# A Low-Power Analog Integrated Deep Spatio-Temporal Inference Network with Application to Digit Classification

Vassilis Alimisis, Nikolaos P. Eleftheriou and Paul P. Sotiriadis

Department of Electrical and Computer Engineering
National Technical University of Athens, Greece
E-mail: alimisisv@gmail.com, eleftheriou_nikos@hotmail.com , pps@ieee.org

2023 International Conference on Microelectronics (ICM) | 979-8-3503-8082-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICM60448.2023.10378893

***Abstract*—In the realm of advanced machine learning, a burgeoning paradigm, known as deep machine learning, has emerged to address intricate, high-dimensional data in a structured manner, taking cues from biological inspirations. This study contributes novel findings utilizing a recently introduced deep learning framework, termed the Deep Spatio-Temporal Inference Network. It is a discriminative architecture in deep learning, amalgamates elements from unsupervised learning to establish dynamic pattern representation, concurrently incorporating Bayesian model. The proposed inference is composed of parallel-connected Mahalanobis distance circuits and a distance comparator circuit. As a result, this work proposes a novel low power ($2.43\mu W$), low voltage ($0.6$V) analog architecture of a Deep Spatio-Temporal Inference Network with application to digit classification. Confirmation of the analog classifier's effective functioning is achieved through validation with a real-world dataset ($93.15\%$ accuracy). The implementation of the proposed architecture is executed within the TSMC 90nm CMOS process, and its behavior is simulated utilizing the Cadence IC Suite.**

***Index Terms*—Deep Spatio-Temporal Inference Network, digit classification, low-power design, analog VLSI implementation**

## I. Introduction

Digital imaging technology has revolutionized the conventional use of film by transitioning to a realm of bits and bytes [1]. In this paradigm, the quality of an image is gauged by the pixel count it possesses. The heightened resolution of an image corresponds to an increased abundance of these diminutive yet vividly coloured dots [2]. Unlike the traditional camera, which relies on lenses to focus light onto film for image formation, the digital camera employs an image sensor. This sensor, often a CMOS or a charge coupled device (CCD), undertakes the task of translating light into electric charges [2].

The CMOS image sensor, notably found in smartphones, employs color-filter layers to impart hues, while photodiodes perform the crucial role of converting light into electrical signals [2], [3]. This amalgamation culminates in the creation of a digital image, further refined through on-chip image processing in certain applications such as artificial vision and image recognition [3]. On the other hand, CCD image sensors, a preferred choice in machine-vision systems, embody transistorized light sensors on an integrated circuit [4]. These sensors meticulously integrate received light, transmuting electrons into the electrical signals that ultimately manifest as video or still images in various formats [4]. This diversity in sensor technology underscores their respective contributions to distinct domains of visual technology.

Deep-learning neural networks excel across various applications, from speech recognition to self-driving cars [5]. They succeed at deciphering complex patterns in datasets, often surpassing human capabilities. In camera-related applications, diverse neural network variants enhance image quality by addressing blurriness, enhancing colours, and rectifying pixel issues [6]. They also excel at specific tasks like isolating regions of interest. For instance, in surveillance, these networks create feature maps that highlight crucial parts of an image, such as facial details or pedestrian counts [7]. This focused approach reduces memory and computational demands, crucial for resource-efficient edge applications.

The motivation is based on the power and area efficiency requirements of image sensors [8], [9], this paper proposes a new, power-efficient and analog hardware architecture for deep learning that integrates principles from unsupervised learning for dynamic pattern representation alongside Bayes inference. This is called Deep Spatio-Temporal Inference Network. The implemented network is a promising classifier appropriate for battery dependent image smart sensor classification systems, since it achieves $93.15\%$ accuracy. It is implemented and confirmed on a measured digit recognition dataset [10]. The accuracy of the proposed implementation is confirmed through post-layout simulation results obtained in a TSMC $90$nm CMOS process and simulated using Cadence IC Suite. This validation involves a comparison with a software-based implementation. Furthermore, a comprehensive comparison study between the proposed classifier and analog classifiers is included for the sake of thoroughness.

The rest of this work is ordered in the following manner. In Section II the the characteristics of the implemented network are explained and a clarification of its mathematical foundations is provided. The suggested structure and the fundamental components of the proposed classifier are outlined in Section III. The desired behavior of the implemented network is verified via a digit classification dataset and a comparison

with the software-based counterpart is presented in Section IV. Section V provides a comparison study with related analog classifiers. Section VI concludes this work.

