Non von-Neumann Anomaly Detection in Multi-Channel Time-Series using Charge Trap Transistor Crossbars

Ahish Shylendra, Priyesh Shukla, and Amit Ranjan Trivedi
University of Illinois at Chicago, USA. Email: amitrt@uic.edu.

***Abstract*—Many of the existing techniques to detect anomalies in multi-channel time-series lack flexibility and incur significant processing-overheads; therefore, real-time, flexible anomaly detection in resource-constrained edge devices is still an open problem. Addressing the above challenges, we present an ultra-low-power non von-Neumann framework for statistical modeling based anomaly detection in multi-channel time-series. To disruptively minimize the power dissipation for anomaly detection and enhance scalability to complex time-series statistics, we pursue a *co-designed* approach where the anomaly statistics are modelled using an unconventional Harmonic-Mean of Gaussian-like (HMG) functions. We show that multivariate HMG functions can be implemented simply by exploiting the *short-circuit current* of multi-input inverters. A non-von Neumann crossbar of charge trap inverters implements a mixture of HMG function and stores model weights. On Yahoo *real* time-series dataset, our anomaly detection approach achieves an f1-score higher than 0.85 even in the presence of significant process variation and consumes 181fJ/sensor sample for a three-channel time-series. Compared to baseline digital and analog approaches for anomaly detection, our framework is $\sim$40$\times$ and $\sim$6$\times$ more energy efficient, respectively, while being more scalable to time-series dimension.**

I. Introduction

Over the last decade, internet of things (IoT) has made tremendous progress and continues to evolve to create an evermore connected world by coalescing technology and human interaction. It is predicted that by 2025 the number of networked IoT devices will grow to more than 30 billion [1]. Unfortunately, such a massive network of heterogeneous IoT devices poses open serious security threats [2]. In a hostile environment, under cyber-attacks or tampering, sensed data from IoT devices can be aberrant and erroneous. Operating a system with compromised sensing peripherals leads to failure [3].

Since IoTs use thousands of sensors to continuously monitor different attributes of a physical process, sensed data is often high-dimensional and large scale. High dimensionality and large scale of IoT data worsens the complexity of anomaly detection (AD) algorithms. Existing supervised, semi-supervised and unsupervised algorithms are infeasible for AD as it's impractical to have a complete knowledge of root-cause of various anomalies and/or they lack flexibility in handling dynamic operating conditions [3]–[8]. Statistical modeling-based AD learns the statistical features of IoT data and the learned statistics is used for AD. Unlike classification and NN-based approaches, anomaly detection using statistical models can be computationally lightweight as well as easy to update. Therefore, in this work, we pursue multivariate, unsupervised statistical modeling approach for anomaly detection in time-series data.

Gaussian-Mixture models (GMM) are widely used for unsupervised statistical modeling. Although GMM based statistical modeling is relatively light-weight compared to NN-based approaches, it is still computational intensive as its conventional digital implementation requires multiplication, addition, subtraction and accessing look-up table for each dimension [9]. Therefore, it is infeasible to use conventional GMM statistical modeling approach for on-the-edge anomaly detection. Addressing above challenges, our key contributions are:

- We present a novel *co-designed* approach for ultra low-power multivariate statistical modeling. We show that, unlike multivariate Gaussian function, harmonic-mean of Gaussian (HMG) functions can be much easily implemented by multi-input inverters by exploiting their short-circuit current for computing. A mixture of HMG also provides high fidelity of statistical modeling and shows comparable accuracy to the conventional approach, yet with a fraction of energy overhead.
- By exploiting threshold voltage (V_{TH}) programmability of charge trap (CT) transistors, we discuss the design of *non von-Neumann crossbars for statistical modelling*. Each column of the crossbar hosts an HMG mixture function. All mixture functions are computed in parallel. For d-dimensional time-series, crossbar requires d-rows of CT transistors whereas number of columns are governed by time-series complexity. Therefore, even for high-dimensional, complicated time-series statistics, computing overheads are minimal.
- We demonstrate the efficacy of our approach on Yahoo time-series dataset [10]. Our model parameters are learned by Expectation-Maximization (EM) for HMG mixtures. Our approach achieves an f1-score higher than 0.85 for anomaly detection on Yahoo *real* time-series dataset [10] even in the presence of significant process variation. Our approach consumes 181fJ with 150 HMG mixture components

Sec. II discusses the background on CT transistors and non von-Neuann crossbar architecture using multi-input CT inverters for HMGM based statistical modeling and anomaly detection. Sec. III discusses TCAD modeling of CT. Sec. IV presents accuracy results on benchmark dataset. Sec. VI concludes.

