

DGEMM without FP64 Arithmetic – Using FP64 Emulation and FP8 Tensor Cores with Ozaki Scheme

Daichi Mukunoki

mukunoki@cc.nagoya-u.ac.jp

Information Technology Center, Nagoya University
Nagoya, Aichi, Japan

Abstract

As the demand for AI computation rapidly increases, more hardware is being developed to efficiently perform the low-precision matrix multiplications required by such workloads. However, these operations are generally not directly applicable to scientific computations due to accuracy requirements. The Ozaki scheme – an accurate matrix multiplication method proposed by Ozaki et al. in 2012 – enables FP64 matrix multiplication (DGEMM) using low-precision matrix multiplication units, such as FP16 Tensor Cores. This approach has since been extended to utilize integer arithmetic, offering lower computational cost compared to floating-point-based implementations. In fact, it has achieved higher performance than hardware FP64 operations on GPUs equipped with fast INT8 Tensor Cores designed for AI workloads. However, recent AI-oriented processors trends have shifted toward improving the performance of low-precision floating-point operations, such as FP8, rather than integer operations. Motivated by this shift, this study revisits the use of low-precision floating-point operations in the Ozaki scheme. Specifically, we explore the use of FP8 Tensor Cores. In addition, for processors that support very slow or no hardware-based FP64 operations, we also consider FP64 arithmetic emulation based on integer arithmetic. This completely eliminates hardware FP64 instructions. Furthermore, we explore the use of blocking in the inner-product dimension to accelerate FP16-based implementations. We demonstrate the effectiveness of these methods by evaluating the performance on an NVIDIA RTX Blackwell architecture GPU.

Keywords

Ozaki scheme, matrix multiplication, GEMM, Tensor Cores, FP8

ACM Reference Format:

Daichi Mukunoki. 2018. DGEMM without FP64 Arithmetic – Using FP64 Emulation and FP8 Tensor Cores with Ozaki Scheme. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 9 pages. <https://doi.org/XXXXXX.XXXXXXXX>

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.

Conference acronym 'XX, Woodstock, NY

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/2018/06
<https://doi.org/XXXXXX.XXXXXXXX>

1 Introduction

With the rapid advancement of AI technology, the demand for computing resources has surged. Since low-precision matrix multiplication is at the core of many AI workloads, processors with dedicated units like NVIDIA Tensor Cores have been developed. For example, on NVIDIA Blackwell architecture GPUs (2024), the 5th generation Tensor Core supports the following floating-point formats¹: FP64 (IEEE binary64), TensorFloat-32 (TF32), FP16 (IEEE binary16), bfloat16 (BF16), FP8 (E4M3² and E5M2), FP6 (E3M2 and E2M3), FP4 (E2M1), and the INT8 integer format. Furthermore, processors specifically designed for AI processing that do not support FP64 are also emerging.

However, traditional scientific computations that rely on FP64 cannot directly benefit from such hardware. The differing precision requirements between general-purpose and AI workloads increase hardware design complexity and raise development costs. Moreover, as AI workloads become more dominant in general-purpose systems, the improvement of FP64 performance is stagnating relative to that of low-precision operations. For instance, the performance ratio of FP64 to FP16, when using Tensor Cores and without sparsity, is 1:16 on the NVIDIA A100 (2020), approximately 1:29.5 on the H100 SXM (2022), and 1:125 on the GB200 NVL72 (2024). This stagnation challenges sustainable performance growth in scientific computing.

Matrix multiplication is a fundamental operation in scientific computing, widely used across various applications. It is computationally intensive, and its performance is highly dependent on FP64 arithmetic throughput. Consequently, matrix multiplication is among the operations most affected by the stagnation of FP64 performance improvements. In this paper, we refer to matrix multiplication as GEMM (general matrix multiplication), following the terminology of the Basic Linear Algebra Subprograms (BLAS) [6].

The Ozaki scheme [28], proposed by Ozaki et al. in 2012, is a method that enables high-precision matrix multiplication to be computed using low-precision matrix multiplications. For example, it allows FP64 GEMM (also known as DGEMM) to be performed using Tensor Core operations on FP16 data [23]. Furthermore, this method has been extended to utilize integer GEMM [26], which reduces computational cost compared to floating-point-based approaches. This technique has garnered significant attention due to its ability to exploit the rapidly improving performance of INT8 operations for AI workloads.

However, as of 2025, a noteworthy trend in AI processors is the shift from INT8 to lower-bit floating-point formats such as FP8. While INT8 is efficient for inference tasks [14], it is not well-suited

¹FP32 (IEEE binary32) is supported only on the CUDA core.

²'E' and 'M' indicate the bit lengths of the exponential and mantissa parts, respectively.

Table 1: Theoretical peak performance of GB300 NVL72 and GB200 NVL72 systems (Tensor Core performance does not consider sparsity).

	GB300 NVL72	GB200 NVL72
INT8 Tensor Core	12 POps/s	360 POps/s
FP4 Tensor Core	1,080 PFlops/s	720 PFlops/s
FP8/FP6 Tensor Core	360 PFlops/s	360 PFlops/s
FP16/BF16 Tensor Core	180 PFlops/s	180 PFlops/s
TF32 Tensor Core	90 PFlops/s	90 PFlops/s
FP32	5760 TFlops/s	5760 TFlops/s
FP64/FP64 Tensor Core	100 TFlops/s	2880 TFlops/s

for training. FP8, on the other hand, can be used for both inference and training [19], driving the development of unified FP8-based hardware. For example, as shown in Table 1³, in the GB200 (2024) equipped with Blackwell architecture GPUs, FP8 and INT8 achieve the same throughput. However, in the planned specifications for the GB300, INT8 throughput is only about one-thirtieth that of FP8. Although the GB200 and GB300 may target different workloads, the emergence of a product that prioritizes FP8 over INT8 is noteworthy. For low precision INT, The Hopper architecture (2022) removed INT4 support from the Tensor Cores⁴, which had been introduced in the Ampere architecture (2020). As for AMD, the Instinct MI350 series (2025) supports INT8 with the same throughput as FP8. However, in the Keynote presentation at Advancing AI 2025⁵, there was no mention of INT8 performance; instead, only the performance of low-precision FP was highlighted.

Against this background, we revisit the use of low-precision floating-point operations in the Ozaki scheme, targeting modern AI hardware. This paper first evaluates the performance of DGEMM using FP8 Tensor Cores on a Blackwell GPU, in comparison with FP16. Then, considering AI processors with limited or no FP64 support, we develop an implementation that eliminates FP64 instructions entirely and analyze its performance. Furthermore, we propose a new blocking strategy to further enhance performance.

