

CUDA IMPLEMENTATION OF AN LBG METHOD FOR VQ COMPRESSION

AKIYOSHI WAKATANI (KONAN UNIVERSITY, JAPAN)

Summary: We implement VQ compression algorithms on GPUs by using CUDA (Compute Unified Device Architecture) in two ways(SIMD type programming and SPMD type programming) and evaluate them on NVIDIA's GTS 250 and 8600 GT. We achieves the speedup of over 25 compared with the execution with only a CPU. Our results also show that the key to maximizing the performance is determine the optimal ratio of the loop sizes within the GPU and out of the GPU in consideration of the thread creation cost and the loop cost of the GPU.

Background and motivation

Main loop of LBG method for VQ compression

```
for(i=0; i<T; i++){
  min=distance(i,0);
  jmin=0;
  for(j=1; j<K; j++){
    x=distance(i,j);
    if(x<min)
      min=x; jmin=j;
  }
  code[i]=jmin;
}
```

This loop is divided into two loops; the **outer loop** invokes GPU repeatedly and the **inner loop** is within the GPU code.

Method 1

n= the number of processors
SPMD type programming

- process invocation cost is low
- Frequently used for conventional parallel computers

VS

Method 2

n= T
SIMD type programming

- This is recommended for CUDA programming

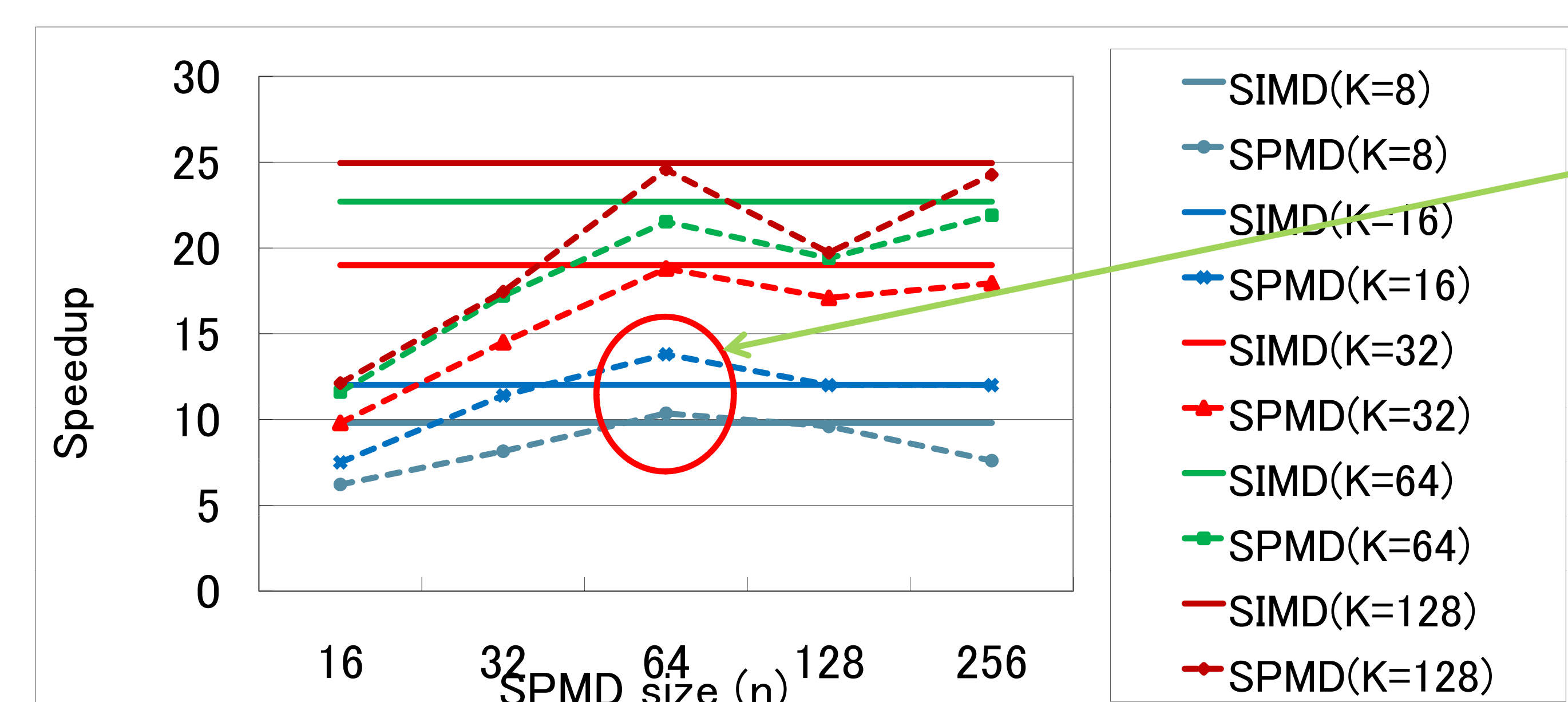
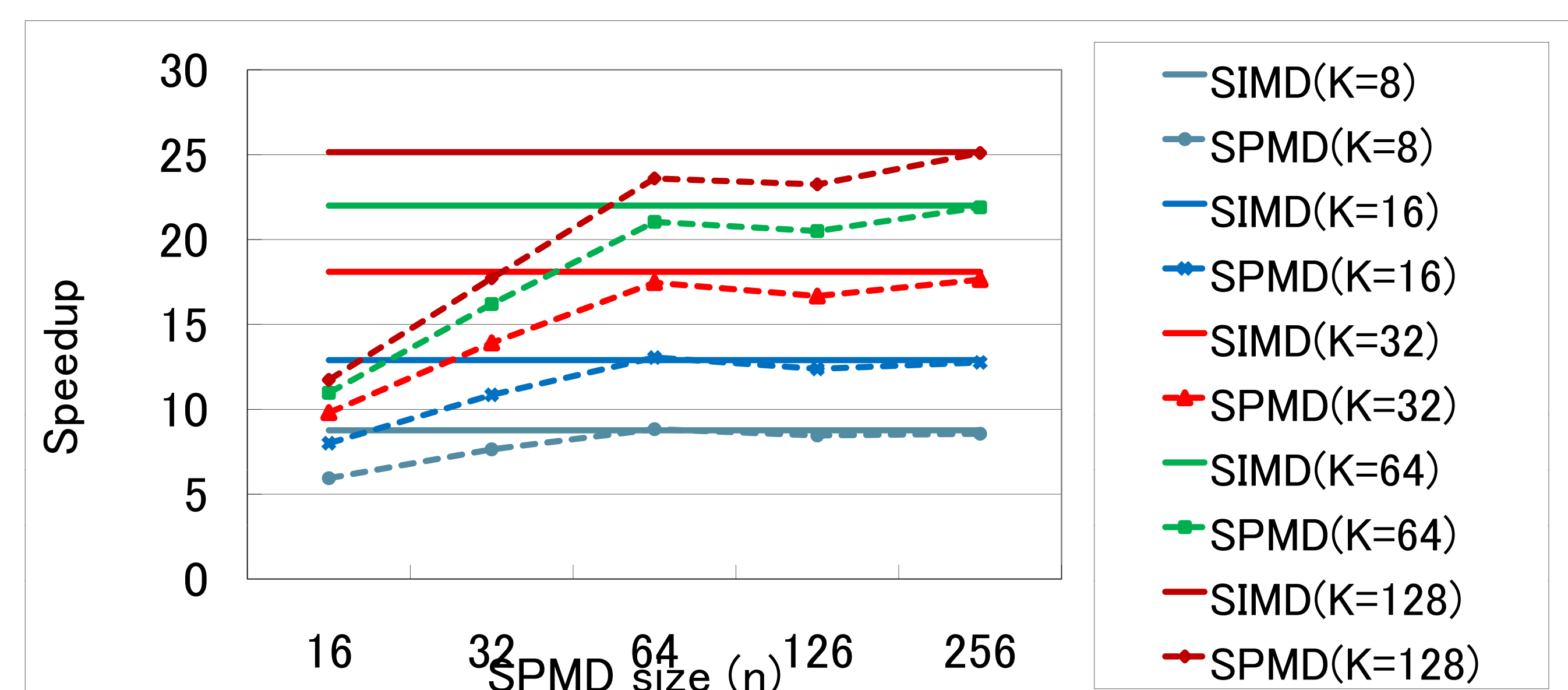
Pseudo CUDA code

```
__global__ kernel(){
  for(i=0; i<T/n; i++){
    task(K x V);
  }
}

int main(){
  for(j=0; j<n; j++){
    kernel();
  }
}
```

Results and discussion

The speedups of the **SPMD type programming** are close to those of the **SIMD type programming** when the **SPMD value (n)** is appropriately chosen.



In some cases, the **SPMD type programming** outperforms the **SIMD type programming**.

Therefore, by selecting SPMD size appropriately, SPMD type programming can be an alternative method to SIMD type programming on GPGPU systems to achieve a high performance computing. Moreover, when a different combination of the CPU and the GPU is adopted, better results may be obtained by using SPMD type programming.

Contact:

Akiyoshi Wakatani (Faculty of Intelligence and Informatics, Konan University, JAPAN)

Email: wakatani @konan-u.ac.jp FAX +81-78-435-2540