Machine Learning on Telecommunication Satellite

Michael Petry
Airbus Defence & Space GmbH
Technical University of Munich
+49 89 31794285
Michael.Petry@airbus.com

Andreas Koch
Airbus Defence & Space GmbH
Technical University of Munich
+49 89 31797915
Andreas.c.Koch@airbus.com

Martin Werner
Technical University of Munich

+49 89 289 555 50
Martin.Werner@tum.de

Uwe Hoch
Airbus Defence & Space GmbH
+49 89 31799059
Uwe.Hoch@airbus.com

Tim Helfers*
Airbus Defence & Space GmbH
+49 89 31799045
Tim.Helfers@airbus.com
* Presenting author

Richard Wiest
Airbus Defence & Space GmbH
+49 89 31799474
Richard.Wiest@airbus.com

ABSTRACT

In this paper, we describe interesting findings and research results of the Airbus project MaLeTeSa (Machine Learning on Telecommunication Satellite). We provide a summary of the next-generation satellite processing hardware platform including benchmark results with respect to computing power and energy consumption based on a selected deep learning application. Moreover, we outline use-cases in the anomaly detection and telecommunications domain in an in-orbit environment that are suitable to be tackled using machine learning approaches. Lastly, we summarize our findings and provide an outlook towards future work in the project.

Conference Topics Covered

• **5. Advances in processors/processing** → Next generation processing: new technologies, new products • **3. Data handling and processing for data-hungry/data-critical missions and applications**→ 3.1. Data management aspects for remote sensing missions - on-board processing, AI, ML; 3.2. Data management for telecommunications missions - high performance switching, high speed AD/DA conversion, high data rates.

Keywords

Machine learning, FPGA, Xilinx Versal, Anomaly Detection, Telecommunications, Radio Frequency

1. INTRODUCTION

Artificial intelligence is very widespread today in the terrestrial field and offers completely new possibilities for solving many problems quickly and efficiently: This includes image recognition, speech recognition, but also behavioral observations and autonomous reactions. Recently, initial application studies and in-orbit demonstrations for artificial intelligence - here in particular machine learning - in the field of earth observation missions have been promoted by ESA, DLR, and others. This paper states the progress of an Airbus project dedicated to this research field with a particular focus on in-orbit edge processing powered by artificial intelligence. Typical applications here are data reduction (detection of empty image areas such as clouds), identification of objects (e.g., ships at sea), anomaly and change detection of satellite system or payload data, and radio frequency applications (e.g. dynamic spectrum utilization or signal classification) to name a few.

Ultimately, the goal of this project is to perform a detailed analysis, optimization and characterization of a neural network-based implementation of telecommunications payload related applications, as well as the proof of function in the selected example applications. The findings form the technological basis for a future development of a qualification model of an artificial intelligence (AI) hardware and later flight hardware development for future telecom payloads. This pre-development is intended to facilitate future use of artificial intelligence in the payload of telecommunications satellites. The technology developed here is intended to find application in the next generation of telecom payload computers.

The project is divided into the following four phases:

1. *Consolidation of the fields of application and requirements*: Possible use-cases for machine learning in telecommunications are investigated whereas the focus is set on three distinct topics:
   a. "Edge computing" capabilities (i.e. for distributed processing of data from the telecommunications data stream)
   b. Anomaly detection and prediction

This work has been supported by the DLR under project name „Maschinelles Lernen in Telekommunikations-Satelliten“, Grant number: 50 YB 2103.

c. Telecommunications (radio frequency)

2. *AI Application development:*
    a. Definition of neural network architectures
    b. Selection of datasets for training
    c. Detailed definition of the testing device
3. *Construction of a laboratory sample and test device*
4. *Testing and demonstration*

As of now, phase 1 has been finished, whereas the aim of this paper is to summarize its main results.

The structure of this paper is given in the following:

Since deployment is ultimately targeted for an in-orbit environment, the selected FPGA hardware platform suitable for the space environment, i.e., the Xilinx Versal AI Core Series, is briefly described and benchmarked regarding machine learning inference performance in section 2. Section 3 provides a detailed description of the performed use-case analysis and their respective requirements regarding computational complexity and dataset availability as well as implementation challenges. Following this, two distinct use-cases for further implementation with respect to the prior mentioned topics (see Phase 1) are selected. Lastly, section 4 concludes this paper and provides an outlook to future work.