It should be noted that the primary objective of this study is not to outperform DGEMM implementations that utilize hardware FP64 units on specific processors. Rather, the goal is to offer new insights into the application of low-precision operations such as FP16 and FP8 in the Ozaki scheme for DGEMM emulation.

The remainder of this paper is organized as follows. Section 2 introduces related work of this study. Section 3 outlines the overview of the Ozaki scheme. Section 4 presents extended implementations of the Ozaki scheme. Section 5 presents the experimental results on an RTX Blackwell GPU. Finally, the conclusion is presented in Section 6.

2 Related Work

The concept of performing high-precision computations using low-precision operations has existed for a long time. Early work includes double-word arithmetic proposed by Dekker in 1971 [4]. It realizes

³<https://resources.nvidia.com/en-us-blackwell-architecture>

⁴CUDA cores, not Tensor Cores, support INT4[17].

⁵<https://www.amd.com/content/dam/amd/en/documents/corporate/events/advancing-ai-2025-distribution-deck.pdf>

double-length mantissa using floating-point arithmetic of a certain precision. This technique has been extended to triple-word [7] and quadruple-word [13] arithmetic. QD⁶ is known as an implementation of double- and quadruple-word arithmetic for Fortran and C++. These multi-word arithmetic techniques have been applied to enable FP64-equivalent computations on processors that either lack FP64 support or deliver significantly higher performance for lower-precision operations. For example, the use of double-word arithmetic to enhance precision on GPUs without native FP64 support has been explored [30]. Later, with the introduction of FP16 and BF16 [2] for AI workloads, similar approaches were investigated. For instance, matrix multiplication using double-word arithmetic with FP16 [8] and double- or triple-word arithmetic with BF16 [12] have been studied.

In addition to these floating-point arithmetic-based methods, high-precision (and arbitrary-precision) arithmetic libraries using integer arithmetic have been developed. For example, the GNU Multiple Precision Arithmetic Library (GMP)⁷ [11] is a C library that supports arbitrary-precision integer and floating-point arithmetic. The GNU MPFR Library (MPFR)⁸ [9] builds upon GMP and provides arbitrary-precision floating-point arithmetic compliant with the IEEE 754 standard, guaranteeing correct rounding. Also, GNU and Intel compilers provide FP128 (IEEE binary128, 128-bit floating-point, also known as quadruple-precision) arithmetic operations.

For GEMM, methods to achieve FP32-equivalent accuracy for GEMM using FP16 Tensor Cores [18][27] have been proposed. Later, the Ozaki scheme [28] has gained prominence as a method for computing high-precision GEMM using low-precision GEMM. It was originally proposed to compute accurate GEMM on matrices stored in a given floating-point format, using arithmetic of the same format. The method was later extended to allow different formats for data storage and computation. For example, it enables DGEMM to be computed using FP16 Tensor Cores [23]. A major advantage of the Ozaki scheme is that it can utilize highly optimized GEMM routines provided for specific architectures, whereas previous methods require implementing GEMM from scratch. The scheme has also been applied in various contexts, such as high-precision GEMM (e.g., FP128 [24] and double-/triple-/quadruple-word arithmetic [15]), computational methods ensuring reproducibility [21][22][25], and quantum chemistry calculations [3].

A notable extension of the Ozaki scheme is its use of integer arithmetic proposed by Ootomo et al.[26] in 2024, which reduces computational cost compared to the floating-point-based approach. This is because the low-precision GEMMs in the Ozaki scheme are computed in fixed-point form and do not require handling the exponent part. This approach is well suited to the GPU architectures, as INT8 Tensor Core performance is rapidly improving. Subsequently, the integer-based method has been further refined [1][31]. The latest advancement is Ozaki Scheme II [29], which leverages the Chinese Remainder Theorem (CRT) to further reduce computational cost.

Beyond GEMM, there are also efforts to apply low-precision operations to other computations. Examples include the use of FP16

⁶<https://github.com/BL-highprecision/QD>

⁷<https://gmplib.org>

⁸<https://www.mpfr.org>

Tensor Cores for Fast Fourier Transformations [16] and the application of low-precision formats such as FP8 and FP16 in sparse matrix computations [10][32]. Additionally, a comprehensive discussion on the effectiveness of AI-oriented matrix multiplication units, such as Tensor Cores, in various HPC applications has been presented [5].

3 Ozaki Scheme

This section outlines the mixed-precision version [23] of the Ozaki scheme, which constitutes the core technique used in this study.

The Ozaki scheme is an error-free transformation for matrix multiplication that decomposes a single floating-point matrix multiplication into a sum of several matrix multiplications, each of which can be computed without rounding errors. Although this scheme was originally proposed for matrix multiplication, its fundamental concept can be regarded as an error-free transformation of an inner product. Therefore, we explain its principle here on the basis of inner products.

The scheme consists of the following three steps.

- Step.1. **Slicing**⁹: This step slices the input vectors into multiple vectors of the same length on an element-wise basis. This is done to ensure that the subsequent computation can be performed without rounding errors.
- Step.2. **Computation**: This step computes the products of all combinations of sliced vectors. At this stage, Tensor Cores can be utilized.
- Step.3. **Accumulation**¹⁰: This step accumulates the results obtained in the previous step to produce the final result.

Hereafter, TypeX denotes a floating-point data type, such as FP64 or FP16. $\mathbb{F}_{\text{TypeX}}$ represents the set of numbers representable in TypeX. $\text{fl}_{\text{TypeX}}(\cdot)$ and $\text{cvt}_{\text{TypeX}}(\cdot)$ denote the TypeX arithmetic operation and type-conversion to TypeX with round-to-nearest-even rounding, respectively. u_{TypeX} denotes the unit round-off of TypeX. m_{TypeX} is the number of mantissa bits of TypeX, including the hidden bit: $m_{\text{TypeX}} = -\log_2 u_{\text{TypeX}}$.

