Structure Discovery for Gene Expression Networks with Emerging Stochastic Hardware

Sourabh Kulkarni, Sachin Bhat, Csaba Andras Moritz
Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA
skulkarni@umass.edu, andras@ecs.umass.edu

***Abstract*— Gene Expression Networks (GENs) attempt to model how genetic information stored in the DNA (Genotype) results in the synthesis of proteins, and consequently, the physical traits of an organism (Phenotype). Deciphering GENs plays an important role in a wide range of applications from genetic studies of the origins of life to personalized healthcare. Probabilistic graphical models such as Bayesian Networks (BNs) are used to perform learning and inference of GENs from genetic data. Current techniques of generating BNs of GENs from data, which are mostly approximate in nature, involve searching and scoring of multiple probabilistic graphical structures. However, while search algorithms can be efficiently implemented in software, the same is not true for scoring. Scoring of probabilistic models with inherent parallelism is inefficient when performed sequentially over conventional architectures comprising of deterministic devices. In this paper, we introduce a new nanoscale hardware acceleration framework, enabling fast and efficient Bayesian inference operations, significantly accelerating the scoring aspect of the BN learning of GENs using a combination of emerging stochastic devices and CMOS technology. The stochasticity of the devices is utilized to efficiently perform approximate inference on probabilistic networks, and the circuit framework constituting these devices is designed to exploit the inherent parallelism in these models. We demonstrate approximate inference operation over a small BN. We estimate the performance benefits of five orders of magnitude in performing inference operations using this architecture over software-only approaches.**

Keywords—Genetics, Genomics, Gene Expression Networks, Bayesian Networks, Bayesian Learning, Approximate Inference, Structure Discovery, PMA-STT MTJ, Nanoscale Architecture.

I. INTRODUCTION

The DNA of an organism stores all the information required for its creation, sustenance and development. Genomics is the study of gene expression, the natural process occurring in all lifeforms, in which the information in the DNA is translated to physical traits of an organism. Deciphering gene expression in lifeforms provides great insight into the workings of the machinery of life. Applications of genomics range from species studies to personalized healthcare. Recent advances in genetic engineering [1] make genome sequencing more accessible for everyone. This has led to a plethora of possibilities in the domain of personalized medicine, as a person's genome potentially contains information about various aspects of the person's health, for example, allergies, predisposition to several diseases [2][3] and compatibility with several medications [4]. Although all this data lies in the sequenced genome of a person, Gene Expression Networks (GENs) are required to be learnt per-person to be able to access this information. Research [5] has shown that Bayesian networks [6] (BNs), which are a type of probabilistic graphical models, are excellent formalisms to model the GENs. The process of learning the BN which most accurately depicts the GEN from genetic data, though, turns out to be complicated.

Current implementations of GENs are limited to few thousand nodes as the time complexity is exponential in number of nodes [7]. Furthermore, a GEN comprising of a reference plant genome (with size comparable to human genome), which has approximately 20,000 genes, requires use of supercomputers, and could take days to complete [8]. These limitations mean that performing these learning tasks at a per-person basis is prohibitive in current software implementations. Performing Bayesian learning and inference using software is inefficient, as it involves performing probabilistic computations, over several layers of abstraction, on circuits and architectures that are inherently deterministic. Also, software implementations generally perform computations in a node-by-node traversal fashion, while the structure of BNs usually allow for parallel computation in a level-by-level traversal. This means that a potentially exponential speedup in computation remains unutilized.

In [9], emerging magneto-electric devices called Perpendicular Magnetic Anisotropy Spin Transfer Torque Magnetic Tunnel Junctions (PMA-STT MTJs), are detailed, which exhibit stochastic behavior, i.e., their switching operation is probabilistic, and the probability of switching is controllable. We utilize this unique property to design circuits which perform probabilistic operations by directly controlling the switching probability of the devices. We also design an architecture based on these devices in conjunction with CMOS support circuitry which exploits the parallelism enabled by BNs to the fullest. We demonstrate the functionality of the architecture by performing inference operations on an example BN. The potential benefits of using this architecture to accelerate BN learning of GENs are discussed.