2. FPGA Hardware Platform Benchmark

The first part of this section gives a short overview of the hardware platform utilized in this paper. Following this, an initial benchmark regarding machine learning inference performance based on an edge-processing application in the telecommunications domain is summarized.

2.1 Xilinx Versal AI Core Series

The Versal presents a powerful hardware platform, targeting a range from very demanding embedded tasks up to networking, communications and data center applications [1]. Its heterogeneous architecture, which Xilinx refers to as an adaptive compute acceleration platform (ACAP), extends the combination of processing system (PS) and programmable logic (PL) with a third class of compute engines. A coarse overview of the Versal's computing resources is shown in Fig. 1. These novel engines are particularly designed for accelerating signal processing and AI inference computations, and thus are known as AI Engines in the context of the Versal [2].

In the following subsection we will describe the machine learning application used for benchmarking the Versal.

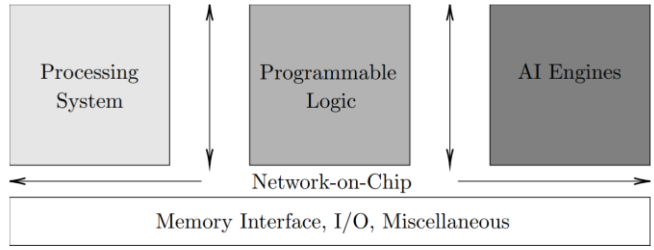

Figure 1: Xilinx Versal computing resources. Adapted from [1].

2.2 Deep-Learning-based Edge-Processing Application

Providing a radio terminal with more intelligence, for example in terms of signal analysis and spectrum awareness capabilities, is a key objective for space- and ground-based communications alike, and culminates in a concept that in the literature is commonly referred to as a cognitive radio [3]. The application selected for benchmarking performs a signal localization task in the wide-band domain based on a batch of measured I/Q samples. The problem is formulated as an image segmentation task which is a well-studied technique from the computer vision domain. A spectrogram is generated from input data blocks of 512x512 IQ-symbols and is fed into a convolutional neural network based machine learning model, also known as the U-NET. The U-Net architecture used here is by default almost fully compatible with the hardware inference accelerators provided by Xilinx. The only necessary modification, compared to the implementation in the original paper, is the introduction of a zero-padding in the convolutional layers for avoiding the asymmetry in image resolution on both sides of the network. With this modification, the resulting CNN implementations are entirely executed on the generic inference accelerator, without the necessity to delegate incompatible processing steps from the FPGA fabric to the embedded CPU.

Lastly, we want to note the necessity to perform a fixed-point quantization of all model parameters prior to deployment, since the Xilinx-based hardware accelerator is restricted to 8-bit quantized values. This is due to much lower required hardware complexity as well as lower energy consumption for realizing fixed-point units compared to floating-point units. Furthermore, quantized values with a reduced bit width require less memory space, which is important for storing parameters and intermediate results directly in the on-chip memory in order to avoid data transfers to the external memory [4].

2.3 Benchmark Results

We evaluated the Versal ACAP in two different configurations that are described in Table 1. In addition, various clock speeds of the programmable logic and the AI-Engines were analyzed. The implementation of the U-Net as described in subsection 2.2 makes use of around 500,000 trainable parameters. A forward pass through the quantized network requires more than 3 billion 8-bit Multiply-Accumulate (MAC)-operations.

Table 1. Hardware Utilization Configurations of the Hardware Inference Accelerator on Xilinx Versal

Configuration	C32B1	C64B1
Flip-Flops	110k / 1.8M (6%)	132k / 1.8M (7%)
LUTs	81k / 0.9M (9%)	92k / 0.9M (10%)
Block RAM	0 / 32 Mb	0 / 34 Mb
DSP Slices	139 / 1,968 (7%)	139 / 1,968 (7%)
UltraRAM	57 / 130Mb (44%)	57 / 130Mb (44%)
AI Engines	32 / 400 (8%)	64 / 400 (16%)

The benchmark results are summarized in Tab. 2 and will be explained in detail in the following two subsections.