We now consider computing the inner product of $\mathbf{x}, \mathbf{y} \in \mathbb{F}_{\text{Type1}}^k$ using a Tensor Core operation, which computes matrices stored in Type2 with accumulation in Type3. First, we determine constants γ, ξ , and ρ as defined below.

$$\gamma := \left\lceil m_{\text{Type1}} - \frac{m_{\text{Type3}} - \log_2 k}{2} \right\rceil \quad (1)$$

$$\xi := m_{\text{Type1}} - m_{\text{Type2}} \quad (2)$$

$$\rho := \max(\gamma, \xi) \quad (3)$$

Then, \mathbf{x} in Type1 is recursively sliced into $\underline{\mathbf{x}}_i^{(p)}$ in Type2 as follows. Equations (4)–(8) are processed recursively from $p = 1$, starting with $\mathbf{x}^{(1)} = \mathbf{x}$, until $\mathbf{x}^{(p)} = 0$. The constant 0.75 in Equation (5) was proposed by [20]. Equations (6)–(8) are applied element-wise for $1 \leq i \leq k$. Here, we define that Equation (4) yields $c_x^{(p)} := 0$

when $\max_{1 \leq i \leq k} |\mathbf{x}^{(p)}_i| = 0$.

$$c_x^{(p)} := \left\lceil \text{fl}_{\text{Type1}} \left(\log_2 \left(\max_{1 \leq i \leq k} |\mathbf{x}_i^{(p)}| \right) \right) \right\rceil \quad (4)$$

$$\sigma := 0.75 \cdot 2^{\rho + c_x^{(p)}} \quad (5)$$

$$\mathbf{v}_i := \text{fl}_{\text{Type1}} \left((\mathbf{x}_i^{(p)} + \sigma) - \sigma \right) \quad (6)$$

$$\underline{\mathbf{x}}_i^{(p+1)} := \text{fl}_{\text{Type1}} \left(\mathbf{x}_i^{(p)} - \mathbf{v}_i \right) \quad (7)$$

$$\underline{\mathbf{x}}_i^{(p)} := \text{cvt}_{\text{Type2}} \left(\text{fl}_{\text{Type1}} \left(2^{-c_x^{(p)}} \mathbf{v}_i \right) \right) \quad (8)$$

For Equation (5), if the condition $m_{\text{Type1}} - \lfloor \log_2 \sigma \rfloor > 0$ is not satisfied, the slicing process fails because m_{Type2} is insufficient to retain the information of a sliced vector.

Consequently, \mathbf{x} is transformed as follows.

$$\mathbf{x} = \sum_{p=1}^{s_x} 2^{c_x^{(p)}} \underline{\mathbf{x}}_i^{(p)} \quad (9)$$

$c_x^{(p)}$ corresponds to the exponent part of \mathbf{x} , and $\underline{\mathbf{x}}_i^{(p)}$ corresponds to the mantissa part of \mathbf{x} .

The vector \mathbf{y} is processed in the same manner as \mathbf{x} . Consequently, the inner product $\mathbf{x}^T \mathbf{y}$ is transformed as follows:

$$\mathbf{x}^T \mathbf{y} = \sum_{p=1}^{s_x} \sum_{q=1}^{s_y} 2^{c_x^{(p)} + c_y^{(q)}} \underline{\mathbf{x}}_i^{(p)}^T \underline{\mathbf{y}}_j^{(q)} \quad (10)$$

In the above, the computation of the right-hand side can be performed using the Type3 arithmetic without rounding errors. Therefore, the following equation is valid.

$$\underline{\mathbf{x}}_i^{(p)}^T \underline{\mathbf{y}}_j^{(q)} = \text{fl}_{\text{Type3}} \left(\underline{\mathbf{x}}_i^{(p)}^T \underline{\mathbf{y}}_j^{(q)} \right) \quad (11)$$

It is important to note that in Equation (10), the accumulation does not involve floating-point arithmetic; the equation holds exactly when performed without rounding errors. However, to achieve at least the accuracy of Type1 arithmetic, accumulation can be performed using Type1 arithmetic.

Although the above computation is described for inner product, it can be naturally extended to matrix multiplication. In this case, the computation in Equation (10) can be implemented using GEMM with Tensor Cores, as provided by cuBLAS¹¹.

The computational cost of this scheme depends on the number of slices, with memory consumption increasing accordingly. The number of slices is influenced by the following factors:

- i. The inner dimension k as in Equation (1).
- ii. The range of absolute values of the input elements, corresponding to the maximum in Equation (4).
- iii. The maximum number of significant bits in the input elements.

4 Extended Implementation of Ozaki Scheme

This section proposes extended implementation methods for the existing Ozaki scheme in light of recent AI-oriented hardware developments. Two notable trends are observed: the introduction of

⁹Some papers refer to this as “splitting”.

¹⁰Some papers refer to this as “summation”.

¹¹<https://docs.nvidia.com/cuda/cublas/>

low-precision floating-point formats such as FP8 replacing low-precision integer types like INT8, and the decline in FP64 performance, including the emergence of AI processors that lack FP64 support entirely.

Our implementation builds upon OzBLAS¹², a DGEMM implementation using FP16 Tensor Cores developed in prior work [23]. In addition, we incorporate several optimizations inspired by the FP128 GEMM implementation [24]. First, we improve accuracy by optimizing the order of accumulation. Second, since the Ozaki scheme performs matrix slicing along the inner-product direction, we transpose one of the input matrices in advance to enable continuous memory access, using cublasDgemm for the transposition. The sliced matrices are then processed by specifying the transposition mode in the GEMM routine.

4.1 Using FP8 Tensor Cores

This study targets NVIDIA RTX Blackwell architecture GPUs¹³ equipped with the 5th Generation Tensor Cores, specifically the GeForce RTX 5060 Ti. The theoretical floating-point performance is summarized in Table 2.

This study focuses exclusively on FP8, since FP4 is too small to be practical in the Ozaki scheme and the target GPUs exhibit identical performance for FP6¹⁴ and FP8. However, for example, the AMD Instinct MI350X delivers FP6 performance that is twice that of FP8, suggesting that FP6 may be effective in certain environments. Among the FP8 formats, we adopt E4M3, which allows each slice to retain more mantissa bits than E5M2. The accumulation precision can be set to either FP32 or FP16, and we use FP32 accumulation, as higher accumulation precision tends to reduce the number of required slices. Computations are performed with cublasGemmEx in cuBLAS for FP16 and cublasLtMatmul in cuBLASLlt for FP8. FP8 operations require 16-element memory alignment, and our implementation currently supports only problem sizes that are multiples of 16. This limitation could be addressed by introducing zero-padding, which we leave as future work.

Table 3 shows the minimum number of GEMMs required to compute DGEMMs using various Tensor Core operations, with respect to the inner dimension k . These values are for the case where all the FP64 mantissa bits are fully filled. The number of GEMMs corresponds to the square of the number of matrix slices. Columns marked with ‘–’ indicate that slicing is no longer feasible due to insufficient mantissa precision, making the computation impossible. The case of FP8 (E5M2) is not shown here, but it leads to the same situation as FP6 (E3M2). In the Ozaki scheme, GEMM computation ideally dominates the overall execution time. Therefore, if a Tensor Core operation with FP8 input achieves twice the throughput of FP16 input while requiring twice as many GEMMs, the expected performance would be comparable between FP8 and FP16. However, it is important to note that when FP32 is used as the accumulation format, the difference in the number of GEMMs between FP16 and FP8 decreases as k increases, giving FP8 the potential to outperform FP16.