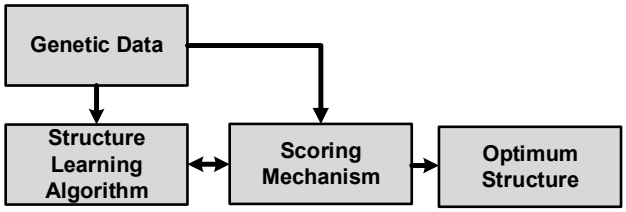

Figure 1. Design flow for structure learning in BNs. Learning algorithm generates graphs from data, which are scored by the scoring mechanism. The algorithm iteratively improves upon the graph structure until an optimum structure is selected.

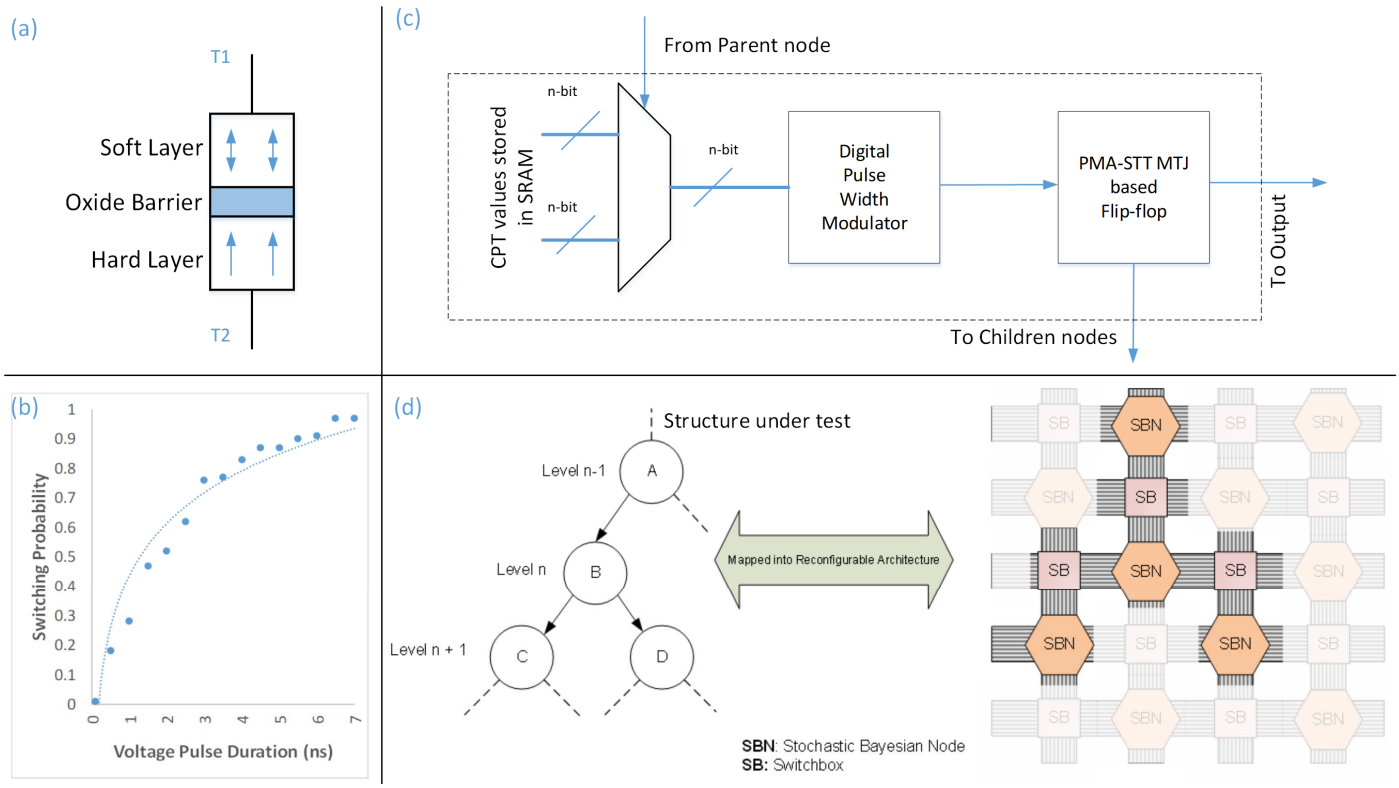

Figure 2. Framework components; (a) Schematic of PMA-STT MTJ. This two-state, two-terminal device consists of a soft magnetization layer, an oxide barrier and a hard magnetization layer; (b) MTJ stochastic switching results based on simulations; (c) Schematic of modules in each SBN; and (d) Reconfigurable SBN-SB framework mapping every node in a BN structure under test to a SBN. The edges of the BN are implemented with metal routing layers typically used in FPGAs, with connectivity configured through switch boxes.