II. Non von-Neumann Charge Trap Crossbars for Ultra-low-power Statistical Modeling

A. Charge Trap (CT) Transistor

Charge Trap (CT) transistors exploit oxygen vacancy defects created in HfO_2 due to metal deposition for charge-trapping based embedded non-volatile memory (eNVM) effect. Key advantages of CT-based eNVM is technology scalability to sub-30 nm and FinFET dimensions as well as low programming voltages [11]. In [11], for 30-nm channel length CT transistors, short-duration pulses of magnitude $\sim$1.5 V were applied to the gate (V_G) while biasing the drain at $\sim$1.3 V and grounding the source/body to program. Programming pulses induce vertical field assisted hot-electron-injection to gate-dielectric traps; thereby, V_{TH} of the transistor increases. Conversely, V_{TH} decreases due to charge detrapping when negative pulses is applied to the gate and the source, drain, and body are grounded. Notably, CTs can be programmed/erased with voltages under 2V [11], allowing on-chip low-power programming using standard input/output (I/O) transistors. Moreover, as discussed in [11], CT-eNVM requires minimal changes in the standard fabrication flow.

B. Exploiting CT Inverters for Gaussian-like Kernel Function

In a CMOS inverter, short-circuit current flows during input signal transition when both pull-up and pull-down devices are turned on simultaneously. Prior work [12] has shown that inverter's short-circuit current can be modeled using a "Gaussian-like" function as

$$I_{\text{INV}} \approx I_{0,\text{INV}} \exp\Big(- \frac{(V_{\text{IN}} - \mu)^2}{\sigma^2(\alpha + |V_{\text{IN}} - \mu|)} \Big). \quad (1)$$

Here, μ is the V_{IN} at which the inverter current peaks. μ depends on threshold voltage of NMOS and PMOS, i.e., $V_{\text{TH,n}}$ and $V_{\text{TH,p}}$. α

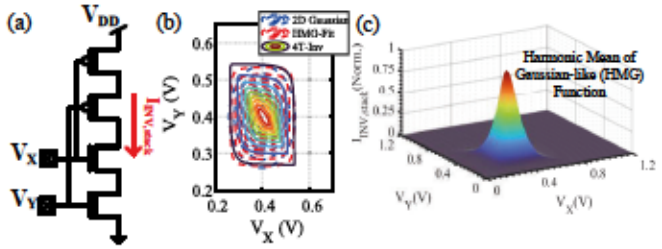

Figure 1: (a) Circuit diagram of two-input inverter, (b) Short-circuit current in a two-input inverter closely matches harmonic-mean of Gaussian-like (HMG) function, and (c) Two-input inverter's short-circuit current as a function of V_X and V_Y. Simulations are performed using 22-nm technology models [13].

is a technology-dependent fitting parameter. σ models the variance of Gaussian-like I_{INV} and it depends on the thermal voltage and transistor's ideality factor, η. μ can be modulated by a Δ amount by programming NMOS and PMOS threshold voltages as $V_{TH,n} \rightarrow V_{TH,n} + \Delta$ and $V_{TH,p} \rightarrow V_{TH,p} - \Delta$, respectively. Importantly, with the above relation, only μ changes without affecting the peak current.

Interestingly, inverter's short circuit current is well-suited to physically emulate a Gaussian-like kernel function for statistical modeling. The current is unimodal and with appropriate transistor sizing can be made symmetric. By designing inverter with CT technology, V_{TH} of NMOS and PMOS can be programmed, thereby the mean (μ) of Gaussian-like kernel function can be set as desired.

C. Co-designing Kernel with Current Conduction Physics

Although inverter's short-circuit current in (1) greatly simplifies the implementation of a univariate kernel for statistical modeling, multivariate kernels are still expensive by requiring the product of the 1-D kernels. Particularly, analog multiplication is needed since the output of inverter-based realization is in the current-mode. Hence, with increasing dimensionality of statistical model, even with simplified univariate kernel, the necessary analog multiplications for multi-variate kernels become a critical bottleneck.