Table 2: Specification of GeForce RTX 5060 Ti GPU used in the evaluations. TC: Tensor Cores. Without sparsity. Performance is calculated based on the boost clock.

Specification	Value
Clock (boost)	2.57 GHz
Num of CUDA / Tensor Cores	4608 / 144
Memory	GDDR7, 16 GB, 448 GB/s
FP4TC (FP32 accum)	189.48 TFlops/s
FP6TC (FP32 accum)	94.74 TFlops/s
FP8TC (FP32 accum)	94.74 TFlops/s
FP16TC (FP32 accum)	47.37 TFlops/s
FP32	23.69 TFlops/s
FP64TC	0.3701 TFlops/s
FP64	0.3701 TFlops/s

Table 3: Minimum number of GEMMs required for DGEMM using Tensor Cores. Type2 is the input format and Type3 is the accumulation format of Tensor Cores.

Type2	FP16	FP16	FP8 (E4M3)	FP8 (E4M3)	FP6 (E3M2)	FP6 (E3M2)
Type3	FP32	FP16	FP32	FP16	FP32	FP16
$k = 8$	25	121	121	121	196	196
16	25	196	121	196	196	196
32	36	196	121	196	196	196
64	36	324	121	324	196	324
128	36	324	121	324	196	324
256	36	729	121	729	196	729
512	49	729	121	729	196	729
1024	49	2809	121	2809	196	2809
2048	64	2809	121	2809	196	2809
4096	64	–	121	–	196	–
8192	81	–	121	–	196	–
16384	81	–	121	–	196	–
32768	121	–	121	–	196	–
65536	121	–	121	–	196	–
131072	196	–	196	–	196	–
262144	196	–	196	–	196	–

4.2 FP64 Arithmetic Emulation

When computing DGEMM using the Ozaki scheme, some FP64 operations are required for the slicing and accumulation processes. Specifically, it requires addition, multiplication, max, and a comparison operator ($a < b$).

We develop an implementation that excludes hardware FP64 instructions, assuming processors with no hardware support for FP64 operations. Our approach emulates FP64 operations using 32-bit and 64-bit integer arithmetic, in a manner similar to MPFR. Since integer operations are typically required for address calculations, we assume that performance degradation is unlikely even on AI-oriented processors.

The emulation algorithms for addition and multiplication are conceptually related to Dekker’s double-length floating-point algorithm. Specifically, the mantissa of an FP64 value is split into two

¹²<https://github.com/mukunoki/ozblas>

¹³<https://images.nvidia.com/aem-dam/Solutions/geforce/blackwell/nvidia-rtx-blackwell-gpu-architecture.pdf>

¹⁴As of cuBLAS 12.9, no GEMM routines supporting FP6 are available.

32-bit integer values, and the upper and lower parts are computed in a manner analogous to pen-and-paper arithmetic with two-digit numbers. The results are then computed using 64-bit integer arithmetic. Listing 1 shows part of the FP64 multiplication code. The exponent is computed separately. Our implementation employs round-to-nearest-even rounding, producing results that are bitwise identical to hardware FP64 operations. However, for simplicity, it omits handling overflow, underflow, and subnormal numbers.

An alternative approach is to use a three-word extension of Dekker’s algorithm with FP32 arithmetic. This approach has been used to speed up FP128 GEMM with FP64 using the Ozaki scheme [24]. However, this technique still requires some FP64 operations, and the exponent range is limited to that of FP32. Therefore, this study explores the integer-based approach.

Listing 1: Part of the FP64 multiplication code

```

typedef struct {
    uint32_t parts[4];
} uint32x4;

__device__ inline uint32x4 mul_mantissa (uint64_t a, uint64_t b) {
    uint32_t a_low = a & 0xFFFFFFFF;
    uint32_t a_high = a >> 32;
    uint32_t b_low = b & 0xFFFFFFFF;
    uint32_t b_high = b >> 32;

    uint64_t p00 = (uint64_t)a_low * b_low;
    uint64_t p01 = (uint64_t)a_low * b_high;
    uint64_t p10 = (uint64_t)a_high * b_low;
    uint64_t p11 = (uint64_t)a_high * b_high;

    uint64_t middle = p01 + p10;
    uint64_t carry = (middle < p00) ? (1ULL << 32) : 0;

    uint64_t low_result = p00 + ((middle & 0xFFFFFFFF) << 32);
    carry += (low_result < p00) ? 1 : 0;
    uint64_t high_result = p11 + (middle >> 32) + carry;

    uint32x4 result;
    result.parts[0] = low_result & 0xFFFFFFFF;
    result.parts[1] = low_result >> 32;
    result.parts[2] = high_result & 0xFFFFFFFF;
    result.parts[3] = high_result >> 32;
    return result;
}

```

4.3 Inner-product-wise blocking

As shown in Table 3, the smaller the precision gap between Type2 and Type3, the more rapidly the number of GEMMs increases with respect to the inner dimension k . To address this issue, the matrix can be partitioned along the inner product dimension, the Ozaki scheme can be applied to each block, and the results can then be accumulated. In this paper, we refer to this strategy as inner-product-wise blocking. Since this approach applies error-free transformation on a block-by-block basis, numerical errors are introduced during the accumulation of the block results, leading to lower accuracy compared to the non-blocking case. However, if the accumulation is performed using FP64 arithmetic, the resulting accuracy is comparable to that of a GEMM computed in FP64 arithmetic. Therefore, this loss of accuracy is not problematic when the goal is to emulate DGEMM.

In addition, this blocking approach contributes to a reduction in the working memory required to store slice matrices, which is one of the drawbacks in the Ozaki scheme. Prior studies have addressed this issue by applying blocking along the outer dimension, a strategy we refer to in this paper as outer-product-wise blocking. In practice,

it is feasible to combine both inner-product-wise and outer-product-wise blocking to execute computations within constrained memory budgets.

Furthermore, inner-product-wise blocking has a minor advantage that it avoids redundant slicing of the same region in one matrix with outer-product-wise blocking. However, this overhead is practically negligible, as slice processing accounts for only a small fraction of the total execution time, as demonstrated in Figure 4 in Section 5. Nonetheless, this issue should still be taken into account depending on the shape and size of the matrices.