## II. Deep Spatio-Temporal Inference Network Mathematical Model

Deep Spatio-Temporal Infernce Network consists of multiple instances of an identical cortical circuit, referred to as nodes [11]. Each node is a parameterized model that learns through unsupervised learning. These nodes exist across all hierarchy layers, aiming to grasp significant spatiotemporal patterns shown in presented data [11]. The lowest layer nodes take raw sensory input, e.g., image pixels, and continually develop a belief state to characterize observed sequences. Higher layers receive belief states from lower corresponding layers. These beliefs across the hierarchy are utilized as valuable features given to a classifier or regression learner, which can be trained through supervised learning.

Firstly, the selected winning centroid relies exclusively on the Euclidean distance [11], [12]. The distance $d_x$ between a centroid $x$ and an observed input o is represented as follows:

$$d_x = ||\boldsymbol{o} - \boldsymbol{\mu_x}||\psi_x. \tag{1}$$

The clustering algorithm uses the starvation trace $\psi_x$ to involve centroids initially positioned far from dense areas in the observation space. This helps centroids that might otherwise never be chosen for updates due to their distant location. This enables inactive or starved centroids to gradually adjust their perceived distance to input vectors over time. When not selected as the centroid, they decrease this apparent distance; conversely, their apparent distance increases when they are the selected centroid. The mean estimate of the winning centroid, $\mu_x$, is adjusted towards the present input along with the estimated variance $\sigma_x^2$ in a combined manner so that:

$$\mu_x \leftarrow \mu_x + \alpha(o - \mu_x), \tag{2}$$

$$\sigma_x^2 \leftarrow \sigma_x^2 + \beta[(o - \mu_x)^2 - \sigma_x^2], \tag{3}$$

where $\alpha$ and $\beta$ are positive numbers close to 0. Then, the posterior distribution $Pr(o|s)$ is derived by normalizing the Euclidean distances between the input and every centroid s, such that :

$$n_s = \sum_{i=1}^{d} \frac{(o_i - \mu_{i,s})^2}{\sigma_{i,s}^2}, \tag{4}$$

$$p_s = \frac{{n_s}^{-1}}{\sum_{s' \in S}({n_{s'}}^{-1})}, \tag{5}$$

## III. Proposed Architecture

In this section, we analyze the proposed analog implementation of the Network. The architecture presented is versatile, accommodating various numbers of classes, centroids, and input dimensions. For the implementation of the Euclidean distance function in equation (1), we employ a current-mode Mahalanobis distance circuit [13], shown in Fig. 1. This

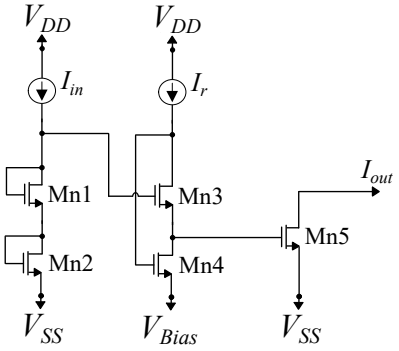


Fig. 1: The Mahalanobis distance is approximated by the translinear circuit which computes the $\frac{Iin^2}{I_r}$.

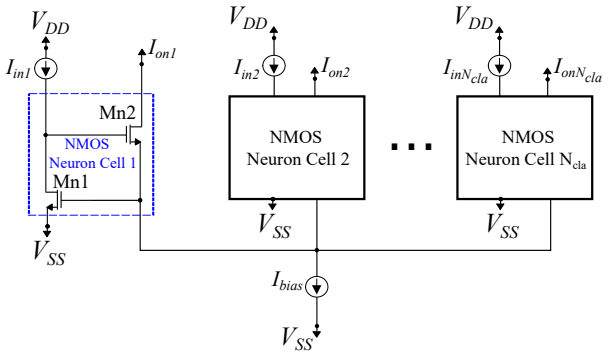


Fig. 2: A $N_{cla}$-neuron Standard Lazzaro NMOS Winner-Take-All (WTA) circuit.

circuit, which operates in a translinear fashion, embodies the mathematical expression: ${I_{in}}^2$ divided by $I_r$. Summation within the current domain is straightforward, achieved by connecting wires containing the currents to be summed.

Moving forward, the nearest centroid is determined through a distance comparison circuit, specifically a Winner-Take-All (WTA) circuit [14]. In a classification problem with $N_{cla}$ classes, the typical Lazzaro WTA circuit consists of $N_{cla}$ neurons. These neurons share a shared bias current, as depicted in Fig. 2. Each sub-circuit in the WTA circuit corresponds to an individual class. The WTA circuit effectively identifies the class with the highest input current and assigns a non-zero output current to the corresponding neuron. Simultaneously, the remaining neurons receive an output current of zero.