Table 2. Benchmark Results for the Wideband Signal Localization Application

Config	Frequency [MHz]	Latency [ms]	TP [FPS]	P_{idle} [W]	P_{active} [W]
C32B1	333/1250	12.34	79	17.1	19.6
C32B1	200/800	18.05	55	15.5	17.0
C32B1	200/400	26.75	37	14.6	15.9
C64B1	333/1250	12.10	81	19.3	22.3

In the configuration with 32 AI Engines operating at the maximum recommended frequency of 333 MHz for the PL and 1250 MHz for the AIEs, the Versal DPU achieves a throughput of around 79 frames per second. Consequently, the latency for performing inference on a single data sample is 12.34 ms. However, as depicted in Tab. 2, the Versal consumes a tremendous amount of power for achieving this performance. In the idle state, the system already consumes 17.1 W of power, which is increased to 19.6 W during the computations, a change of 2.5 W. Equipping the hardware accelerator with more AI Engines increases the power consumption even more, as might be expected. In this configuration the device consumes 19.3 W in the idle state, a plus of 2.2 W compared to the configuration with half the AI Engines. As indicated by the last row in Table 2, doubling the amount of AI Engines for this specific application results only in slightly increased performance, but not to an extent that would justify the increase in power consumption.

3. Use-case analysis

In this section we will summarize the two application domains analyzed during this project and show-case selected use-cases and their machine learning approach.

3.1 Anomaly Detection

The Failure Detection, Isolation and Recovery (FDIR) subsystem is a critical function on board all spacecraft since it is vital for ensuring the safety, autonomy and availability of the system during the mission lifetime. As spacecraft telemetries sent to ground increase, the monitoring burden on ground operations is starting to reach critical levels in terms of operations. Spacecraft monitoring systems require expert knowledge that is challenging to develop and maintain, meaning that algorithm assisted anomaly detection is more and more employed. Moreover, due to bandwidth limitations, only a subset of low frequency telemetries can be downlinked or are checked only sporadically. The inflight TM budget is limited by hardware and mission resources in terms of on-board TM packet generation rate, storage capacity in the satellite memory and the downlink data volume considering the TTC bandwidth available during the contact between spacecraft and ground station. As anomalies require potentially quick responses from operators, many companies are currently working on developing on-board anomaly detection and prediction systems. The task of anomaly detection is key with regards to the Failure Detection component of FDIR. Anomalies found in data usually come in three types:

- Point anomalies – an individual data instance is considered as anomalous with respect to the rest of the data
- Contextual anomalies – an individual data instance is anomalous in a specific context, but not otherwise anomalous if occurring at a certain time or certain region
- Collective anomalies – a collection of related data instances is anomalous with respect to the entire dataset, but not individual values. They have two variations:
    - Events in unexpected order
    - Unexpected value combinations

Figure 2 provides an exemplary visualization of those types.

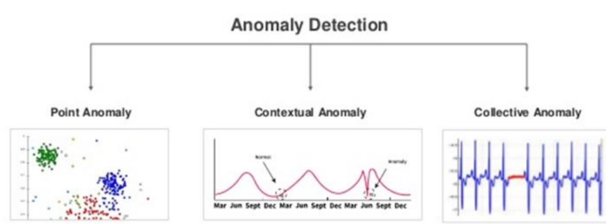

Figure 2: Anomaly types in telemetries.

In the context of FDIR, it is mandatory to detect any type of anomaly that could potentially present a threat to a system/subsystem of the spacecraft or the spacecraft as a whole, thus the analysis of the spacecraft telemetry in the context of time becomes more interesting as opposed to detecting single isolated outliers in the data. Furthermore, usually the monitoring of several telemetries at a time results in improved accuracy in detecting anomalies due to the additional context information that is available from monitoring several variables at once.

Multiple algorithms have been proposed to tackle this issue, however, in this paper, we want to restrict our focus to machine-learning strategies only.