The paper is organized as follows; Section II discusses Bayesian learning and the process of structure discovery; Section III details the proposed architecture and its role in acceleration of structure discovery in BNs; Section IV demonstrates a BN Inference example using the proposed architecture; Section V concludes the paper.

II. Structure Discovery in BNs

BNs are probabilistic graphical models, which encode the knowledge of a domain in a Directed Acyclic Graph (DAG). A node in the graph represents a random variable, while an edge represents causal relation between nodes. The strength of these causal relations is represented by conditional probability parameters, stored in the Conditional Probability Table (CPT) of a node. Hence, a BN has two major aspects – the structure of the graph and the parameters in the CPT. The process of learning a BN from data typically involves individually learning both aspects of the BN. Learning structure of a BN is known as structure discovery, while learning the correct values to populate the CPTs is known as parameter learning. The work proposed in this paper shall focus on structure discovery.

Structure discovery in BN from data is known to be NP-complete. It means that the only way to obtain an optimal BN structure is to iteratively search over the state space of all possible graph structures for a given dataset. This process is shown in Figure 1. Based on the data, an algorithm starts with an initial graph (usually random), which is then 'scored' over the data by a scoring mechanism. The scoring mechanism indicates the quality of a BN structure. To put it simply, it determines how likely is the structure to explain the data. Based on the score the algorithm iteratively updates the graph structure. If the modified graph obtains a better score, it is accepted and is further improved upon.

Most of the scoring methods in BN learning are analytical, and require complex mathematical computations. These methods are currently done using software and significantly impact the time required for structure discovery. There are two major types of scoring mechanisms - Bayesian scoring and Information-theoretic scoring [10]. In this work, we use stochastic-sampling based [11] scoring, which is a type of Bayesian scoring mechanism.

Bayesian scoring applies the Bayes' rule to determine the quality of a given BN structure. The Bayes score for a structure given the data is calculated by using the Bayes rule:

$$\boldsymbol{P(Structure|Data)} \propto \boldsymbol{P(Data|Structure).P(Structure)} \quad (1)$$

Intuitively, the equation indicates that the score of a structure is proportional to how closely it describes observed data (and on the prior probability of the structure, which could be provided by a domain expert). This scoring computation must be done over several iterations of the graph until the best is selected. The number of iterations is usually a design choice,

TABLE I. INFERENCE ACCURACY AND PERFORMANCE BENEFITS OF THE PROPOSED ARCHITECTURE

Experiments	Conditioned Inferences				Inference Runtime	
	P(Da|Gm1=T)	P(Db|Gm1=T)	P(Da|Gm1=F)	P(Db|Gm1=F)	Gm=T	Gm=F
Proposed Framework	0.77	0.18	0.28	0.62	2µs	2µs
Baseline	0.8	0.2	0.3	0.6	132ms	168ms

but above a certain critical number, the iterations provide diminishing returns.

The Bayesian score of a graph is computed using a method known as stochastic sampling. This method involves generating a functional model of the graph with the CPT values set, and performing sampling over each node of the graph for several iterations. For example, a node with a probability value of 0.5 (calculated using the CPT values and the result from the parent), when sampled for 10 iterations, is expected to return 'True' in 5 of those iterations. The number of iterations required to calculate a proper score is known as the sufficient statistic. The inference data obtained from the stochastic sampling process is used to compute the Bayesian score of the graph, which is the degree of correlation between the sampled data and learning data.

