

Toward Matrix Multiplication for Deep Learning Inference on the Xilinx Versal

Citation

Lei, J., Flich, J. and Quintana-Ortí, E.S. (2023) 'Toward Matrix Multiplication for Deep Learning Inference on the Xilinx Versal', in Proceedings of the 31st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), pp. 227-234.

Year

2023

Version

Authors' camera-ready version

Link to publication

<https://ieeexplore.ieee.org/document/10136983>

Published in

IEEE

DOI

10.1109/PDP59025.2023.00043

License

This publication is copyrighted. You may download, display and print it for Your own personal use. Commercial use is prohibited.

Take down policy

If you believe that this document breaches copyright, please contact the authors, and we will investigate your claim.

BibTex entry

```
@INPROCEEDINGS{10136983,  
  author={Lei, Jie and Flich, José and Quintana-Ortí, Enrique S.},  
  booktitle={2023 31st Euromicro International Conference on Parallel, Distributed and Network-Based  
Processing (PDP)},  
  title={Toward Matrix Multiplication for Deep Learning Inference on the Xilinx Versal},  
  year={2023},  
  volume={},  
  number={},  
  pages={227-234},  
  doi={10.1109/PDP59025.2023.00043}}
```

Toward Matrix Multiplication for Deep Learning Inference on the Xilinx Versal

Jie Lei, José Flich, Enrique S. Quintana-Ortí
Depto. de Informática de Sistemas y Computadores
Universitat Politècnica de València, Valencia, Spain
{jlei,jflich,quintana}@disca.upv.es

Abstract—The remarkable positive impact of Deep Neural Networks on many Artificial Intelligence (AI) tasks has led to the development of various high performance algorithms as well as specialized processors and accelerators. In this paper we address this scenario by demonstrating that the principles underlying the modern realization of the general matrix multiplication (GEMM) in conventional processor architectures, are also valid to achieve high performance for the type of operations that arise in deep learning (DL) on an exotic accelerator such as the AI Engine (AIE) tile embedded in Xilinx Versal platforms. In particular, our experimental results with a prototype implementation of the GEMM kernel, on a Xilinx Versal VCK190, delivers performance close to 86.7% of the theoretical peak that can be expected on an AIE tile, for 16-bit integer operands.

Index Terms—Matrix Multiplication, Deep Learning, Xilinx Versal Artificial Intelligence Engine (AIE), High Performance.

I. INTRODUCTION

In the last two decades, the slow-down of Moore’s law and the end of Dennard scaling has resulted in the adoption of multicore processors, followed by the raise of domain-specific accelerators (e.g., NVIDIA tensor cores and Google TPUs) and asymmetric multicore designs (e.g., ARM big.LITTLE; Apple M1, M2; and Intel Alder Lake, Raptor Lake) [1]–[4]. Xilinx is not oblivious to the performance-power advantages of specialized hardware and has responded by introducing the Versal AI Engine (AIE) Core, a design that comprises a set of compute engines, advanced I/O, and integrated DDR controllers, targeting an ample range of workloads [5], [6].

In response to this, we analyze how to map the general matrix multiplication (GEMM) on an *AIE-enabled* Xilinx Versal Adaptive Compute Accelerated Platform (ACAP). Our main motivation for targeting this computational kernel is that GEMM is the cornerstone upon which many scientific and engineering applications are built. Furthermore, in deep learning (DL) training and inference with the popular CNNs (convolutional neural networks) and the recent transformers, most arithmetic can be cast in terms of GEMM [7], [8].

In addressing the efficient realization of GEMM on the Xilinx Versal, we make the following two major contributions:

- We demonstrate that the ideas underlying the modern realization of GEMM on conventional processor architectures, equipped with a hierarchical multilayered memory and single-instruction multiple-data (SIMD) arithmetic units, carry over to the AIE-enabled Xilinx Versal ACAP. For this purpose, we map the matrix operands to the

distinct levels of the memory hierarchy as well as develop an architecture-specific *micro-kernel* for the AIE tile.

- We customize our general design of the algorithm for the particular case of the Xilinx Versal VCK190. Furthermore, we perform a thorough experimental analysis on this platform, exposing the performance caveats and the key role of the cache configuration parameters.

At this point, we emphasize that our design is presented for the Versal VCK190, but it is straight-forward to adapt them for other AIE-enabled Versal ACAPs.

The rest of the paper is structured as follows: In Section III we briefly review the Versal VCK190; and in Section IV, we revisit the modern implementation of GEMM. Next, in Section V we describe the design of a GEMM micro-kernel for the Versal AIE tile, and the mapping of the GEMM algorithm into the memory organization of the AIE-enabled ACAP. In Section VI we perform a complete experimental evaluation of the proposed algorithm; and in Section VII we discuss the insights gained from our study, and sketch a roadmap for ongoing and future work. After that, we close the paper with a glossary of acronyms used during the text.

II. RELATED WORK

Multiply-accumulate (MAC) operations account for roughly 80% of the arithmetic in machine learning inference [9]. These MACs operations can often be aggregated into a GEMM kernel, for example in the case of (CNNs) [10], transformers [11], and long short-term memory (LSTM) networks [12]. As a result, extensive research has focused on accelerating GEMM, by optimizing memory accesses and/or speeding up the kernel using hardware accelerators. BLIS [13] provides significantly higher memory efficiency and has been combined with approximating computing [14], targeting CNN and working with SIMD based accelerators [10], [15], [16]. In this paper, we map the BLIS design into Xilinx Versal platform.

Matrix multiplication is a fundamental kernel for deep learning inference and training. For the particular case of CNNs, the IM2COL transform [8] casts the expensive convolution operators in terms of a GEMM by transforming the input activation tensor into a much larger matrix. For many convolutional layers, the GEMM that results from the application of this transform presents one large dimension and two small ones. From that point of view, it often offers substantial performance

gains to develop specialized implementations of GEMM to tackle these cases particular [16].

