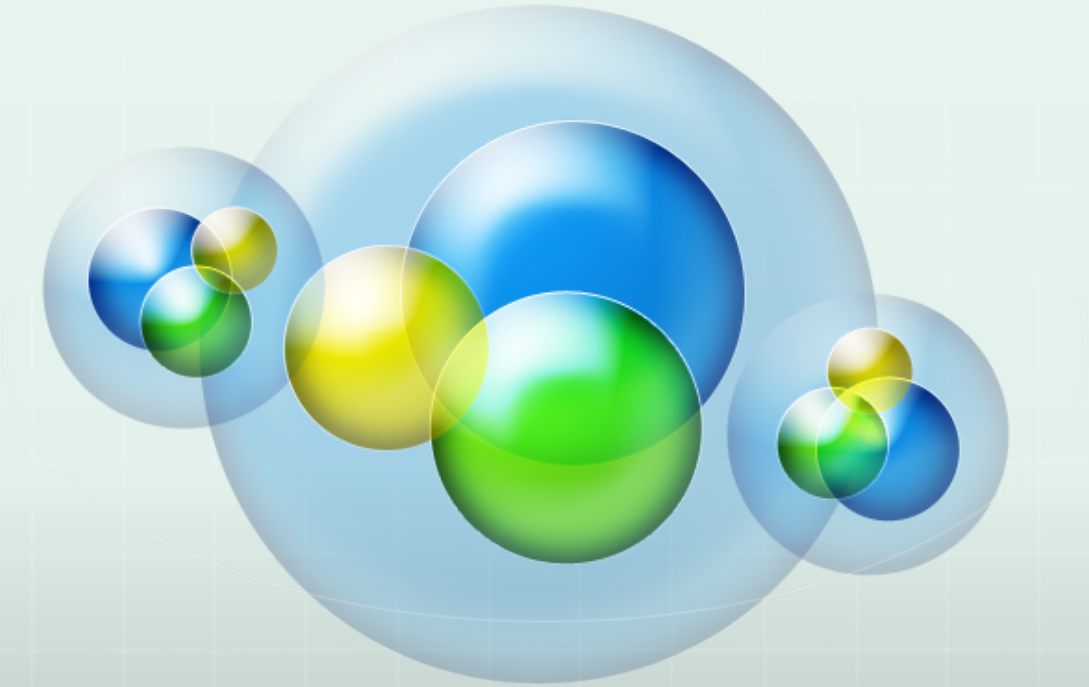


CS475 data dependence

Data dependence

Consider two statement instances x and y, where x executes after y in the sequential version of the program, that we want to parallelize

- x depends on y if x and y access (read or write) the same memory location, notation: y←x
- there are different kinds of dependences:
  - y:write←x:read **RAW** read after write, **true** dependence
  - y:read ←x:write **WAR** write after read, **anti** dependence
  - y:write←x:write **WAW** write after write, **output** dependence
  - y:read←x:read **RAR** read after read, **input** dependence

For the first 3, order matters (changing the order changes the outcome of the program).

For the 4th it does not, so why do we talk about it? (memory)

Parallelization

- When true, anti, or output dependences occur in a sequential program, their order cannot be changed when parallelizing the program. WHY?

- changing the order changes the outcome of the program

Examples

```
EX1: for (i=1; i<N; i++)
       for (j=1; j<M; j++)
        A[i,j] = f(A[i,j-1], A[i-1,j])

EX2: for (i=1; i<N; i++)
       for (j=1; j<M; j++)
        B[j] = f(B[j-1], B[j])

 EX3: for (i=1; i<N; i++)
       for (j=1; j<M; j++)
        C[i] = f(C[i-1], C[i])
```

