

ParaBase: A Configurable Parallel Baseband Processor for Ultra-High-Speed Inter-Satellite Optical Communications

Seungkyu Choi¹, Huanshihong Deng², Kuan-Yu Chen², Yufan Yue², David Blaauw², Hun-Seok Kim² Kyung Hee University, South Korea¹, University of Michigan, Ann Arbor, MI, USA²

ABSTRACT

This paper presents ParaBase, a configurable baseband processing architecture that efficiently handles parallel sample streams targeting ultra-wide bandwidth inter-satellite optical communications for the target data rates exceeding 100 Gbps. We propose a parallelogram-style systolic accelerator, specifically designed for parallel processing for correlation kernels, preserving the hardware efficiency inherent in the systolic-array architecture. ParaBase supports end-to-end baseband processing through a set of heterogeneous configurable accelerators customized to their respective parallel processing requirements. It outperforms the SIMD-style architecture and surpasses previous baseband processors by $\sim\!7.0\times$ in terms of energy efficiency for FIR filtering. The overall architecture reports energy efficiency of up to 2.9 TOPS/W and 121.8 Gbits/J while supporting a wide range of data rates from $1\sim\!128$ Gbps via fast reconfiguration for various modulation schemes.

CCS CONCEPTS

• **Hardware** → Digital signal processing.

KEYWORDS

Baseband processor, systolic array, digital signal processing

1 INTRODUCTION

In light of the rapid evolution of optical communication, the need for ultra-high bandwidth exceeding 100 Gbps has become imperative for emerging applications. Optical communication, especially Free Space Optics (FSO) has become an enabling technology to realize high-speed multi-Gbps connectivity between low Earth orbit (LEO) satellites. It offers exceptional security and low error rates, effectively meeting the high bandwidth demands of future communication systems available anywhere on Earth [1–3].

In a typical communication system, a high-performance processor is required to perform computationally demanding workloads in baseband digital signal processing. A recent trend is to utilize a Software-Defined Radio (SDR) architecture as a programmable platform to execute these computation-intensive baseband kernels. Several SDR processor designs have been proposed to demonstrate the feasibility of programmable datapaths to perform various baseband workloads with high hardware efficiency. In particular, [4, 5]

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ISLPED '24, August 5-7, 2024, Newport Beach, CA, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-0688-2/24/08...\$15.00 https://doi.org/10.1145/3665314.3673174

propose custom SIMD execution models to compute in parallel with low complexity, and [6,7] propose systolic array architectures to efficiently handle multiple kernels via programmable datapaths.

FSO inter-satellite communications impose a challenging requirement for SDRs to support data rates beyond 100 Gbps to fully leverage the ultra-high bandwidth of tens of GHz available to FSO. This necessitates high-speed baseband processing whose throughput significantly exceeds what can be attained through a single sample stream as the processor runs at a clock frequency much lower than the available bandwidth. It implies that the processor must concurrently deal with parallel streams of samples within a single computing clock cycle. This technical challenge was not fully addressed in prior architectures [5–7] as they were mostly designed for terrestrial wireless communications with relatively low throughput (bandwidth), ranging from a few Mega to Giga (G) samples/s.

In this paper, we propose ParaBase, a parallel baseband processing architecture tailored to accommodate parallel sample streams achieving data rates exceeding 100 Gbps. Our contribution to this research is twofold. We first propose a correlation accelerator dedicated to processing correlated output streams in parallel. The major challenge in designing the architecture lies in handling the correlation between adjacent parallel streams to produce parallel output streams every clock cycle. Therefore, We introduce a novel systolic architecture capable of effectively managing this parallel-to-parallel inter-dependency. The proposed accelerator exhibits superior performance (64G complex samples per sec) and energy efficiency (26.4G complex samples per joule) compared to prior designs.

Secondly, we propose a reconfigurable datapath to support multiple FSO standards by partitioning the workloads into correlation and non-correlation kernels, providing dedicated support for each through heterogeneous accelerators customized to their respective parallel processing requirements. The overall system can be configured at the accelerator level with seamless parallel streams between accelerators without degrading the overall performance, while providing excellent configurability and flexibility. As a result, it can support various baseband demodulation workloads defined in a variety of FSO standards. To the best of our knowledge, this is the first study to address multi-standard flexible FSO baseband processing featuring a reconfigurable datapath with parallel streams.

