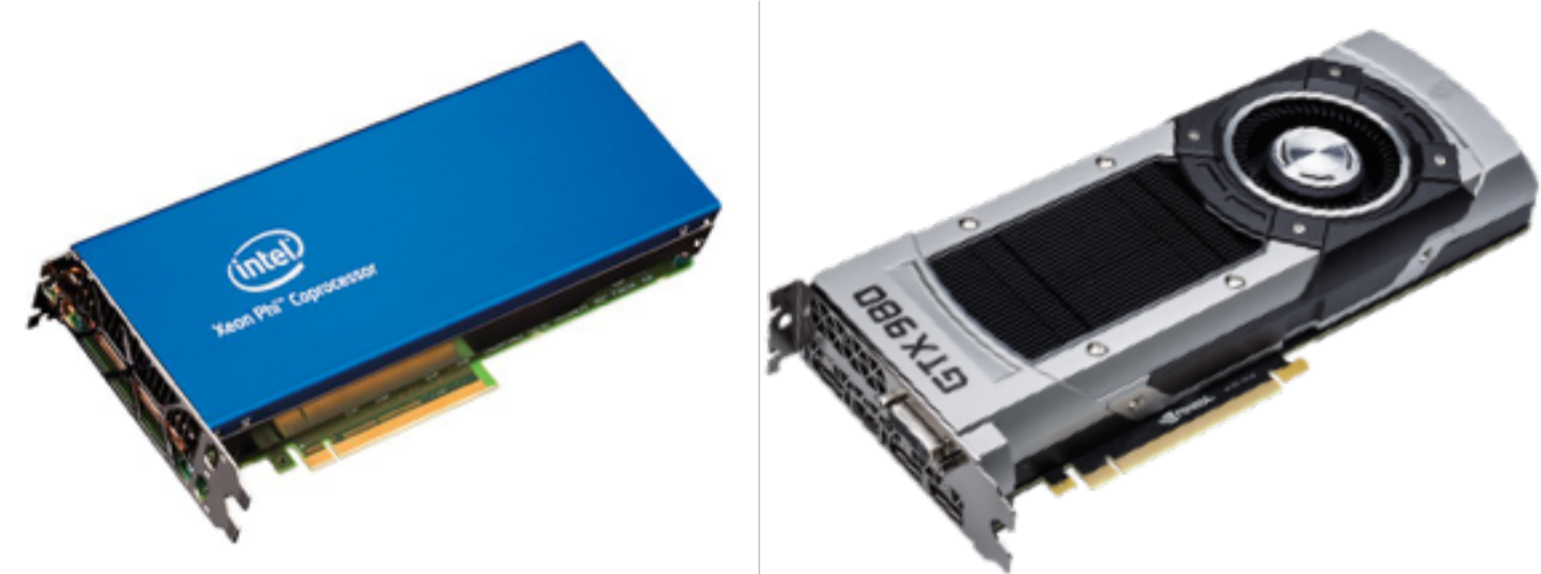




A Novel Heterogeneous Algorithm for Multiplying Scale-Free Sparse Matrices



Heterogeneous Many-Core Platforms

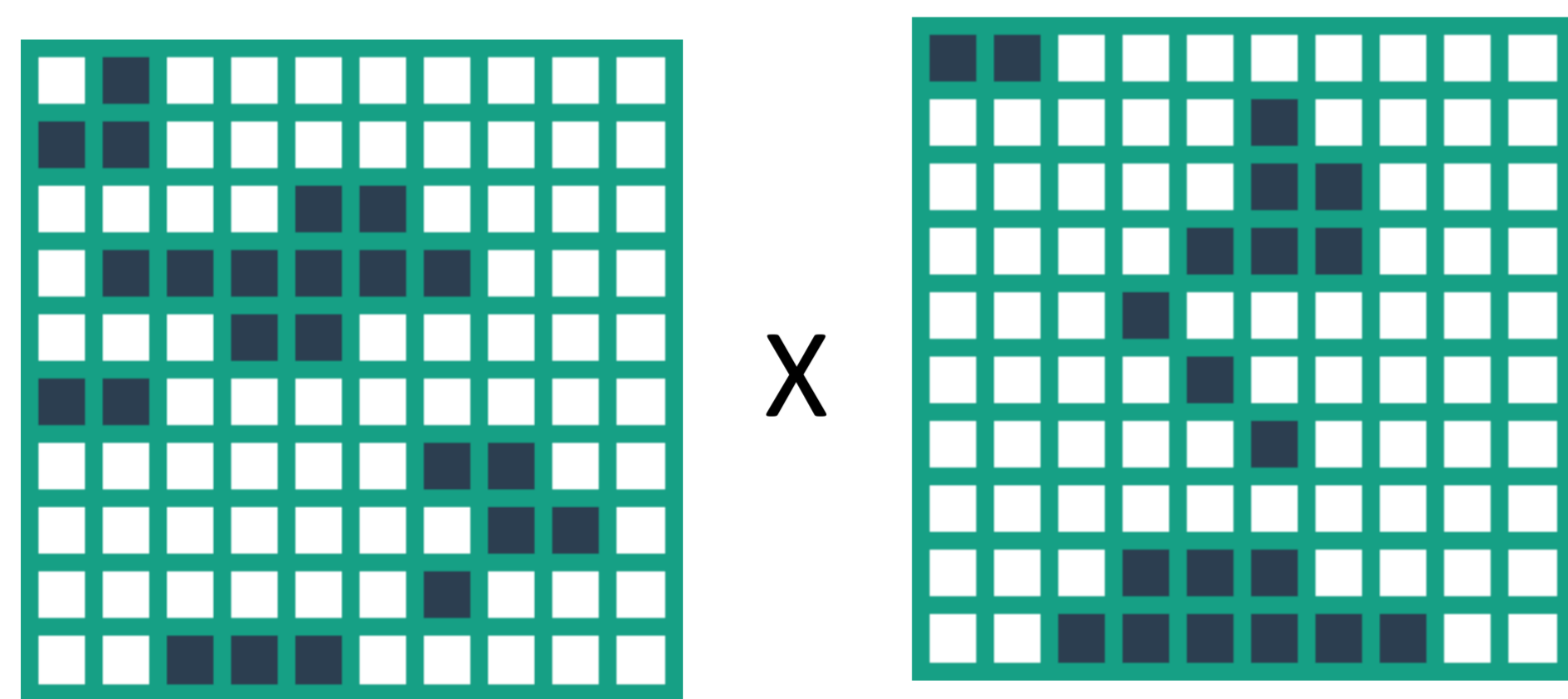


Speed of CPU comparable to GPU.

It is therefore a bad idea to keep CPU idle while GPU is working hard. We could give some portion of input to CPU.