Iteration & Data Space

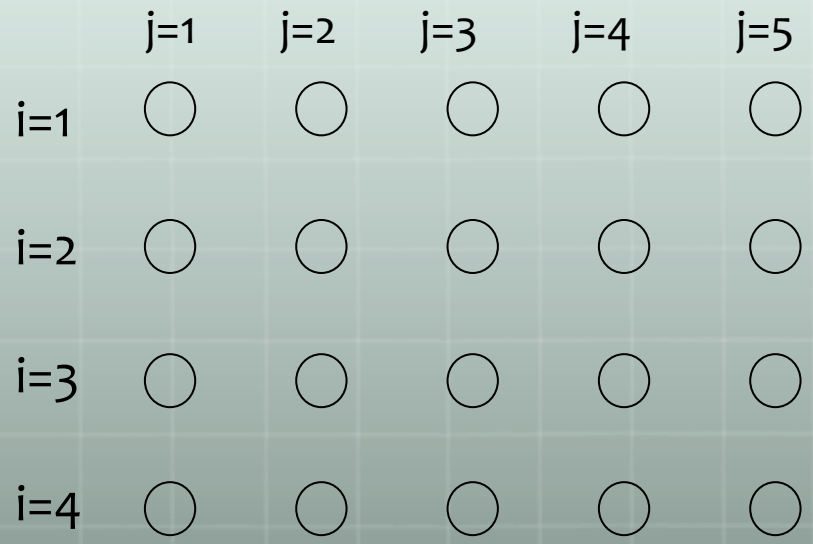
- Iteration Space: set of values that the loop iterators can take, in our 3 examples:
  - a rectangular region, with "corners" [1,1] and [N-1, M-1]
- Data Space: set of values of array indices accessed by the statements in the program
  - Ex 1: 2-D array, similar to the iteration space
  - Ex 2: 1-D array, bounded by [0, M-1]
  - Ex 3: 1-D array, bounded by [0, N-1]
- We will concentrate on the iterations space here

Drawing Spaces

There are many ways to draw a space

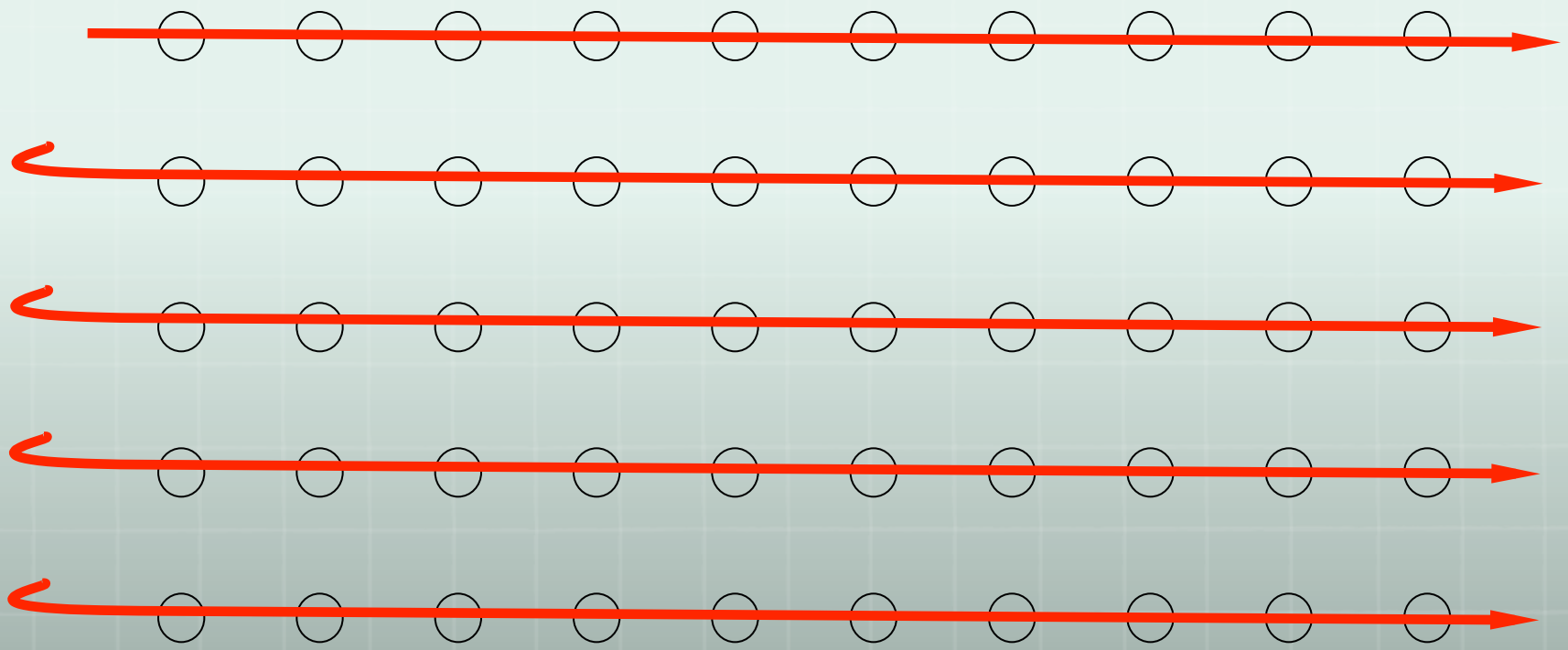
We will draw iteration and data spaces like this:

row index i goes down, column index j goes right




```
for (i=1; i<N; i++)
    for (j=1; j<M; j++)
        ...A[i,j]...
```