III. RECONFIGURABLE FRAMEWORK FOR BAYESIAN LEARNING

Once a BN structure is generated by structure learning algorithm, it is mapped onto the proposed framework (Figure 2(d)), which has a reconfigurable architecture with Stochastic Bayesian Nodes (SBNs) representing the nodes of the BN, and the metal routing and switch boxes implementing the edge connectivity of the BN structure. The mapped structure is sampled for the number of iterations required by the sufficient statistic. The samples obtained from these iterations are then processed ex-situ to obtain the score.

A. PMA-STT MTJs

Perpendicular Magnetic Anisotropy Spin Torque Transfer MTJs are emerging magneto-electric devices which exhibit probabilistic switching dynamics. The stochasticity in switching characteristics of these devices is governed by the magnitude and duration of current applied. Figure 2(a) shows the device schematic and Figure 2(b) shows the graph demonstrating the switching behavior of the device. This unique property, along with the ease of integration with CMOS, enables its use in our framework.

B. Stochastic Bayesian Node

SBN models a node in a Bayesian network to perform stochastic sampling described in earlier sections. The node consists of multiplexers, a digital pulse width modulator [12], and a PMA-STT MTJ based flip-flop as shown in Figure 2(c). The property of the MTJ discussed above enables precise control of the switching probability of the device by modulating the duration of the applied current. A digital Pulse width modulator is used to generate precise duration voltage pulses. It maps the digital input obtained from the SRAM to a PWM signal which is representative of the probability value. For example, a voltage pulse of duration 1.5ns corresponds to a switching probability of 50%. The digital inputs to the DPWM represent the CPT values which are pre-configured in the SRAM based on experimental data. These appropriate CPT values to be fed to the DPWM is selected based on the outcome of the parent SBN using multiplexers. Once the pulse proportionate to the CPT value is applied to the MTJ, the outcome (whether the MTJ switched) is stored using a flip flop. This latched outcome is fed to the next level and is available for read out.

C. Reconfigurable Framework

To perform scoring, the framework should be able to map arbitrary BNs, therefore reconfigurability is a necessity. The proposed architecture adopts an FPGA-like reconfigurable architecture. It consists of programmable switchboxes and SBN. These two units are used to map BN structures (see Figure 2(d)). A Bayesian structure to be scored is mapped onto the framework by mapping a node into a SBN while the connectivity between nodes is enabled by switchboxes. The CPT values are stored in the corresponding SBN SRAM cells. The configured Bayesian structure is stochastically sampled until the sufficient statistic is obtained. The sampled data is then used to compute the Bayesian score of that structure, as described in equation (1). The sampled data is extracted from the network for scoring using a scan-chain based system, and this can be done in parallel with the sampling process.

IV. EVALUATION METHODOLOGY AND RESULTS

We demonstrate the functionality of this framework using an example BN shown in Figure 3. The example demonstrates a simple GEN, which consists of three nodes; a gene marker which is an evidence node, and two types of a disease which the gene marker can indicate, which are the output nodes. Although the CPT values were chosen arbitrarily for demonstration purposes, they are representative of the values

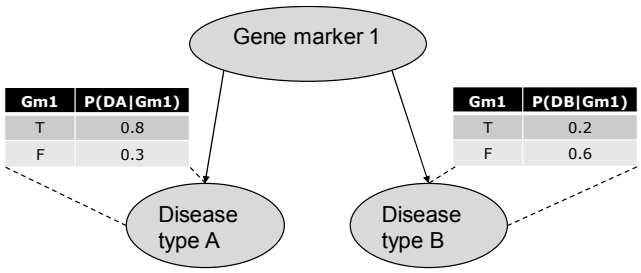

Figure 3. Example GEN for framework evaluation.

typically found in GENs.

Experimental setup for proposed framework: The device model was obtained from [9]. The stochastic switching properties were verified using Cadence Spectre tool. The MTJ device exhibits stochastic behavior up to a pulse duration of 7ns, as seen in Figure 2(b). After the stochastic switching phase is passed, another pulse is applied to reset the MTJ state. The sampling duration was set to 20 ns to ensure sufficient time for reset. The propagation delay thorough the SBN corresponds to a delay through a couple of CMOS multiplexers, which is negligible compared to the sampling duration, hence did not affect the overall performance. Stochastic inference over the example BN was performed as follows – the BN was sampled 100 times each with the evidence of gene marker being true and false, and the outcomes were recorded.