III. OVERVIEW OF THE XILINX VERSAL VCK190 ACAP

A. Heterogeneous architecture

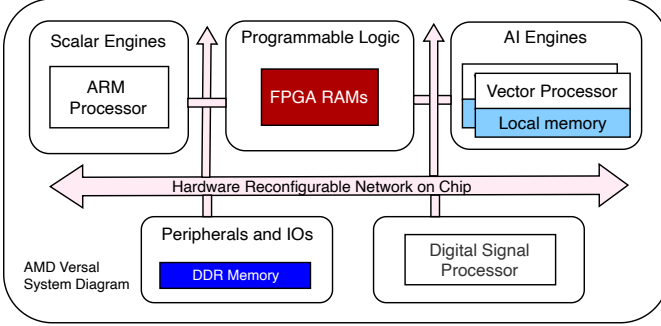


Fig. 1. Block diagram of the Versal AI Core.

The Versal AI Core comprises a collection of *heterogeneous* silicon designs that aim to deliver high performance for both edge and cloud computing workloads; see Figure 1. For the particular case of Versal VC1902 processor [5] targeted in this work, the architecture contains:

- 1) A dual-core ARM Cortex-A72 processor for compute-intensive workloads, plus a dual-core ARM Cortex-R5F processor for time-critical tasks.
- 2) 1,968 DSPs for accelerated arithmetic operations.
- 3) A customizable FPGA (or PL) with 899,840 LUTs.
- 4) 400 standalone “vector cores”, referred to as AIE tiles, and organized as a bidimensional array.

Among other components, each AIE tile contains a SIMD arithmetic unit and, altogether, these vector cores conform the VC1902’s key feature for compute-intensive (DL) workloads. Concretely, they provide up to 128 MAC operations per clock cycle (per AIE tile) for integer 8-bit (INT8) arithmetic [17]. Furthermore, the Versal VCK190 also contains a reprogrammable NoC for high-bandwidth communication between the modules.

B. Versatile memory architecture

The Versal VCK190 features a flexible memory architecture basically consisting of 1) 32 KB of “local” memory per AIE tile; 2) distributed Block and Ultra RAMs with capacity for 4.3 MB and 16.3 MB, respectively; and 3) a global 2-GB DDR4 memory. Furthermore, each AIE tile can directly access the 32 KB of its “neighbors” in the 2D grid, while the local memory of the remaining AIE tiles is reachable via DMA. The dedicated interconnects along with the NoC deliver non-blocking, deterministic communication for the AIEs.

The reconfigurability of the FPGA allows to leverage the Block RAM and Ultra RAM as temporary buffers for the AIE, with the data being directly fed to the AIE tiles via streaming interfaces for low latency access. The AIE tiles access the global DDR4 via the NoC for higher throughput.

IV. HIGH PERFORMANCE GEMM IN CONNECTION WITH DEEP LEARNING INFERENCE

A. GEMM in conventional processor architectures

Consider the GEMM $C \leftarrow C + AB$, where matrix A is $m \times k$, matrix B is $k \times n$, and matrix C is $m \times n$. Modern implementations of this kernel (e.g., those in AMD AOCL, OpenBLAS, BLIS and, possibly, Intel oneAPI) follow the ideas underlying GotoBLAS2 [18] to apply blocking (tiling) [19] to the matrix operands using five nested loops around two packing routines and an architecture-dependent *micro-kernel*; see Figure 2 top-right, and the loops labeled as L1, L2, ..., L5 there.

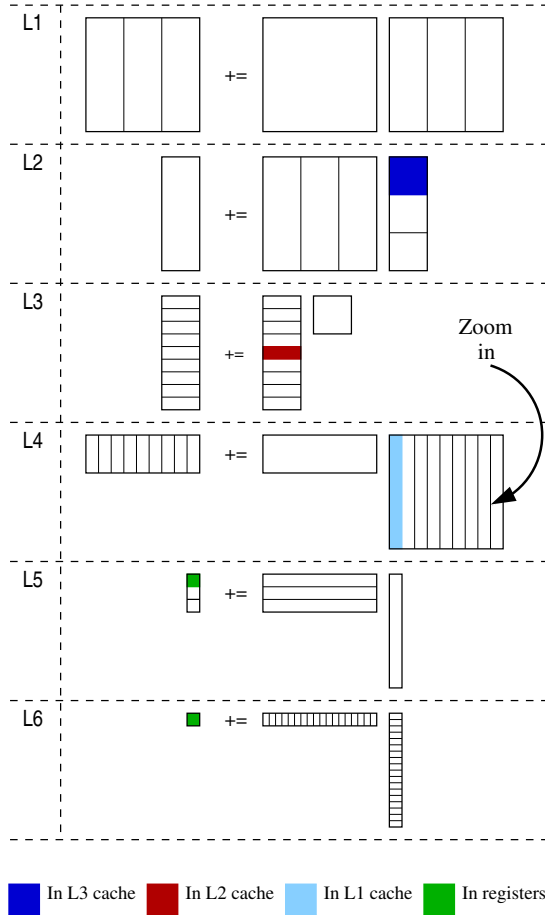
For this baseline algorithm, a proper selection of the strides (also known as the *cache configuration parameters*) for the three outermost loops, given by m_c, n_c, k_c , combined with a careful arrangement of the matrix inputs into two buffers, A_c, B_c (respectively, of dimensions $m_c \times k_c$ and $k_c \times n_c$; see Figure 2, left and top-right), orchestrated via the packing routines, substantially reduces the number of cache misses [20], [21]. For simplicity, hereafter we assume that m, n, k are integer multiples of m_c, n_c, k_c , respectively.

In addition, the micro-kernel comprises one more loop (labeled as L6 in Figure 2, middle-right) that repeatedly updates an $m_r \times n_r$ *micro-tile* of C , known as C_r , via a sequence of k_c rank-1 transforms, each involving a column of an $m_r \times k_c$ micro-panel of A_c and a row of a $k_c \times n_c$ micro-panel of B_c . These micro-panels correspond to the blocks of A_c, B_c denoted as A_r, B_r in Figure 2, bottom-right. In current processors with SIMD arithmetic units, the micro-kernel “dimensions”, given by $m_r \times n_r$, are chosen to accommodate vectorization inside the micro-kernel loop. In this line, the specialized arrangement introduced by the packing routines ensures accessing the data in the micro-panels with unit stride from the micro-kernel which, in turn, enables loading their data using SIMD instructions.

