3P3-10

Applicability of Single Precision GPU for Fast 2D FEM Simulation of SAW Devices Using Hierarchical Cascading Technique
階層的縦続法を用いた SAW デバイスの高速 2D FEM 解析に対する単精度 GPU の適用性

Naoto Matsuoka[1,2‡], Luyan Qiu[2], Xinyi Li[3,2], Tatsuya Omori[2] and Ken-ya Hashimoto[2]
([1]Nihon Dempa Kogyo; [2]Chiba Univ.; [3]Univ. of Elect. Sci. and Tech. of China)
松岡　直人[1,2‡], 邱　魯岩[2], 李　昕熠[3,2], 大森　達也[2], 橋本　研也[2]
([1]日本電波工業, [2]千葉大学, [3]電子科技大学)

1. Introduction

Nowadays, surface acoustic wave (SAW) devices structure become complex to improve device performances for temperature compensation and/or loss reduction.[1-3] These SAW devices employ multilayered structrures, and their optimal design is essential. For the prupose, FEM is widely used, and its drastic speed up is demanded.

Recently the hierarchical cascading technique (HCT) was proposed.[4] It acceralates computing speed drastically without spoiling advantages of FEM provided that the target structure is mainly periodic like SAW devices. Another merit of this technique is ability to reuse intermediate calculation results[5]. This can acceralate the speed drastically when a sturtucture is simulated successively by scanning design parameters.

In this Symposium, the authors demonstrate further acceralation of the HCT calculation using a high-end general purpose graphic processing unit (GP-GPU)[6]. The acceralation is obvious when the FEM model is huge such as full 3D.

Currently only expensive high-end GPUs support floating-point double precision (FP64) computation. In contrast, many GPUs support single precision (FP32) one, and its usage instead of FP64 enables us to double the calculation speed and halve the memory usage although accuracy may be somewhat degraded.

This paper discusses applicability of FP32 for GPU-based 2D FEM simulation of SAW devices using HCT. A one-port synchronous SAW resonator and a double-mode SAW (DMS) filter are designed on 42°YX-LiTaO$_3$, and their frequency responses are simulated using GPU-based HCT. It was concluded that the calculation error is not obvious even when FP32 is used.

2. Model setup

Fig. 1 shows a 2D SAW device structure under concern. In HCT, the whole structure is sliced into small pieces such as an electrode period, and the FEM matrix (A-matrix) of each cell is converted to a small matrix (B-matrix) where only degrees of feedom (DOFs) at left and right boundaries and those of surface charge are chosen and those inside of the cell are eliminated. Then B-matrices are cascaded. Finally, total charge on electrodes is obtained by termination of the total B-matrix by damping lines, which are created by cascading cells with small damping for more than 2^{20} times.[5]

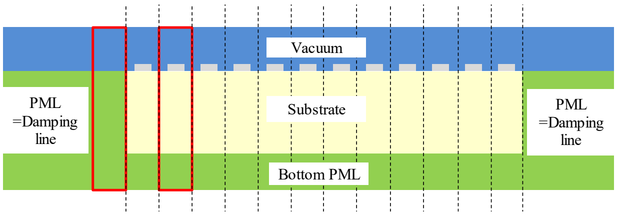

Fig.1. Schematic of device structure

This HCT/FEM operation was implemented in MATLAB and COMSOL. In this study, NVIDIA Quadro P5000 is chosen as the GPU, and is implemented in a workstation with Intel Xeon W-2123 having 4 cores (clock 3.6 GHz). The GPU has 2560 cuda cores (clock 1.607 GHz) and the design is focused on accelerating FP32 computing.

3. Comparison CPU/ GPU -One-port resonator-

First, a simple one-port SAW resonator is simulated. The resonator design is given in **Table I**. In FP32 calculation, required memory sizes is 2 GB and 4.5 GB for CPU and GPU, respectively, while the value is 2 GB when only the CPU is employed under FP64 operation.

Table I. Resonator design

Substrate	42°YX-LiTaO$_3$
IDT Pitch	4 μm
Al thickness	300 nm
Metallization ratio	0.5
IDT	129 fingers
Reflector	32 fingers
DOFs/period	6729

Fig. 2 shows calculated input adimttances of the resonator for both FP32 and FP64. It is seen that two results are almost identical. The maximum

‡matsuoka@ndk.com

error is about 3% in the susceptance near the anti-resonance frequency, and is less than 1% in the absolute value of the admittance.