The main contributions of our work are summarized as follows:

- A parallelogram-style systolic array accelerator is proposed to execute correlated operations among parallel streams.
- A flexible baseband architecture consisting of heterogeneous configurable accelerators is proposed for ultra-high performance optical communications exceeding 100 Gbps.
- Extensive analysis is provided to characterize the throughput and energy efficiency of end-to-end baseband processing for various optical communication standards.

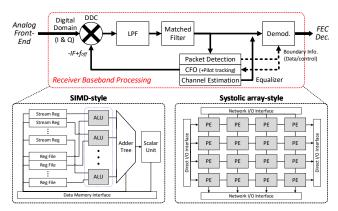


Figure 1: Functional block diagram of baseband processing and its architectures.

2 BACKGROUND

noise introduced by the channel.

In baseband processing, the receiver processing complexity dominates the transmit counterpast [5,8]. Thus we focus on the receiver baseband in this work. The received optical signature in the receiver baseband in this work. The received optical signature is stream, regardless of the optical signature in the analog frontend. The baseband processor takes complex-valued samples centered at IF and performs additional processing to demodate the information/coded bits. It is noteworthy that the receiver baseband processing is intricate for two reasons: 1) the united Manny Intricate signal/packet/symbol arrival time and carrier frequency offsets, and 2) the presence of

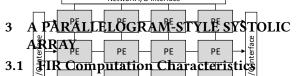
Figure 1 presents a high-level functional block diagram of the receiver baseband processing. This process involves computationally intensive operations of Digital Down-Conversion (DDC) followed by Low Pass Filter (LPF) and matched/equalization filtering. Additionally, the baseband processing requires a synchronization process to obtain the timing and frequency offset of the desired signal. The proposed datapath leverages a widely used data-aided method [9] for synchronization, where cross-correlation between the known symbol patterns and the received signals detects the timing of the signal and estimates the channel (and phase/frequency offset) at the peak of the correlation. The synchronization process via correlations as well as LPF and matched/equalization Finite Impulse Response (FIR) filtering are the most computationally intensive kernels in the baseband processing kernels.

2.2 Baseband Processor Architectures

Prior SDR architectures [4–7,10] facilitate parallel computing through the utilization of multiple processing units, each equipped to perform operations such as Multiply-and-Accumulate (MAC), division, and logical operations. They can be categorized with two datapath models: the Single Instruction Multiple Data (SIMD) datapath and the systolic array datapath as depicted in Figure 1. The SIMD structure streams multiple operands from the streaming registers to multiple arithmetic/logical units (ALUs). The SIMD architecture's scalability to enhance parallelism is facilitated by its straightforward datapath design. However, a notable drawback lies in the

potential congestion within the streaming data control pathways. The systolic array architecture [6,7], on the other hand, operates by streaming data between adjacent processing units and passing partial results to the next unit for further computations. This approach effectively mitigates wiring and memory/register access congestions and promotes efficient computation with deterministic execution scheduling. However, systolic array programming for specific kernels is a high-effort process and it often limits its scalability for enhanced performance.

The motivation for this work stems from the inefficiencies observed in previous architectures when tasked with high-throughput (beyond 100 Gbps) parallel data processing. Both architectures show poor performance when executing correlated operations between multiple parallel streams, which constitute the majority of computationally intensive workloads in optical communications. As a solution, we propose a novel systolic-array accelerator explicitly designed for the parallel streams of correlated operations as elaborated in Section 3.



The computation of FIR filters is ${}^{p}E$ convolution between the input signal \overline{S} and the coefficients c as shown in Eq. (1). The filter length (number of tabs) is indicated by ${}^{p}E$, and ${}^{p}E$ is the ${}^{p}E$ -th coefficient of the filter. Note that the signal ${}^{p}E$, coefficient ${}^{p}E$ and output ${}^{p}E$ are all complex valued in general.

$$o[i] = \sum_{j=0}^{N-1} s[i-j] \cdot c[j]$$
 (1)

The output o[i] is obtained through complex-valued MAC (C-MAC) operations with the streaming input s[i-j] and its corresponding coefficient c[j]. Considering this streaming nature of computation in correlation and FIR filtering, a systolic array architecture is a natural choice to efficiently manage the continuous stream of samples and to produce outputs through a pipelined accumulation path [7].

Figure 2(a) presents a typical streaming dataflow of FIR filtering in a systolic array processor. As an input sample continuously streams to the processor, it is sent to the MAC units every cycle for multiplication. Subsequently, the multiplied results are added and then forwarded to the next MAC unit. With a pipelined datapath for partial sums, the dataflow generates an output every cycle at the end of the processing path.

