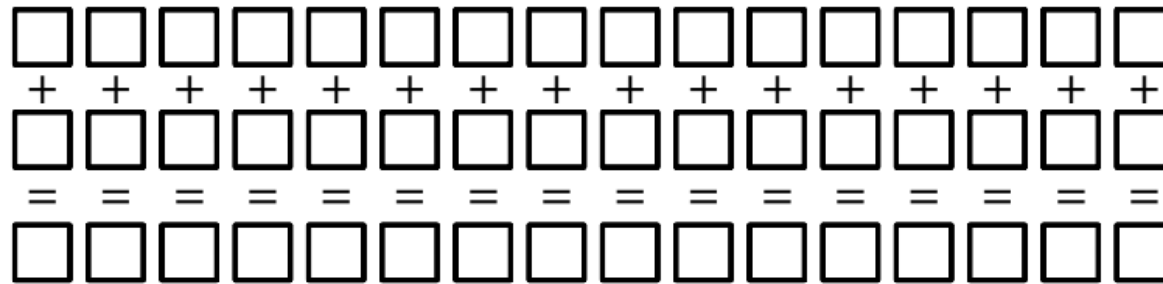


**An example of vectorization** – Vector instructions improve the performance by processing multiple data items concurrently.

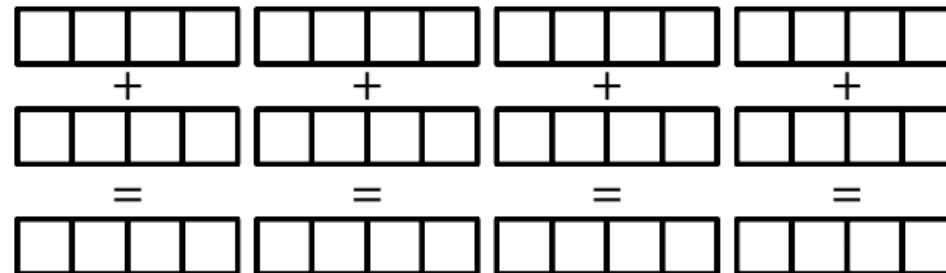
```
for (int i=0; i<16; i++)  
    a[i] = b[i] + c[i];
```