Overcoming this limitation, we propose a *co-design* approach where a multivariate kernel is realized through the harmonic-mean of 1-D kernels instead of their product. Fig. 1(a) shows four transistor CT inverter controlled by two input voltages, V_X, V_Y. Fig. 1(b-c) show contour and surface plots respectively of $I_{INV,stack}$ while varying V_X and V_Y. Since the column current shows a Gaussian-like sensitivity to each input channel (V_X or V_Y), the overall column current emulates a harmonic mean of Gaussian-like (HMG) function, following the Kirchhoff's law. Fig. 1(b) shows a good correlation between SPICE-simulated characteristics and modelled characteristics using the harmonic mean of Gaussian (HMG). Notably, in Fig. 1(b), the co-designed 2D kernel function matches quite well with a 2D Gaussian except at the tail of the distribution. Unlike multivariate Gaussian, HMG functions are realized much simply. For d-dimensional HMG, d-NMOS and d-PMOS transistors need to be stacked in a column. A d-dimensional HMG function can be analytically defined as

$$I_{INV,stack} \approx \frac{I_{0,INV}}{\sum_{i=1,\ldots,d} \exp\left(\frac{(V_i-\mu_i)^2}{\sigma^2(\alpha+|V_i-\mu_i|)}\right)} \quad (2)$$

where μ_i are mean of constituent Gaussian-like functions and can be programmed by controlling the V_{TH} of corresponding transistors in the inverter-stack. $I_{0,INV}$ is the peak current and depends on the transistor sizing, technology, and supply voltage.

D. Non-von-Neumann CT Crossbars for Statistical Inference

Using the above multi-input CT inverters as building blocks, a mixture of HMG functions can be implemented by connecting inverters in Fig. 1(a) in parallel, where each inverter implements one mixture component. A d-dimensional input voltage is applied row-wise to all the inverters in the array. Resulting total current through

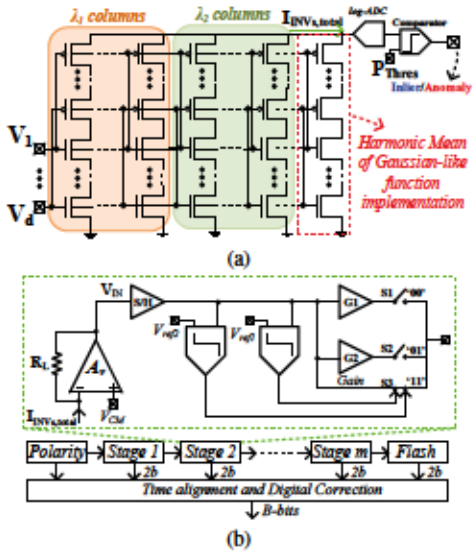

Figure 2: Multi-input inverter-stack columns (with each column implementing a multi-variate HMG function) can be connected in parallel to implement multivariate mixture model, where overall $I_{INVs,total}$ is proportional to the likelihood estimate of the input (V_1,..,V_d).

parallel inverters will be proportional to the likelihood of the input voltage. The log-likelihood of a d-dimensional input vector, $\mathbf{V}^{[1-d]}$, can be computed using HMGM as

$$\ell(\mathbf{V}^{[1-d]}|\boldsymbol{\Theta}) = \sum_{i=1}^{M} \ln \sum_{j=1}^{N} \lambda_j I_{INV,stack}(V_i^{[1-d]}; \mu_j^{[1-d]}, \sigma, \alpha) \quad (3)$$

where $V_i^{[1-d]}$ is a d-dimensional input sample, $\boldsymbol{\Theta}$ denotes the HMG mixture model parameters including λ_j (mixing proportion) and $\mu_j^{[1-d]}$ (mean) for the respective mixture component j. The parameters $\boldsymbol{\Theta}$ are learned using the Expectation-Maximization algorithm (EM) [12]. Although in general EM also learns the variance σ for each mixture component respectively, we chose a fixed σ, achieved by multi-input CT inverter in Fig. 1, for all mixture components. This simplifies our crossbar implementation, however, also requires relatively more mixture components to achieve a matching accuracy to generalized EM-based learning. For a crossbar with N columns, λ_j for a mixture component is implemented by dedicating round($\lambda_j \times N$) columns for the component j. $\mu_j^{[1-d]}$ is encoded by programming V_{TH} of respective CT inverters for the component j.