The limitation of this conventional dataflow to extend it to multiple parallel sample streams arises from its requirement for global streams sent to all MAC units. These global sample streams produce multiple results that are accumulated in their corresponding MAC units, serving as the partial sums of the consecutive outputs. Because of inter-sample dependency, broadcasting of multiple parallel streams to processing elements creates wiring congestion and consumes excessive power. Hence, extending this conventional systolic architecture to multiple parallel stream inputs is a non-trivial task.

Figure 2: The (a) conventional and (b) proposed dataflow.

3.2 The Proposed Dataflow

i+2R-1 i+R-1

We overcomethe limitation of the conventional FIR filtering / correlation structure by stibiliting the global input streams with sequential streams flowing within the same. MAC row to accomplish a 2D extension of the systolic array dataflow. We provide the parallel streaming samples as depicted in Figure 2(b) sent to each designated MAC row and then passed to the next column every cycle. The same input sample is multiplied with different coefficients when it passes through a row of the array. At the same time, the array offers a diagonal path to add the product of the adjacent sample and coefficient, and to accumulate output values to generate the partial/final sum of the filtering / correlation.

Figure 2(b) illustrates the datapath of the proposed architecture which accommodates R parallel sample streams and supports up to *N*-tap FIR filtering. While the samples stream horizontally, the coefficients in the weight buffer are loaded vertically into the complex-valued MAC (C-MAC) units, programming the same coefficient value within the same column. The C-MAC outputs propagate to the upper row units of the next column, forming a parallelogramshaped array where N diagonal C-MAC units are involved in generating one output. This diagonal C-MAC row is duplicated R times to facilitate parallel processing of R samples/streams, resulting in a total of RN C-MAC units. We obtain R output samples $(out[i] \sim out[i+R-1])$ after $N \times cycle_{MAC}$ cycles where $cycle_{MAC}$ is the latency of C-MAC. Meanwhile, as can be observed from Eq. (1) and Figure 2(b), the computation requires additional N-1 samples of $s[i-N+1] \sim s[i-1]$ in addition to the current samples entering the array. Therefore, we implement a delayed streaming buffer to store the last N-1 samples so that they can be streamed together with the current samples. Note that the ith output is actually an N-1 sample-shifted version of the FIR filtering / correlation output. Hence, we employ a relocating buffer at the end so that the input

samples for the next accelerator are aligned with the indices of the output samples arrived from FIR filtering / correlation. The cross-correlation required in the synchronization can be executed by storing the expected symbol pattern data instead of the FIR filter coefficients in the weight buffer.

3.3 Low Power Techniques

1) Configurable Clock Gating is applied to execute the correlated operations with a subset of C-MAC units when the accelerator is programmed with different filter lengths. The number of active columns corresponds to the tap number of an FIR filter or correlation. Our accelerator can support any integer number of taps less than or equal to N. Unused C-MAC columns are clock-gated to propagate accumulated results without any computations to the next column. Clock gating is only applied to idle C-MAC units while output registers are active for signal propagation.