In summary, for high performance this formulation of GEMM relies on three parameters (m_c, n_c, k_c) which can be adjusted to fit the hierarchical memory system of current architectures (see Figure 2 left), plus an architecture-dependent micro-kernel that can be tuned, for example, to accommodate vectorization.

B. GEMM and DL inference

A significant part of the arithmetic costs of recent transformers for natural language processing is due to GEMM of the form $C \leftarrow C + AB$, where A is the weight matrix and B/C are the input/output activation matrices. In addition, for the popular CNNs leveraged in computer vision and general signal processing tasks, the lowering approach [8] casts the convolution operator in terms of a large GEMM where A is the filter matrix, C corresponds to the output activation matrix and B is the augmented matrix obtained by applying the IM2COL (or IM2ROW) transform to the input activation matrix. To tackle its large memory costs, this transformation can be blocked, as part of the packing done inside the realization of the GEMM [16], or applied on-the-fly, as part of the direct convolution algorithm [7].



```

1 for (jc=0; jc<n; jc+=nc)           // Loop L1
2 for (pc=0; pc<k; pc+=kc) {         // Loop L2
3   // Pack B
4   Bc := B(pc:pc+kc-1, jc:jc+nc-1);
5   for (ic=0; ic<m; ic+=mc) {       // Loop L3
6     // Pack A
7     Ac := A(ic:ic+mc-1, pc:pc+kc-1);
8     for (jr=0; jr<nc; jr+=nr)      // Loop L4
9       for (ir=0; ir<mc; ir+=mr)    // Loop L5
10        // Micro-kernel
11        C(ic+ir:ic+ir+mr-1,
12          jc+jr:jc+jr+nr-1)
13        += Ac(ir:ir+mr-1, 0:kc-1)
14        * Bc(0:kc-1, jr:jr+nr-1);
15  }}

```

```

1 for (pr=0; pr<kc; pr++)           // Loop L6
2   C(ic+ir:ic+ir+mr-1,
3     jc+jr:jc+jr+nr-1)
4   += Ac(ir:ir+mr-1, pr)
5   * Bc(pr, jr:jr+nr-1);

```

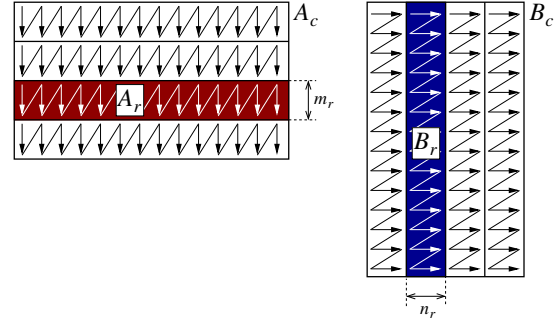


Fig. 2. Baseline high performance algorithm for GEMM. Left: data transfers across the memory hierarchy; Top-Right: blocked algorithm; Middle-Right: Micro-kernel; Bottom-Right: Packing of input matrix operands.

V. MAPPING GEMM TO XILINX VERSAL VCK190 FOR DEEP LEARNING INFERENCE

A. Distributing the data across the memory hierarchy

The Versal VCK190 offers a variety of possibilities for distributing the problem data across its memory resources in order to attain high data reusability. In particular, since the distinct memory levels in the Versal VCK190 present different bandwidth rates and capacities, (see Table I,) depending on the arithmetic intensity of the target algorithm, this feature can be exploited to improve efficiency via cache-friendly programming techniques, such as blocked algorithms [19], combined with a careful distribution of the problem data across the levels of the memory hierarchy.

For the Versal VCK190, on the one hand, the DDR4 memory provides sufficient space to accommodate the global program data which, in our problem, corresponds to the GEMM matrix operands B, C comprising the input/output activations in the case of DL inference. On the other hand, the RAMs in the programmable logic provide high memory throughput, and can be employed as a scratchpad to temporarily store the full pre-packed matrix A for the weight/filter matrix. In practice, this implies that there is no need to keep a copy of

the unpacked A in global memory. Finally, the fast and near-processor local memory delivers high throughput, and can be used to keep a packed micro-panel B_r . These three levels, global memory, FPGA RAMs and local memory, will play the role of the L1, L2 and L3 cache memories in a conventional processor (see Figure 2 left), as discussed next.

TABLE I
MULTI-LEVEL MEMORY HIERARCHY IN THE VERSAL VCK190.

Level	Capacity
AIE vector registers	2 KB
AIE tile local memory	32 KB
FPGA RAMs	20 MB
DDR4 (global) memory	2 GB

In Figure 3, we graphically illustrate the distribution of the matrix operands and packed buffers adopted for mapping the GEMM kernel on the Versal VCK190. For simplicity, hereafter we consider a CNN model, and the GEMM that results from applying the IM2ROW transform to a single convolution operator, taking into account the special usage of GEMM in a DL inference scenario to reduce the communication overhead:

- The operand A contains the read-only filters for the convolution. The full matrix can be thus pre-packed (off-

line) into a collection of buffers A_c and kept in the FPGA RAMs. Here we exploit that, once a DL model is deployed, it is used to perform inference with many input samples. Furthermore, for CNN models, the FPGA RAMs are usually large enough to contain the filter parameters (weights and biases) for all the model. Therefore, we can pre-pack A and pre-load the result into the FPGA storage, with a null cost for this during the inference process.