Machine learning approaches can be split into supervised and unsupervised approaches, based on whether labels are required. In order to create anomaly labels for telemetry data, actual anomalies need to be detected and confirmed by hand. As this activity is very time consuming, only a few labeled anomaly datasets are available for spacecraft telemetry. Furthermore, the amount of labels to be added is limited to the number of anomalies that occurred on spacecrafts. A better alternative is to insert simulated anomalies. Although this is usually accomplished by ensuring that inserted anomalies have the same statistical properties as actual anomalies, there is no guarantee that the labels are representative. Due to the limitation in availability of labeled training data, the complexity of supervised approaches as a whole is limited. When considering unsupervised approaches, available training data is not only limited to annotated datasets. Therefore, model complexity can continue to increase with training data volume, which is beneficial for the performance of data-driven machine learning approaches. Accordingly, we have identified the following neural network architectures to be suitable for this task:

- Recurrent Neural Networks (including LSTM and GRU), working in the time domain.
- Convolutional Neural Networks and Autoencoder approaches, that focus on feature detection.
- Mixed topologies, like Temporal Convolutional Network that can store local data in an input memory, but are fully unrolled unlike RNN.
- Gradient Boosting Decision Trees for predicting telemetry data directly
- Attention blocks incorporated into an LSTM network
- Generative Adversarial Networks for failure detection and isolation
- Support Vector Machines for estimating anomaly scores of sparse feature vectors
- Autoregression models predicting telemetry data based on previous data only

3.2 Telecommunications (RF)

In this subsection we want to describe two selected use-cases in the radio frequency domain that are suitable to be tackled using a machine-learning approach.

3.2.1 Autoencoder-based communication

It is possible to utilize auto-encoders to define a channel-adaptive modulation scheme to transmit signals over a noisy channel. This way the modulation scheme is adapted in order to reduce the symbol error. A traditional autoencoder is an unsupervised neural network that learns how to efficiently compress data, which is also called encoding. The autoencoder also learns how to reconstruct the data from the compressed representation such that the difference between the original data and the reconstructed data is minimal.

The auto-encoder jointly optimizes the transmitter and the receiver as a whole. This joint optimization has the potential of providing a better performance than the traditional systems. An architectural example is shown in Fig. 3.

The following properties have been shown to be suitable to be implemented inside the auto-encoder structure:

- MIMO functionality
- OFDMA functionality
- signal detection
- channel estimation
- equalization
- parameter estimation, e.g., signal to noise ratio, etc.
- transmitter and receiver synchronization
- detecting and mitigating hardware impairments, e.g., phase offset, carrier frequency offset, sample frequency offset, etc.

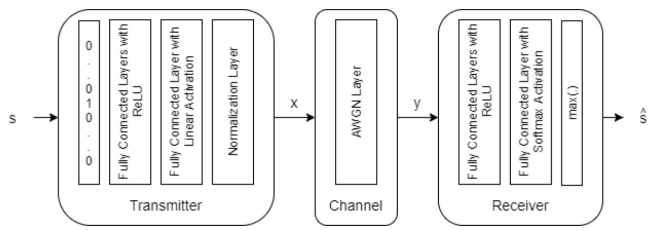

Figure 3: Autoencoder for Adaptive Symbol Modulation. [5]

3.2.2 RF Fingerprinting

Hardware imperfections in RF transmitters impart a unique property to the transmitted signal, similar to a fingerprint, which can be exploited by the communication system. A strategy, called RF fingerprinting, recognizes different transmitters using device level differences in their RF frontends. Hardware impairments are unavoidable due to several reasons: In the last decades, extensive focus has been set on digitizing every aspect possible of the communication system. This stems from the fact, that the probability of error in the digital domain can practically be reduced to an arbitrary low number with limited effort, due to various methods like error detection and error correction. However, since radio signals are electronmagnetic radiation, (digital) communication systems have no alternative but to emulate this analog quantity using some kind of electric circuit, terminated with an antenna. Not only does this domain change subjects the signal to noise, but also to non-idealities in the digital-to-analog hardware block, which is the key-enabler for RF fingerprinting.