Scalar instructions



32 loads  
16 adds  
16 stores

vector length  
↔

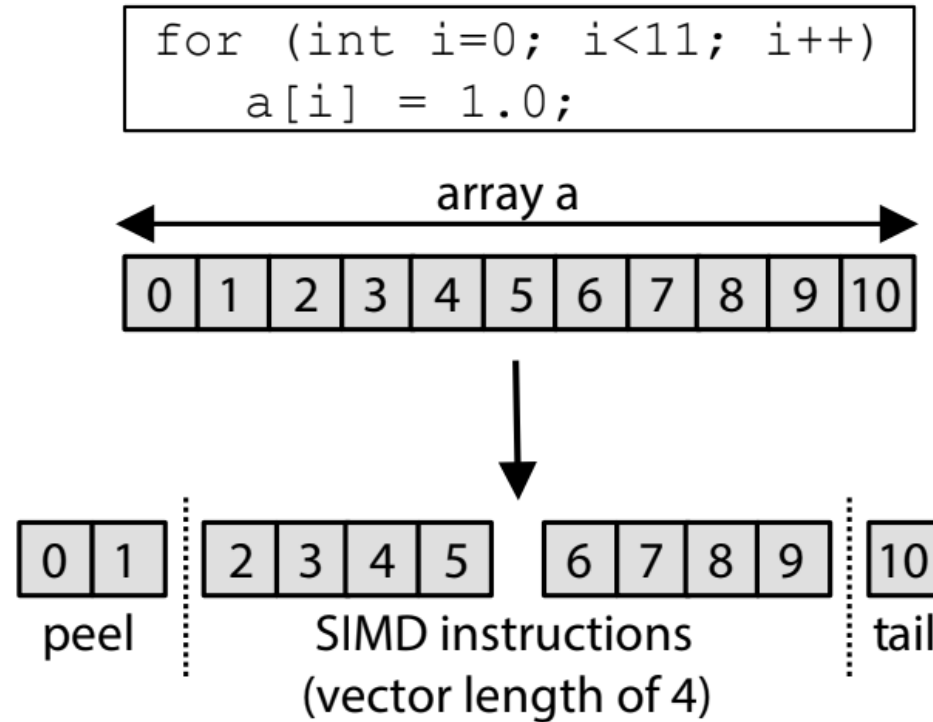
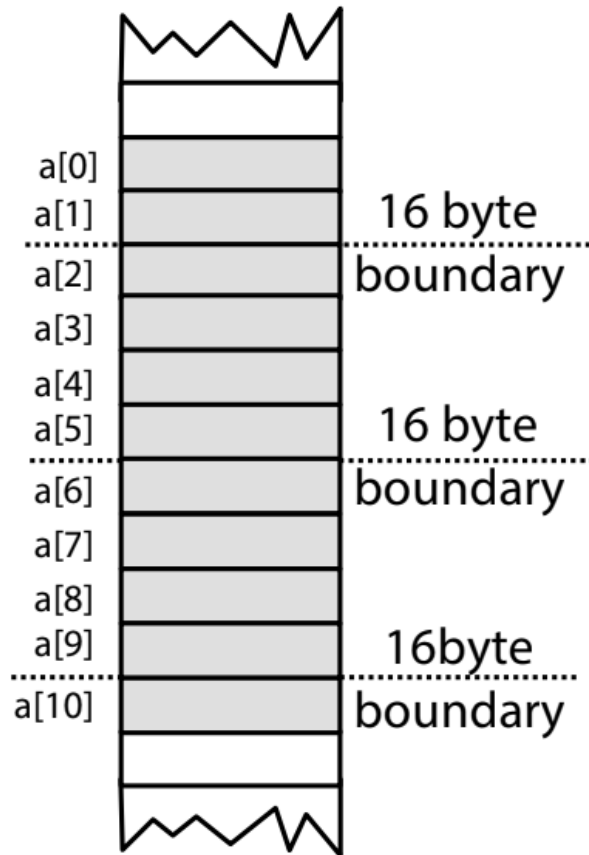