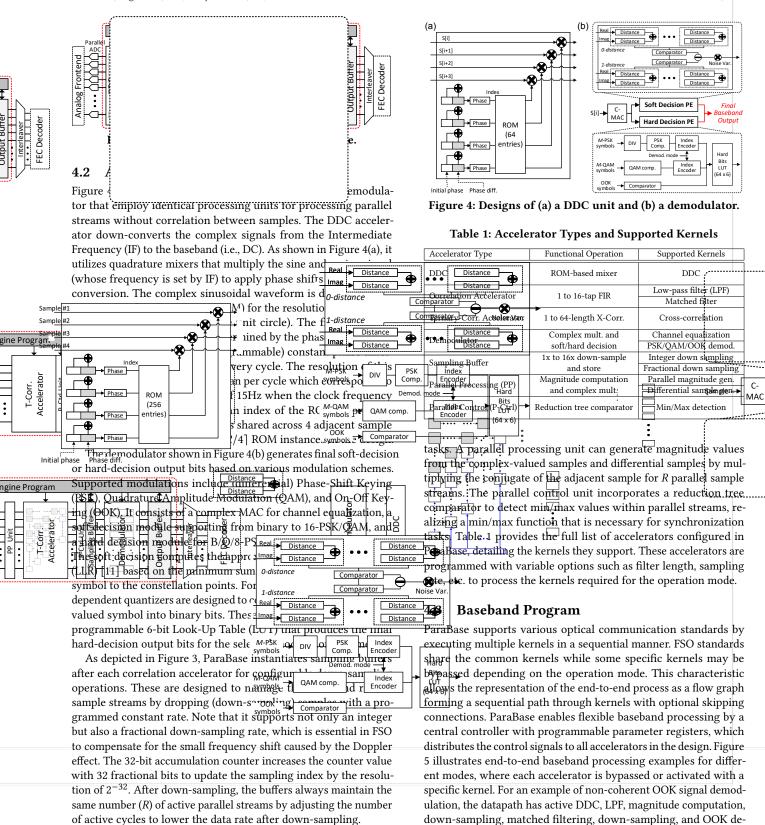
2) Ternary Weight Cross-Correlation is utilized to provide power-and area-efficient implementation of cross-correlation-based synchronization. Many FSO standards define the sync word or preamble with ternary complex values of $\{0, \pm 1, \pm i\}$ eliminating the need for high-cost C-MAC operations during synchronization, replacing them with simpler MUXes and additions. This observation allows us to implement a longer-length parallelogram array without significant area and power overhead for longer cross-correlation with ternary-valued patterns. Our design supports N=16 for general complex-valued correlation/filtering, and the length is extended to N=64 for ternary-valued correlations by adding additional columns specifically designed for ternary-valued correlations.

4 PARABASE IMPLEMENTATION

4.1 Overall Architecture

Supporting full programmability for the target throughput > 100 Gbps is challenging and unnecessary for optical communications where multiple standards define similar modulation schemes. Hence we made a design choice to implement ParaBase using a set of heterogeneous configurable accelerators where each accelerator is dedicated to a specific subset of kernels. This configuration forms a datapath seamlessly connecting parallel data streams between kernels. Each accelerator can be individually configured to execute kernels in various modes and can be bypassed when not needed within the overall datapath. Further details regarding the methodology for datapath configuration are discussed in Section 4.3.

Figure 3 depicts the overall datapath of ParaBase. The architecture comprises parallel computing accelerators including a DDC accelerator, correlation accelerators, a demodulator, sampling buffers, and auxiliary units such as a reduction tree comparator and data stream controllers. Each accelerator is equipped with parallel I/O paths and supports parallel computations for multiple streams of sample data. Intra-accelerator parallelism is supported in two different ways: 1) correlation accelerators employ the proposed parallelogram-style systolic array architecture, and 2) the other accelerators that do not exhibit inter-sample correlations utilize a SIMD-style implementation for the parallel stream size of R. By providing heterogeneous parallelism architectures for those accelerators, we offer a power- and area-efficient processor design to process R samples at every clock cycle.



modulation while the remaining accelerators are bypassed as shown

in Figure 5.

ParaBase employs additional hardware components, including

parallel processing units and a parallel control unit for auxiliary

ParaBase: A Configurable Parallel Baseband Processor for Ultra-High-Speed Inter-Satellite Optical Communications ISLPED '24, August 5-7, 2024, Newport Beach, CA, USA

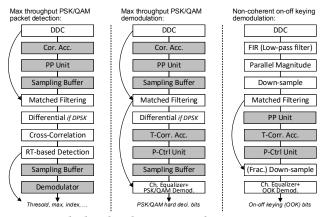


Figure 5: The baseband program and operation mapping.

5 EVALUATION

5.1 Experimental Methodology

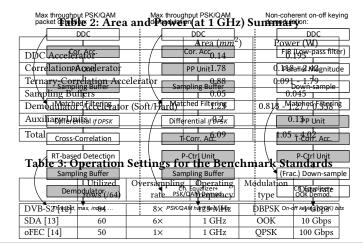
We evaluate ParaBase using a parallel stream size of R =64 operating at a clock frequency of 1 GHz to attain a processing rate of 64 Giga (G) complex samples/s. This corresponds to the sampling frequency (bandwidth) of 64 GHz of complex-valued baseband signals. We fully implemented a Verilog RTL design of ParaBase and synthesized it using Synopsys Design Compiler with Global Foundry 12 nm design kit. Area and power consumption numbers shown in Table 2 are estimated from post-synthesis simulations and Primetime PX (PTPX) analysis. The precision of the datapath and computation is set to 16-bit fixed-point complex numbers. (N =)16-length correlation accelerators (with a 64-length ternary correlation mode) are implemented in the evaluated ParaBase.