The following gives an overview about different categories of hardware impairments:

- Power amplifier imperfections:
  - Non-linearity
  - Memory effect
- Local oscilator imperfections:
  - Carrier frequency offset
  - Sample frequency offset
  - Phase noise
- IQ/Baseband-Path
  - Gain imbalance
  - Phase imbalance
- Digital-to-Analog Converter:
  - Non-linearities
  - Quantization

A major benefit can be realized in the area of information security as shown in [6], [7], and [8]. Also, RF fingerprinting was shown to perform well under different modulation techniques. Further benefits are a reduction in protocol overhead, minimization of the impact of hardware impairments, and the possibility of transmitter classification and identification.

3.2.3 Modulation classification

Lastly, we want to briefly mention the popular use case of modulation classification in the RF domain. In order to achieve high capacity on a dynamically shared channel, it is possible to utilize predictive and monitoring models oriented towards a machine-learning approach with the goal of identifying the current modulation scheme used by other channel participants [9]. By using this knowledge, the device's modulation scheme can be adapted according to the current utilization of the channel and the impact of interference can be reduced. Especially in conjunction with the wideband spectrum utilization application introduced in subsection 2.2 this use-case provides a powerful tool towards robust communication.

4. Conclusion and Outlook

In this paper we summarized the findings of phase 1 of the MaLeTeSa project. In detail, we described and benchmarked the next-generation FPGA-based hardware platform Xilinx Versal. Moreover, we described relevant use-cases for telecommunications satellites that show promising performance when solved using machine learning strategies.

Future work will be oriented towards implementing machine learning approaches of selected use-cases on the Versal hardware platform. The performance and reliability can then be demonstrated and proven using a test device to be developed.

The hardware demonstrator (laboratory sample) is already being developed and set up under aerospace aspects, so that nothing stands in the way of later further development of the breadboard. This means that the use of electronic components, the mechanical and thermal structure, the analysis and production already take into account the known space boundary conditions. With this, the further path towards the qualification and then flight model is sufficiently prepared.

A successful completion can demonstrate the feasibility and advantages to potential customers and show the feasibility for future products and missions - in order to enable project acquisition.

5. REFERENCES

[1] Advanced Micro Devices, Inc. 2022. Versal Architecture and Product Data Sheet: Overview. https://docs.xilinx.com/v/u/en-US/ds950-versal-overview. [Online: accessed 25-January-2023].

[2] Xilinx Inc. 2022. System-Level Benefits of the Versal Platform. https://www.xilinx.com/content/dam/xilinx/support/documents/white_papers/wp539-versal-system-level-benefits.pdf. [Online; accessed 25-January-2023].

[3] Friedrich K. Jondral. 2007. Cognitive Radio: A Communications Engineering View. IEEE Wireless Communications 14, 4 (2007), 28–33. https://doi.org/10.1109/ MWC.2007.4300980

[4] Xilinx Inc. 2022. DPUCVDX8G for Versal ACAPs. https://www.xilinx.com/content/dam/xilinx/support/documents/ip_documentation/dpucvdx8g/v1_1/pg389-dpucvdx8g.pdf. [Online; accessed 24-January-2023]

[5] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," in IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563-575, Dec. 2017, doi: 10.1109/TCCN.2017.2758370.

[6] Morin, C., Cardoso, L.S., Hoydis, J., Gorce, JM., Vial, T. (2019). Transmitter Classification with Supervised Deep Learning. In: , et al. Cognitive Radio-Oriented Wireless Networks. CrownCom 2019. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 291. Springer, Cham. https://doi.org/10.1007/978-3-030-25748-4_6

[7] Cyrille Morin, Leonardo Cardoso, Jakob Hoydis, Jean-Marie Gorce. Deep Learning-based Transmitter identification on the physical layer. 2020. ffhal-03117090.

[8] S. S. Hanna and D. Cabric, "Deep Learning Based Transmitter Identification using Power Amplifier Nonlinearity," 2019 International Conference on Computing, Networking and Communications (ICNC), 2019, pp. 674-680, doi: 10.1109/ICCNC.2019.8685569

[9] T. J. O'Shea, T. Roy and T. C. Clancy, "Over-the-Air Deep Learning Based Radio Signal Classification," in IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 1, pp. 168-179, Feb. 2018, doi: 10.1109/JSTSP.2018.2797022.