Row major execution order



References and Dependences

- Reference: an occurrence of an (array) variable on either
  - left hand side of an assignment (write)
  - right hand side of an assignment or in an expression (read)

  of a statement in the loop body
- Dependences specify which iteration points depend on which others

Finding dependences

Very hard problem (undecidable in general) but we have special decidable cases, e.g. when the dependences are expressed by linear expressions in loop indices. e.g.,

i-1, j+1, 2*i+3, 3*j + n

For example, finding true dependences:

An iteration point [i, j] reads a memory location, that one or more iterations may have written.

- Find the writers as a function of [i,j].
- Find the "most recent writer" in this set (again, as a function of [i,j])

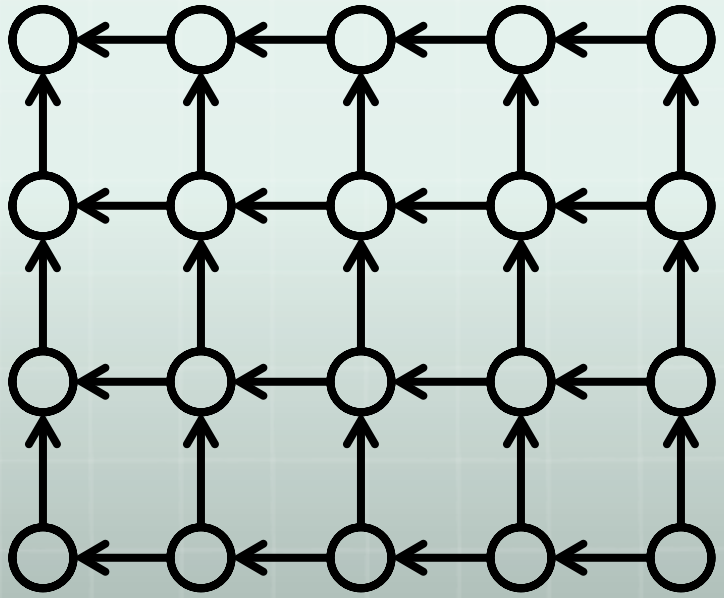
Examples

```
EX1: for (i=1; i<N; i++)
       for (j=1; j<M; j++)
        A[i,j] = f(A[i,j-1], A[i-1,j])

EX2: for (i=1; i<N; i++)
       for (j=1; j<M; j++)
        B[j] = f(B[j-1], B[j])

 EX3: for (i=1; i<N; i++)
       for (j=1; j<M; j++)
        C[i] = f(C[i-1], C[i])
```

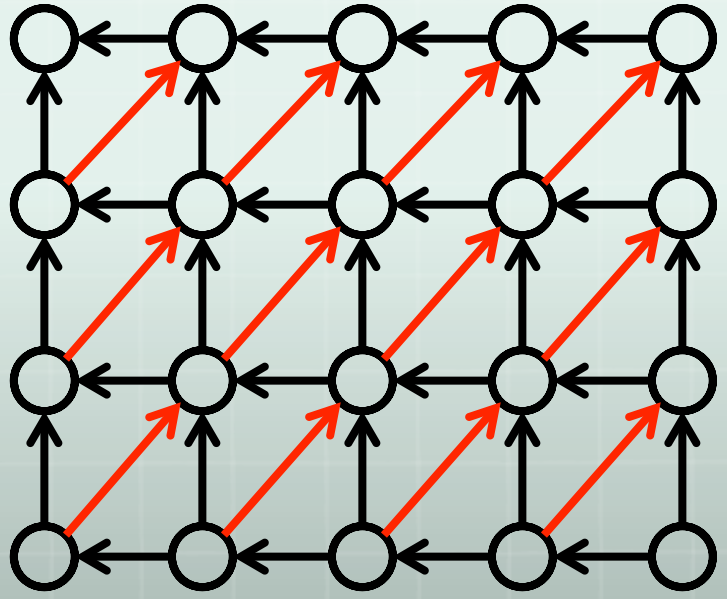
Dependence Graph (Ex 1)



```
Ex1: for (i=1; i<N; i++)
       for (j=1; j<M; j++)
         A[i,j] = f(A[i,j-1], A[i-1,j])
```