- Matrix B corresponds to the activation inputs of the convolution operator. This operand varies from one inference sample to another and, therefore, we need to pack it into buffer B_c during the execution of GEMM. Following the approach in high performance realizations of this operation, we perform this packing in the DDR4 global memory (as there is no equivalent of an L3 cache in the Versal ACAP). Also during the execution, the individual micro-panels B_r , are copied into the AIE local memory via GMIO. For this purpose, the AIE tile directly accesses the global memory through the NoC. Here we amortize the cost of transferring B_r by re-using the entries of this block multiple times; see loop L5 in Figure 2; top-right.
- Finally, matrix C contains the activation outputs of the convolution operator. At each execution of the micro-kernel, a small micro-tile C_r , of dimension $m_r \times n_r$, is first loaded directly from the DDR4 global memory into the AIE tile vector registers, and written back at the end of the micro-kernel; see Figure 2, left and middle-right.

At this point, it is worth remarking some differences between our solution and the approach taken when implementing GEMM in a conventional processor:

- From the point of view of hardware, a conventional processor integrates a number of cache levels (usually, between two and three), and relies on a memory controller to orchestrate the data movements across them. In contrast, in the Versal VCK190 the transfers need to be explicitly encoded into the algorithm. This puts an extra burden into the programmer’s shoulders, but offers a strict control over the data transfers as the intermediate memory levels behave as scratchpads.
- The Versal system provides (window and) streaming access methods for communication with the “outside world.” Streaming provides high throughput and lower latency for the PL to AIE communication, but is also fundamentally different from the type of communication that occurs between the arithmetic units and the cache/memory levels of a conventional processor.
- From the algorithmic point of view, given that the target is to implement GEMM for DL inference, we can pre-pack the weight/filter matrix A into a collection of buffers residing into the FPGA RAM. Therefore, there is no need for the packing that occurs within loop L3 of the baseline algorithm for GEMM.

B. Design of the micro-kernel

Each AIE tile has four accumulator registers that can be used to store the results of the vector data path. An

accumulator register is 768-bit wide and can be viewed as 16 accumulator lanes of $(768/16=)$ 48 bits each. For high performance, the width of the accumulator registers affects the practical dimensions of the micro-kernel. For the implementation of a DL-oriented GEMM in the Versal VC1902, we adopt INT16 as the baseline datatype (to be discussed later, in subsection VI-A), and set the dimensions of the micro-kernel to $m_r \times n_r = 16 \times 4$; that is, the micro-tile C_r updated by loop L6 of the micro-kernel comprises 16 entries (capacity of an accumulator register) for 4 columns (number of accumulator registers) of C .

Figure 4 displays our micro-kernel for the VC1902. After some initial declarations, the code comprises a loop (Line 19), corresponding to L6, that iterates over the `kc` dimension of the micro-panels A_r, B_r . At each iteration, the loop body multiplies the entries in one column of A_r (16 elements, in `ar0` or `ar1`) with those in one row of B_r (4 elements in `br`, accessed through the macro `Bref`), accumulating the intermediate results on four different registers (`mac` operations in Lines 27–49).

The Artificial Intelligence Engine (AIE) intrinsic `mac16()` computes 32 INT16 multiply-and-accumulate (MAC) operations in one cycle. For the INT16 datatype, each MAC operation can involve a vector with (up to) 32 elements and a second vector with (up to) 2 elements. However, our micro-kernel ($m_r \times n_r = 16 \times 4$) does not match the input dimensions of the intrinsic. To accommodate the dimension $n_r = 4$ to the requirements of `mac16()`, we therefore divide the 16×4 MAC operation into two parts; see Lines 33 and 42. In this manner, the additional call to the `mac16()` intrinsic increases the input by additional four times to fully utilize the function.

Each iteration of the loop body retrieves four rows of A_r from the FPGA memory (64 elements, in Lines 27–30) and 4 columns of B_r from the local memory (16 elements, in Line 24). Therefore, a loop iteration retrieves 64+16 elements from memory levels that are close to the arithmetic units, performing $64 \cdot 16 \cdot 2 = 2,048$ arithmetic operations with those. This helps to amortize the cost of memory transfers with enough arithmetic computation, in principle yielding a compute-bound micro-kernel.

Splitting the computation into two parts delivers a high computation-to-communication ratio. Furthermore, the high utilization of the accumulator and vector registers (respectively, 100% and 75% of the total resources) along with compiler optimization arguments facilitate to overlap the MAC operation with data transfer, thus improving performance.

After the loop is complete, the code “transfers” these intermediate results to matrix C . For this purpose, each execution of the micro-kernel loads a 16×4 micro-tile C_r (Lines 53–56) from global memory and, after updating its contents, stores the results back to global memory (Lines 59–63). The cost of these final data transfers can be amortized provided k_c is sufficiently large (to be discussed later).

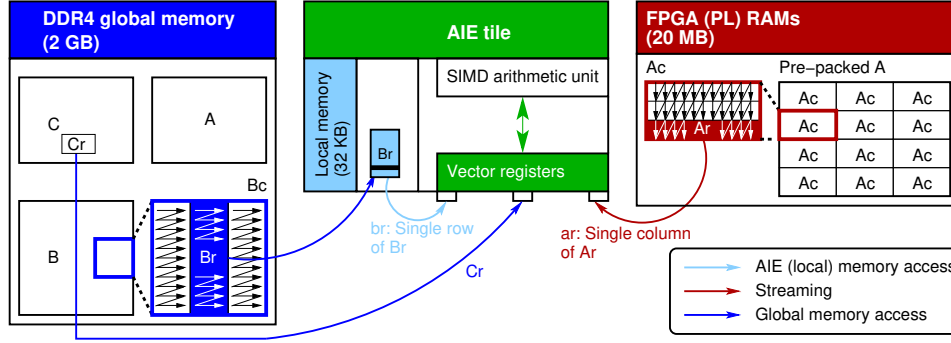


Fig. 3. Data mapping and transfers between different memories of Versal VCK190. The data transfers from the micro-tile C_r in the global memory, the micro-panel A_r in the FPGA memory, and the micro-panel B_r in the local memory move data directly to the vector registers. The data copy of a micro-panel of B_C in the global memory to B_r in the local memory is carried out by one of the scalar engines (ARM processors).

