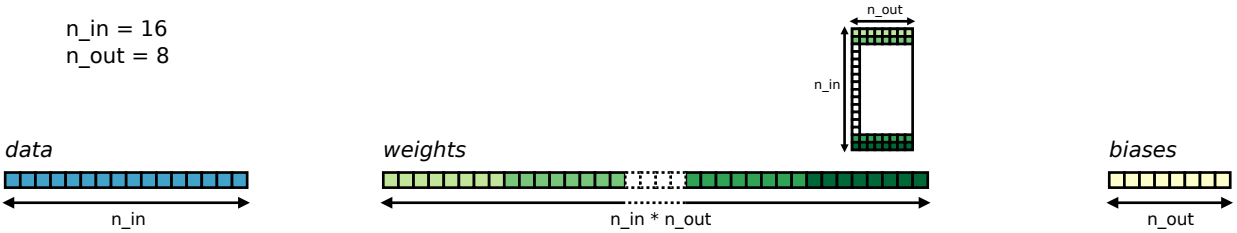
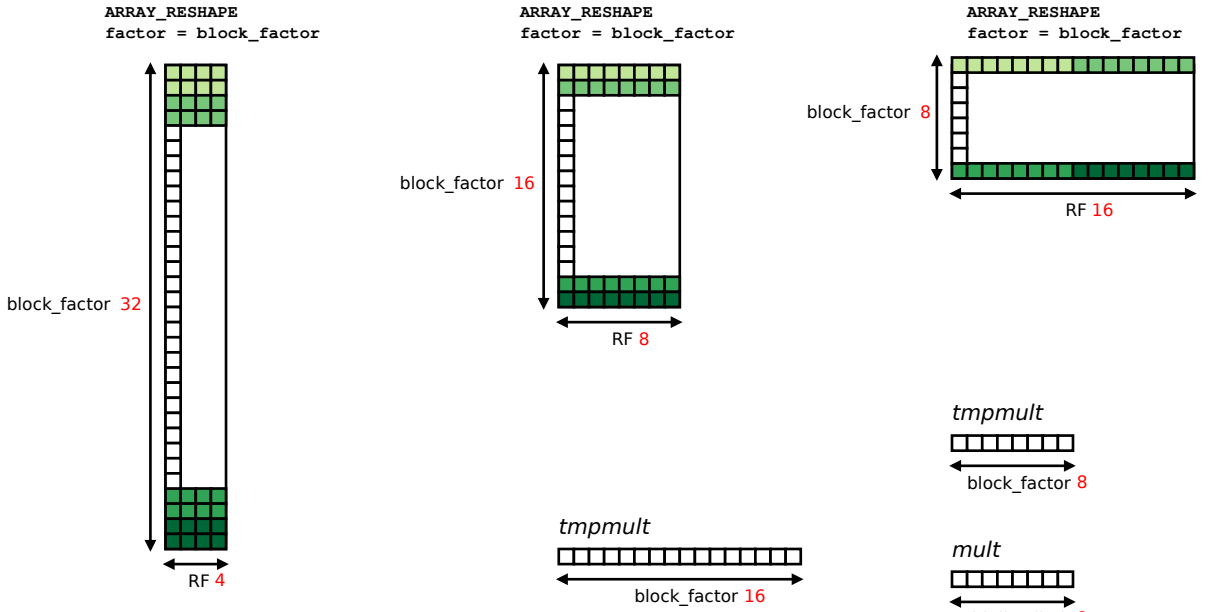
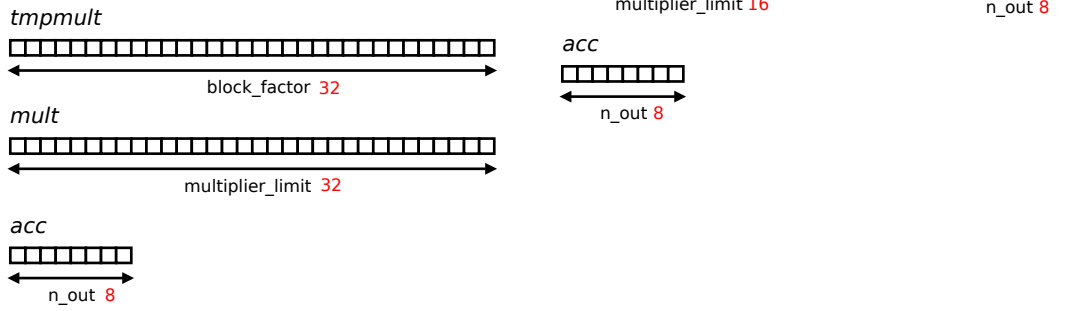
n_in = 16
n_out = 8