However, the primary concern with inner-product-wise blocking is that each block computation updates the entire result matrix, leading to increased memory accesses. Furthermore, a common drawback of both blocking strategies is the potential degradation in execution efficiency (throughput) as the block size decreases.

Considering the above, the trade-off between advantages and disadvantages must be considered in determining the block sizes, which are tuning parameters. In this study, we do not implement an automatic mechanism for determining optimal block sizes; instead, we manually set a suitable block size for the sole purpose of demonstrating the concept of inner-product-wise blocking.

5 Evaluation

This section demonstrates the performance of the implementation described in the previous section.

For the performance of DGEMM with the Ozaki scheme, the following abbreviations are used hereafter.

- **DGEMM-FP8TC:** DGEMM using FP8 Tensor Cores with FP32 accumulation.
- **DGEMM-FP16TC:** DGEMM using FP16 Tensor Cores with FP32 accumulation.
- **FP64emu:** using the FP64 arithmetic emulation in slicing and accumulation.

5.1 Experimental Setting

We conducted evaluations on Geforce RTX 5060 Ti (RTX Blackwell architecture) with 16 GB of memory, with CUDA 12.9 (nvcc V12.9.86). Table 2 summarizes the specification of the GPU. We used the best value from three executions. Work memory is allocated all at once before the computation, and that time is not included in performance.

We compute $C = AB$, where matrix A is $m \times k$ and matrix B is $k \times n$. Those matrices are square matrices ($m = n = k$). We initialized the input FP64 matrices with pseudo-uniform random numbers of (1, 10) to have the same number of GEMMs shown in Table 3. Considering that GEMM is $O(n^3)$ and the other processes are $O(n^2)$, this experimental condition is considered to be the case where the cost of GEMMs as a percentage of overall execution time is minimized.

5.2 Performance of cuBLAS GEMMs

First, as a preliminary evaluation, we measured the performance of the GEMM routines used in the DGEMM emulation of the Ozaki scheme, along with the standard DGEMM and SGEMM routines for reference. Figure 1 shows the throughput of GEMM for each format. The evaluated GEMM routines are as follows.

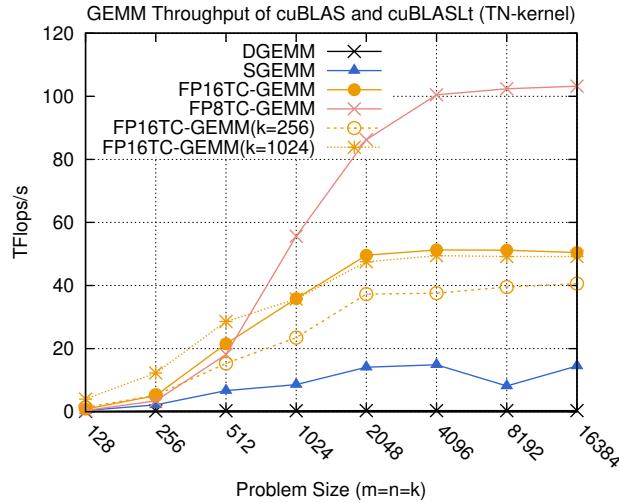


Figure 1: GEMM throughput of cuBLAS and cuBLASLlt using various formats. Matrix A is transposed (TN-kernel). If k is specified, it is fixed across all problem sizes.

- **DGEMM:** `cublasDgemm`
- **SGEMM:** `cublasSgemm`
- **FP16TC-GEMM:** `cublasGemmEx` using FP16 Tensor Cores with FP32 accumulation
- **FP8TC-GEMM:** `cublasLltMatmul` using FP8 Tensor Cores with FP32 accumulation

To optimize memory access in the slicing process, matrix A is transposed, as mentioned in Section 4; thus, this performance evaluation corresponds to the case where matrix A is transposed (the so-called TN kernel). When the problem size is sufficiently large, the performance approaches the theoretical peak performance¹⁵. The results with fixed k are intended to evaluate throughput when the block size is set to k in inner-product-wise blocking. For FP16, as shown in Table 3, the number of GEMMs is 36 when $k = 256$, but increases to 49 for $k = 512$ and 1024, and to 64 for $k = 2048$. There is almost no drop in throughput for $k = 1024$, but a decrease is observed for $k = 256$. Therefore, from the perspective of GEMM throughput and the number of GEMMs, adopting a block size of 1024 is optimal. For FP8, the number of GEMMs remains constant within this problem size range.

5.3 Accuracy of DGEMM using Ozaki Scheme

Our implementation in this paper achieves an error level equivalent to or lower than that of standard triple-loop GEMM with FP64 arithmetic, regardless of whether FP8 or FP16 Tensor Cores are used or not. This is not specific to our implementation, but a general nature of the Ozaki scheme when using all slices of matrices and using the GEMM results with the square of the number of slices (e.g., as shown in [24]). In the Ozaki scheme, no rounding error occurs during the GEMM computation phase. Rounding error arises only during the accumulation phase with FP64 arithmetic.

¹⁵Depending on runtime conditions, the GPU may operate at frequencies exceeding the boost clock, resulting in observed performance surpassing the theoretical peak.

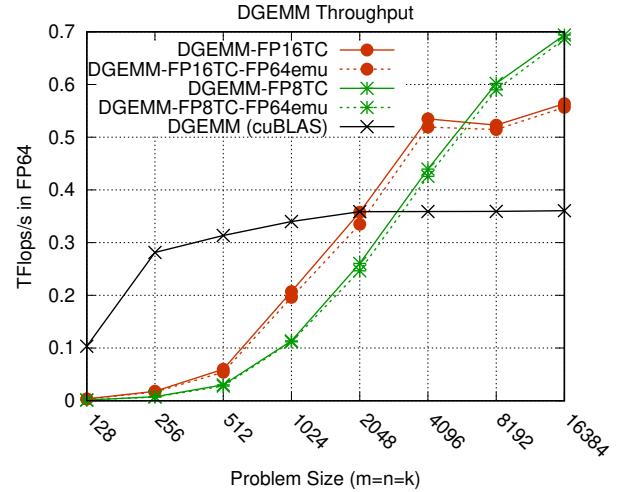


Figure 2: DGEMM throughput in FP64-equivalent Flops/s. Inner-product-wise blocking is not applied. Work-memory size is 8 GB.

