

# CUDA. Сумма на префиксе

Ахтямов Павел

МФТИ

# Мотивация

- Мы можем посчитать сумму в массиве

# Мотивация

- Мы можем посчитать сумму в массиве
- Но можем ли мы решить обратную задачу?

# Задача Scan

## Задача Scan

Есть массив  $A$ :  $A[0]$ ,  $A[1]$ , ...

# Задача Scan

Есть массив  $A$ :  $A[0]$ ,  $A[1]$ , ...

Хотим посчитать:

# Задача Scan

Есть массив  $A$ :  $A[0]$ ,  $A[1]$ , ...

Хотим посчитать:

- $A[0]$

# Задача Scan

Есть массив  $A$ :  $A[0]$ ,  $A[1]$ , ...

Хотим посчитать:

- $A[0]$
- $A[0] + A[1]$



# Задача Scan

Есть массив  $A$ :  $A[0]$ ,  $A[1]$ , ...

Хотим посчитать:

- $A[0]$
- $A[0] + A[1]$
- $A[0] + A[1] + A[2]$

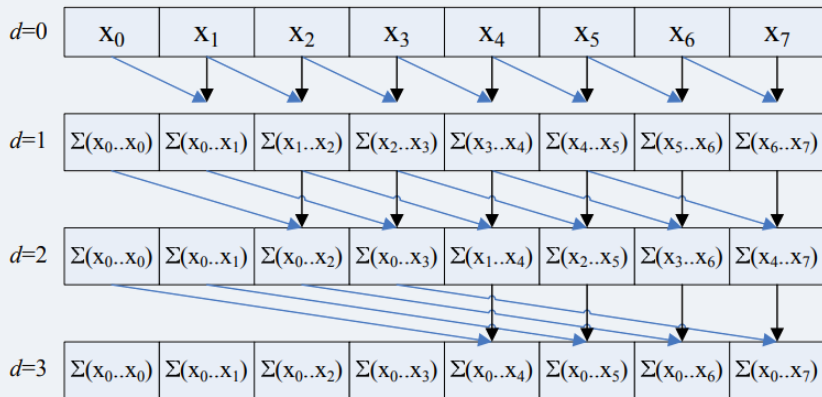
# Задача Scan

Есть массив  $A$ :  $A[0]$ ,  $A[1]$ , ...

Хотим посчитать:

- $A[0]$
- $A[0] + A[1]$
- $A[0] + A[1] + A[2]$
- ...

# Алгоритм решения!



# Проблемы алгоритма

# Проблемы алгоритма

- $O(N \log N)$  действий

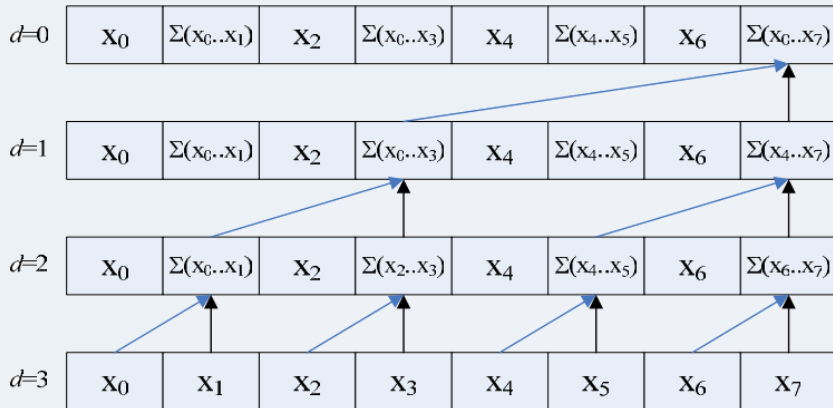
# Проблемы алгоритма

- $O(N \log N)$  действий
- $O(N \log N)$  warp-операций!