VI. EXPERIMENTAL RESULTS

The evaluation of the GEMM algorithm in this section was carried out using the Xilinx Vitis 2022.1 developing tool. The AIE transaction-level System C simulator was used to profile the timing, resource requirements, and assembly instructions of the designs, enabling an accurate performance analysis [22].

A. Approximate computing with different datatypes

Precision scaling is a prominent approximate computing technique that is often leveraged in quantized DL inference on edge devices in order to reduce energy consumption while maintaining acceptable precision in the results. In this line, the Versal VC1902 supports arithmetic for multiple datatypes, which can be exploited to improve the cost-effectiveness ratio. Concretely, the Versal provides a variety of real datatype precision formats. Concretely, the AIE supports INT8, INT16 and FLOAT32, yet the performance of the different formats varies significantly, with lower precision requiring fewer AIE cycles. For example, in one clock cycle, the acAIE can perform 128 UINT8 MACs operations; 32 for INT16, and 8 for FP32.

These figures motivate us to choose INT16 as the baseline datatype for our DL-oriented GEMM kernel on the Versal VCK190, as this datatype provides fair precision, reducing memory utilization and accelerating computation. In addition, the `mac16()` intrinsic for the INT16 datatype accommodates a flexible implementation. The general design of the micro-kernel will be adapted for lower precision in future work.

B. Evaluation of the micro-kernel for the AIE tile

In subsection VI-B we motivated the selection of an $m_r \times n_r = 16 \times 4$ micro-kernel (for the INT16 datatype). While there exist other micro-kernel dimensions that will also produce correct results, choosing a smaller micro-kernel results in 1) using fewer accumulator registers/lanes which will likely render idle cycles for the accumulator arithmetic units yielding lower performance; and 2) a worse arithmetic-memory ratio for the body of loop L6. Furthermore, the compiler can transparently deal with a micro-kernel that exceeds the maximum number of registers/lanes by temporarily saving

certain data to memory (and eventually restore it) during the micro-kernel execution. Unfortunately, this register spilling can significantly degrade performance when occurring inside loop L6.

To illustrate this, we performed the following two experiments. In the first case, we profiled the execution of our 16×4 micro-kernel with $k_c = 256$. The number of MAC operations performed inside the micro-kernel for this configuration is thus 16,384, and the design took 596 cycles to complete, which corresponds to 27.5 MACs per cycle. The theoretical peak for the AIE tile when operating with INT16 and accumulating on the 16 lanes per accumulator register is 32 MACs per cycle. Here, the small performance drop due to the cost of loading the micro-tile C_r from the global memory, updating it, and storing the results back in memory.

In the second configuration, we employed a 32×4 micro-kernel which, in theory, features a higher re-use rate of C_r while this micro-kernel resides in the local memory. However, this approach exceeds the capacity of the accumulator registers, producing register spilling. Concretely, with $k_c = 256$, this design performs twice as many MAC operations as the previous one, and should thus offer a theoretical cost that roughly corresponds to $596 \cdot 2 = 1,192$ cycles. In practice though, it required 1,429 cycles to complete, that is, about 20% longer than expected. The second configuration thus only achieves 23 MACs per cycle, which is far from the AIE peak.

We close the previous discussion with two final remarks:

- The accumulator registers in the Versal VC1902 can be viewed as four 768-bit accumulators. Our micro-kernel involves 8 `mac16()` intrinsics that operate with these 4 independent accumulator registers, therefore using 100% of this type of resource. Maintaining the INT16 data type while augmenting the size of micro-kernel could increase the utilization of the register space, but this would eventually lead to register spilling. The current design utilizes 75% of the space from the vector register resources. Indeed, the second scenario presented in this section (with a 32×4 micro-kernel) exceeds the maximum vector register resources. In consequence, during the execution of the micro-kernel loop, some of these extra


```

1 #define Bref(i,j) Br[i*nr+j]
2 void micro_kernel( input_window_int16 * __restrict DDR_IN,
3                   input_window_int16 * __restrict PL_IN,
4                   int16 *Br,
5                   output_window_int16 * __restrict out) {
6
7     // Vectors, Accumulator for ar, br, C
8     v32int16 ar0, ar1;
9     v16int16 br;
10    v16acc48 Cacc0, Cacc1, Cacc2, Cacc3;
11
12    // Parameters for mac16() intrinsics
13    unsigned int xoffsets      = 0x73727170;
14    unsigned int xoffsets_hi   = 0x77767574;
15    unsigned int xsquare       = 0x3120;
16    unsigned int zoffsets      = 0x0;
17    unsigned int zoffsets_hi   = 0x0;
18
19    for (unsigned int i=0; i<Kc/4; i+=1)
20        // Unroll loop to overlap transfers and computations
21        chess_prepare_for_pipelining
22        chess_loop_range(Kc/4, ) {
23            // Read a vector of 16 INT16 into br
24            br = *(v16int16*) Bref[i];
25
26            // Read into ar
27            ar0 = upd_w( ar0 ,0, window_readincr_v16(PL_IN));
28            ar0 = upd_w( ar0 ,1, window_readincr_v16(PL_IN));
29            ar1 = upd_w( ar1 ,0, window_readincr_v16(PL_IN));
30            ar1 = upd_w( ar1 ,1, window_readincr_v16(PL_IN));
31
32            // Compute two mac16()
33            Cacc0 = mac16(Cacc0, ar0 ,0, xoffsets, xoffsets_hi,
34                          xsquare, br ,0 , zoffsets, zoffsets_hi,4);
35            Cacc1 = mac16(Cacc1, ar0 ,0, xoffsets, xoffsets_hi,
36                          xsquare, br ,1 , zoffsets, zoffsets_hi,4);
37            Cacc2 = mac16(Cacc2, ar0 ,0, xoffsets, xoffsets_hi,
38                          xsquare, br ,2 , zoffsets, zoffsets_hi,4);
39            Cacc3 = mac16(Cacc3, ar0 ,0, xoffsets, xoffsets_hi,
40                          xsquare, br ,3 , zoffsets, zoffsets_hi,4);
41
42            //
43            Cacc0 = mac16(Cacc0, ar1 ,0, xoffsets, xoffsets_hi,
44                          xsquare, br ,8 , zoffsets, zoffsets_hi,4);
45            Cacc1 = mac16(Cacc1, ar1 ,0, xoffsets, xoffsets_hi,
46                          xsquare, br ,9 , zoffsets, zoffsets_hi,4);
47            Cacc2 = mac16(Cacc2, ar1 ,0, xoffsets, xoffsets_hi,
48                          xsquare, br ,10, zoffsets, zoffsets_hi,4);
49            Cacc3 = mac16(Cacc3, ar1 ,0, xoffsets, xoffsets_hi,
50                          xsquare, br ,11, zoffsets, zoffsets_hi,4);
51        }
52
53    // Read Cr from global memory
54    v16int16 C0 = window_readincr_v16(DDR_IN);
55    v16int16 C1 = window_readincr_v16(DDR_IN);
56    v16int16 C2 = window_readincr_v16(DDR_IN);
57    v16int16 C3 = window_readincr_v16(DDR_IN);
58
59    // Convert result, add it to Cr and write back to memory
60    C0=operator+(srs(Cacc0,0),C0);window_writeincr(out,C0);
61    C1=operator+(srs(Cacc1,0),C1);window_writeincr(out,C1);
62    C2=operator+(srs(Cacc2,0),C2);window_writeincr(out,C2);
63    C3=operator+(srs(Cacc3,0),C3);window_writeincr(out,C3);
64 }