Figure 3 shows the maximum relative error compared to the GEMM computed with 128-bit arithmetic by MPFR. The MPFR result is rounded to FP64. Using FP8 increases the number of slices compared to FP16, resulting in more additions being performed. Consequently, the accumulation of rounding errors may be slightly larger. Furthermore, utilizing inner-product-wise blocking increases the number of accumulations, further increasing a larger accumulation of rounding errors. The FP64 arithmetic emulation is performed using round-to-nearest-even rounding, producing results that are bitwise identical to those of hardware FP64 operations.

5.4 Performance of DGEMM using Ozaki Scheme

Figure 2 shows the performance results. Inner-product-wise blocking is not applied in this experiment. Note that the Flops/s values are expressed in FP64-equivalent units, calculated based on the number of floating-point operations performed by DGEMM in FP64 arithmetic ($2m^3$), rather than the actual number of operations executed by Tensor Cores. Figure 4 presents the performance breakdown.

DGEMM-FP8TC outperforms DGEMM-FP16TC for $m \geq 8192$. This is because the throughput of FP8 Tensor Cores is twice that of FP16 Tensor Cores, whereas the number of GEMMs at $m = 8192$ is only approximately $121/81 \approx 1.49$ times greater for DGEMM-FP8TC than for DGEMM-FP16TC, as shown in Table 3. At $m = 16384$, DGEMM-FP8TC is approximately 1.26 times faster, which is consistent with the expected speedup, $(94.74/121)/(47.37/81) \approx 1.34$, based solely on the number of GEMMs. Although DGEMM-FP8TC is theoretically expected to be faster even at $m = 2048$ – where the number of GEMMs is about 1.89 times greater – the breakdown reveals that non-GEMM overhead becomes significant at this point, and GEMM count alone no longer determines overall performance.

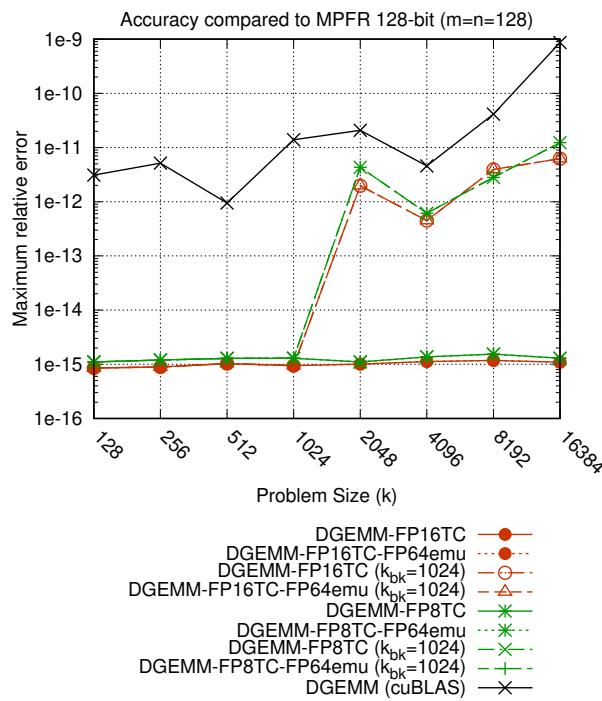


Figure 3: Accuracy comparison with MPFR 128-bit arithmetic.
If k_{bk} is specified, the inner-product-wise blocking with k_{bk} is used.

When the problem size is small, the performance of DGEMM emulation using the Ozaki scheme falls significantly below that of cuBLAS DGEMM. As shown in Figure 1, this occurs because FP8TC-GEMM and FP16TC-GEMM require larger problem sizes to reach their theoretical peak performance. The underlying reason is that the Bytes/Flop ratio decreases as precision decreases. This issue may be alleviated to some extent by using batched GEMM [21].

For the FP64 arithmetic emulation, the performance degradation is modest. The slicing and accumulation processes are memory-intensive operations. Moreover, when the problem size is sufficiently large, the time spent on slicing and accumulation constitutes only a small fraction of the total execution time. Consequently, the cost of FP64 arithmetic operations does not become apparent.

Figure 5 shows the execution time and its breakdown for various inner-product-wise block sizes in DGEMM-FP16TC when the working memory is 8 GB or 1 GB. For the block size k_{bk} , the number of GEMMs is 49 for $k_{bk} = 1024$, 64 for $k_{bk} = 4096$, and 81 for $k_{bk} = 16384$, as listed in Table 3. While the GEMM cost clearly decreases with decreasing k_{bk} , the accumulation cost increases. Consequently, the optimal block size is $k_{bk} = 4096$ in both cases.

6 Conclusion

In this paper, we examined extensions to an existing Ozaki scheme implementation to compute DGEMM using low-precision floating-point GEMMs, motivated by recent trends in AI-oriented processors.

We evaluated the following points on the RTX 5060 Ti GPU based on the RTX Blackwell architecture.

First, we considered the use of FP8 Tensor Cores, which are becoming standard in AI computation and replacing INT8. We demonstrated that the use of FP8 Tensor Cores is feasible. Their performance is competitive with that of FP16 Tensor Cores when FP8 achieves roughly twice the throughput of FP16. Moreover, for processors that support very slow FP64 operations or do not support them at all, we considered the use of the FP64 arithmetic emulation based on integer arithmetic. Only modest performance degradation was observed, because the slicing and accumulation processes are memory-intensive. We also explored the inner-product-wise blocking for performance improvement when using Tensor Core operations where the input and computation precision are close. Specifically, this was introduced with the expectation that it would mitigate the weakness of FP16 Tensor Cores with FP32 accumulation compared to FP8 Tensor Cores with FP32 accumulation. Although its effectiveness was limited due to the significant impact of increased accumulation costs, performance improvements were observed under certain conditions.

A variety of improvements for the Ozaki scheme have been proposed, and performance evaluation using combinations of these methods remains future work. Several ideas have various advantages-disadvantages tradeoffs, and comprehensive performance evaluation is required.

The growing demand for AI computing has led to the introduction of AI-oriented designs in general-purpose processors as well as AI-specialized processors. However, the requirements for computational precision differ between AI and traditional scientific computations, making the integration of these hardware architectures increasingly challenging. Moreover, as demand for AI computing continues to expand, there is concern that the role of FP64 will decline further relative to the growing reliance on low-precision operations. Achieving FP64-equivalent precision using low-precision operations through the Ozaki scheme or classical high-precision arithmetic methods is expected to mitigate these difficulties to some extent. At the same time, the computational precision and hardware requirements for AI computing are changing rapidly, necessitating the development of emulation technologies that address these trends. To sustain performance improvements in scientific computing, a multifaceted approach encompassing AI technology trends, hardware design, and FP64 emulation techniques will be required.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP25K24387.

