europass

Marco Angioli

Abitazione : Via San Pietro Parenzo,8, 00138, Roma, Italia

E-mail: angioli.marco@gmail.com ☎ **Telefono:** (+39) 3283598617

Data di nascita: 27/11/1997 **Luogo di nascita:** Roma, Italia **Nazionalità:** Italiana

ISTRUZIONE E FORMAZIONE

[09/2016 – 10/12/2019]

Laurea Triennale in Ingegneria Elettronica

Sapienza Università di Roma

Paese: Italia | | **Voto finale:** 100/110 | **Tesi:** Stato dell'arte nella computazione neuromorfica (neuromorphic computing)

[09/2019 – 15/07/2022]

Laurea Magistrale in Ingegneria Elettronica

Sapienza Università di Roma

| **Voto finale:** 110/110L | **Tesi:** Analisi ed implementazione di algoritmi Contextual Bandits su processori RISC-V dotati di accelerazione vettoriale riconfigurabile

[01/11/2022 – Attuale]

Dottorato in Tecnologie dell'Informazione e delle Comunicazioni (ICT), Curriculum: Ingegneria Elettronica

Sapienza University Of Rome https://phd.uniroma1.it/web/MARCO-ANGIOLI_nP1773057_IT.aspx

Città: Roma | **Paese:** Italia | **Campi di studio:** Digital Hardware Design for Artificial Intelligence, Hyperdimensional Computing, Embedded AI

ESPERIENZA LAVORATIVA

Sapienza Università di Roma

Città: Roma | **Paese:** Italia

[01/01/2022 – 31/12/2024]

Tutor universitario

Tutor del corso "Digital System Programming" per il Corso di Laurea Magistrale in Ingegneria Elettronica.

Città: Örebro | **Paese:** Svezia

[03/2025 – 07/2025]

Visiting Researcher

Attività svolta nell'ambito del Dottorato (Sapienza) presso AASS Research Centre; collaborazione con Prof. Denis Kleyko per la realizzazione di acceleratori hardware per Hyperdimensional Computing (HDC).

COMPETENZE LINGUISTICHE

Lingua madre: Italiano

Altre lingue:

Inglese

ASCOLTO B2 **LETTURA** C1 **SCRITTURA** C1

PRODUZIONE ORALE B2 **INTERAZIONE ORALE** B2

Francese

ASCOLTO A1 **LETTURA** A2 **SCRITTURA** A2

PRODUZIONE ORALE A1 **INTERAZIONE ORALE** A1

Livelli: A1 e A2: Livello elementare B1 e B2: Livello intermedio C1 e C2: Livello avanzato

COMPETENZE

Linguaggi di programmazione

C | Python | C++ | Bash | VHDL | System Verilog

Software

ModelSim | Vivado | Synopsis Fusion Compiler | Synopsis Verdi | Stm32cubeMX | Anaconda | Matlab e Simulink | Autocad | Overleaf | Inkscape | Office | draw.io

Altre competenze

Brain-Inspired Algorithms | ASIC design | FPGA design | Hyperdimensional Computing/Vector Symbolic Architecture | Arithmetic Unit | Low-power design | Hardware Accelerators | Algorithmic Optimization for AI on the edge | Communication Protocols | Machine Learning | AI and Deep Neural Networks | Data Processing | Reinforcement Learning | Embedded Linux (Petalinux, Buildroot, Yocto), SoC (FPGA) | Memory DDR3,DDR4,LPDDR4 Interfaces

Sistemi Operativi

Windows | Unix Linux

PROGETTI

Design VHDL di un'unità FMA (Fused Multiply Add) ad 8 bit

Progetto sviluppato durante il corso Digital Integrated System Architectures della Laurea Magistrale in Ingegneria Elettronica. Design dell'unità in VHDL, sintesi su Vivado, valutazione delle performance ed ottimizzazioni.

Implementazione della funzione di Motion Detection sulla board Zybo Z7-20 ed ottimizzazione tramite accelerazione hardware

Progetto svolto durante il corso Digital System Programming della Laurea Magistrale in Ingegneria Elettronica. Utilizzo dell'evaluation board Zybo Z7-20 e della telecamera esterna Pcam 5C per la realizzazione di una funzione di motion detection ad alte performance grazie all'accelerazione hardware.

Manutenzione predittiva tramite reti neurali ricorrenti

Progetto svolto durante il corso di Machine Learning for Signal Processing della Laurea Magistrale in Ingegneria Elettronica. Utilizzo di reti neurali ricorrenti su Python per la predizione di possibili malfunzionamenti delle turbine degli aerei. Utilizzo di LSTM, Bi-LSTM e GRU.

