

Parallel Skyline for GPU Computing

Wei-Mei Chen

Department of Electronic and Computer Engineering
National Taiwan University of Science and Technology

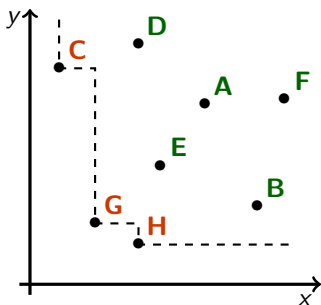


September 2, 2016

[joint with Yi-Wen Peng]

Skyline

- ▶ A point is a **skyline point** if no point in the same dataset that dominates it.
- ▶ The skyline of a multidimensional set: the points for which no other point exists that is at least as good along every dimension.



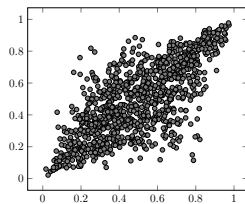
{C, G, H}

Skyline Algorithms

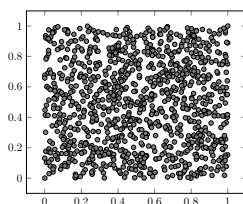
- ▶ **Dominance:**
 - ▶ Sorting-based: BNL(2001), SFS(2003), SaLSa(2008)
 - ▶ Partitioning-based: D&C(2001), NN(2002), LS(2007)
- ▶ **Incomparability:**
SkyTree(2009), OSPS(2009), TwoPhase(2012),
BSkyTree(2014), Hybrid(2015)

Data Distribution

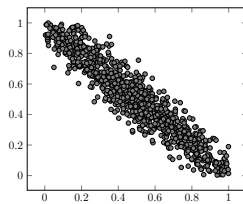
Correlated



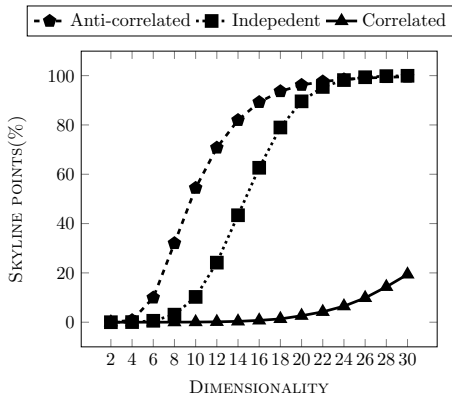
Independent



Anti-correlated



Number of Skyline Points of Synthetic Data



Graphics Processing Units (GPU)

GPU is a specialized multiprocessor that discharge and accelerates 2D or 3D graphics.

- ▶ Beginning: the first 3D add-in cards in 1995
- ▶ Modern GPUs:
 - ▶ NVIDIA: GeForce series
 - ▶ AMD: Radeon series
 - ▶ Intel: HD/Iris Graphics series
- ▶ GPUs are good at parallel, arithmetically intense, and streaming-memory problems.

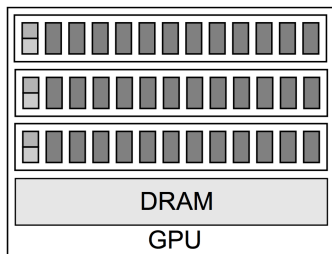
GPU Programming

- ▶ General-purpose computing on graphics processing units (**GPGPU**):
 - ▶ Using GPUs to perform computation in applications traditionally handled by CPUs.
- ▶ **Programming frameworks**
 - ▶ **OpenCL** is an open standard that can be used to program CPUs, GPUs, and other devices from different vendors.
 - ▶ **CUDA** is specific to NVIDIA GPUs.

Performance vs. Portability

Parallelism

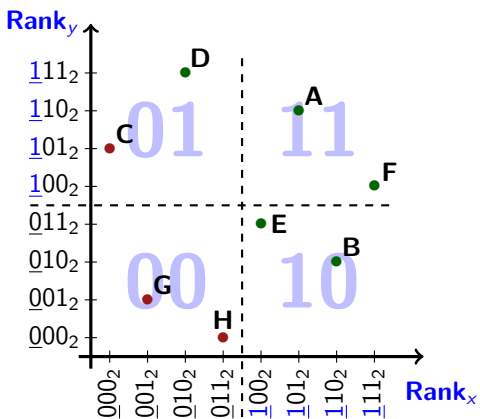
- ▶ **Task parallelism** is the simultaneous execution on multiple cores of many different functions across the same or different datasets.
- ▶ **Data parallelism** (SIMD) is the simultaneous execution on multiple cores of the same function across the elements of a dataset.