For benchmarking, we evaluate the optical communication standards of DVB-S2 [12], SDA [13], and oFEC [14]. The optical links include 1 Gbps Differential Binary Phase Shift Keying (DBPSK), 10 Gbps On-Off Keying (OOK), and 100 Gbps Quadrature PSK (QPSK), where detailed settings are specified in Table 3. Test vectors for validation are generated from a golden model developed in Matlab.

5.2 FIR Filtering and Correlation Comparison

5.2.1 Area and Power Analysis. We examine the area and power characteristics of the proposed correlation accelerator. For comparison, we customize and synthesize a scaled SIMD accelerator based on the streaming dataflow demonstrated in [5]. However, due to the heightened congestion within the streaming datapath, the SIMD architecture directly scaled from [5] cannot operate with the same target frequency of 1 GHz. We therefore design two versions of the SIMD accelerator to match the throughput of our accelerator. One variant (SIMD-low) operates at 500 MHz, incorporating 2× processing units with the same configuration, while the other (SIMD-high) has additional pipeline stages to operate at 1 GHz.

The comparison of the estimated area and power is presented in Figure 6. The SIMD-high model exhibits an area that is approximately on par with our model, yet consumes 1.4× more power. Conversely, the SIMD-low model shows a comparable power consumption but occupies 60% more area than our design. In summary, the proposed parallelogram-style systolic array demonstrates superior hardware efficiency over SIMD models.



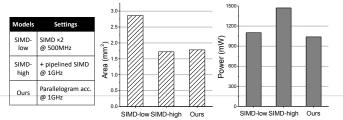


Figure 6: Area and power comparison with the scaled SIMD

SIMD- SIMD-

chieved when running a short-tap FIR using the simps aime accelerator. Additionally, we analyze the 64-length ternary low correlation accelerator in Figure 7(b), where we compare the area simple two compares the area of the comparent of the constitution of the const

efficiency of the ternary correlator is due to the significant reduction in combinational logic, whereas the sequential logic to register samples does not reduce as significantly. The ternary correlator exhibits $> 4 \times$ less power consumption for the $4 \times$ longer correlation compared to the full correlator as shown in Figure 7(a) (i.e., per tap power consumption is $< 16 \times$ lower). The configurable clock gating is applied with a resolution of 16 columns in the ternary correlator.

5.2.3 Energy Efficiency over SOTA Works. Our design is compared with prior state-of-the-art (SOTA) SDR accelerators. Systolic array architectures [6, 7] are programmable for longer-tap FIR filters but they face performance scaling limitations to meet the target performance of tens of Giga samples/s as their performance is constrained by the inability to handle parallel correlated input streams. Table 4 shows that our design attains a remarkable energy efficiency (in addition to the throughput gain) with an improvement of $3.9\times \sim 7.0\times$ compared to previous SDR architectures.

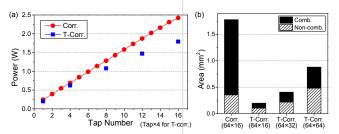
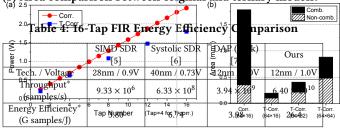


Figure 7: (a) Estimated power of clock gated corr. accelerators, (b) Area comparison between original and ternary models.



^{*} Throughput and efficiency are scaled to our 12nm process and 1.0V supply voltage.

5.3 End-to-End Simulation Results

Three FSO standards ranging from 1 Gbps to 100 Gbps are evaluated to estimate the energy efficiency of the end-to-end baseband processing. The baseband processor is configured to operate for each standard as specified in Section 4.3. For DVB-S2 DBPSK demodulation, for example, our baseband processor is configured to sequentially perform DDC, LPF, down-sampling, matched filtering, differential sample calculation, and BPSK soft decision.