Manutenzione predittiva sui motori dei treni

Progetto sviluppato durante il corso di Computational Intelligence della Laurea Magistrale in Ingegneria Elettronica. Realizzazione di una rete che, a partire dall'analisi di campioni di pressione raccolti dai compressori dei treni, permette di prevedere guasti o possibili condizioni fallimentari del motore stesso. Utilizzo di un classificatore SVM con kernel RBF ed algoritmo di ottimizzazione genetico.

Valutazione del grado rotazionale di cifre scritte a mano tramite rete neurale convolutiva

Progetto svolto durante il corso di Circuiti e algoritmi per il Machine Learning della Laurea Magistrale in Ingegneria Elettronica. Utilizzo di una rete neurale convolutiva per la risoluzione di un problema di regressione mirato alla valutazione del grado rotazionale delle cifre scritte a mano contenute nel dataset MNIST.

PUBBLICAZIONI

[2022] **Contextual Bandits Algorithms for Reconfigurable Hardware Accelerators**

Riferimento: Accepted at ApplePies,Genova, Italy, 26-27 september 2022, published by Springer, indexed in Scopus

Authors: *Marco Angioli, Marcello Barbirotta, Abdallah Cheikh, Antonio Mastrandrea, Francesco Menichelli, Saeid Jamili, and Mauro Olivieri*

Abstract: *Reconfigurable processing cores for IoT and edge computing applications are emerging topics to calibrate costs, energy consumption and area occupation with performance and reliability on Commercial Off the Shelf (COTS) devices. This work analyzes how to take advantage of Machine Learning to potentially automate the reconfiguration process of an Hardware accelerator inside the Klessydra Vector Cooprocessor Unit (VCU) [1][2][3], choosing the best configuration according to the workload. The problem is modeled with a contextual bandits approach using the Linear UCB algorithms and validated with offline Python simulations.*

[2022] **Implementation of Dynamic Acceleration Unit Exchange on a RISC-V Soft-Processor**

Riferimento: Accepted at ApplePies,Genova, Italy, 26-27 september 2022, published by Springer, indexed in Scopus

Authors: *Saeid Jamili, Abdallah Cheikh, Antonio Mastrandrea, Marcello Barbirotta, Francesco Menichelli, Marco Angioli, Mauro Olivieri*

Abstract: Reconfigurable computing, also known as adaptive computing, exploits the reconfigurability of reprogrammable logic devices like FPGAs to perform runtime hardware reconfigurations, enabling the system to better adapt to the underlying application. By using reconfigurable computing, parts of the logic implemented on an FPGA can be dynamically changed according to the task demands during runtime. The underlying hardware can be changed to trade off performance/power or can be modified to perform functional reconfiguration, reprogramming the behavior of a functional unit. By exploiting this flexibility, we can significantly scale the performance and power efficiency of a system. We present a dynamic acceleration unit exchange on a RISC-V soft-processor, based on the open-source Klessydra-T13 RISC-V core. We demonstrate reconfiguration for functional versatility or for improving the hardware accelerator performance, providing, as a case study, an example of how a deep neural network like VGG16 can be accelerated by using runtime reconfiguration techniques.

[2023] **Automatic Hardware Accelerators Reconfiguration through LinearUCB Algorithms on a RISC-V Processor**

Riferimento: 2023 18th Conference on Ph. D Research in Microelectronics and Electronics (PRIME)

Authors: Marco Angioli, Marcello Barbirotta, Antonio Mastrandrea, Saeid Jamili, Mauro Olivieri

Abstract: Reconfigurable processors are hardware architectures that allow for the dynamic configuration of processing resources to optimize performance and power consumption, using partial reconfiguration to modify a portion of the design or update it without affecting the entire system. In this work, we present an automatic reconfiguration technique that leverages machine learning (ML) algorithms to automatically select the optimal configuration of a general-purpose hardware accelerator according to the workload and reconFigure the architecture at run-time. The problem is formulated as a Contextual Bandit (CB) case using the Linear Upper Confidence Bound (LinearUCB) algorithms and verified using the RISC-V Klessydra family cores as a case of study.