Although similar non-von-Neumann statistical model inference was considered in [14], [15], it required unconventional technologies such as III-IV SOHTFET whereas the current work achieves similar capacity with CT transistors that follow conventional fabrication schemes and can be integrated with typical CMOS modules. In [16], [17], bump circuits were used for a similar Gaussian mixture model-based inference, however, high power consumption of each kernel circuit limits the scalability of the approaches.

E. Anomaly Detection (AD) with Non-von-Neumann CT Crossbar

In Fig. 2, total crossbar current, $I_{INV,total}$, is proportional to the likelihood of input voltage $\mathbf{V}_{sample}^{[1-d]}$ applied at their gates. Subsequently, $I_{INV,total}$ is digitized using logarithmic ADC to obtain log-likelihood in digital domain. Further, the computed log-likelihood is compared against pre-defined likelihood threshold (P_{TH}) to detect if $\mathbf{V}_{sample}^{[1-d]}$ is an inlier or an anomaly, following the rule below

$$V_{sample}^{[1-d]} = \begin{cases} \text{Outlier}, & \ell(V_{sample}^{[1-d]}|\boldsymbol{\Theta}) < P_{TH} \\ \text{Inlier}, & \ell(V_{sample}^{[1-d]}|\boldsymbol{\Theta}) > P_{TH} \end{cases} \quad (4)$$

For log-ADC, transimpedance amplifier (A_v) with load-resistance R_L converts $I_{INVs,total}$ to an analog voltage. R_L is chosen depending number of mixture components, to provide full-range output. Moreover,

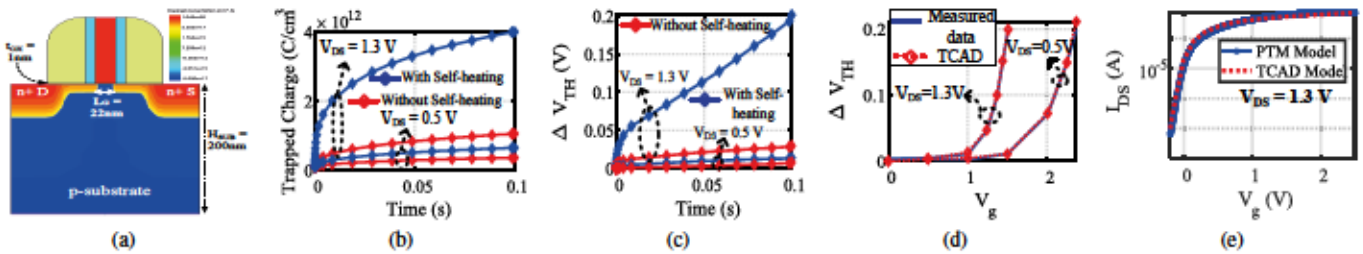

Figure 3: Device simulation study of eNVM Charge-Trap Transistors. (a) Structure of 22nm NMOS planar device used in TCAD simulation, (b) Hot-electron injection along with self-heating enhances tunneling of electrons from channel to gate-oxide, (c) Device V_{TH} increases with time at V_{GS} = 1.5V and V_{DS} = 1.3 V, due to charge-trapping, (d) TCAD model calibration with measurement results from [11] and (e) PTM model is [13] is tuned to match TCAD model characteristics.

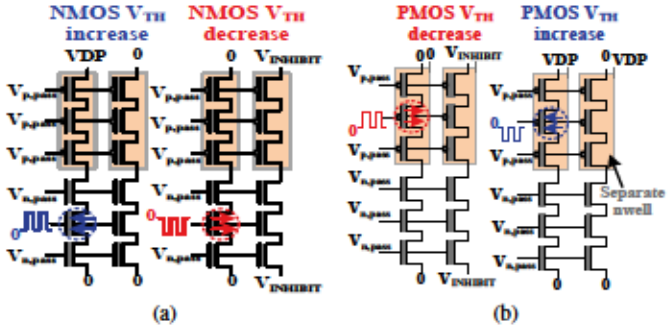

Figure 4: Circuit configuration to program/erase (a) NMOS devices, (b) PMOS devices, in CT crossbar.