The architecture of the proposed network, as depicted in Fig. 3, is developed for a classification problem involving $N_{cla}$ classes and $N_d$ features (input dimensions). The quantity of centroids within each class is a hyperparameter, typically determined through exploratory data analysis. In this generalized schematic, each class comprises one centroid and $N_d$ input dimensions, illustrated in Fig. 3. The output of each Mahalanobis distance circuit (MDC) describes each input dimension. The mathematical model, as described by equations (4) and (5), involves the summation of Euclidean distances. This summation is executed within each circuit, leveraging current mirrors (CM) to minimize potential distortions in cal-

culations that might arise from undesired effects on the output currents of the Mahalanobis circuits. The classifier's prediction is denoted by the resulting output currents, characterized by high or low values. The dimensions of the transistors are equal to $W/L = 3.2\mu m/1.6\mu m$. Notably, all transistors in the mentioned designs operate in the sub-threshold region, with voltage source rails set as $V_{DD} = -V_{SS} = 0.3$ V and $I_r = 3nA$.

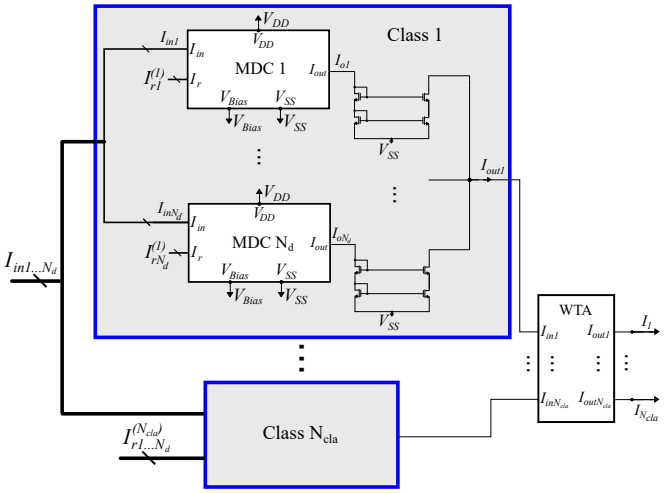


Fig. 3: Block diagram of a generic Analog Deep Spatio-Temporal Inference Network with $N_{cla}$ classes and $N_d$ features. These classes perform the summation of currents generated by the $N_d$ MDC circuits associated with each input. The final output is determined through a WTA circuit, employing a current-mode representation.

## IV. Digit Recognition and Simulation Outcomes

In this section, the proposed network is validated through a testing on a digit recognition problem [10]. The architecture presented herein is realized utilizing the TSMC $90nm$ CMOS process in conjunction with the Cadence IC suite. The power supply rails for the entire classifier are established at $V_{DD} = -V_{SS} = 0.3V$. All simulation results are derived from the layout, as depicted in Fig. 4, through post-layout simulations. This classification problem considers a handwritten digit recognition task, which consists of $N_{cla} = 10$ classes and $N_d = 64$ inputs. For comparison purposes, we have reduced the number of classes to $N_{cla} = 2$, which consists of 5 centroids per class (binary classifier, odd/even). The dataset used is provided by Python's Sklearn package [15] and consists of $8x8$ pixel images of digits. Each pixel consists of a grayscale value between 0 and 16. The classifier receives all relevant metrics directly. The required parameters for the system are determined through the calculation of the mean value, variance, and prior probability for each class.

To test the proposed classifier both in terms of classification specificity and circuit's behavior over PVT variations, two separate tests are conducted on the layout. To address the experimental variability, the results from 20 different training-test iterations are presented in Fig. 5. The sensitivity of the circuit is further validated through a Monte Carlo analysis. More specifically, Fig. 6 illustrates the Monte Carlo Histogram for N = 100 points.

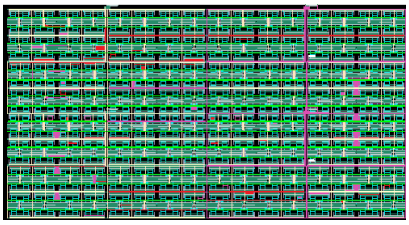

Fig. 4: Layout of the proposed Deep Spatio-Temporal Inference Network architecture based on the design methodology.