[2023] **Fault-Tolerant Hardware Acceleration for High-Performance Edge-Computing Nodes**

Riferimento: Electronics 2023

Authors: Marcello Barbirotta, Abdallah Cheikh, Antonio Mastrandrea, Francesco Menichelli, Marco Angioli, Saeid Jamili, Mauro Olivieri

Abstract: High-performance embedded systems with powerful processors, specialized hardware accelerators, and advanced software techniques are all key technologies driving the growth of the IoT. By combining hardware and software techniques, it is possible to increase the overall reliability and safety of these systems by designing embedded architectures that can continue to function correctly in the event of a failure or malfunction. In this work, we fully investigate the integration of a configurable hardware vector acceleration unit in the fault-tolerant RISC-V Klessydra-fT03 soft core, introducing two different redundant vector co-processors coupled with the Interleaved-Multi-Threading paradigm on which the microprocessor is based. We then illustrate the pros and cons of both approaches, comparing their impacts on performance and hardware utilization with their vulnerability, presenting a quantitative large-fault-injection simulation analysis on typical vector computing benchmarks, and comparing and classifying the obtained results. The results demonstrate, under specific conditions, that it is possible to add a hardware co-processor to a fault-tolerant microprocessor, improving performance without degrading safety and reliability.

[2024] **Design, Implementation and Evaluation of a New Variable Latency Integer Division Scheme**

Riferimento: IEEE Transactions on Computers 2024

Authors: Marco Angioli, Marcello Barbirotta, Abdallah Cheikh, Antonio Mastrandrea, Francesco Menichelli, Saeid Jamili, Mauro Olivieri

Abstract: Integer division is key for various applications and often represents the performance bottleneck due to its inherent mathematical properties that limit its parallelization. This paper presents a new data-dependent variable latency division algorithm derived from the classic non-performing restoring method. The proposed technique exploits the relationship between the number of leading zeros in the divisor and in the partial remainder to dynamically detect and skip those iterations that result in a simple left shift. While a similar principle has been exploited in previous works, the proposed approach outperforms existing variable latency divider schemes in average latency and power consumption. We detail the algorithm and its implementation in four variants, offering versatility for the specific application requirements. For each variant, we report the average latency evaluated with different benchmarks, and we analyze the synthesis results for both FPGA and ASIC deployment, reporting clock speed, average execution time, hardware resources, and energy consumption, compared with existing fixed and variable latency dividers.

[2024] **AeneasHDC: An Automatic Framework for Deploying Hyperdimensional Computing Models on FPGAs**

Riferimento: IEEE WCCI 2024 - The IEEE World Congress on Computational Intelligence

Authors: Marco Angioli, Saeid Jamili, Marcello Barbirotta, Abdallah Cheikh, Antonio Mastrandrea, Francesco Menichelli, Antonello Rosato, Mauro Olivieri

Abstract: Hyperdimensional Computing (HDC) is a bioinspired learning paradigm, that models neural pattern activities using high-dimensional distributed representations. HDC leverages parallel and simple vector arithmetic operations to combine and compare different concepts, enabling cognitive and reasoning tasks. The computational efficiency and parallelism of this approach make it particularly suited for hardware implementations, especially as a lightweight, energy-efficient solution for performing learning tasks on resource-constrained edge devices. The HDC pipeline, including encoding, training, and comparison stages, has been extensively explored with various approaches in the literature. However, while these techniques are mainly oriented to improve the model accuracy, their influence on hardware parameters remains largely unexplored. This work presents AeneasHDC, an automatic and open-source platform for the streamlined deployment of HDC models in both software and hardware for classification, regression and clustering tasks. AeneasHDC supports an extensive range of techniques commonly adopted in literature, automates the design of flexible hardware accelerators for HDC, and empowers users to easily assess the impact of different design choices on model accuracy, memory usage, execution time, power consumption, and area requirements.

[2024] **Exploring Variable Latency Dividers in Vector Hardware Accelerators**

Riferimento: IEEE Proceedings - 19th Conference on Ph.D Research in Microelectronics and Electronics (PRIME)

Authors: Marco Angioli, Marcello Barbirotta, Abdallah Cheikh, Antonio Mastrandrea, Mauro Olivieri

Abstract: Efficient hardware implementation of integer division remains one of the most significant challenges in the design of vector hardware accelerators, particularly in achieving a balance between performance and hardware overhead. This work uses the RISC-V Klessydra-T13 Vector Coprocessor Unit as a case study to explore the impact of variable latency division architectures for vector hardware accelerators. We analyze existing designs in the literature to identify the most effective approach, detailing the whole integration process with new instructions for supporting vector divisions in the RISC-V custom instruction set and optimizing the implementation for the target accelerator by exploiting hardware reuse. We also provide real-world computation kernels and Monte Carlo simulations to demonstrate how a variable latency divider can significantly improve the division time over traditional methods with negligible hardware overhead.