Q. How to partition the input and distribute among the processing units (CPUs, GPUs, Accelerators etc).

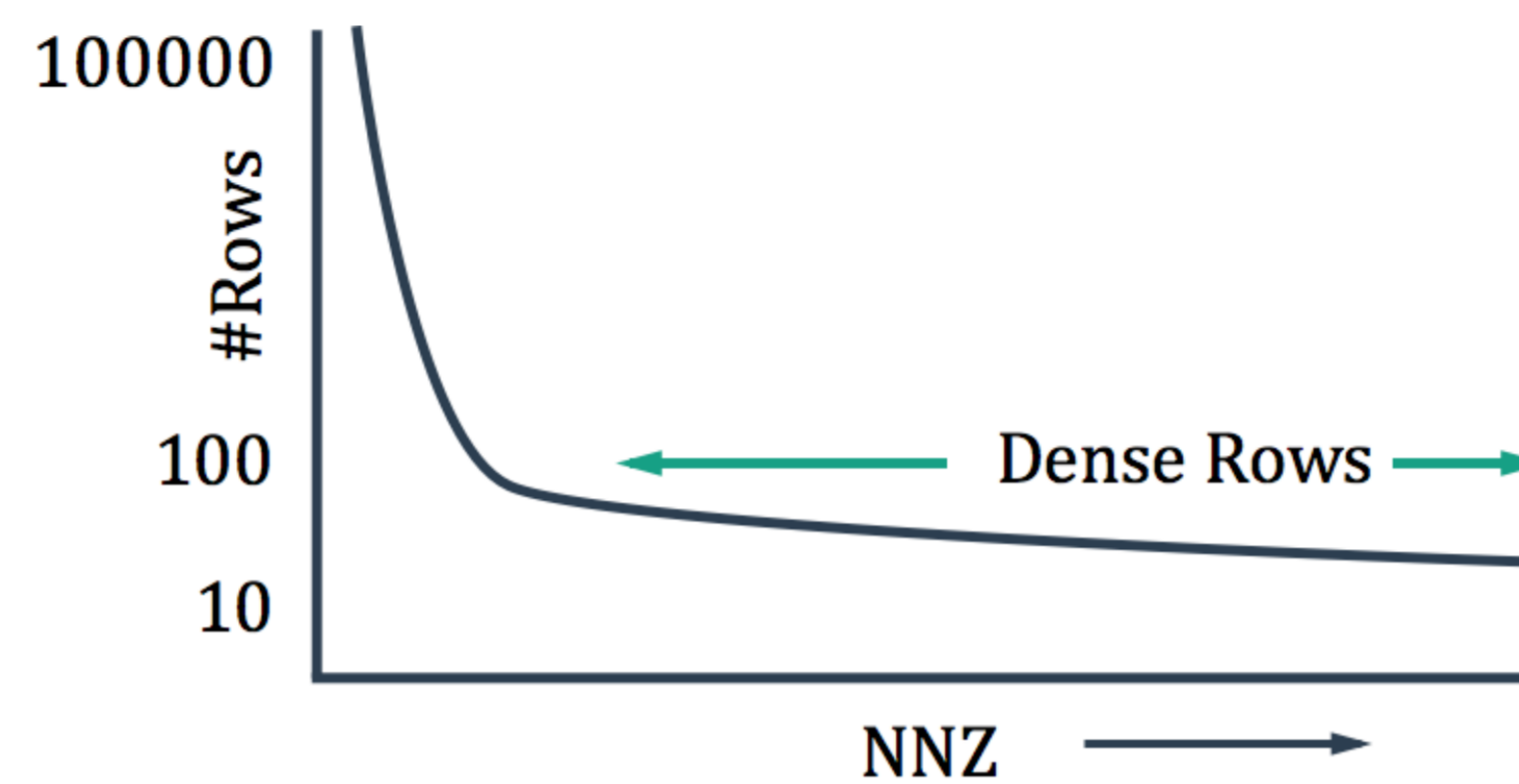
Sparse Matrix-Matrix Multiplication



Sparse matrix-matrix multiplication (SPMM) is highly irregular, hence tough to optimize for GPU's.

Q. Can we exploit the characteristics of input to achieve better performance ?

Real-world Graphs (Scale-free)