# Проблемы алгоритма

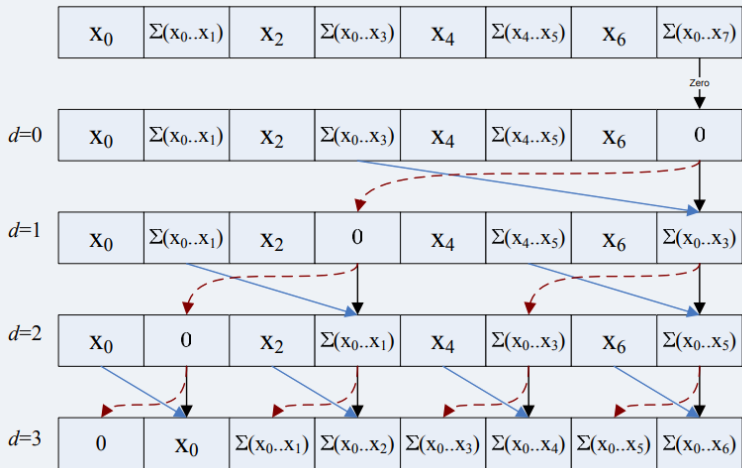
- $O(N \log N)$  действий
- $O(N \log N)$  warp-операций!
- Необходима синхронизация!

# Сделаем две стадии!





# Вторая стадия!



# Проблемы алгоритма

# Проблемы алгоритма

- $O(N)$  действий

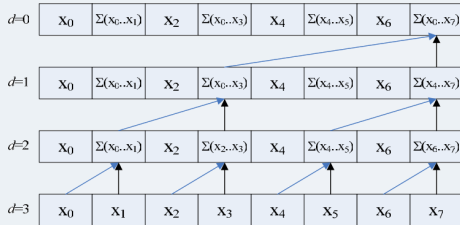
# Проблемы алгоритма

- $O(N)$  действий
- Сложная реализация

# Проблемы алгоритма

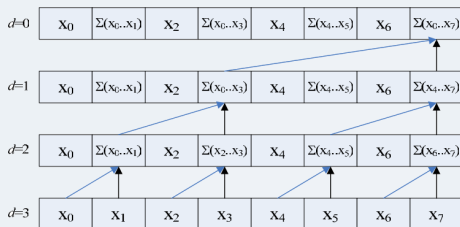
- $O(N)$  действий
- Сложная реализация
- Банк конфликты!

# Ищем конфликт!

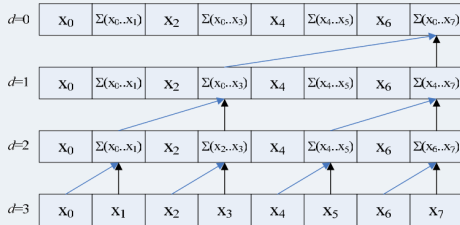


# Ищем конфликт!

- 0-й поток:  $0 + 1$



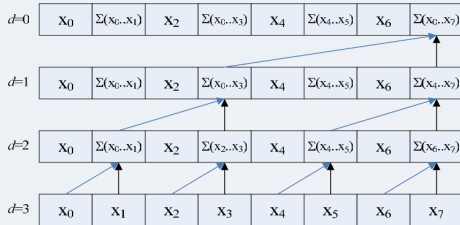
# Ищем конфликт!



- 0-й поток:  $0 + 1$
- 0-й поток:  $1 + 3$

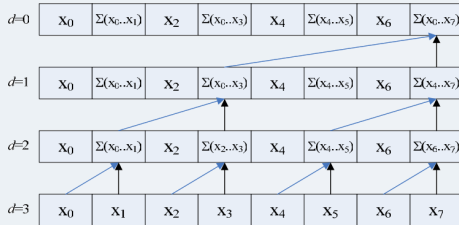


# Ищем конфликт!



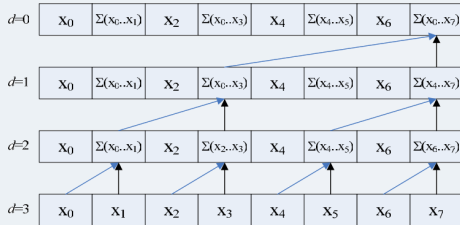
- 0-й поток:  $0 + 1$
- 0-й поток:  $1 + 3$
- 0-й поток:  $3 + 7$

# Ищем конфликт!



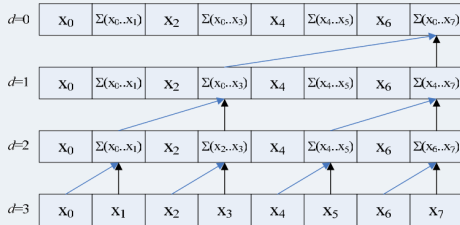
- 0-й поток:  $0 + 1$
- 0-й поток:  $1 + 3$
- 0-й поток:  $3 + 7$
- 0-й поток:  $7 + 15$