SIMD instructions

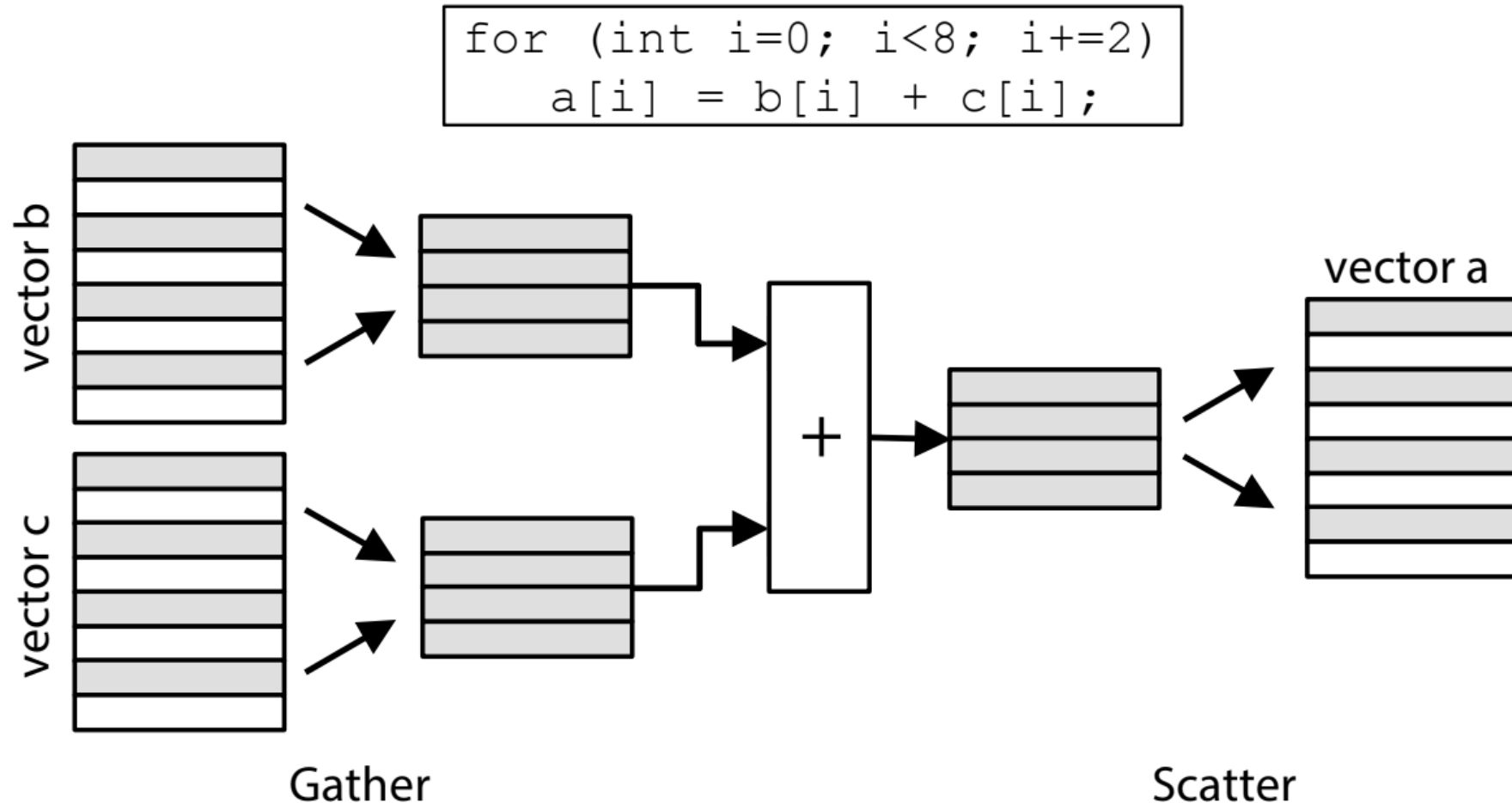
8 loads  
4 adds  
4 stores

## Loop modifications needed to vectorize a loop using SIMD

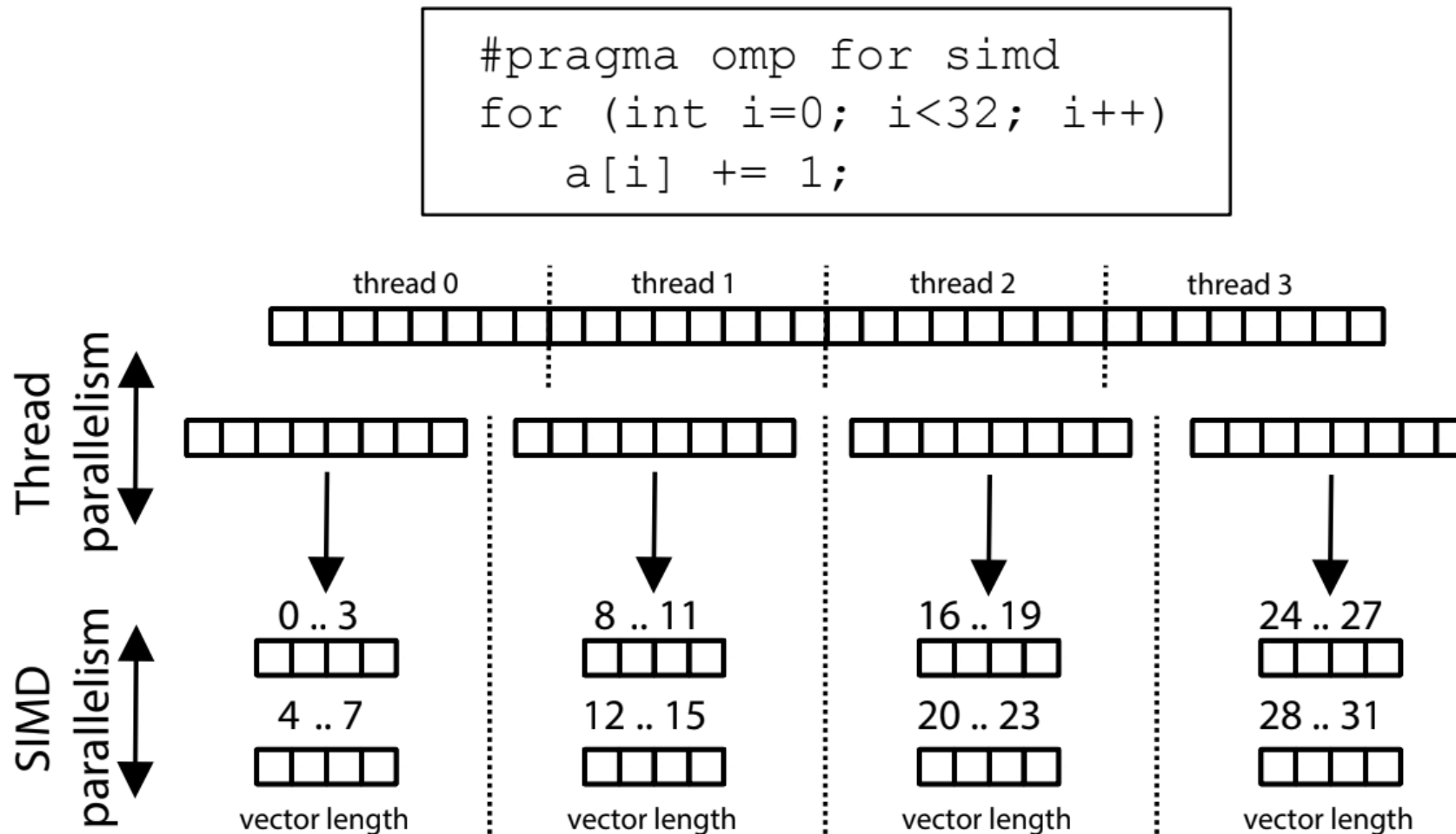
**instructions** – To enforce alignment and ensure the remaining loop length is a multiple of the vector length, the compiler may need to peel off iterations and also treat the tail end separately.