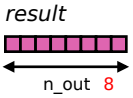


$$block\_factor = \left\lceil \frac{n\_in * n\_out}{RF} \right\rceil \qquad multiplier\_limit = \left\lceil \frac{n\_in * n\_out}{min(n\_in, RF)} \right\rceil$$

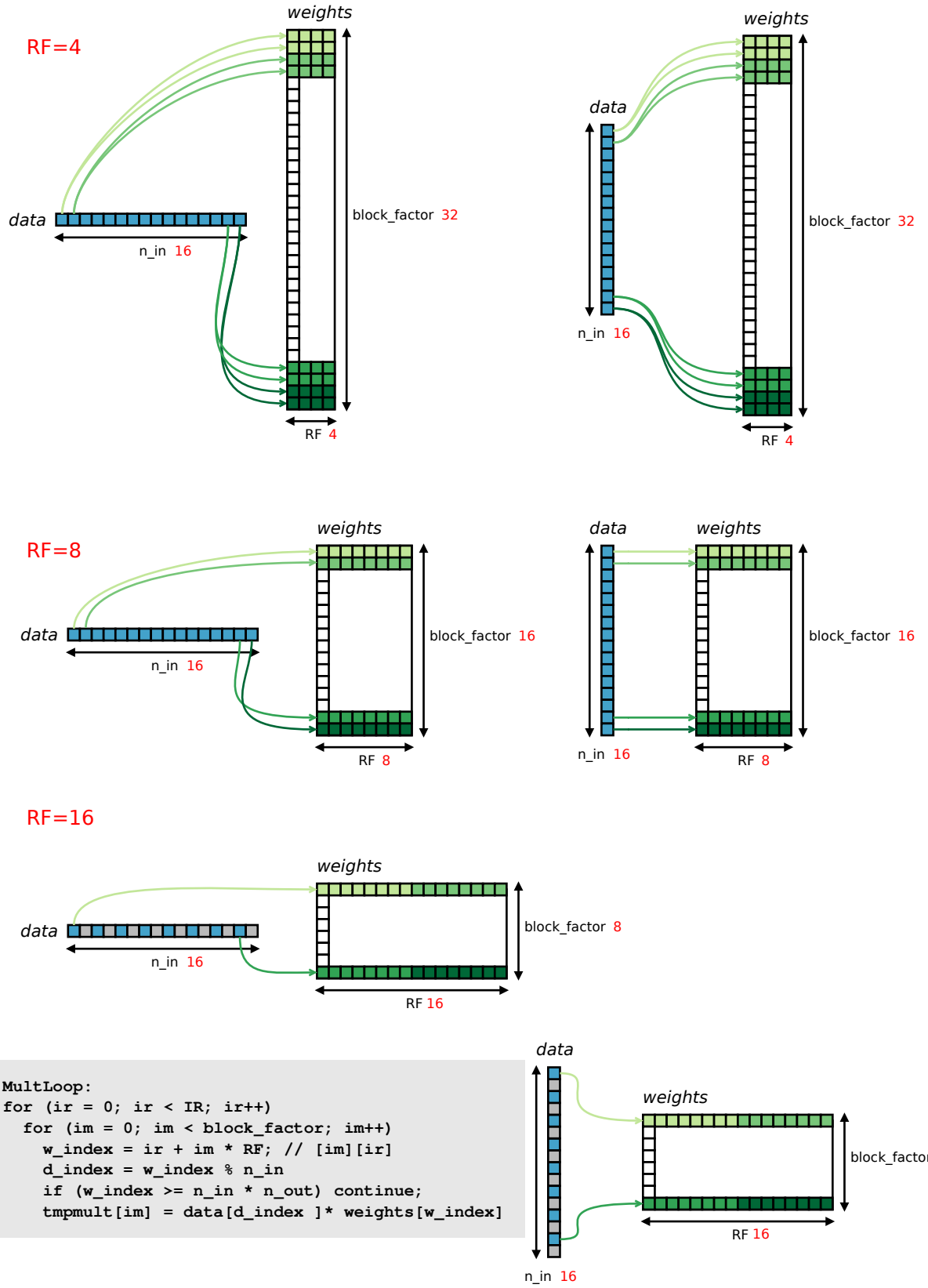
RF=4 RF=8 RF=16







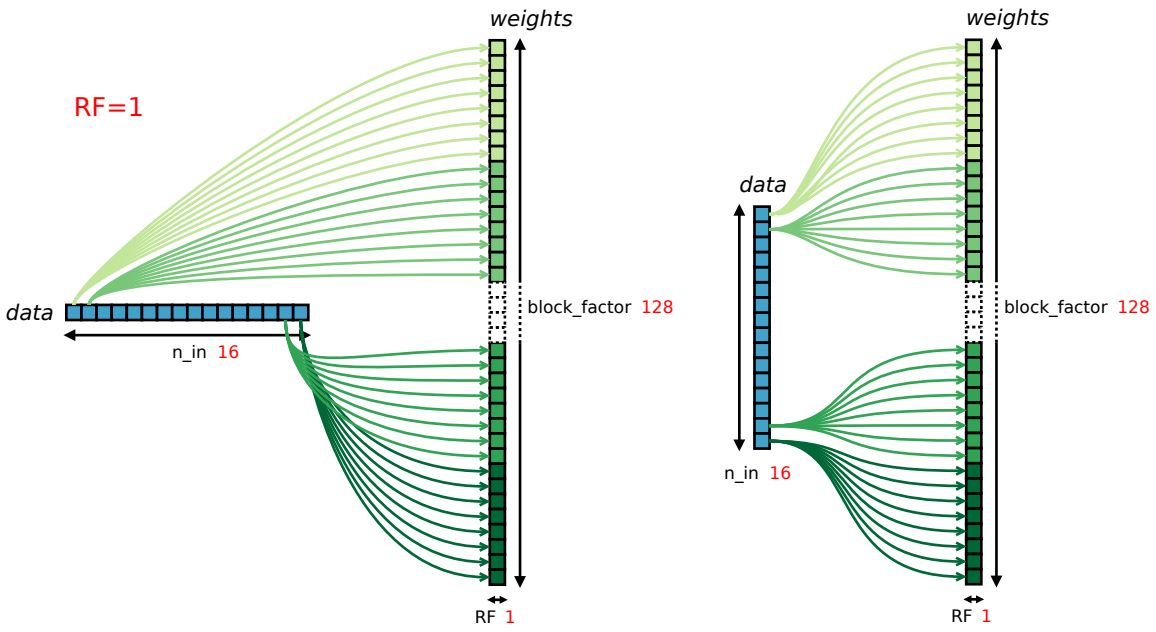
Two different visualizations of MultLoop




```
MultLoop:
for (ir = 0; ir < IR; ir++)
  for (im = 0; im < block_factor; im++)
    w_index = ir + im * RF; // [im][ir]
    d_index = w_index % n_in
    if (w_index >= n_in * n_out) continue;
    tmpmult[im] = data[d_index ]* weights[w_index]
```

Two extreme situations: (RF = 1, block_factor = 128) and (RF = 128, block_factor = 1)

RF=1




```
MultLoop:
for (ir = 0; ir < IR; ir++)
  for (im = 0; im < block_factor; im++)
    w_index = ir + im * RF; // [im][ir]
    d_index = w_index % n_in
    if (w_index >= n_in * n_out) continue;
    tmpmult[im] = data[d_index] * weights[w_index]
```

RF=128