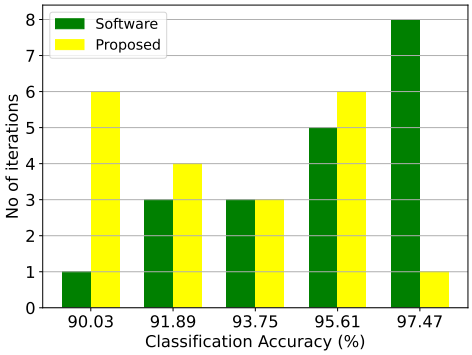


Fig. 5: Classification results of the proposed architecture and the equivalent software model on the digit recognition dataset over 20 iterations.

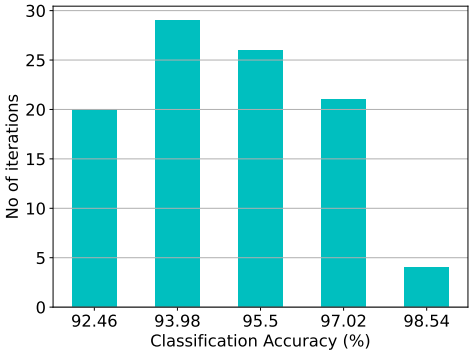


Fig. 6: Post-layout Monte-Carlo simulation results of the proposed architecture on the digit recognition dataset with $\mu_M = 94.87\%$ and a standard deviation of $\sigma_M = 1.73\%$

## V. Performance Summary and Discussion

This section aims to present a comparative analysis of various analog classifiers developed by our research team. By adjusting these classifiers to the same application as the one tested in this work a fair and unbiased comparison can be conducted. In Table I a performance summary is illustrated for a Fuzzy [16], a Gaussian Mixture Model (GMM) [17], a Bayesian [18], a Threshold [19], a Support Vector Machine (SVM) [20] and a centroid-based [21] classifier.

Firstly, our work surpasses the performance of the related analog classifiers in mean accuracy, processing speed, and

TABLE I: Analog classifiers' comparison on the Digit Recognition

| | Classifier | Min accuracy | Mean accuracy | Max accuracy | Power consumption | Processing speed | Energy per classification | No. of Dimensions |
|---|---|---|---|---|---|---|---|---|
| **This work** | Spatio | 89.10% | 93.15% | 97.20% | $2.43\mu W$ | $430K\frac{\text{classifications}}{\text{s}}$ | $\frac{5.65\text{ pJ}}{\text{classification}}$ | 64 |
| **[16]** | Fuzzy | 85.20% | 90.82% | 95.10% | $2.71\mu W$ | $4.55K\frac{\text{classifications}}{\text{s}}$ | $\frac{595{,}6\text{ pJ}}{\text{classification}}$ | 13 |
| **[17]** | GMM | 77.70% | 83.72% | 88.90% | $3.38\mu W$ | $100K\frac{\text{classifications}}{\text{s}}$ | $\frac{33.8\text{ pJ}}{\text{classification}}$ | 13 |
| **[18]** | Bayes | 73.40% | 81.75% | 84.20% | $2.08\mu W$ | $100K\frac{\text{classifications}}{\text{s}}$ | $\frac{20.8\text{ pJ}}{\text{classification}}$ | 13 |
| **[19]** | Threshold | 78.60% | 82.55% | 86.40% | $1.21\mu W$ | $100K\frac{\text{classifications}}{\text{s}}$ | $\frac{12.1\text{ pJ}}{\text{classification}}$ | 13 |
| **[20]** | SVM | 84.40% | 85.74% | 86.90% | $82.12\mu W$ | $140K\frac{\text{classifications}}{\text{s}}$ | $\frac{586.57\text{ pJ}}{\text{classification}}$ | 13 |
| **[21]** | Centroid | 86.30% | 91.32% | 95.40% | $3.42\mu W$ | $100K\frac{\text{classifications}}{\text{s}}$ | $\frac{34.2\text{ pJ}}{\text{classification}}$ | 13 |

energy consumption per classification. It is important to emphasize that, for this specific application, we have a high input dimension number. The proposed topology offers a distinct advantage in that it obviates the need for Principal Component Analysis (PCA), enabling the utilization of all 64 input dimensions without any loss of information. To attain optimal accuracy, the remaining topologies should truncate the dimensions to 13. This is the main limitation of the previous related works (specific number of input dimensions). While our network demonstrates the ability to accurately classify all 10 classes, we transformed the problem into a binary classification scenario to facilitate a meaningful comparison with binary analog classifiers [16], [19], [20].

