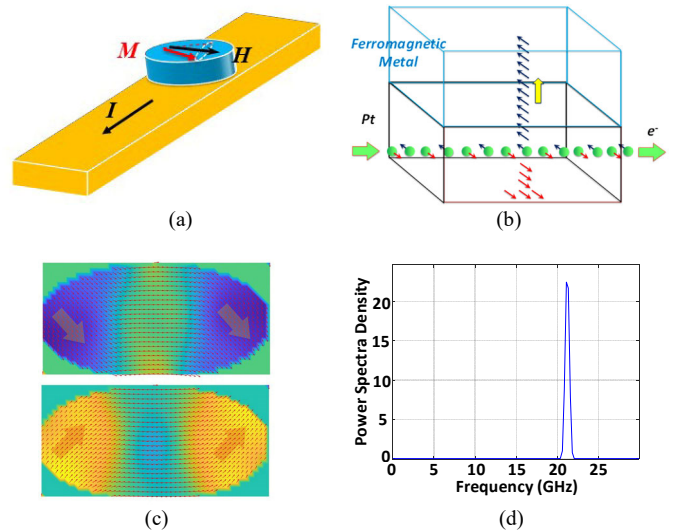


Fig. 2 Spin torque devices as wireless synapses: (a) schematic illustration of the spin torque oscillator with the oscillation magnetic element fabricated directly on top of a heavy metal nanowire; (b) illustration of the spin Hall effect that yields pure spin current injection into the magnetic element placed above; (c) the magnetization states during magnetic oscillation obtained with micromagnetic simulation; (d) the power spectrum density and the oscillation frequency of the spin torque oscillator provides Gaussian approximation.

in Fig. 2(b). The heavy metal can be a choice of Pt, Ta, W, or Ir while the range of the choices of the materials for the magnetic element can be broad. The injected spin current generates a spin transfer torque that facilitates steady magnetization oscillation at the Kittel oscillation frequency under a given external magnetic field. The utilization of spin orbital torque in this special design eliminates the need of a top electrode for the magnetic element that enables closer distance with the DNA strand for stronger radio-frequency (RF) magnetic field. Fig. 2(c) shows two magnetic states of the magnetic elements with roughly 180° phase difference during an oscillation cycle. Fig. 2(d) shows the spectrum of the spin torque oscillation corresponding to an external magnetic field of 0.5T. The oscillation frequency and its spectrum distribution provide Gaussian variational inference for the WBNN, which are sensitive to the current magnitude and external magnetic fields set by a nearby magnetized ferromagnetic nanoparticle. These modulation schemes provide complex learning mechanisms, equivalent to synaptic strength in neurons and synaptic weights in cognitive computational models.

IV. ARCHITECTURE

The wireless neuron for the WBNN consists of a DNA actuator, a ferromagnetic nanoparticle, and an STO. Dynamic DNA actuators permit precise placement of nanocomponents within a few nano meters and the length of the single-stranded DNA (ssDNA) can be altered in response to electromagnetic inputs [12]. Hence, the learning mechanism can be achieved by dynamic tuning of ferromagnetic (FM) nanoparticle separation distance using DNA origami actuators responsive to magnetic inputs. Networks are assembled on DNA origami templates, and pruned using DNA or electromagnetic inputs processed through nanoparticle tuned transformers.

In Fig. 3, one logic cell consists of an STO with an external confining magnetic field applied by an external magnetized nanoparticle. The oscillator is powered by a constant DC current I_{DC} . The output of the oscillator is a small modulation of the current (10^{-11} W) that radiates out as a dipole field. The first cell, indexed 0, has that modulation at a center frequency as f_0 with a standard deviation as σ_0 ($\mu=f_0$, $\sigma=\sigma_0$). This frequency is programed by the intensity of the local field, which is set by the distance between the magnetized FM particle and the STO, as shown in Fig. 3(b). The distance is itself set by the extension of a DNA molecule (DNA₀), as shown in Fig. 3(c).

The logic cell indexed 0 radiates its microwave power to a set of downstream cells, indexed 1, ..., i, ..., n. These cells have exactly the same structure, and are powered by weighted DC currents I_{DC1} to I_{DCn} to set the oscillation frequency distributions during the learning process. Under the condition that DNA absorbs the microwaves and heats up, it will in steady state reach a temperature T_i that is determined by how much radiating energy is harvested by the DNA actuator. Given that the DNA is with an aspect ratio of $10\text{nm} \times 1\text{nm} \times 1\text{nm}$ and a thermal conductivity of 0.1 W/m-K , the DNA has a conductance of $K=10^{-10}\text{ m} \times 0.1\text{ W/m-K}=10^{-11}\text{ W/K}$. With a heat load of 10^{-11} W , the temperature rise will be of the order of 1 K. This temperature