Iteration [i,j] depends on:
[i, j-1] and [i-1, j]
neighbors on west and north
and writes A[i,j]
(read by [i+1,j] and [i,j+1])

Dependence Graph (Ex 2)

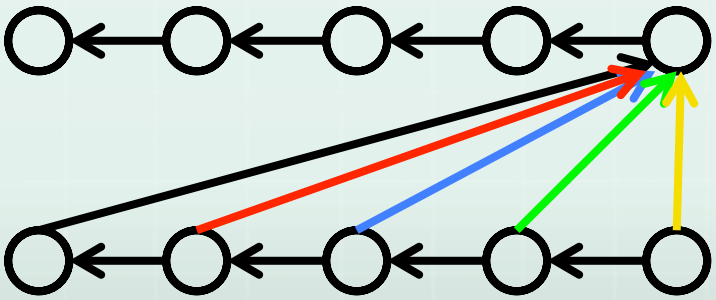


```
Ex2: for (i=1; i<N; i++)
       for (j=1; j<M; j++)
         B[j] = f(B[j-1], B[j])
```

Iteration [i,j] true-depends on:
[i, j-1] and [i-1, j]
neighbors on west and north

Iteration [i-1, j+1] reads a memory location that iteration [i, j] will overwrite, so [i,j] anti-depends on [i-1,j+1], and thus, [i-1,j+1] must be executed before [i, j] , and hence, [i,j] and [i-1,j+1] **cannot execute in parallel**

Dependence graph (ex 3)



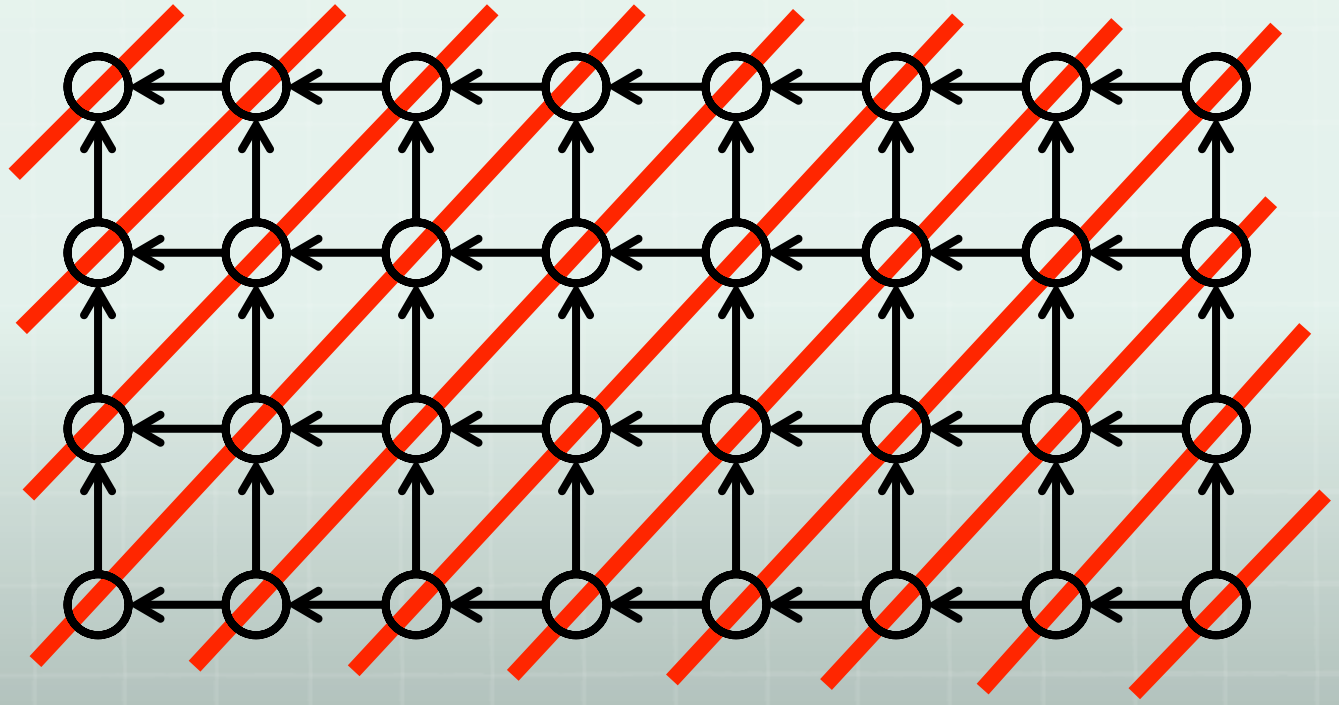
```
for (i=1; i<N; i++)
    for (j=1; j<M; j++)
        C[i] = f(C[i-1], C[i])
```