References

- [1] Ahmad Abdelfattah, Jack Dongarra, Massimiliano Fasi, Mantas Mikaitis, and Françoise Tisseur. 2025. Analysis of Floating-Point Matrix Multiplication Computed via Integer Arithmetic. arXiv:2506.11277 [math.NA] <https://arxiv.org/abs/2506.11277>
- [2] Neil Burgess, Jelena Milanovic, Nigel Stephens, Konstantinos Monachopoulos, and David Mansell. 2019. Bfloat16 Processing for Neural Networks. In *2019 IEEE 26th Symposium on Computer Arithmetic (ARITH)*. 88–91. doi:10.1109/ARITH.2019.00022
- [3] William Dawson, Katsuhisa Ozaki, Jens Domke, and Takahito Nakajima. 2024. Reducing Numerical Precision Requirements in Quantum Chemistry Calculations. *Journal of Chemical Theory and Computation* 20, 24 (2024), 10826–10837. doi:10.1021/acs.jctc.4c00938 PMID: 39644230

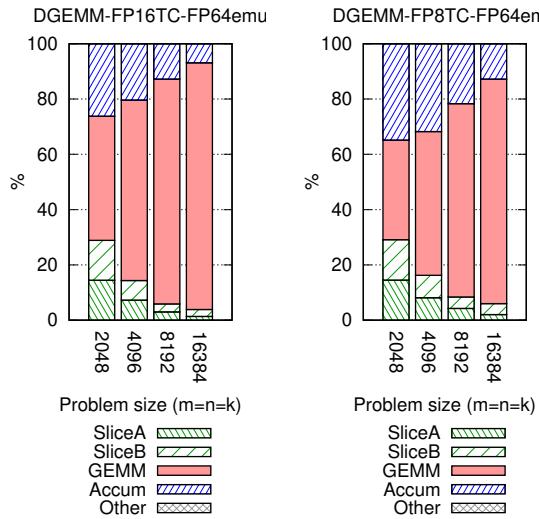


Figure 4: Execution time breakdown. Inner-product-wise blocking is not applied. Work-memory size is 8 GB.

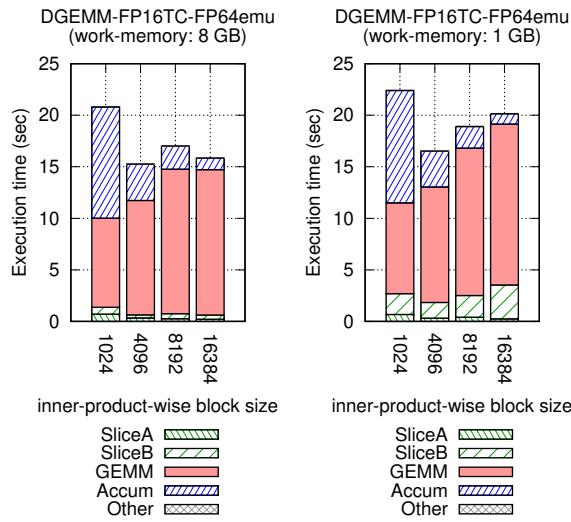


Figure 5: Execution time when $m = n = k = 16384$ with various inner-product-wise block size. Work-memory size is 8 GB (left) or 1 GB (right).

- [4] T. J. Dekker. 1971. A Floating-Point Technique for Extending the Available Precision. *Numer. Math.* 18 (1971), 224–242.
- [5] Jens Domke, Emil Vatai, Aleksandr Drozd, Peng ChenT, Yosuke Oyama, Lingqi Zhang, Shweta Salaria, Daichi Mukunoki, Artur Podobas, Mohamed WahibT, and Satoshi Matsuoka. 2021. Matrix Engines for High Performance Computing: A Paragon of Performance or Grasping at Straws?. In *2021 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. 1056–1065. doi:10.1109/IPDPS49936.2021.00114
- [6] J. J. Dongarra, C. J. Du, S. Hammarling, and I. S. Duff. 1990. A Set of Level 3 Basic Linear Algebra Subprograms. *ACM Trans. Math. Softw.* 16, 1 (March 1990), 1–17.
- [7] Nicolas Fabiano, Jean-Michel Muller, and Joris Picot. 2019. Algorithms for Triple-Word Arithmetic. *IEEE Trans. Comput.* 68, 11 (Nov. 2019), 1573–1583. doi:10.1109/TC.2019.2918451