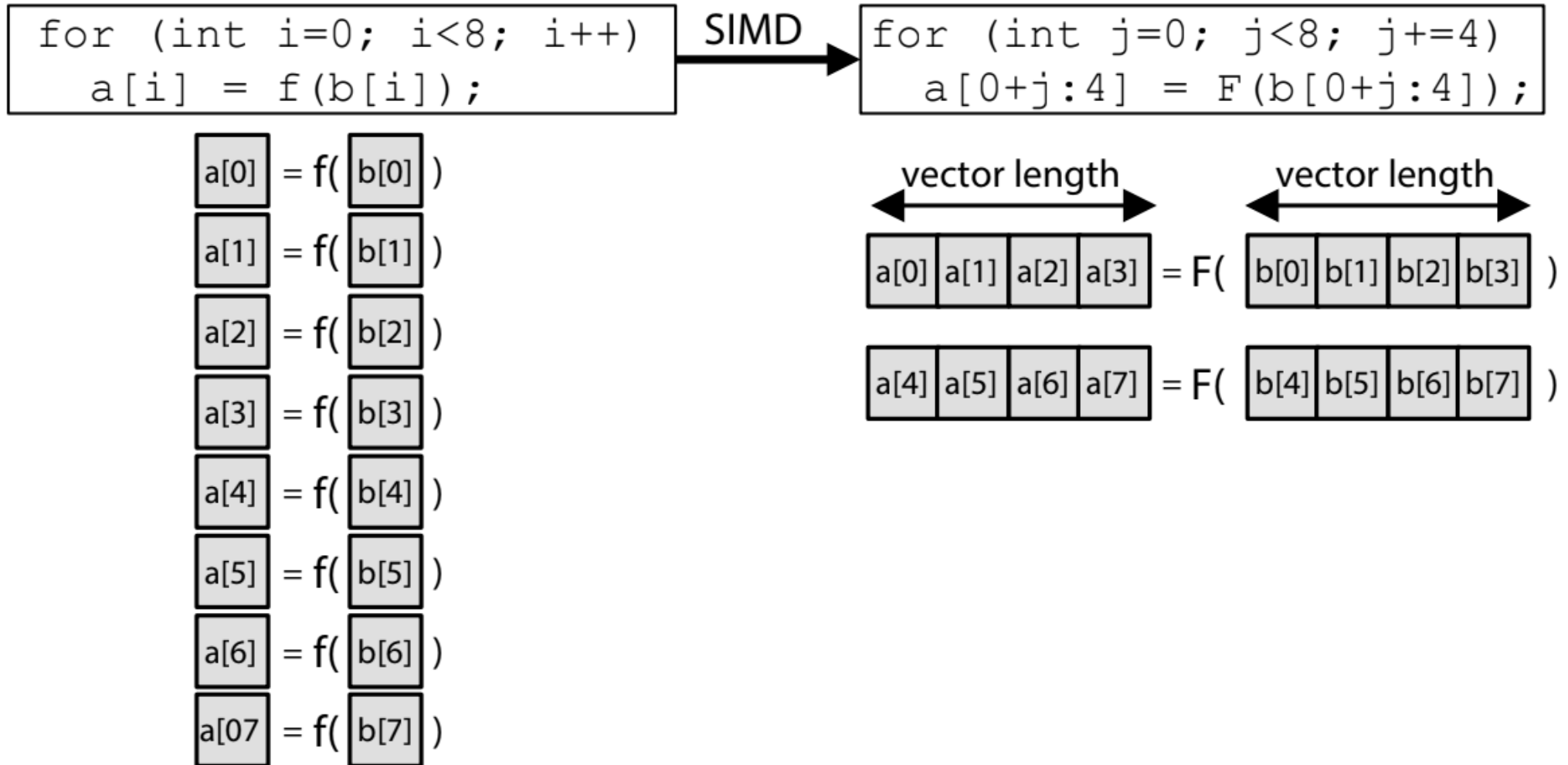
**Gather and scatter scalar data elements in and out of a vector** – Scalar data elements are gathered into vectors, operated on as a vector, and then scattered back to their destination locations.



## Combining thread and SIMD parallelism – Thread and SIMD parallelism are used to execute a loop.



**Illustration of function calls with SIMD** – The scalar function  $f()$  is modified and renamed to  $F()$  in this example. This function supports vector input arguments and returns an entire vector.



**Conditional control flow converted to masked vector instructions** – A vector mask predicate is used to enable or disable an operation on a given vector element. If the indexed mask is 1, the operation occurs. When it is 0, the operation is masked off.

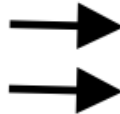
```
for (int i=0; i<8; i++)  
    if (b[i] < 0) a[i] += 1;
```

vector b

4
-2
-3
5

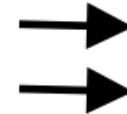
mask

a[0]
a[1]
a[2]
a[3]



0
1
1
0

+
+
+
+



a[0]
a[1]+1
a[2]+1
a[3]

---

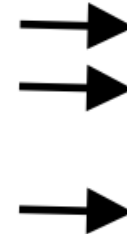
-6
-4
8
-9

a[4]
a[5]
a[6]
a[7]



1
1
0
1

+
+
+
+



a[4]+1
a[5]+1
a[6]
a[7]+1