

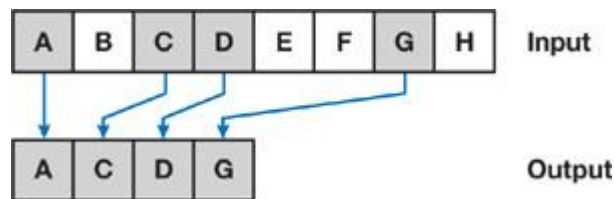
HW 2: Implementace filtru - CUDA

Bodů: 6

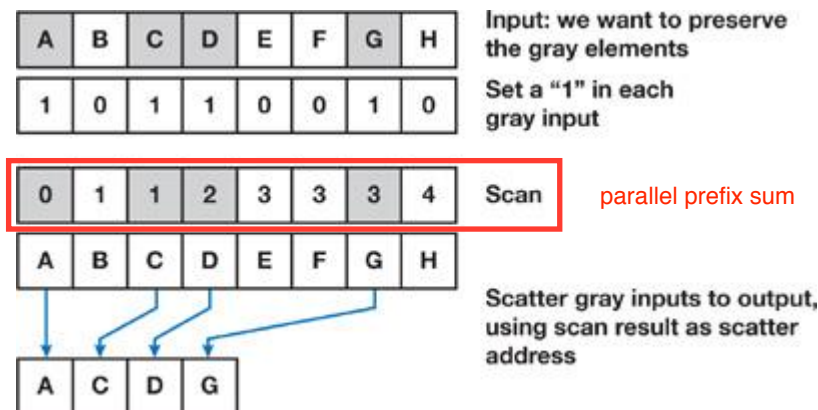
Termín odevzdání: do čtvrtka 24. 11. 2015 (12:00)

Naimplementujte jednoduchý filtr, který ze vstupního pole o velikosti N vybere prvky vyhovující zadané podmínce a umístí je do nového, výstupního pole (viz. obrázek níže, kde jednotlivá písmena reprezentují celé záznamy a šedá políčka reprezentují prvky vyhovující zadané podmínce). Vstupní pole se bude skládat ze záznamů obsahujících klíč typu int a hodnotu typu float:

```
struct Data {  
    int key;  
    float data;  
};
```



Vstupní pole vygenerujte náhodně tak, aby klíče prvků ležely v rozsahu 0 až 100. Pro filtrování prvků použijte jednoduchou funkci (nazvanou např. filter), které zadáte hodnotu klíče a meze intervalu a výsledkem bude 1 (klíč leží v daném intervalu) nebo 0 (klíč neleží v daném intervalu).

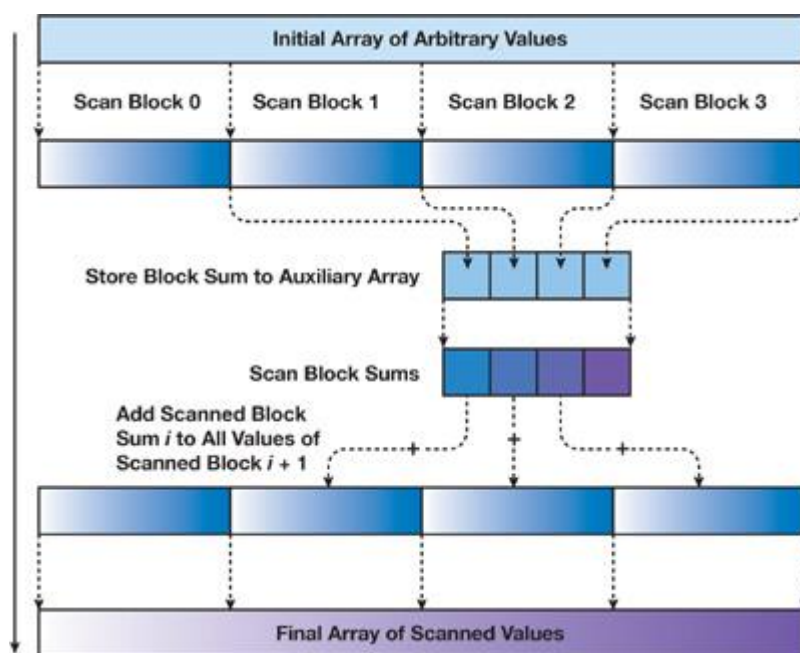


Celé filtrování bude probíhat v několika krocích (viz. obrázek výše):

- 1) Rozdělení zpracovávaného pole délky N mezi bloky (velikost bloku nechť je dána **THREADS_PER_BLOCK**). Každý blok vláken načte svoji část vstupního pole a vytvoří pole ve sdílené paměti, které bude pro každý prvek určovat, zda bude prvek ležet ve výstupním poli nebo ne (hodnoty 0/1 získáte zavoláním funkce filtru na daný klíč).
- 2) Výpočet pořadí prvků ve výstupním poli. Pořadí prvku ve výstupním poli je dáno součtem prvků předcházejících (mínus jedna – indexy začínají od 0). Pro výpočet pořadí prvků využijte paralelní algoritmus známý pod názvem **parallel prefix sum** (nebo také scan, podrobnější informace o implementaci algoritmu najdete na URL: http://http.developer.nvidia.com/GPUGems3/gpugems3_ch39.html - podkapitola

nazvaná "A Work-Efficient Parallel Scan"). Po podrobném prostudování algoritmu zjistíte, že výpočet pořadí prvků bude mít dva kroky:

- Výpočet pořadí prvků ve výstupním poli v rámci bloku vláken (scan).
- Výpočet výsledného pořadí (ke každému bloku se musí přičíst počty prvků, které uloží předchozí bloky do výstupního pole, což vede na další scan, tentokrát nad délkami částí výstupních polí, které vytvoří jednotlivé bloky – viz. popis algoritmu scanu – podkapitola "Arrays of Arbitrary Size").



3) Zápis prvků do výstupního pole na pozice vypočtené v kroku 2.

Implementace musí být schopna zpracovat pole libovolné délky N (které se vejde do paměti GPU), stejně tak musí být možné měnit meze intervalu pro filtrování prvků vstupního pole.

Forma odevzdání: emailem na adresu sloup@fel.cvut.cz se subjectem: "A4M39GPU – HW2".