- [8] Massimiliano Fasi, Nicholas J. Higham, Florent Lopez, Theo Mary, and Mantas Mikaitis. 2023. Matrix Multiplication in Multiword Arithmetic: Error Analysis and Application to GPU Tensor Cores. *SIAM Journal on Scientific Computing* 45, 1 (2023), C1–C19. doi:10.1137/21M1465032
- [9] L. Fousse, G. Hanrot, V. Lefèvre, P. Léplièvre, and P. Zimmermann. 2007. MPFR: A Multiple-precision Binary Floating-point Library with Correct Rounding. *ACM Trans. Math. Software* 33, 2 (2007), 13:1–13:15.
- [10] Stéf Graillat, Fabienne Jézéquel, Theo Mary, Roméo Molina, and Daichi Mukunoki. 2024. Reduced-Precision and Reduced-Exponent Formats for Accelerating Adaptive Precision Sparse Matrix–Vector Product. In *Euro-Par 2024: Parallel Processing*, Jesus Carretero, Sameer Shende, Javier García-Blas, Ivona Brandic, Katzin Olcoz, and Martin Schreiber (Eds.). Springer Nature Switzerland, Cham, 17–30.
- [11] Torbjörn Granlund and Gmp Development Team. 2015. *GNU MP 6.0 Multiple Precision Arithmetic Library*. Samurai Media Limited, London, GBR.
- [12] Greg Henry, Ping Tak Peter Tang, and Alexander Heinecke. 2019. Leveraging the bfloat16 Artificial Intelligence Datatype For Higher-Precision Computations. In *2019 IEEE 26th Symposium on Computer Arithmetic (ARITH)*. 69–76. doi:10.1109/ARITH.2019.00019
- [13] Y. Hida, X.S. Li, and D.H. Bailey. 2001. Algorithms for quad-double precision floating point arithmetic. In *Proceedings 15th IEEE Symposium on Computer Arithmetic*. ARITH-15 2001. 155–162. doi:10.1109/ARITH.2001.930115
- [14] Norman P Jouppi, Cliff Young, Nishant Patil, David Patterson, Gaurav Agrawal, Raminder Bajwa, Sarah Bates, Suresh Bhatia, Nan Boden, Al Borchers, et al. 2017. In-datacenter performance analysis of a tensor processing unit. In *Proceedings of the 44th Annual International Symposium on Computer Architecture*. ACM, 1–12. doi:10.1145/3079856.3080246
- [15] Tomonori Kouya and Taiga Utsugiri. 2023. Optimization of Multiple-Precision LU Decomposition Using Ozaki Scheme. In *Computational Science and Its Applications – ICCSA 2023 Workshops*, Osvaldo Gervasi, Beniamino Murgante, Ana Maria A. C. Rocha, Chiara Garau, Francesco Scorzà, Yeliz Karaca, and Carmelo M. Torre (Eds.). Springer Nature Switzerland, Cham, 529–545.
- [16] Binrui Li, Shenggan Cheng, and James Lin. 2021. tcFFT: A Fast Half-Precision FFT Library for NVIDIA Tensor Cores. In *2021 IEEE International Conference on Cluster Computing (CLUSTER)*. 1–11. doi:10.1109/Cluster48925.2021.00035
- [17] Weile Luo, Ruibo Fan, Zeyu Li, Dayou Du, Qiang Wang, and Xiaowen Chu. 2024. Benchmarking and Dissecting the Nvidia Hopper GPU Architecture . In *2024 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE Computer Society, Los Alamitos, CA, USA, 656–667. doi:10.1109/IPDPS57955.2024.00064
- [18] Stefano Markidis, Steven Wei Der Chien, Erwin Laure, Ivy Bo Peng, and Jeffrey S. Vetter. 2018. NVIDIA Tensor Core Programmability, Performance & Precision. In *2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. 522–531. doi:10.1109/IPDPSW.2018.00091
- [19] Paulius Micikevicius, Dusan Stosic, Neil Burgess, Marius Cornea, Pradeep Dubey, Richard Grisenthwaite, Sangwon Ha, Alexander Heinecke, Patrick Judd, John Kamalu, Naveen Mellempudi, Stuart Oberman, Mohammad Shoeybi, Michael Siu, and Hao Wu. 2022. FP8 Formats for Deep Learning. arXiv:2209.05433 [cs.LG] <https://arxiv.org/abs/2209.05433>
- [20] A. Minamihata, K. Ozaki, T. Ogita, and S. Oishi. 2016. Improved extraction scheme for accurate floating-point summation. In *The 35th JSST Annual Conference International Conference on Simulation Technology*.
- [21] Daichi Mukunoki, Takeshi Ogita, and Katsuhisa Ozaki. 2020. Reproducible BLAS Routines with Tunable Accuracy Using Ozaki Scheme for Many-core Architectures. In *Proc. 13th International Conference on Parallel Processing and Applied Mathematics (PPAM2019)*, Lecture Notes in Computer Science, Vol. 12043. Springer Berlin Heidelberg, 516–527. doi:10.1007/978-3-030-43229-4_44
- [22] Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, and Roman Iakymchuk. 2021. Conjugate Gradient Solvers with High Accuracy and Bit-Wise Reproducibility between CPU and GPU Using Ozaki Scheme. In *The International Conference on High Performance Computing in Asia-Pacific Region (Virtual Event, Republic of Korea) (HPC Asia 2021)*. Association for Computing Machinery, New York, NY, USA, 100–109. doi:10.1145/3432261.3432270
- [23] Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, and Toshiyuki Imamura. 2020. DGEMM using Tensor Cores and Its Accurate and Reproducible Versions. In *ISC High Performance 2020, Lecture Notes in Computer Science*, Vol. 12151. Springer International Publishing, 230–248. doi:10.1007/978-3-030-50743-5_12
- [24] Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, and Toshiyuki Imamura. 2021. Accurate Matrix Multiplication on Binary128 Format Accelerated by Ozaki Scheme. In *Proceedings of the 50th International Conference on Parallel Processing (Lemont, IL, USA) (ICPP '21)*. Association for Computing Machinery, New York, NY, USA, Article 78, 11 pages. doi:10.1145/3472456.3472493
- [25] Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, and Toshiyuki Imamura. 2023. Infinite-Precision Inner Product and Sparse Matrix-Vector Multiplication Using Ozaki Scheme with Dot2 on Manycore Processors. In *Parallel Processing and Applied Mathematics*, Roman Wyrzykowski, Jack Dongarra, Ewa Deelman, and Konrad Karczewski (Eds.). Springer International Publishing, Cham, 40–54.
- [26] Hiroyuki Ootomo, Katsuhisa Ozaki, and Rio Yokota. 2024. DGEMM on integer matrix multiplication unit. *The International Journal of High Performance*

- Computing Applications* (2024). doi:10.1177/10943420241239588
- [27] Hiroyuki Ootomo and Rio Yokota. 2022. Recovering single precision accuracy from Tensor Cores while surpassing the FP32 theoretical peak performance. *The International Journal of High Performance Computing Applications* 36, 4 (2022), 475–491. doi:10.1177/10943420221090256
 - [28] K. Ozaki, T. Ogita, S. Oishi, and S. M. Rump. 2012. Error-free transformations of matrix multiplication by using fast routines of matrix multiplication and its applications. *Numer. Algorithms* 59, 1 (2012), 95–118.
 - [29] Katsuhisa Ozaki, Yuki Uchino, and Toshiyuki Imamura. 2025. Ozaki Scheme II: A GEMM-oriented emulation of floating-point matrix multiplication using an integer modular technique. arXiv:2504.08009 [cs.MS] <https://arxiv.org/abs/2504.08009>
 - [30] Andrew Thall. 2006. Extended-precision floating-point numbers for GPU computation. In *ACM SIGGRAPH 2006 Research Posters* (Boston, Massachusetts) (*SIGGRAPH '06*). Association for Computing Machinery, New York, NY, USA, 52–es. doi:10.1145/1179622.1179682
 - [31] Yuki Uchino, Katsuhisa Ozaki, and Toshiyuki Imamura. 2025. Performance enhancement of the Ozaki Scheme on integer matrix multiplication unit. *The International Journal of High Performance Computing Applications* 39, 3 (2025), 462–476. doi:10.1177/10943420241313064
 - [32] Dechuang Yang, Yuxuan Zhao, Yiduo Niu, Weile Jia, En Shao, Weifeng Liu, Guangming Tan, and Zhou Jin. 2024. Mille-feuille: A Tile-Grained Mixed Precision Single-Kernel Conjugate Gradient Solver on GPUs. In *SC24: International Conference for High Performance Computing, Networking, Storage and Analysis*. 1–16. doi:10.1109/SC41406.2024.00064