Table I: Characteristics of Yahoo 3D-datasets

Dataset	1	2	3	4
Number of Data-points	1461	1439	1441	1461
Number of Anomalies	50	43	80	34

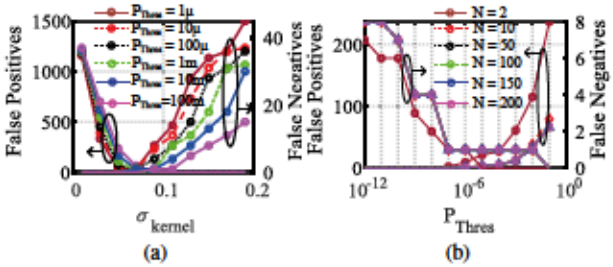

Figure 5: F_P and F_N performance of discussed approach, (a) for different P_{thres} and σ_{kernel}, (b) for different P_{thres} and N.

negative feedback configuration of A_v biases parallel, multi-input, inverters at V_{CM}. An N-bit log-ADC in Fig. 2(b) converts V_{IN} to equivalent digital-bits using the expression below [18]

$$\log_{10}\left(\frac{V_{IN}}{V_{range}}\times 10^{C}\right)=\frac{b_{N-1}2^{N-1}+\ldots+b_0}{2^N}C \qquad (5)$$

where C is code efficiency factor for improving dynamic range.

Interestingly, above non von-Neumann *AD* framework can operate directly on analog inputs $V_{sample}^{[1-d]}$. This is favorable since not only *AD* is pushed close to data source for low-latency detection but also downstream processing overheads (such as digitization and storage) can be averted for anomalous samples. Notably, even though the crossbar processing is in analog-domain, domain conversion overheads (log-ADC in Fig. 2) does not scale with signal dimensions and/or detection model complexity. With higher signal dimension, more rows in the crossbar are needed; likewise, with higher signal complexity, more columns in the crossbar are needed to inhabit more mixture functions in the model. Yet, only one log-ADC at the final digitization step is needed.

III. Simulation Results and Methodology

A. TCAD Modeling of Charge-Trap (CT) Transistors

Fig. 3 shows TCAD modeling of the CT transistor characteristics and their calibration against the measurements reported in [11]. Device simulations were performed using Sentaurus TCAD [19]. In Fig. 3(a), channel-length (L_G) of planar NMOS device is 22nm, HfO_2 thickness is 1nm, and substrate is 200nm thick. Substrate is Boron-doped with a concentration $5\times10^{17}cm^{-3}$ and Source/Drain are Phosphorous doped at $1\times10^{20}cm^{-3}$ concentration. Physics models including self-heating related thermodynamic model, charge-trapping related Fowler-Nordheim, hot-carrier injection, mobility degradation and band-gap narrowing models are used for simulation. In Fig. 3(b), trap-concentration in HfO_2 increases over time due to hot-electron injection when stressed with a high gate-voltage (V_g = 1.5V). Moreover, higher V_{DS} enhances charge-trapping due to self-heating [11]. Consequently, in Fig. 3(c), device's V_{TH} can be modulated with controlled charge-trapping by modulating V_{DS}. Conversely, when negative pulses are applied to gate and drain/source are grounded, device V_{TH} decreases due to charge detrapping. In Fig. 3(d), TCAD model is tuned to match the measurement characteristics presented in [11]. Furthermore, SPICE-level predictive technology model is calibrated to match the TCAD characteristics in Fig. 3(e).

In Fig. 4(a), while programming NMOS V_{TH} to a higher value, a particular column is selected by applying a high voltage, V_{DP}, across the column. The unselected columns are grounded. A programming pulse is applied to the gate of the selected NMOS device and the gate of other devices are held at a passing potential, V_{pass}. V_{pass} is chosen appropriately to inhibit HCI by ensuring the gate-channel potential is not high. Such biasing condition increases V_{TH} of selected device through HCI. On the contrary, to decrease V_{TH} of NMOS, drain/source of the selected NMOS is connected to the ground and a high-voltage negative pulse is applied to the gate. This causes the electrons in the gate-dielectric to tunnel back and detrap.