[2025] **Efficient implementation of linearucb through algorithmic improvements and vector computing acceleration for embedded learning systems**

Riferimento: 2025/1/22

As the Internet of Things expands, embedding Artificial Intelligence algorithms in resource-constrained devices has become increasingly important to enable real-time, autonomous decision-making without relying on centralized cloud servers. However, implementing and executing complex algorithms in embedded devices poses significant challenges due to limited computational power, memory, and energy resources. 2This article presents algorithmic and hardware techniques to efficiently implement two LinearUCB Contextual Bandits algorithms on resource-constrained embedded devices. Algorithmic modifications based on the Sherman–Morrison–Woodbury formula streamline model complexity, while vector acceleration is harnessed to speed up matrix operations. We analyze the impact of each optimization individually and then combine them in a two-pronged strategy. The results show notable improvements in execution time and energy consumption, demonstrating the effectiveness of combining algorithmic and hardware optimizations to enhance learning models for edge computing environments with low-power and real-time requirements.

Autori: Marco Angioli, Marcello Barbirotta, Abdallah Cheikh, Antonio Mastrandrea, Francesco Menichelli, Mauro Olivieri | **Nome della pubblicazione**: ACM Transactions on Embedded Computing Systems | **Editore**: ACM

Link: https://dl.acm.org/doi/pdf/10.1145/3736226

[2025] **Configurable Hardware Acceleration for Hyperdimensional Computing Extension on RISC-V**

Riferimento: Preprint

Hyperdimensional Computing (HDC) is a brain-inspired computing paradigm that models information using high-dimensional distributed representations called hypervectors (HVs). HDC leverages parallel and simple vector arithmetic operations to combine and compare different concepts, emerging as a lightweight alternative for performing AI learning tasks on resource-constrained devices and as an ideal candidate for hardware implementations. In this work, we present a highly flexible hardware acceleration unit designed to optimize the execution time of HDC learning tasks. Integrated into the execution stage of the Klessydra T03 RISC-V core, the unit accelerates the core arithmetic operations on binary HVs and can be configured at synthesis time in terms of hardware parallelism, supported operations and size of the local memories, trading off execution time with hardware resources to meet the demand of different applications. A custom RISC-V Instruction Set Extension is designed to efficiently control the accelerator, with instructions fully integrated into the GCC compiler chain and exposed to the programmer as intrinsic function calls. Dedicated Control Status Registers allow users to specify the characteristics of the high-dimensional space and the target learning tasks at runtime, controlling the hardware loops of the accelerator and enabling the same hardware architecture to be used for various tasks. The dual flexibility coming from hardware configuration and software programmability sets this work apart from application-specific solutions in the literature, offering a unique, versatile accelerator adaptable to a wide range of applications and learning tasks.

Autori: Rocco Martino,Marco Angioli,Antonello Rosato,Marcello Barbirotta,Abdallah Cheikh,Mauro Olivieri | **Nome della pubblicazione**: PREPRINT

Link: https://www.techrxiv.org/doi/full/10.36227/techrxiv.173337827.72919533

HD-CB: The First Exploration of Hyperdimensional Computing for Contextual Bandits Problems

Riferimento: Preprint

Hyperdimensional Computing (HDC), also known as Vector Symbolic Architectures, is a computing paradigm that combines the strengths of symbolic reasoning with the efficiency and scalability of distributed connectionist models in artificial intelligence. HDC has recently emerged as a promising alternative for performing learning tasks in resource-constrained environments thanks to its energy and computational efficiency, inherent parallelism, and resilience to noise and hardware faults. This work introduces the Hyperdimensional Contextual Bandits (HD-CB): the first exploration of HDC to model and automate sequential decision-making Contextual Bandits (CB) problems. The proposed approach maps environmental states in a high-dimensional space and represents each action with dedicated hypervectors (HVs). At each iteration, these HVs are used to select the optimal action for the given context and are updated based on the received reward, replacing computationally expensive ridge regression procedures required by traditional linear CB algorithms with simple, highly parallel vector operations. We propose four HD-CB variants, demonstrating their flexibility in implementing different exploration strategies, as well as techniques to reduce memory overhead and the number of hyperparameters. Extensive simulations on synthetic datasets and a real-world benchmark reveal that HD-CB consistently achieves competitive or superior performance compared to traditional linear CB algorithms, while offering faster convergence time, lower computational complexity, improved scalability, and high parallelism.