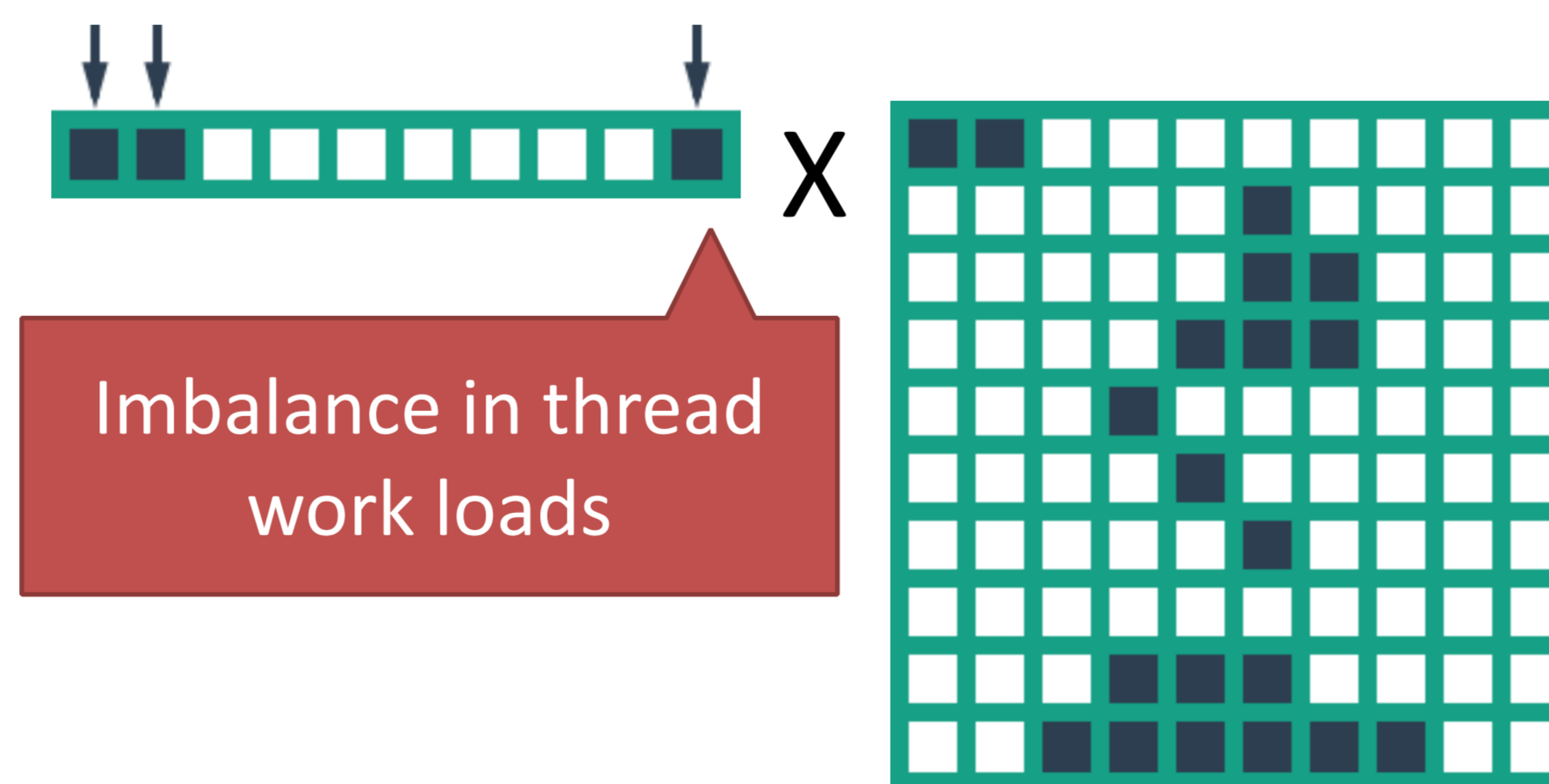


Row-Row Formulation

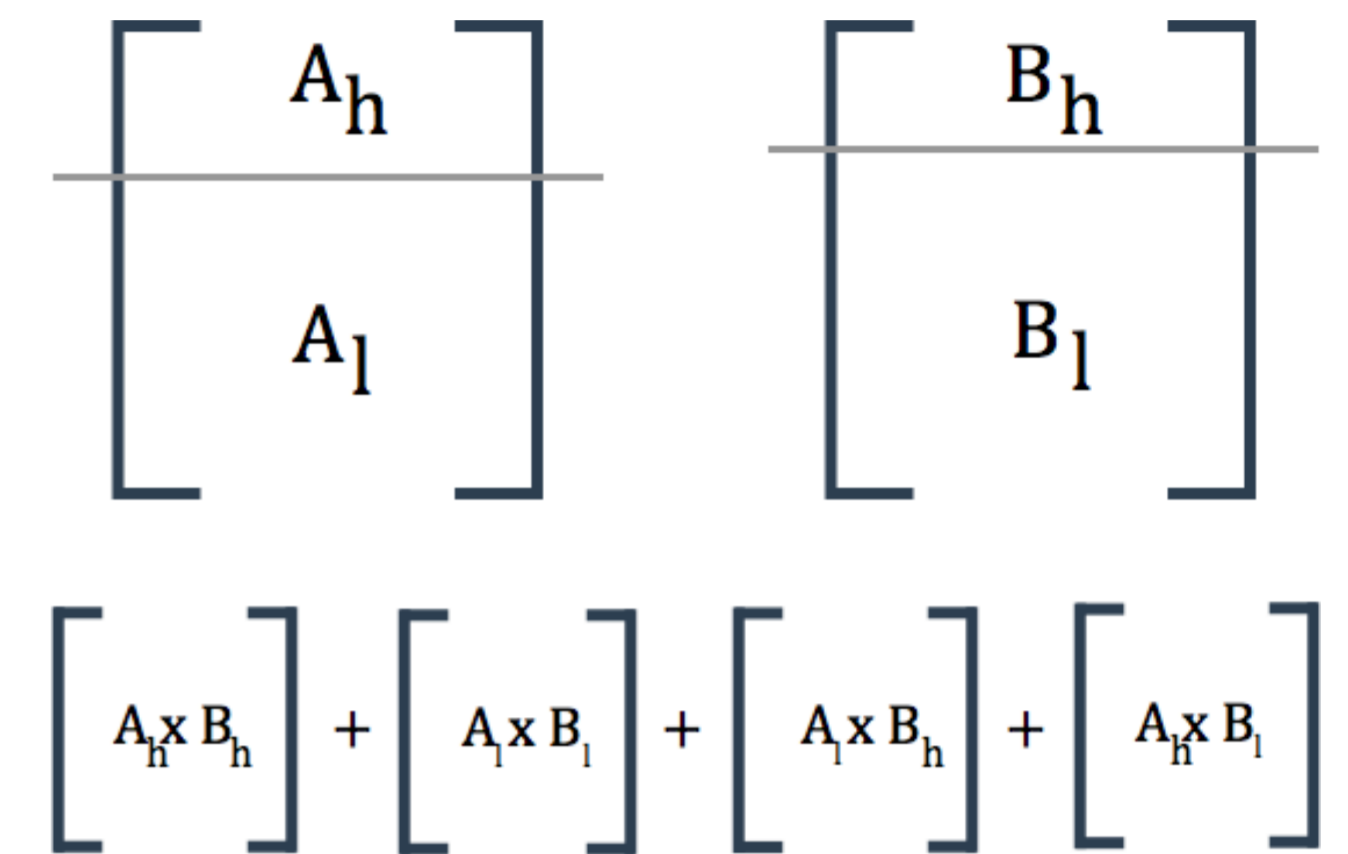
Matam et. al, proved row-row formulation of matrix multiplication out performs usual row-column formulation for SPMM in GPU's.