```

Fig. 4. Simplified version of the micro-kernel for the AIE tile.

vector need to be temporally stored in the local memory to be later retrieved back into vector registers, leading to a performance degradation that we aim to avoid. [17]

- The idea behind the 48-bit accumulators is to have 16-bit multiplication results and accumulate over those results without bit overflows. In the VCK190, 48-bit is the minimum accumulator “size”.

C. Selecting the cache configuration parameter k_c

Having determined the dimensions of the micro-kernel, we next investigate how to select the optimal value for k_c . Note that we proceed through the hierarchy from the processor registers toward the slower levels of the memory system. Thus, in the previous subsection we discussed the values of $m_r \times n_r$ in connection with the number of accumulator registers; in this subsection we choose k_c taking into account the capacity of the local memory; and in the next subsection we will investigate how to set the optimal value of m_c with respect to the capacity of the FPGA RAMs.

Ideally, we would like k_c to be as large as possible because, from the point of view of the micro-kernel, that option reduces the overhead of loading/saving the micro-tile C_r from/to global memory ($m_r \times n_r$ reads plus the same number of writes from that level) compared with the number of arithmetic operations ($2m_r \cdot n_r \cdot k_c$). Also, the micro-panel B_r that resides in the local memory is $k_c \times n_r$ and a larger k_c implies a better re-use of the resources in that level. However, with $n_r = 4$ already set, the largest dimension for k_c is limited by the capacity of the local memory. Furthermore, as k_c is the blocking parameter for the k -dimension of GEMM, from the application perspective we have the constraint that $k_c \leq k$.

In this subsection, we first demonstrate the impact of k_c on the performance of the micro-kernel. For that purpose, we profiled the ratio of cycles spent in MAC operations (useful arithmetic) against the total number of cycles (including the overhead due to loading/storing data) when the micro-kernel is executed in isolation. Figure 5 illustrates the effect of this parameter on performance. For instance, with $k_c = 64$, the micro-kernel required a total of 212 cycles to execute, out of which 150 correspond to MAC operations and the remaining are due to data transfers. Thus the efficiency of this configuration is around 60%. As k_c is increased, it approaches an horizontal asymptote: 75.3%, 85.9%, and 87.6% for $k_c = 128$, 256 and 290, respectively. When $n_r = 4$, at most $k_c = 290$ in order to ensure that the micro-panel B_r fits into the local memory.

This experiment shows that the ratio between arithmetic and data transfers for the micro-kernel strongly depends on the value of the cache configuration parameter k_c . In particular, the copies for C_r from the global memory to local memory can be basically hidden by choosing an appropriate (large) value for k_c . In addition, choosing a proper dimension of the micro-kernel helps to reduce the negative impact of the copies for B_r from the local memory and the micro-panels of A_c from the FPGA memory to about 12% of the cycles.

Here it is important to emphasize that, once k_c is fixed, this sets an upper bound on the performance of the GEMM kernel, since the complete algorithm is assembled using the micro-kernel as the cornerstone building block. The only missing data transfers that we did not consider yet, because they do not occur inside the micro-kernel, correspond to the copies from the buffer B_c , in global memory, to the micro-panel B_r , in the local memory; see Figure 3. These copies are orchestrated

by the scalar engines (ARM processors) in the VCK190 and their cost is discussed in the following subsection.

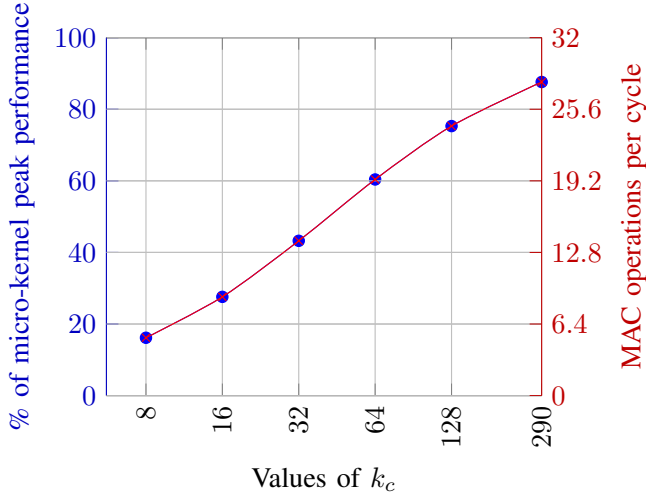


Fig. 5. Impact of k_c on the performance of the $m_r \times n_r = 16 \times 4$ micro-kernel.