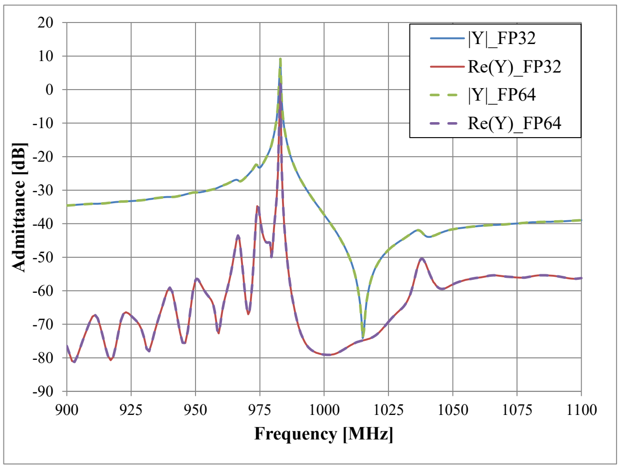

Fig.2. Calculated admittance of Resonator

Table II compares computation time required for each frequency point. The GPU accelerated the computation about 10 times. However, this values is lower than one expected from theoretical peak FLOPS data of CPU (230 GFLOPS) and GPU (8.9 TFLOPS). This is because the model size is small and only small portion of GPU cores are utilized.

Table II. Computation time of Resonator

	Computation Time / frequency point [sec.]	
	FP64-CPU	FP32-GPU
A Matrix to B Matrix	8.1	0.6
Obtain Damping line	0.5	0.2
Cascading B Matrix	0.3	0.1
Solve Charge	0.02	0.01
Total time	8.92	0.91

4. Comparison CPU/ GPU -DMS filter-

Next, a DMS filter with the three IDT configuration is simulated as a complex model. Its design is similar to the previous resonator, and each IDT is composed of three regions with different electrode pitch and reflector has a different electrode pitch, and total number of electrodes and DOFs are 231 and over 1.5 million, respectively. In FP32 calculation, required memory sizes are 2.2 GB and 8 GB for CPU and GPU, respectively, while the value is 2.2 GB when only the CPU is employed under FP64 operation.

Fig. 3 shows calculated results for both FP32 and FP64. Although the model is much more complex, two results are almost identical, and. the maximum error is is less than 1%.

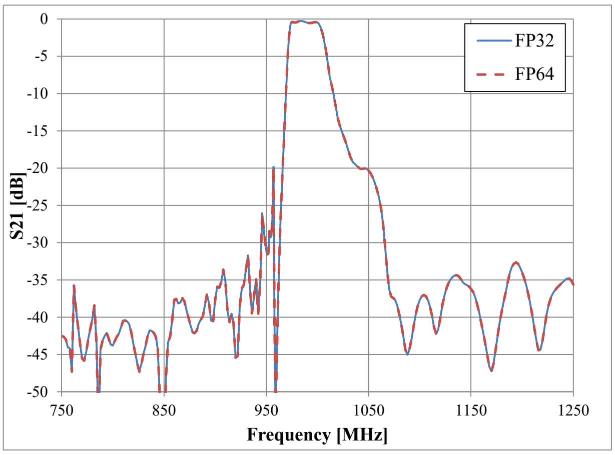

Fig.3. Calculated transfer function of DMS

Table III compares computation time required for each frequency point. The GPU accelerated the computation more than 10 times.

Table III. Computation time of DMS

	Computation Time / frequency point [sec.]	
	FP64-CPU	FP32-GPU
A Matrix to B Matrix	20.2	1.4
Obtain Damping line	0.5	0.2
Cascading B Matrix	0.9	0.3
Solve Charge	0.02	0.01
Total time	21.62	1.91

5. Conclusion

This paper discussed applicability of FP32 GPU to the HCT-based FEM calculation. Computation error was negligible, and computation is about 10 times faster. This means cheaper GPU only accelerating FP32 still applicable for HCT-based FEM computation.

Since so many cores and memories are unused in GPU, further acceleration seems possible by calculating multiple frequency points in parallel.

Acknowledgment

Xinyi Li acknowledges the support of the Japanese Government (MEXT) for the scholarship through the Super Global University Project.

References

1. H.Nakamura, et al., Jpn. J. Appl. Phys. 47, 2008 pp.4052-4055.
2. M.Miura, et al., Proc. IEEE Ultrason. Symp., 2004, pp.1322-1325.
3. T.Takai, et al., Proc. IEEE Ultrason. Symp., 2016, 10.1109/ULTSYM.2016.7728455
4. J.Koskela, et al., Proc. IEEE Ultrason. Symp., 2016, 10.1109/ULTSYM.2016.7728574
5. X.Li, et al., 2018 Jpn. J. Appl. Phys. **57** 07LC08
6. X.Li, et al., to be presented at USE2018