totally sequential dependence

Wavefront parallelization

- When we know the dependences between iterations (the dependence graph)
  - analyze to determine which iterations can happen at which time step (hopefully many iterations can happen at the same time step)
  - rewrite the program to represent this new order

Ex1 wavefront parallelization



Reorder the loop:

Outer loop executes the diagonals sequentially from top left to bottom right

Inner loop executes in parallel all nodes in one diagonal

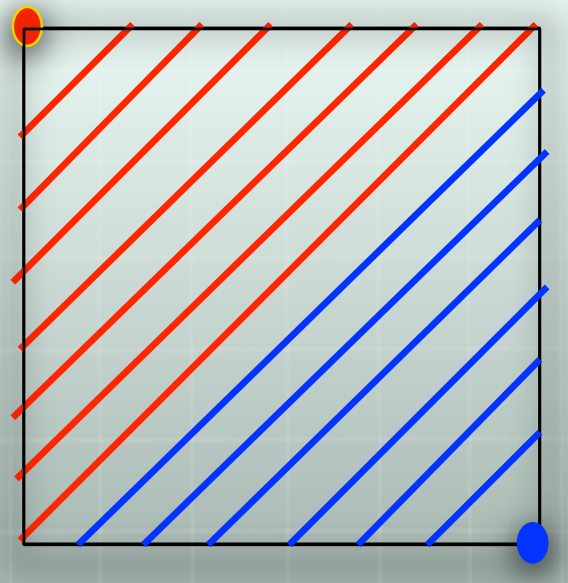
Ex1 wavefront parallelization

```
for (i=1; i<N; i++)
    for (j=1; j<M; j++)
        A[i,j] = f(A[i,j-1], A[i-1,j])
```

case 1: N=M → square grid

```
for d = 0 to N-1
// i walks from top side to left side
   for i = 0 to d

for d = N to 2N-2
// i walks from right side to bottom
   for i =  d-N+1 to N-1
```

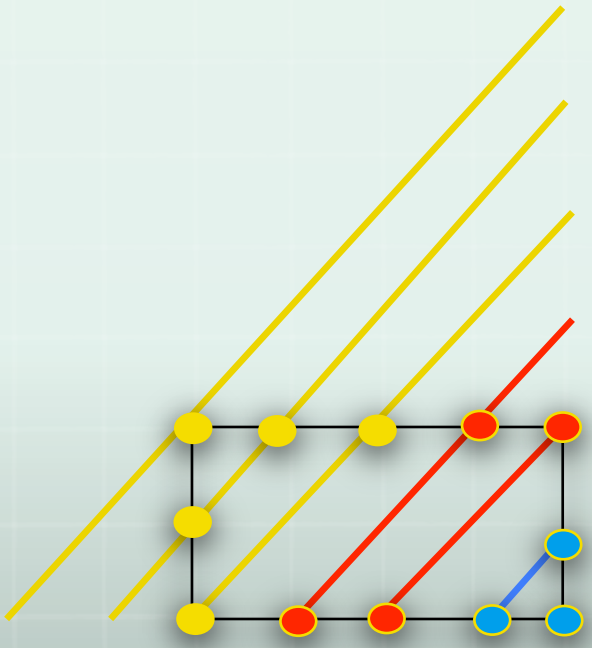


Ex1 wavefront parallelization

```
for (i=1; i<N; i++)
    for (j=1; j<M; j++)
        A[i,j] = f(A[i,j-1], A[i-1,j])
```

case 2: N != M

```
        rectangular grid
        // i starts at top side, right side
        // finishes left side, bottom
        for d = 0 to N+M-2
          for i =  max(0,d-M+1) to min(d,N-1)
```