Autori: Marco Angioli, Antonello Rosato, Marcello Barbirotta, Rocco Martino, Francesco Menichelli, Mauro Olivieri | **Nome della pubblicazione**: Preprint

Link: https://arxiv.org/pdf/2501.16863

[2025] **HD-CB_BIN: A Lightweight Approach for Contextual Bandit Learning in Real-Time Applications**

As the Internet of Things expands, the need to embed artificial intelligence algorithms into resource-constrained devices for real-time applications is growing. These systems require efficient and scalable algorithms to perform rapid and autonomous decision-making without relying on centralized cloud servers. Hyperdimensional Computing (HDC) has recently emerged as a compelling paradigm for learning tasks in such environments, offering computational efficiency, exceptional parallelism, and scalability.

In this work, we present HD-CB$_{BIN}$, a lightweight and efficient implementation of the HD-CB framework for modeling and automating sequential decision-making Contextual Bandits (CB) problems on embedded systems. By introducing modifications to the original algorithm, HD-CB$_{BIN}$ exclusively uses binary hypervectors, significantly reducing computational demands.

We benchmark the performance of HD-CB$_{BIN}$ on synthetic datasets, comparing it to the real-valued counterpart and the traditional state-of-the-art LinUCB algorithm. We also evaluate its execution time, computational complexity, and memory requirements on various embedded platforms and demonstrate additional gains through hardware acceleration.

The results show that our approach achieves linear execution time with respect to the context vector size and up to a $141\times$ speedup over LinUCB while maintaining competitive performance, establishing a new milestone in contextual bandit algorithms for time-critical, resource-constrained applications.

Autori: Marco Angioli, Antonello Rosato, Marcello Barbirotta, Rocco Martino, Andrea Marcelli, Antonio Mastrandrea, Mauro Olivieri | **Nome della pubblicazione**: IEEE International Joint Conference on Neural Networks | **Editore**: IEEE

[2025] **Efficient Hyperdimensional Computing with Modular Composite Representations**

Riferimento: Submitted to IEEE Emerging Topics in Computational Society

The modular composite representation (MCR) is a computing model that represents information with high-dimensional integer vectors using modular arithmetic. Originally proposed as a generalization of the binary spatter code model, it aims to provide higher representational power while remaining a lighter alternative to models requiring high-precision components. However, despite this potential, MCR has received limited attention in the literature. Systematic analyses of its trade-offs and comparisons with other models, such as binary spatter codes, multiply-add-permute, and Fourier holographic reduced representation, are lacking, sustaining the perception that its added complexity outweighs the improved expressivity over simpler models.

In this work, we revisit MCR by presenting its first extensive evaluation, demonstrating that it achieves a unique balance of information capacity, classification accuracy, and hardware efficiency. Experiments measuring information capacity demonstrate that MCR outperforms binary and integer vectors while approaching complex-valued representations at a fraction of their memory footprint. Evaluation on a collection of 123 classification datasets confirms consistent accuracy gains and shows that MCR can match the performance of binary spatter codes using up to 4.0x less memory.

We investigate the hardware realization of MCR by showing that it maps naturally to digital logic and by designing the first dedicated accelerator for it. Evaluations on basic operations and seven selected datasets demonstrate a speedup of up to three orders-of-magnitude and significant energy reductions compared to a software implementation. Furthermore, when matched for accuracy against binary spatter codes, MCR achieves on average 3.08x faster execution and 2.68x lower energy consumption.

These findings demonstrate that, although MCR requires more sophisticated operations than binary spatter codes, its modular arithmetic and higher per-component precision enable much lower dimensionality of representations. When realized with our dedicated hardware accelerator, this results in a faster, more energy-efficient, and high-precision alternative to existing models.

Autori: Marco Angioli, Christopher J. Kymn, Antonello Rosato, Amy Loutfi, Mauro Olivieri, Denis Kleyko | **Nome della pubblicazione**: IEEE Emerging Topics in Computational Society | **Editore**: IEEE

CONFERENZE E SEMINARI

[26/09/2022 – 27/09/2022] **ApplePies: Applications in Electronics Pervading Industry, Environment and Society** Genova

[2023] **PRIME: Ph. D Research in Microelectronics and Electronics** Valencia

[2023] **ACM Summer School - 2023 Edition** Barcellona

[2024] **PRIME: Ph. D Research in Microelectronics and Electronics** Larnaca

[2024] **IEEE WCCI 2024 - The IEEE World Congress on Computational Intelligence** Yokohama

[2025] **IEEE IJCNN 2025 - International Joint Conference on Neural Networks** Roma

Autorizzo il trattamento dei miei dati personali presenti nel CV ai sensi dell'art. 13 d. lgs. 30 giugno 2003 n. 196 - "Codice in materia di protezione dei dati personali" e dell'art. 13 GDPR 679/16 - "Regolamento europeo sulla protezione dei dati personali".

MarcoAngioli