In Figure 8(a), the energy efficiency is reported in Tera operations per second per watt (TOPS/W) for the workloads. Figure 8(a) shows an average of 2.3 TOPS/W and 1.9 TOPS/W achieved when operating synchronization and demodulation for the three standards, respectively. The highest energy efficiency of 2.9 TOPS/W is attained when operating DVB-S2 synchronization workload. We include the peak mode for an additional benchmark that employs QPSK hard decision using the full 64 G complex samples/s input achieving a data rate of 128 Gbps. While our design maintains high energy efficiency for most tasks, the actual TOPS/W varies depending on the total arithmetic operations and data-dependent control logic in the mode-specific datapath. Figure 8(b) shows the energy efficiency with respect to the data rate, which is measured by the number of demodulated information bits for all modes. DVB-S2 energy efficiency is low because it involves significant down-sampling from 8 GHz to 1 GHz for the final relatively low data rate of 1 Gbps (i.e., DDC and FIR filtering run at 8× higher sampling rate before down-sampling the signal to the actual symbol rate). The energy efficiency per information bit is the highest for the peak mode reaching up to 121.8 Gbits/J for the throughput of 128 Gbps. This translates to a total baseband power consumption of 1.05 W at 128 Gbps for FSO inter-satellite communications.

6 CONCLUSION

This paper proposed ParaBase, a configurable parallel baseband architecture to perform ultra-high-speed optical communication baseband processing for data rates exceeding 100 Gbps. It presents

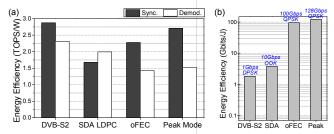


Figure 8: (a) GOPS/W and (b) Gbits/J energy efficiency of the end-to-end baseband process.

a novel parallelogram-style systolic array architecture to manage multiple parallel sample streams effectively for correlation kernels. The proposed architecture leverages multiple heterogeneous configurable accelerators to support a wide range of FSO standards with outstanding performance and efficiency compared to prior SDR accelerators. The proposed solution only consumes 1.05 W for 128 Gbps throughput, showing the feasibility of an energy-efficient, reconfigurable multi-standard SDR for future inter-satellite optical communications.

ACKNOWLEDGMENTS

This research was, in part, funded by the Defense Advanced Research Projects Agency. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

REFERENCES

- Arun K. Majumdar. Chapter 4. In Optical Wireless Communications for Broadband Global Internet Connectivity. Elsevier, 2019.
- [2] Elon musk's unmatched power in the stars. https://www.nytimes.com/interactive/ 2023/07/28/business/starlink.html. Accessed: 2023-07-28.
- [3] L. Ouvry et al. An ultra-low-power 4.7ma-rx 22.4ma-tx transceiver circuit in 65-nm cmos for m2m satellite communications. IEEE Transactions on Circuits and Systems II: Express Briefs, 2018.
- [4] Y. Lin et al. Soda: A low-power architecture for software radio. In 33rd International Symposium on Computer Architecture (ISCA), 2006.
- [5] Y. Chen et al. A low power software-defined-radio baseband processor for the internet of things. In IEEE International Symposium on High Performance Computer Architecture (HPCA), 2016.
- [6] F. Yuan and D. Marković. A 13.1gops/mw 16-core processor for software-defined radios in 40nm cmos. In Symposium on VLSI Circuits Digest of Technical Papers, 2014.
- [7] K. Chen et al. A 507 gmacs/j 256-core domain adaptive systolic-array-processor for wireless communication and linear-algebra kernels in 12nm finfet. In IEEE Symposium on VLSI Technology and Circuits, 2022.
- [8] Hela Belhadj Amor and Carolynn Bernier. Software-hardware co-design of multistandard digital baseband processor for iot. In Design, Automation Test in Europe Conference Exhibition (DATE), 2019.
- [9] J. Huber and W. Liu. Data-aided synchronization of coherent cpm-receivers. IEEE Transactions on Communications, 1992.
- [10] Marco Bertuletti, Yichao Zhang, Alessandro Vanelli-Coralli, and Luca Benini. Efficient parallelization of 5g-pusch on a scalable risc-v many-core processor. In Design, Automation Test in Europe Conference Exhibition (DATE), 2023.
- [11] A.J. Viterbi. An intuitive justification and a simplified implementation of the map decoder for convolutional codes. IEEE Journal on Selected Areas in Communications, 1998.
- [12] The second generation digital video broadcasting by satellite. available at https://www.etsi.org/technologies/dvb-s-s2, 2009. Accessed: 2009-08-01.
- [13] Sda tranche 1 optical communications terminal standard. available at https:// www.sda.mil/l, 2021. Accessed: 2021-06-01.
- [14] Open roadm msa 5.0 w-port digital specification. available at http://www.openroadm.org/, 2021. Accessed: 2021-07-01.