D. Selecting the cache configuration parameter m_c

Consider now the buffer A_c that resides in the FPGA RAMs, of dimension $m_c \times k_c$, and let us follow a reasoning analogous to that applied for k_c . First, there are two upper bounds on m_c : 1) With k_c set in the previous subsection, the largest dimension for m_c can be easily derived from the capacity of the FPGA memory; and 2) from the application point of view, as m_c is the blocking parameter from the problem dimension m , we have that $m_c \leq m$. Let us discuss next the benefits of choosing a large value for m_c by analyzing the overhead due to a data copy from the buffer B_c in the global memory to the micro-panel B_r in the AIE tile local memory. For this purpose, consider for example that $k_c = 290$ and $n_r = 4$. For this configuration, the data copy between global memory and the local storage introduces a significant overhead: 8,309 cycles for the copy versus 8,952 cycles for the execution of a single micro-kernel. The number of MAC operations per cycle is thus reduced to only 2. Fortunately, this type of data copy only occurs once per iteration of loop L4, and the copied micro-panel B_r is then re-used in m_c/m_r executions of the micro-kernel (one per iteration of loop L5). Therefore, with m_r already fixed, this reasoning points in the direction of choosing a large value for m_c , because a large ratio m_c/m_r then paves the road toward amortizing the cost of the copy for B_r with enough arithmetic from inside the micro-kernel.

To close this section, we explore the experimental impact of m_c on the performance for a specific GEMM problem. For that purpose, we chose $(m, n, k) = (4,096, 4,096, 290)$, set $n_c = n, k_c = k$, and varied m_c for different executions of the GEMM kernel. Figure 6 shows the performance benefits from choosing a larger value for m_c , but at the same time exposes a clear horizontal asymptote around 87.6% of the peak performance (which corresponds to 32 INT16 MACs,

or 64 INT16 arithmetic operations, per cycle). Therefore, this is perfectly aligned with the performance of the micro-kernel when considered in isolation (that is, the building block for our GEMM kernel), which in Figure 5 was reported to achieve up to 86.7% of the peak performance. This implies that the overhead due to the copies between B_c and B_r can be hidden with arithmetic provided m_c is chosen to be large enough.

As a rough summary of the previous analysis, our realization of GEMM for the Versal losses 15% of the peak performance because of the copies from A_c and B_r to the vector registers, while an additional 4–5% is lost because of the transfers between B_c and B_r .

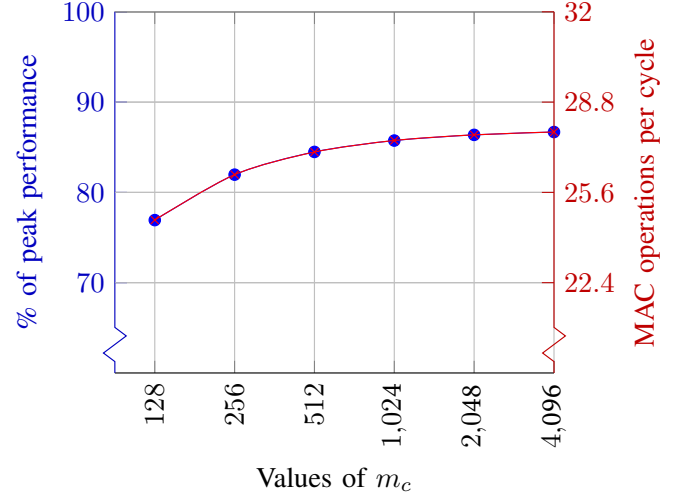


Fig. 6. Impact of m_c on the performance of the GEMM kernel. For this experiment, $m_r \times n_r = 16 \times 4$, $(m, n, k) = (4,096, 4,096, 290)$, $n_c = n$, and $k_c = k$.

VII. FINAL DISCUSSION AND FUTURE WORK

In this work, we have designed a prototype GEMM kernel that follows the ideas of GotoBLAS2, matching the different data transfers of the algorithm across the memory hierarchy of the Versal VCK190 ACAP as well as developing an appropriate micro-kernel (to operate with INT16 input data) on the AIE tile. From the practical point of view, we have demonstrated that, by carefully choosing the “cache” configuration parameters, this design achieves about 28 MACs per cycle, which corresponds to 86.7% of the AIE peak performance *when using the* `mac16()` (advanced level AI Engine Intrinsic) [23]. This low level library provides detailed control of the implementation, as well as close-to-peak performance. In future work, we plan to address the following aspects:

INT8 based Advanced intrinsics. The current INT16 delivers 32 MACs per cycle, yet lower precision datatypes such as INT8 can deliver up to 128 MACs per cycle. Our initial implementation, with the $m_r \times n_r = 16 \times 4$ micro-kernel and INT16 operands, delivers performance close to the theoretical peak for INT16 when k_c sufficiently large. The design of an INT8-based micro-kernel and its integration into the GEMM algorithm is ongoing work.

Exploiting multiple AIE tiles. At this point we recognize that a high performance implementation of GEMM for an AIE-enabled Versal should target all AIE tiles in the system. Doing so for a platform with a very large count of tiles without hitting a hard memory wall due to the data accesses is a very challenging task yet inescapable objectives. For this purpose, we expect that it becomes primordial to exploit the high speed interconnection among tiles to exchange the data in the local memories.

VIII. GLOSSARY

ACAP: Adaptive Compute Acceleration Platform; **AIE:** Artificial Intelligence Engine; **API:** Application Programming Interface; **CNN:** Convolutional Neural Network; **DDR:** Double Data Rate (memory); **DL:** Deep Learning; **DSP:** Digital Signal Processor; **GMIO:** Global Memory Input/Output; **INT8:** Integer 8-bit (arithmetic); **LUT:** Look-Up Table; **MAC:** Multiply-and-Accumulate; **NoC:** Network-on-chip; **PL:** Programmable Logic; **SIMD:** Single-Instruction Multiple-Data; **TOPS:** Tera-operations per second; **TPU:** Tensor Processing Unit; **GEMM:** General Matrix Multiplication; **RAM:** Random Access Memory.