$$C(i,:) = \sum_{j \in I_i(A)} A(i,j) * B(j,:)$$

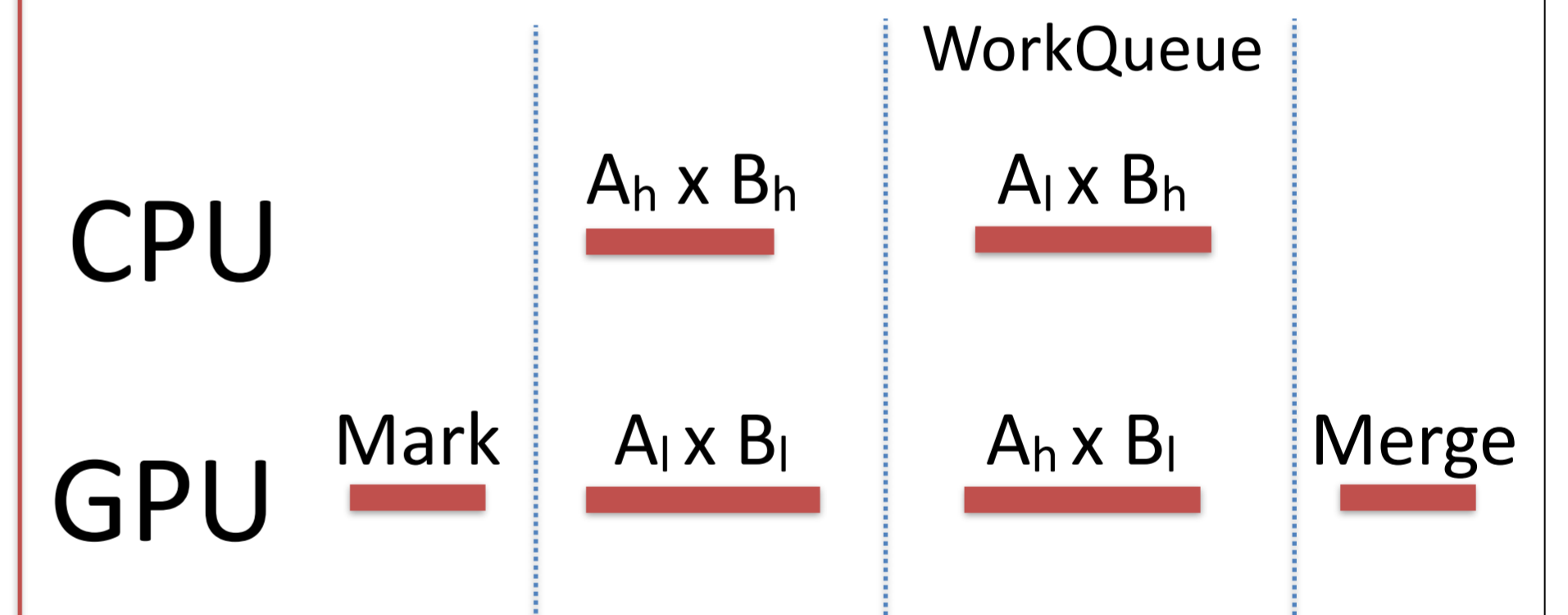
Problem? Thread Load Imbalance



Proposed Solution - HHCPU



Time line Diagram of HH-CPU



Experimental Results