## VI. Conclusion

In this work, a new, power-efficient ($2.43\mu W$), low supply (0.6V) architecture of an analog Deep Spatio-Temporal Inference Network for digit recognition was proposed. The presented architecture consists of Mahalanobis distance circuits and a distance comparator circuit. All post-layout simulation results were obtained using the TSMC 90nm CMOS process and were compared with a software-based implementation and a variety of analog classifiers. The implemented architecture achieves 93.15% classification accuracy and reasonable sensitivity characteristics. It can serve as a fundamental building block in intelligent sensor systems designed for image classification.

## References


[1] H. J. Trussell and M. J. Vrhel, *Fundamentals of digital imaging*. Cambridge University Press, 2008.

[2] G. Petrie and A. S. Walker, "Airborne digital imaging technology: a new overview," *The Photogrammetric Record*, vol. 22, no. 119, pp. 203–225, 2007.

[3] A. El Gamal and H. Eltoukhy, "Cmos image sensors," *IEEE Circuits and Devices Magazine*, vol. 21, no. 3, pp. 6–20, 2005.

[4] M. Bigas, E. Cabruja, J. Forest, and J. Salvi, "Review of cmos image sensors," *Microelectronics journal*, vol. 37, no. 5, pp. 433–451, 2006.

[5] H. Osipyan, B. I. Edwards, and A. D. Cheok, *Deep neural network applications*. CRC Press, 2022.

[6] M. Gruosso, N. Capece, and U. Erra, "Human segmentation in surveillance video with deep learning," *Multimedia Tools and Applications*, vol. 80, pp. 1175–1199, 2021.

[7] M. Di Benedetto, F. Carrara, L. Ciampi, F. Falchi, C. Gennaro, and G. Amato, "An embedded toolset for human activity monitoring in critical environments," *Expert Systems with Applications*, vol. 199, p. 117125, 2022.

[8] F. Tang, D. G. Chen, B. Wang, and A. Bermak, "Low-power cmos image sensor based on column-parallel single-slope/sar quantization scheme," *IEEE Transactions on Electron Devices*, vol. 60, no. 8, pp. 2561–2566, 2013.

[9] M. Maheepala, M. A. Joordens, and A. Z. Kouzani, "Low power processors and image sensors for vision-based iot devices: a review," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1172–1186, 2020.

[10] F. Nelli and F. Nelli, "Recognizing handwritten digits," *Python Data Analytics: With Pandas, NumPy, and Matplotlib*, pp. 473–486, 2018.

[11] D. George and J. Hawkins, "Towards a mathematical theory of cortical micro-circuits," *PLoS computational biology*, vol. 5, no. 10, p. e1000532, 2009.

[12] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[13] R. J. Wiegerink, *Analysis and synthesis of MOS translinear circuits*. Springer Science & Business Media, 2012, vol. 246.

[14] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of o (n) complexity," *Advances in neural information processing systems*, vol. 1, 1988.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[16] E. Georgakilas, V. Alimisis, G. Gennis, C. Aletraris, C. Dimas, and P. P. Sotiriadis, "An ultra-low power fully-programmable analog general purpose type-2 fuzzy inference system," *AEU-International Journal of Electronics and Communications*, vol. 170, p. 154824, 2023.

[17] V. Alimisis, G. Gennis, K. Touloupas, C. Dimas, M. Gourdouparis, and P. P. Sotiriadis, "Gaussian mixture model classifier analog integrated low-power implementation with applications in fault management detection," *Microelectronics Journal*, vol. 126, p. 105510, 2022.

[18] V. Alimisis, G. Gennis, C. Dimas, and P. P. Sotiriadis, "An analog bayesian classifier implementation, for thyroid disease detection, based on a low-power, current-mode gaussian function circuit," in *2021 International conference on microelectronics (ICM)*. IEEE, 2021, pp. 153–156.

[19] V. Alimisis, G. Gennis, E. Tsouvalas, C. Dimas, and P. P. Sotiriadis, "An analog, low-power threshold classifier tested on a bank note authentication dataset," in *2022 International Conference on Microelectronics (ICM)*. IEEE, 2022, pp. 66–69.

[20] V. Alimisis, G. Gennis, M. Gourdouparis, C. Dimas, and P. P. Sotiriadis, "A low-power analog integrated implementation of the support vector machine algorithm with on-chip learning tested on a bearing fault application," *Sensors*, vol. 23, no. 8, p. 3978, 2023.

[21] V. Alimisis, V. Mouzakis, G. Gennis, E. Tsouvalas, C. Dimas, and P. P. Sotiriadis, "A hand gesture recognition circuit utilizing an analog voting classifier," *Electronics*, vol. 11, no. 23, p. 3915, 2022.