

for $i = 1 \text{ to } N$

for $j = 1 \text{ to } M$

do something (i, j)

نحتاج حل مثالي بدون
دايناميك

1D blocks
1D threads

$i = bidx.x$

$j = tid.x$

do something (i, j)

حالة عندما أكثر من بلوك
وشرير وكلها 1D

N*M Blocks
1 thread

$i = bidx.y$

$j = bidx.x$

do something (i, j)

حالة عندما 2D بلوك
وشرير واحد فقط

1 Block
N*M threads

$i = tid.x$

$j = tid.y$

do something (i, j)

حالة عندما 2D شرير
و بلوك واحد فقط

N*M blocks
N*M threads

$i = bidx.x * blockDim.x + tid.x$

$j = bidx.y * blockDim.y + tid.y$

do something (i, j)

حالة عندما 2D بلوك
و 2D شرير


```
for i = 0 to N
  for j = 0 to X[i]
    do something(i, j);
```

احد التالي لازم
يكون بالداينيك
مشان ما يكون فيه
آيدل ثريزر

~~i = 0 to X~~

Parallel Solution

```
-- global -- void callKernel (...) {
  int i = blockIdx.x;
  int j = threadIdx.x;
  if (j < X[i]) {
    do something(i, j);
  }
}

void main() {
  callKernel <<< N, max(X) >>> (X, ...);
```

Bad Solution
because we will
have idle threads

Dynamic Solution

```
-- global -- void childKernel (int i) {
  int j = threadIdx.x;
  do something(i, j);
}

-- global -- void kernel (...) {
  int i = threadIdx.x;
  childKernel <<< 1, X[i] >>> (i);
  cudaDeviceSynchronize();
}
```

Perfect Solution
we don't have
any idle threads

```
void main()
{
  kernel <<< 1, N >>> (X, ...);
}
```


Question



Let's consider the following parallel code:

```
__global__ void Kernel_A(int *data) {
    data[threadIdx.x] = threadIdx.x;
    __syncthreads();
    if (threadIdx.x == 0) {
        Kernel_C<<< 1, 256 >>>(data);
        Kernel_D<<< 1, 256 >>>(data);
        cudaDeviceSynchronize();
    }
    __syncthreads();
}

__global__ void Kernel_C(int *data) {
    data[threadIdx.x] = threadIdx.x;
    __syncthreads();
    if (threadIdx.x == 0) {
        Kernel_E<<< 1, 256 >>>(data);
        cudaDeviceSynchronize();
    }
    __syncthreads();
}

void host_launch(int *data) {
    kernel_A<<< 1, 256 >>>(data);
    kernel_B<<< 1, 256 >>>(data);
    cudaDeviceSynchronize();
}
```

1. Give and explain the order of execution of the given parallel nested kernels.
2. Explain the role of the **__syncthreads()** statements.
3. Explain the role of the **cudaDeviceSynchronize()** statements.

1-ACEDB

__syncthreads() Allows threads of the same block to wait for each other

cudaDeviceSynchronize() the parent thread waits for its child threads to finish,
In HOST the parent waits for its last called kernel

D&C

```
__global__ void sum_Kernel(    int * data, int * res, int parentStartingIndex, int parentNbElements) {  
    int Ai = threadIdx.x;  
    int childNbElements = parentNbElements / 2;  
    int childStartIndex = parentStartingIndex + Ai * childNbElements;  
    int childEndIndex = childStartIndex + childNbElements - 1;  
    int partialResult = 0;  
    if (childNbElements == 2) { // The base case  
        partialResult = data[childStartIndex] + data[childEndIndex];  
    }  
    else {  
        sum_Kernel<<<1,2>>>(data, &partialResult, childStartIndex, childNbElements);  
        cudaDeviceSynchronize();  
    }  
    atomicAdd(res, partialResult)  
}
```

Quadtree

```
__global__ void Quadtree_Kernel(int * data, int R, int C, int W, int level) {  
    int X, Y;  
    X = threadIdx.x / 2;  
    Y = threadIdx.x % 2;  
  
    int Ri, Ci, Wi;  
    Wi = W / 2;  
    Ri = R + X * Wi;  
    Ci = C + Y * Wi;  
  
    if (level == 10) {  
        do_baseCase(data, Ri, Ci, Wi);  
    } else {  
        Quadtree_Kernel<<<1, 4>>>(data, Ri, Ci, Wi, level+1);  
        cudaDeviceSynchronize();  
    }  
}
```

Odd Even Sort

- How many iterations are required to sort an array of size N . $\text{ceiling}(N/2)$
- How many steps are performed per iteration and describe every one of them.

b) Two steps, Odd step and Even step.

In odd step we compare the element of every odd index with the next cell, if they are out of order we swap

In even step ----- even index -----

- Which threads will be involved in every **step i** in case the algorithm is performed in parallel. Don't forget to specify, for every thread, the index of the cells it will process.

In Odd step: Thread T_i | i is an odd number and will process $\text{data}[i]$, $\text{data}[i+1]$

In Even step: Thread T_i | i is an even number and process $\text{data}[i]$, $\text{data}[i+1]$

Bitonic sort

- Give the number of steps that are required to sort elements of an array of size N . $\log_2(N)$
- Give the number of stages that are required in a given step i . $\text{Stages} = i$
- Give the size of bitonic sequences in a given stage j of a step i . 2^{i-j+1}
- Give the condition that should satisfy a thread to participate in the processing of bitonic sequences of a stage j of a step i . $\text{index} \% 2^{i-j+1} < 2^{i-j}$
- Give the condition that should satisfy a thread that participates in the processing of sequences of a stage j of a step i to sort its corresponding bitonic-sequence ascendingly.