Fig. 4(b) illustrates the procedure to program V_{TH} of a PMOS device in CT crossbar. Each column is designed to have separate n-wells and body contacts. While programming, source/body terminals are applied with V_{DP}, drain is grounded and V_{TH} of the selected PMOS is increased by applying negative gate pulses. High gate-source potential causes holes to tunnel to gate-dielectric from the channel. Gate of the unselected PMOS devices are supplied with a pass voltage V_{pass} and source/n-well of the unselected column is connected to $V_{INHIBIT}$ to inhibit tunneling. On the other hand, V_{TH} of a PMOS device is decreased by grounding source/body/drain and applying positive gate pulses, causing holes to detrap. In this case, the unselected columns are biased at $V_{INHIBIT}$ to avoid detrapping.

B. Anomaly Detection (AD) Accuracy on Yahoo Dataset

We have used 3D time-series traces from Yahoo's real production traffic dataset [10] to benchmark the performance of our multi-

Table II: f1-score performance of our approach for different 3D-timeseries datasets with σ_{HMG}=0.07, N=150

Dataset	P_{thres}					
	1μ	10μ	100μ	1m	10m	100m
1	0.98	0.96	0.94	0.88	0.74	0.60
2	1	1	1	1	0.86	0.71
3	0.8648	0.8743	0.8791	0.8791	0.8938	0.8938
4	0.829	0.85	0.8395	0.76	0.459	0.272

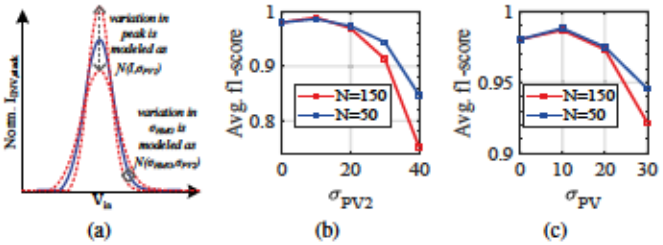

Figure 6: (a) Process-induced variations changes HMG functions' peak-amplitude and σ_{HMG}, (b) At higher σ_{PV2}, AD performance degrades. and (c) AD performance degrades σ_{PV1}=σ_{PV2}=σ_{PV} >20%.

channel AD approach. This dataset consists of 67 univariate time-series traces and captures anomalies including outliers & change-points, i.e., captures typical real-world sensor data attributes. Table I summarizes properties of used 3D time-series datasets.

Meanwhile, for AD, algorithmic parameters such as HMG function standard deviation (σ_{HMG}) and number of mixtures components (N) determines the accuracy of estimated density. Additionally, σ_{HMG} and likelihood-threshold (P_{thres}) parameters define classification hyper-plane for AD. Hence, optimal choice of these parameters is imperative. Fig. 5(a) shows false positives (F_P) and false negatives (F_N) performance of our approach $vs.$ σ_{HMG}, for various P_{thres} and for N=150. F_P increases at low σ_{HMG} due to inaccurately low-likelihood estimates for inliers. Contrarily, at higher σ_{HMG}, F_N increases due to inaccurately higher likelihood estimates for outliers. Meanwhile, at high P_{thres}, increased F_P is due to inaccuracy in classification hyper-plane. At lower P_{thres}, F_N increases due to outliers having sufficiently higher likelihood than P_{thres}.

Although our approach is limited in independently programming σ_{HMG} for each mixture function to retain the simplicity of implementation through multi-input inverters, a large mixture of HMGs can be easily implemented by proportionally increasing number of crossbar columns without significant increase in area/power overheads. Fig. 5(b) shows the F_P and F_N analysis to determine optimal N and P_{thres}. Statistically, higher N provides accurate density estimates in turn improving AD efficiency. However, energy-efficiency of the proposed framework reduces with increasing N. Therefore, optimal choice of N and P_{thres} depends on trade-off between energy-efficiency and AD accuracy. Table II summarizes f1-score of our approach on different datasets. f1-score is computed as

$$f1-score = \frac{TP}{TP + 0.5 \times (F_N + F_P)} \quad (6)$$

here, TP is the number of true positives. An f1-score higher than ≈ 0.85 is achieved through optimal choice of P_{thres} for N=150 and σ_{HMG} = 0.07. Fig. 6 shows the impact of process-induced variation in HMG function on AD efficiency. Process-variation in devices affect HMG function's peak-amplitude and σ_{HMG}. Correspondingly, in Fig. 6(a), σ_{PV1} and σ_{PV2} are used to model variations in implemented HMG. In Fig. 6(b), AD performance degrades for σ_{PV2} >20%. Fig. 6(c) shows combined effect of σ_{PV1} and σ_{PV2} on AD performance. In Fig. 6(c), AD performance degrades for σ_{PV} >20% , where σ_{PV1}=σ_{PV2}= σ_{PV}. Interestingly, using fewer mixture components (N=50) exhibits better resilience against σ_{PV} and σ_{PV2} induced variations.