# Ищем конфликт!



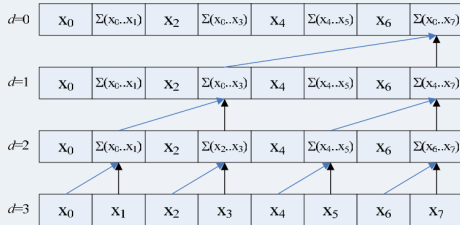
- 0-й поток:  $0 + 1$
- 0-й поток:  $1 + 3$
- 0-й поток:  $3 + 7$
- 0-й поток:  $7 + 15$
- 0-й поток:  $15 + 31$

# Ищем конфликт!



- 0-й поток:  $0 + 1$
- 0-й поток:  $1 + 3$
- 0-й поток:  $3 + 7$
- 0-й поток:  $7 + 15$
- 0-й поток:  $15 + 31$
- 16-й поток:  $32 + 33$

# Ищем конфликт!



- 0-й поток:  $0 + 1$
- 0-й поток:  $1 + 3$
- 0-й поток:  $3 + 7$
- 0-й поток:  $7 + 15$
- 0-й поток:  $15 + 31$
- 16-й поток:  $32 + 33$
- Видим проблему?

Решение: 25-ый кадр!

## Решение: 25-ый кадр!

- Пропускаем каждый 32-й элемент массива

## Решение: 25-ый кадр!

- Пропускаем каждый 32-й элемент массива
- $A[32] \rightarrow A[33]$



## Решение: 25-ый кадр!

- Пропускаем каждый 32-й элемент массива
- $A[32] \rightarrow A[33]$
- $A[64] \rightarrow A[66]$

## Решение: 25-ый кадр!

- Пропускаем каждый 32-й элемент массива
- $A[32] \rightarrow A[33]$
- $A[64] \rightarrow A[66]$
- 0-й поток:  $0 + 1$

## Решение: 25-ый кадр!

- Пропускаем каждый 32-й элемент массива
- $A[32] \rightarrow A[33]$
- $A[64] \rightarrow A[66]$
- 0-й поток:  $0 + 1$
- 16-й поток:  $32 + 33 \rightarrow 33 + 34$  ( $1 + 2$ )

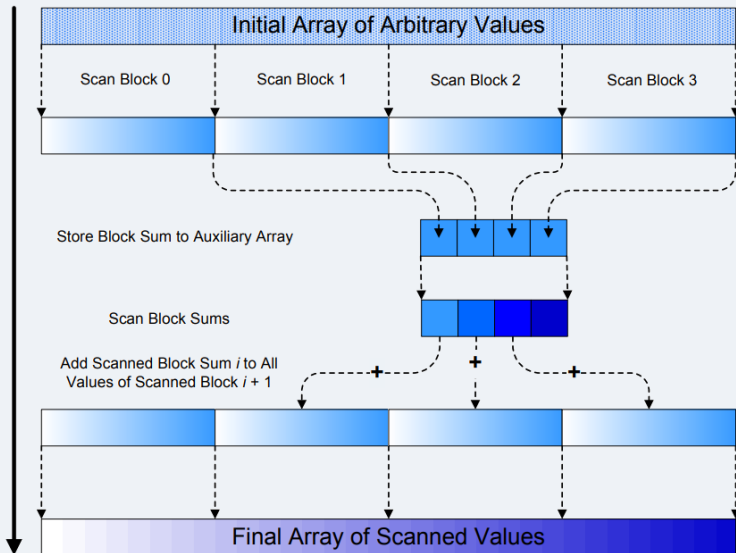
## Решение: 25-ый кадр!

- Пропускаем каждый 32-й элемент массива
- $A[32] \rightarrow A[33]$
- $A[64] \rightarrow A[66]$
- 0-й поток:  $0 + 1$
- 16-й поток:  $32 + 33 \rightarrow 33 + 34$  ( $1 + 2$ )
- Конфликтов нет (если  $WS = 32$ )

## Решение: 25-ый кадр!

- Пропускаем каждый 32-й элемент массива
- $A[32] \rightarrow A[33]$
- $A[64] \rightarrow A[66]$
- 0-й поток:  $0 + 1$
- 16-й поток:  $32 + 33 \rightarrow 33 + 34$  ( $1 + 2$ )
- Конфликтов нет (если  $WS = 32$ )
- Реализация:  $A[i + i / 32]$

# Глобальный Scan



# Что мы сделали?

- Решили задачу Scan

# Что мы сделали?

- Решили задачу Scan
- Нашли ещё один способ решения конфликтов!