Improvement: Skyline Algorithms on GPU

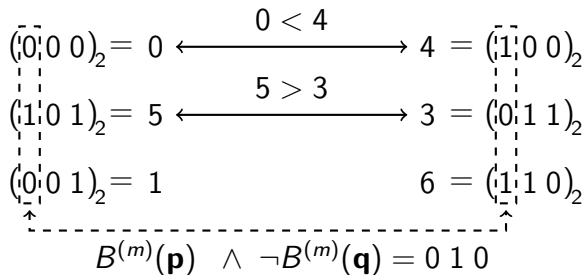
1. Reducing the number of operations:
 - ▣▶ Using the property of incomparability
2. Eliminating the number of branches (if-then statement):
 - ▣▶ Converting to binary operations
3. Rearranging the data:
 - ▣▶ Grouping the given datasets by "strength"

Example [Strength]



Binary Representation of Rank

Consider $\mathbf{p} = (0, 5, 1)$ and $\mathbf{q} = (4, 3, 6)$



one comp $\Rightarrow p_1 < q_1$ and $p_3 < q_3$

Compare bits of all attributes simultaneously

ParallelSkyline(\mathcal{D})

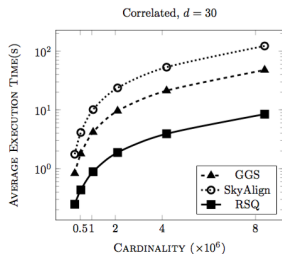
```

1:  $m \leftarrow \lceil \log_2 n \rceil$ 
2: for  $i \leftarrow 1$  to  $d$  do
3:   Sort  $\mathcal{D}$  in non-ascending order by  $i$ -th dimension.
4:   for all  $\mathbf{p} \in \mathcal{D}$  do
5:     Extract bits from  $R^{(i)}(\mathbf{p})$  and save to corresponding characteristic
       bitmaps
6: for  $i \leftarrow 1$  to  $m$  do
7:   Sort  $w(B^{(i)}(\mathbf{p}))$  for  $\mathbf{p} \in \mathcal{D}$ 
       {Now the points in  $\mathcal{D}$  is listed as the sequence  $\langle \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \rangle$ }
8: for  $i \leftarrow 1$  to  $n$  do
9:   for  $j \leftarrow 1$  to  $i - 1$  do
10:     $U \leftarrow 0$ 
11:     $M \leftarrow I_d$ 
12:    for  $\ell \leftarrow m$  to  $1$  do
13:      if  $U = 0$  and  $M \neq 0$  then
14:         $U \leftarrow (B^{(\ell)}(\mathbf{a}_i) \wedge \neg B^{(\ell)}(\mathbf{a}_j)) \wedge M$ 
15:        Update  $M$  to record the bits processed.
16:      if  $U = 0$  and  $M \neq I_d$  then
17:        Mark  $\mathbf{a}_j$  as dominated
18:  $S \leftarrow$  Collect all non-dominated points
19: return  $S$ 

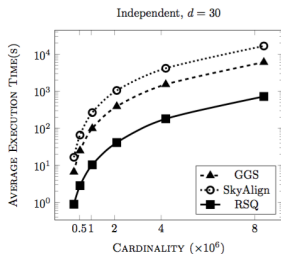
```

Performance: Varying Cardinality

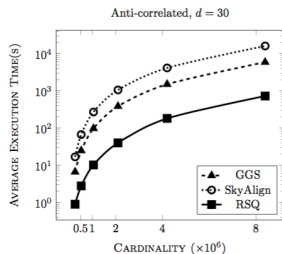
$d = 30$



(a) Correlated



(b) Independent



(c) Anti-correlated

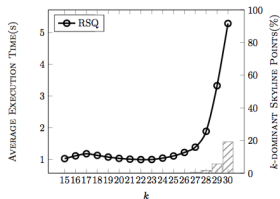
k -Dominant Skyline

$\mathbf{p} = (p_1, \dots, p_d)$ and $\mathbf{q} = (q_1, \dots, q_d)$

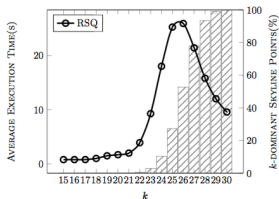
► \mathbf{p} **k -dominate** \mathbf{q}

if there are at least k attributes of \mathbf{p} such that $p_i \leq q_i$ for $i \in \{1, 2, \dots, d\}$ and there exists $j \in \{1, 2, \dots, d\}$ such that $p_j < q_j$, written $\mathbf{p} \prec_k \mathbf{q}$.

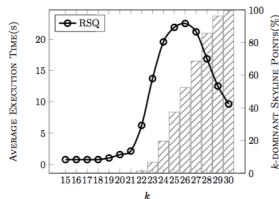
Varying k with $n = 10^6$, $d = 30$



(a) Correlated



(b) Independent



(c) Anti-correlated