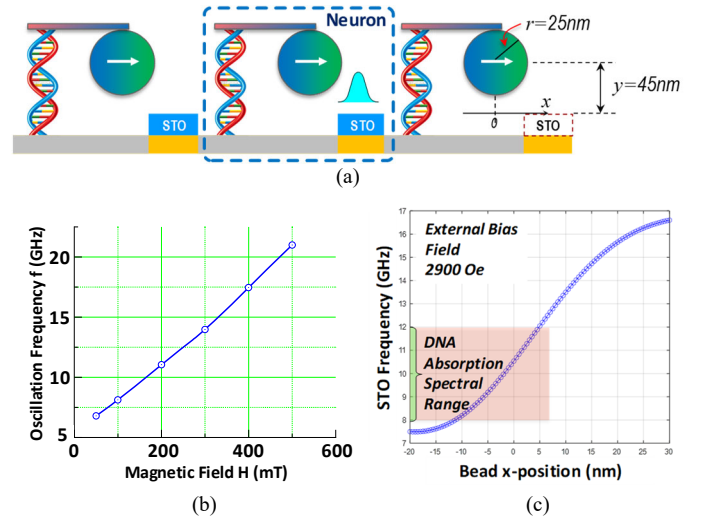


Fig. 3 (a) A single neuron consists of a DNA actuator that carries an FM particle and an STO. (b) The oscillation frequency and spectrum distribution are programmed with an external magnetic field applied by the FM particle. (c) The oscillation frequency changes with the distance, y , between the STO and the FM particle.

rise changes the length of molecule DNA_i and the distance between the FM particle i and the Pt. Fig.4 (a)(b) depicts how the single neuron operates to perform the MAC function in the frequency domain. If the cells 0...n generate resonate frequencies, $f_0 \dots f_n$, outside of the microwave absorption range of the DNA. No radiating energy is integrated by the DNA, so the FM particle does not apply enough field to the local oscillators, stopping them from oscillating. Hereby, the position

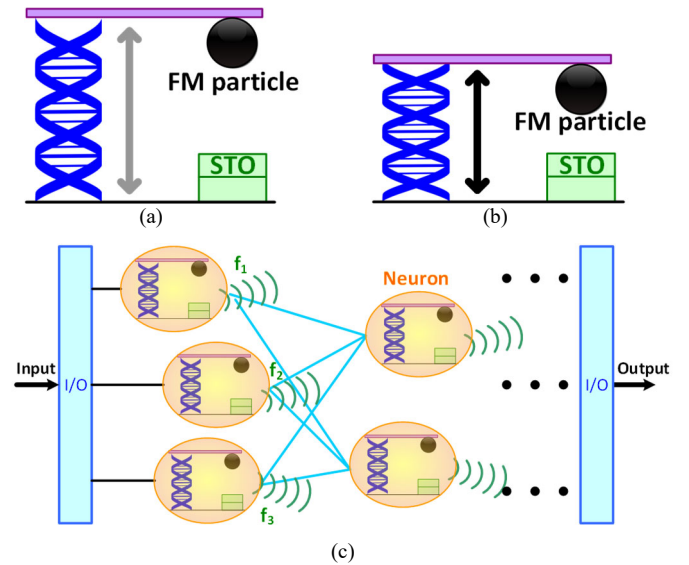


Fig. 4 (a)(b) The length of the DNA changes with the radiation energy harvested from the STOs at the previous layer so that it can alter the distance between the FM particle and that in turn changes the oscillation frequency of the neuron. (c) The overall architecture consists of multiple-layer wireless neurons. Each neuron integrates wireless signals from neurons at the previous layer to perform signal multiplication-and-transformation that generates corresponding wireless signals to inform neurons at the next layer.

of the FM nanoparticle determines the oscillation frequency, f_0 , and the weighted frequency distribution is set by the electrical field of the STO. Because interactions are distance-dependent, physically altering the FM nanoparticle separation distance would tune the oscillation frequency and the radiation energy to inform the neurons at the next layer, constituting a wireless Bayesian neural network, as shown in Fig. 4 (c).

V. SIMULATION RESULTS

The proposed wireless Bayesian neural network of various sizes has been developed to learn the MNIST digits database [13] that is composed of 60,000 training and 10,000 testing pixel images of size 28 by 28. Each digit image is labelled with the corresponding number from 0 to 9. The neural network consists of two hidden layers with 100 neurons for each layer. Fig. 5(a) shows the learning curves on the testing and training set on the MNIST database. The simulation results show that the WBNN converges quickly after 20 epochs. Fig. 5(b) shows density estimates of the weights sampled from the variational posterior. The simulated power consumption for 200 STOs to operate at 20 GHz is 625 μ W. Results from the wireless Bayesian neural network using a Gaussian prior are summarized and compared to other state-of-the-art interfacing classifiers in Table I.