Experimental setup for baseline software simulation: AgenaRisk software [13], widely used for Bayesian inference applications. The software was run on a system with quad-core Intel i7 processor clocked at 3.6GHz. The software performed analytical inference over the example BN. The computation time required for the inference was measured.

The results of both experiments are shown in Table I. Results show that the inference values obtained by stochastic sampling in the proposed framework is within 5% of the expected values. As these inference values are used for a scoring function, the precision obtained is sufficient for that use case. The proposed framework has an estimated runtime of 2µs (100 runs of 20ns). This gives a speedup of five orders of magnitude (~84000x) for inference operation when compared to software baseline. As discussed in Section II, a structure discovery algorithm involves scoring of multiple BN structures. The scoring mechanism involves multiple inference operations at each step. This repeated nature of inference computation makes it the most critical aspect in structure discovery. While BN structure discovery involves several other aspects like mapping BN structures onto the framework and extracting samples from the nodes, the speedup in inference alone, obtained using this framework, will result in a faster structure discovery.

V. Conclusion and Future Work

GENs show promise in the domain of personalized medicine. The abundance of genetic data enables Bayesian learning to model GENs into BNs. This learning mechanisms are currently bottlenecked due to the limitations of software running solely on conventional von-Neumann machines. The proposed framework uses stochastic behavior of emerging magneto-electric devices to accelerate the structure discovery process of Bayesian learning by reducing the runtime for BN inference by five orders of magnitude. The performance benefits provided by the proposed architecture could enable learning a GEN of the size of human genome, thus bringing about a paradigm shift in personalized medicine. Future work shall involve evaluating this framework for larger GENs, which currently take hours vs. milliseconds to compute in software using contemporary computing platforms, and to observe the performance benefits.

References

[1] H. Mathkour and M. Ahmad, "A comprehensive survey on genome sequence analysis," 2010 Int. Conf. Bioinforma. Biomed. Technol., pp. 14–18, 2010.

[2] T. O'Mara et al., "A genome-wide association study to identify genetic markers associated with endometrial cancer grade," *Hered. Cancer Clin. Pract.*, vol. 10, no. Suppl 2, p. 2012, 2012.

[3] D. Bartrés-Faz et al., "APOE and APOC1 genetic polymorphisms in age-associated memory impairment," *Neurogenetics*, vol. 3, no. 4, pp. 215–219, 2001.

[4] Garnett, Mathew J., et al. "Systematic identification of genomic markers of drug sensitivity in cancer cells." *Nature* 483.7391 (2012): 570-575.

[5] Tamada, Yoshinori, et al. "Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection." *Bioinformatics* 19. suppl 2 (2003): ii227-ii236.

[6] J. Pearl, Probabilistic reasoning in intelligent systems: Networks of plausible inference, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.

[7] K. Murphy and S. Mian, "Modelling gene expression data using dynamic Bayesian networks," *Graph. Models*, p. 12, 1999.

[8] S. Misra, M. Vasimuddin, K. Pamnany, S. P. Chockalingam, Y. Dong, M. Xie, M. R. Aluru, and S. Aluru, "Parallel Bayesian Network Structure Learning for Genome-Scale Gene Networks," *SC14 Int. Conf. High Perform. Comput. Networking, Storage Anal.*, pp. 461–472, 2014.

[9] Y. Zhang et al., "Compact modeling of perpendicular-anisotropy CoFeB/MgO magnetic tunnel junctions," *IEEE Trans. Electron Devices*, vol. 59, no. 3, pp. 819–826, 2012.

[10] Chickering, D.M., Heckerman, D. and Meek, C., 1997, August. A Bayesian approach to learning Bayesian networks with local structure. In *Proceedings of the 13th conference on UAI* (pp. 80-89). Morgan Kaufmann Publishers Inc.

[11] Gamerman, Dani, and Hedibert F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press, 2006.

[12] A. P. Dancy and A. P. Chandrakasan, "Ultra low power control circuits for PWM converters," *Power Electron. Spec. Conf. 1997. PESC '97 Rec. 28th Annu. IEEE*, vol. 1, pp. 21–27, 1997.

[13] AgenaRisk software, available: "http://agenarisk.com/"