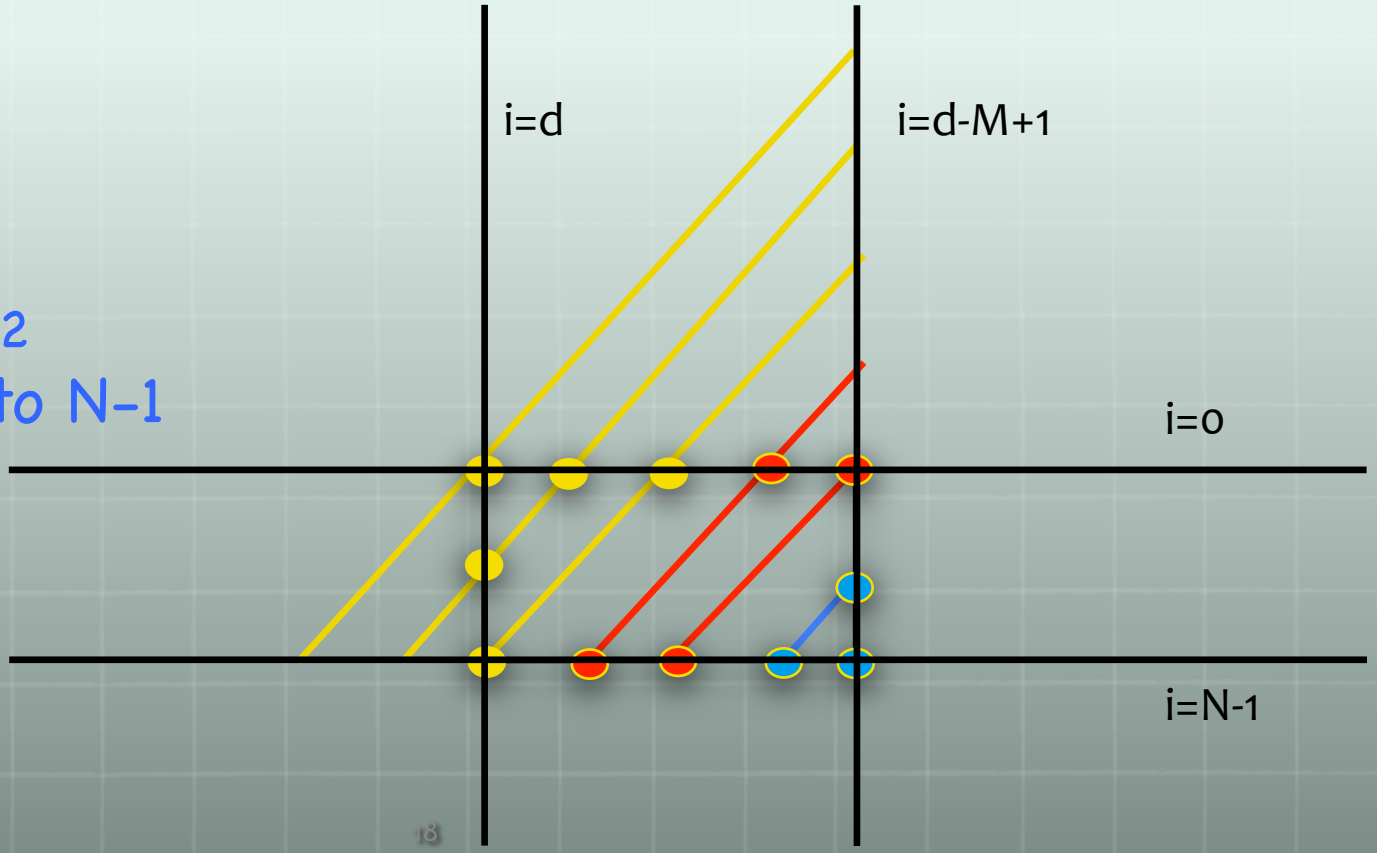
i bounds on line intersections with diagonal

```
for d = 0 to N+M-2
     for i =  max(0,d-M+1) to min(d,N-1)
// can also be written as 3 loops without mins and maxes
```

```
1: d = 0 to N-1
     i = 0 to d

2: d = N to M-1
     i = 0 to N-1

3: d = M to N+M-2
     i = d-M+1 to N-1
```



Ex2 wavefront parallelization