Uzochi Dimkpa
Performance Modeling – Convolution
ITCS 5182-001
High Performance Computing
Prof. Erik Saule
October 9, 2023

Convolution Formula

$$out[i][j] = \sum_{x=0}^{k-1}\sum_{y=0}^{k-1} C[x][y] \cdot in\left[i - \left\lfloor\frac{k}{2}\right\rfloor + x\right]\left[j - \left\lfloor\frac{k}{2}\right\rfloor + y\right]$$

Flops Benchmark

$$8.92079 \times 10^{11}\ \frac{\text{Flops}}{\text{sec}}$$

DRAM Memory Bus Bandwidth Benchmark (*WikiChip*)

$$34.13 \times 10^{9}\ \frac{\text{GB}}{\text{sec}}$$

2.

a. # Flops: $\boxed{2nmk^2}$
b. # Memory moves: 2 moves across bottleneck bandwidth, $\boxed{\text{sizeof(float)} \cdot (2\text{nm} + \text{k}^2)}$
c. $n = 1024, m = 768, k = 3$, 32 bits = 4 bytes,

$$\frac{1024 \cdot 768 \cdot 3^2}{8.92079 \times 10^9} = 0.00079 \text{ s},$$

$$4\left(\frac{2(1024)(768) + (3)^2}{34.13 \times 10^9}\right) = 0.00018433905 \text{ s},$$

$$0.00079 + 0.00018433905 = \boxed{0.00097433905 \text{ s}}$$

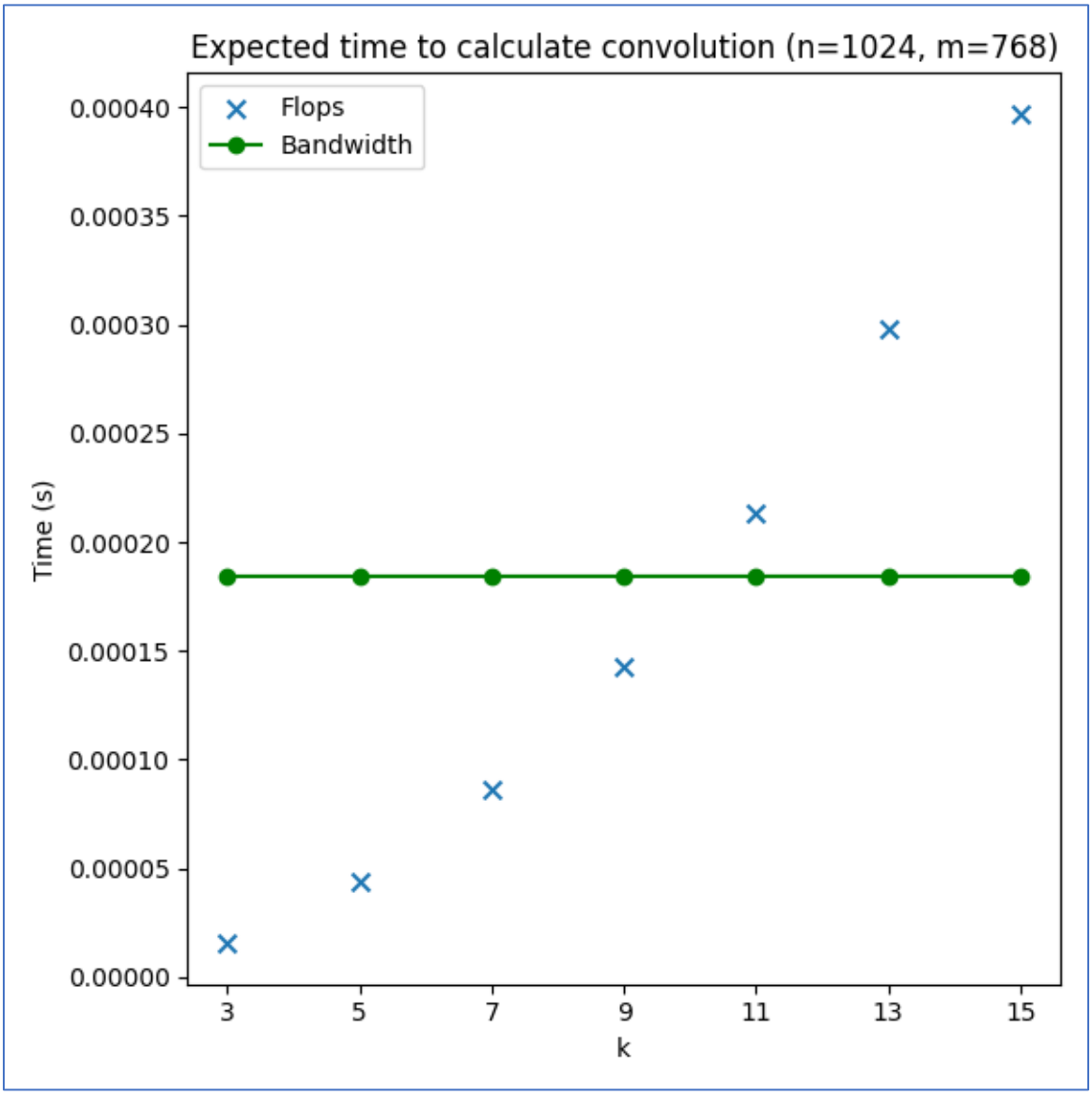
d. $n = 1024, m = 768, k = 11$, 32 bits = 4 bytes,

$$\frac{1024 \cdot 768 \cdot 11^2}{8.92079 \times 10^9} = 0.01066 \text{ s},$$

$$4\left(\frac{2(1024)(768) + (11)^2}{34.14 \times 10^9}\right) = 0.00018435218,$$

$$0.01066 + 0.00018435218 = \boxed{0.01084435218 \text{ s}}$$

e.



3. The way that the data is cached when loaded from memory is likely heavily affecting the performance of the program. For the original implementation, the data is loaded sequentially, line by line across the entire image, so data that was previously loaded into cache and used for calculations would be evicted and replaced by new data, only for that same data to be reloaded into cache and used again.

4. After improving the code with parallel code, I managed to get within a factor of 20 of the model.