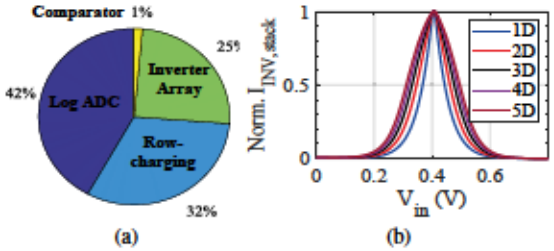

Figure 7: (a) Energy consumption of different modules in non-von-Neumann AD framework, (b) Impact of dimensionality on σ_{HMG}.

C. Power-Performance Characterization

Fig. 7(a) shows energy breakdown for modules in discussed non von-Neumann AD architecture. Our AD consumes an overall energy of 181.2fJ while operating at ∼300MHz on a 3D time-series. Log-ADC consumes 42% of the total energy, as its design includes comparators and Opamps. The energy consumption of log-ADC in [18] is projected to 22nm technology using following assumptions: Energy $\propto$ $(22nm/TechNode_{REF})^2$, Energy $\propto$ $(0.8V/VDD_{REF})^2$, and Energy $\propto$ $2^{(4-Precision_{REF})}$. Energy spent to apply inputs to the crossbar accounts for 32%. HMGM implementation using 150 multi-input inverters and a comparator accounts for 26% of the total energy. Fig. 7(b) considers scalability of our approach to increasing input dimensionality. In the figure, implemented σ_{HMG} with multi-input-inverter increases with time-series dimension. This is due to a decrease in gain at each input channel due to series connection. A consequence of increasing σ_{HMG} is that a proportionally higher programming and input range will be needed for multi-input-inverters at increasing time-series dimensionality.

D. Comparison to State-of-the-Art

Table III compares performance of our non von-Neumann AD approach with an equivalent digital and analog baseline considered in [16]. For Table III, AD in 3D time-series data is considered. Notably, multiplier is required for analog baseline design for 3D time-series processing, therefore, multiplier in [20] is used. Performance parameters of baseline designs are projected using the assumptions mentioned in previous section. ADC overheads in baseline design and in non-von-Neumann AD approach will be similar, therefore, it is not included in energy comparison. Notably, proposed approach is at the least ∼6× energy efficient compared to baseline AD approaches. Moreover, our AD approach combines positive aspects of both analog and digital reference designs in [16].

Table III: Comparison of our approach with state-of-the-art

Design	Analog KDE	Digital KDE	CT Analog (This work)
Technology Scalability	Limited	Higher	Higher
AD Latency	Low	High	Low
Energy/3D-sample	79.8fJ	536.41fJ	13.44fJ
Detection Model Scalability	Limited	Scalable	Scalable

IV. Conclusion

We presented an ultra-low-power framework for anomaly detection in multi-channel time-series data. Our approach leverages short circuit current of inverters based on CT to realize a HMG function for statistical modeling of time-series. Even more, multivariate HMG function is co-designed with kernel model against current conduction principles in a multi-input inverter. Specifically, we showed that a harmonic mean of Gaussian-like function, instead of multivariate Gaussian, is easier to implement as well as provides high accuracy AD. Many HMG functions can be evaluated in parallel through the proposed CT crossbars, enabling high-performance. The proposed non-von-Neumann crossbar architecture for AD consumes 181fJ. Moreover, proposed AD framework is our framework is ∼40× and ∼6× more energy efficient compared to baseline designs.

REFERENCES

[1] S. Zhang, Q. Wu, S. Xu, and G. Y. Li, "Fundamental green tradeoffs: Progresses, challenges, and impacts on 5g networks," *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 33–56, 2017.

