

## ЛАБОРАТОРИЙН АЖИЛ №6

### CUDA алгоритмууд

**Зорилго:** Лекцийн хичээл дээр үзсэн ойлголтуудаа батагана. CUDA архитектурыг ойлгох, grid, block ашиглан параллел тооцоолол хийж сурах, дадлагажих

#### Лабораторийн ажлын даалгавар.

Элементүүд нь  $A_{ij} = B_{ij}^T = B_{ji}; i, j \in 0, \dots, n-1$  байх  $A, B \in \mathbb{R}^{n \times n}$  гэсэн 2 хэмжээст матрицууд өгөгдсөн. Тэгвэл  $A, B$  нь глобал санах ойд байгаа үед  $A$  -аас  $B$  -г үүсгэж болно. Бидний зорилго бол  $A$ -ын эргүүлсэн хувилбарыг  $B$ -д хадгална.  $O(n^2)$  хэмжээний CUDA thread ашиглана. Алгоритм нь хэд хэдэн блок ашиглаж, дурын  $n$  урттай матрицтай ажиллах ёстой.

- Хоёр хэмжээст grid дээрх  $n^2$  thread-тэй  $A$ -г хөрвүүлэх (эрүүлэх) CUDA кернел бич.
- $A$ -ын дэд матриц буюу  $n = TILE * k; k \in N$  байх  $TILE \in \{1, 2, 4, 8, 16\}$  -ын хувьд нэг thread нь  $TILE^2$  хэмжээтэй дэд матрицыг хөрвүүлдэг  $n^2 / TILE^2$  thread-тей кернел бич. Энд  $N$  нь натурал тоон олонлог. Алгоритмыг нь дээр дурдсан  $TILE$ -ийн бүх сонголтууд дээр ажиллах ёстой. Аль хувилбар нь хамгийн сайн гүйцэтгэлтэй вэ? Ажиглалтаа тайлбарла.
- $TILE = 4$  буюу нэг дэд матриц 16 элементтэй байна. 16 элементийг CUDA thread дотор хадгалахын дөрвөн float4 бүтцийг ашиглана уу. Дараа нь float утга солих зургаан үйлдэл хэрэглэн дөрвөн float4 бүтцэд хадгалсан  $4 \times 4$  дэд матрицыг эргүүл. Эцэст нь эргүүлсэн дэд матрицыг өгөгдлийг  $B$ -ийн харгалзах хэсэгт float4 бүтцийн дөрвөн ширхэг бичилтээр хадгална. Индексжүүлэлтийн схемийг дүрслэн үзүүлнэ.

Дээрх гурван алгоритмын ажиллагаан хугацаа CPU clock-оор хэмжинэ. Аль нь илүү сайн байна вэ? Ажиглалтаа тайлбарла.

**Жич:** Даалгаврыг CoLab дээр гүйцэтгээд Moodle системээр pdf-ийг явуулна. Даалгаврыг гүйцэтгэхдээ үндсэн сурах бичгийг ашиглаарай.