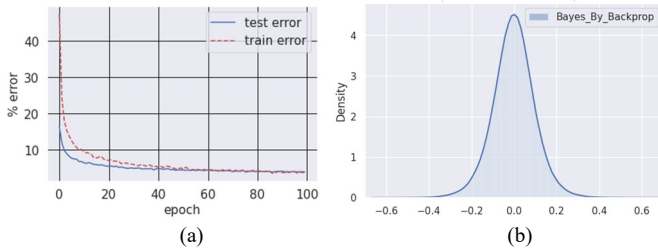


Fig. 5 (a) The learning curve of testing and training errors on MNIST database, and (b) the histogram of the trained weights of the Bayesian neural network.

VI. CONCLUSION

The proposed wireless Bayesian neural networks utilizing magnetically-tuned spin oscillators with thermally-activated DNAs to perform machine learning tasks with Gaussian variational inference. The signals are transformed into the frequency domain based on the weighted spectral distributions of the oscillation energy. Simulation results show that a 2-layer WBNN with 100 neurons for each layer can carry out handwritten digit classification based on the MNIST database with an accuracy rate higher than 96%. The wireless architecture enables energy-efficient computing and adaptive memory, massively expanding computational density and enabling rapid analysis of data sets in energy-constrained IoT devices and sensors.

REFERENCES

[1] P. Shenoy, K. Miller, J. Ojemann, and R. Rao, "Generalized features for electrocorticographic BCIs," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 273-280, Jan. 2008.

TABLE I. PERFORMANCE SUMMARY

	ISSCC 15' [14]	VLSI 16' [15]	This work
Method	Matrix-multiplying ADC	Noise-shaping analog-digital interface	Wireless Bayesian neural network
Power Consumption	663.6 nW	243.7 μ W	625 μ W
Function	Matrix-Multiplying	Matrix-Multiplying	Classification

- [2] T. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Scholkopf, "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003-1010, Jun. 2004.
- [3] G. Blancharda and B. Blankertz, "BCI competition 2003—Data set IIa: Spatial patterns of self-controlled brain rhythm modulations," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1062-1066, Jun. 2004.
- [4] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991-2000, Aug. 2008.
- [5] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, and M. Pregenzer, "Current trends in Graz brain-computer interface (BCI) research," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 216-219, Jun. 2000.
- [6] J. Kronegg, G. Chanel, and S. Voloshynovskiy, "EEG-based synchronized brain-computer interfaces: A model for optimizing the number of mental tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 50-58, Mar. 2007.
- [7] K. Bong, S. Choi, C. Kim, S. Kang, Y. Kim, and H.-J. Yoo, "A 0.62mW ultra-low-power convolutional-neural-network face-recognition processor and a CIS integrated with always-on haar-like," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, San Francisco, CA, Feb. 2017, pp. 248-249.
- [8] M. Roth and F. Gustafsson, "Computation and visualization of posterior densities in scalar nonlinear and non-Gaussian Bayesian filtering and smoothing problems," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5-9 March 2017, New Orleans, LA, USA.
- [9] D. Barber and C. M. Bishop, "Ensemble Learning for Multi-layer Networks," in *Advances in Neural Information Processing Systems*, pp. 395-401, 1998.
- [10] A. Graves, "Practical Variational Inference for Neural Networks," in *Advances in Neural Information Processing Systems*, pp. 2348-2356, 2011.
- [11] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight Uncertainty in Neural Networks," in *International Conference on Machine Learning*, pp. 1613-1622, 2015.
- [12] J. A. Johnson, A. Dehankar, A. Robbins, P. Kabtial, E. Jergens, K. H. Lee, E. Johnston-Halperin, M. Poirier, C. E. Castro, and J. O. Winter, "The path towards functional nanoparticle-DNA origami composites," *Materials Science and Engineering: R: Reports*, vol. 138, pp. 153-209, Oct. 2019.
- [13] Y. LeCun, C. Cortes and C. Burges, MNIST handwritten digit database, [Online]. Available: <http://yann.lecun.com/exdb/mnist>
- [14] J. Zhang, Z. Wang, and N. Verma, "A matrix-multiplying ADC implementing a machine-learning classifier directly with data conversion," *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, pp. 332-333, Feb. 2015.
- [15] F. N. Buhler, A. E. Mendrela, Y. Lim, J. A. Fredenburg, and M. P. Flynn, "A 16-channel noise-shaping machine learning analog-digital interface," in *Symposium on VLSI Circuits Digest of Technical Papers*, Honolulu, HI, USA, June 2016, pp. C30-C31.