[2] A. Ephremides, "Energy concerns in wireless networks," *IEEE Wireless Communications*, vol. 9, no. 4, pp. 48–59, 2002.

[3] A. A. Cook, G. Mısırlı, and Z. Fan, "Anomaly detection for iot time-series data: A survey," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481–6494, 2020.

[4] Y. Guo, W. Liao, Q. Wang, L. Yu, T. Ji, and P. Li, "Multidimensional time series anomaly detection: A gru-based gaussian mixture variational autoencoder approach," in *ACML*, 2018.

[5] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier Detection Techniques for Wireless Sensor Networks: A Survey," *IEEE Communications Surveys Tutorials*, vol. 12, no. 2, pp. 159–170, Feb. 2010.

[6] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient Algorithms for Mining Outliers from Large Data Sets," in *Proc. ACM SIGMOD International Conference on Management of Data*, 2000, pp. 427–438.

[7] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "High-Dimensional and Large-Scale Anomaly Detection Using a Linear One-Class SVM with Deep Learning," *Pattern Recognition*, vol. 58, pp. 121–134, 2016.

[8] E. Elnahrawy and B. Nath, "Context-Aware Sensors," in *Wireless Sensor Networks*. Springer Berlin Heidelberg, 2004, pp. 77–93.

[9] A. Gianelli, N. Iliev, S. Nasrin, M. Graziano, and A. R. Trivedi, "Low power speaker identification using look up-free gaussian mixture model in cmos," in *2019 IEEE Symposium in Low-Power and High-Speed Chips (COOL CHIPS)*. IEEE, 2019, pp. 1–3.

[10] YahooLabs, "S5 - a labeled anomaly detection dataset," https://webscope.sandbox.yahoo.com/, accessed: 11/08/2018.

[11] F. Khan, E. Cartier, J. C. S. Woo, and S. S. Iyer, "Charge trap transistor (ctt): An embedded fully logic-compatible multiple-time programmable non-volatile memory element for high- k -metal-gate cmos technologies," *IEEE Electron Device Letters*, vol. 38, no. 1, pp. 44–47, 2017.

[12] P. Shukla, A. Muralidhar, N. Iliev, T. Tulabandhula, S. B. Fuller, and A. R. Trivedi, "Ultralow-power localization of insect-scale drones: Interplay of probabilistic filtering and compute-in-memory," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, pp. 1–13, 2021.

[13] Y. Cao and W. Zhao, "Predictive technology model for nano-cmos design exploration," in *2006 1st International Conference on Nano-Networks and Workshops*, 2006, pp. 1–5.

[14] A. R. Trivedi and A. Shylendra, "Ultralow power acoustic feature-scoring using gaussian iv transistors," in *Proceedings of the 55th Annual Design Automation Conference*, 2018, pp. 1–6.

[15] A. Trivedi, R. Pandey, H. Liu, S. Datta, and S. Mukhopadhyay, "Gate/source overlapped heterojunction tunnel fet for non-boolean associative processing with plasticity," in *2015 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2015, pp. 17–8.

[16] A. Shylendra, P. Shukla, S. Mukhopadhyay, S. Bhunia, and A. R. Trivedi, "Low power unsupervised anomaly detection by nonparametric modeling of sensor statistics," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, no. 8, pp. 1833–1843, 2020.

[17] A. Shylendra, S. H. Alizad, P. Shukla, and A. R. Trivedi, "Non-parametric statistical density function synthesizer and monte carlo sampler in cmos," in *2020 33rd International Conference on VLSI Design and 2020 19th International Conference on Embedded Systems (VLSID)*. IEEE, 2020, pp. 19–24.

[18] J. Lee, J. Kang, S. Park, J. Seo, J. Anders, J. Guilherme, and M. P. Flynn, "A 2.5 mw 80 db dr 36 db sndr 22 ms/s logarithmic pipeline adc," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 10, pp. 2755–2765, 2009.

[19] "Tcad sentaurus device manual, usa, ca, mountain view:synopsys, inc., 2017."

[20] A. S. A. Naim and S. H. Ruslan, "130 nm low power CMOS analog multiplier," vol. 1049, p. 012071, jul 2018. [Online]. Available: https://doi.org/10.1088/1742-6596/1049/1/012071