IX. ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from European Union's Horizon2020 Research and Innovation programme under the Marie Skłodowska Curie Grant Agreement No. 956090 (APROPOS, <http://www.apropos-itn.eu/>). This work also received funding in Spain from the research project PID2020-113656RB-C22 of MCIN/AEI/10.13039/501100011033, y por FEDER *Una manera de hacer Europa*, as well as from European High-Performance Computing Joint Undertaking (JU) under grant agreement No. 955558 (eFlows4HPC project). The JU receives support from the European Union's Horizon 2020 research and innovation programme, and Spain, Germany, France, Italy, Poland, Switzerland, Norway.

REFERENCES

- [1] G. E. Moore, "Cramming more components onto integrated circuits, reprinted from electronics, volume 38, number 8, april 19, 1965, pp.114 ff." *IEEE Solid-State Circuits Society Newsletter*, vol. 11, no. 3, pp. 33–35, 2006.
- [2] R. Dennard, F. Gaensslen, H.-N. Yu, V. Rideout, E. Bassous, and A. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, 1974.
- [3] J. L. Hennessy and D. A. Patterson, "A new golden age for computer architecture," *Commun. ACM*, vol. 62, no. 2, p. 48–60, 2019.
- [4] N. P. Jouppi *et al.*, "In-datacenter performance analysis of a tensor processing unit," *SIGARCH Comput. Archit. News*, vol. 45, no. 2, p. 1–12, jun 2017. [Online]. Available: <https://doi.org/10.1145/3140659.3080246>
- [5] S. Ahmad *et al.*, "Xilinx first 7nm device: Versal AI Core (VC1902)," in *2019 IEEE Hot Chips 31 Symposium (HCS)*, 2019, pp. 1–28.
- [6] M. Voogel *et al.*, "Xilinx versal premium," in *2020 IEEE Hot Chips 32 Symposium (HCS)*, 2020, pp. 1–46.
- [7] S. Barrachina *et al.*, "Reformulating the direct convolution for high-performance deep learning inference on ARM processors," *J. of Systems Architecture*, 2022, to appear.
- [8] K. Chellapilla, S. Puri, and P. Simard, "High performance convolutional neural networks for document processing," in *International Workshop on Frontiers in Handwriting Recognition*, 2006, available as INRIA report inria-00112631 from <https://hal.inria.fr/inria-001126>.
- [9] A. Castelló, S. Barrachina, M. F. Dolz, E. S. Quintana-Ortí, P. S. Juan, and A. E. Tomás, "High performance and energy efficient inference for deep learning on multicore arm processors using general optimization techniques and BLIS," *Journal of Systems Architecture*, vol. 125, p. 102459, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1383762122000509>
- [10] C. Ramírez, A. Castelló, and E. S. Quintana-Ortí, "A BLIS-like matrix multiplication for machine learning in the RISC-V ISA-based GAP8 processor," *The Journal of Supercomputing*, vol. 78, no. 16, pp. 18 051–18 060, may 28 2022.
- [11] H. Peng, S. Huang, T. Geng, A. Li, W. Jiang, H. Liu, S. Wang, and C. Ding, "Accelerating transformer-based deep learning models on fpgas using column balanced block pruning," in *2021 22nd International Symposium on Quality Electronic Design (ISQED)*, 2021, pp. 142–148.
- [12] W. Zhang, F. Ge, C. Cui, Y. Yang, F. Zhou, and N. Wu, "Design and implementation of lstm accelerator based on fpga," in *2020 IEEE 20th International Conference on Communication Technology (ICCT)*, 2020, pp. 1675–1679.
- [13] F. G. Van Zee and R. A. van de Geijn, "BLIS: A framework for rapidly instantiating blas functionality," *ACM Trans. Math. Softw.*, vol. 41, no. 3, jun 2015. [Online]. Available: <https://doi.org/10.1145/2764454>
- [14] F. G. Van Zee, D. N. Parikh, and R. A. V. D. Geijn, "Supporting mixed-domain mixed-precision matrix multiplication within the BLIS framework," *ACM Trans. Math. Softw.*, vol. 47, no. 2, apr 2021. [Online]. Available: <https://doi.org/10.1145/3402225>
- [15] A. Castelló, E. S. Quintana-Ortí, and F. D. Igual, "Anatomy of the BLIS family of algorithms for matrix multiplication," in *2022 30th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, 2022, pp. 92–99.
- [16] P. San Juan, A. Castelló, M. F. Dolz, P. Alonso-Jordá, and E. S. Quintana-Ortí, "High performance and portable convolution operators for multicore processors," in *2020 IEEE 32nd International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*, 2020, pp. 91–98.
- [17] "AI Engine kernel coding best practices guide (UG1079)," <https://docs.xilinx.com/r/en-US/ug1079-ai-engine-kernel-coding>, accessed: 2022-05.
- [18] K. Goto and R. A. van de Geijn, "Anatomy of a high-performance matrix multiplication," *ACM Transactions on Mathematical Software*, vol. 34, no. 3, pp. 12:1–12:25, May 2008.
- [19] K. Dowd and C. R. Severance, *High Performance Computing*, 2nd ed. O'Reilly, 1998.
- [20] F. G. Van Zee and R. A. van de Geijn, "BLIS: A framework for rapidly instantiating BLAS functionality," *ACM Trans. Math. Softw.*, vol. 41, no. 3, pp. 14:1–14:33, 2015.
- [21] T. M. Low *et al.*, "Analytical modeling is enough for high-performance BLIS," *ACM Trans. on Mathematical Software*, vol. 43, no. 2, pp. 12:1–12:18, Aug. 2016.
- [22] Xilinx, "AI Engine tools and flows user guide (UG1079)," <https://docs.xilinx.com/r/en-US/ug1079-ai-engine-kernel-coding/Tools>, 10 2022, accessed 2022-10.
- [23] "AIE intrinsics," https://www.xilinx.com/htmldocs/xilinx2022_1/aiengine_intrinsics/intrinsics/group_vect_mult_8x8.html#ga792254d404b81ed033a0b3b4158f43fa, accessed: 2022-11.