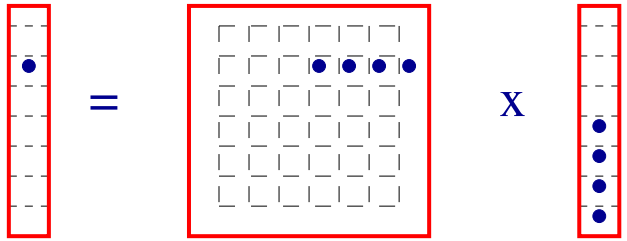
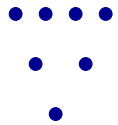
Unifying CPU and GPU programming models

- Memory layout: both SIMD and SIMT need a vector direction
- Optimised stencil code will see lower-level interfaces
- SIMD and SIMT becomes exposed

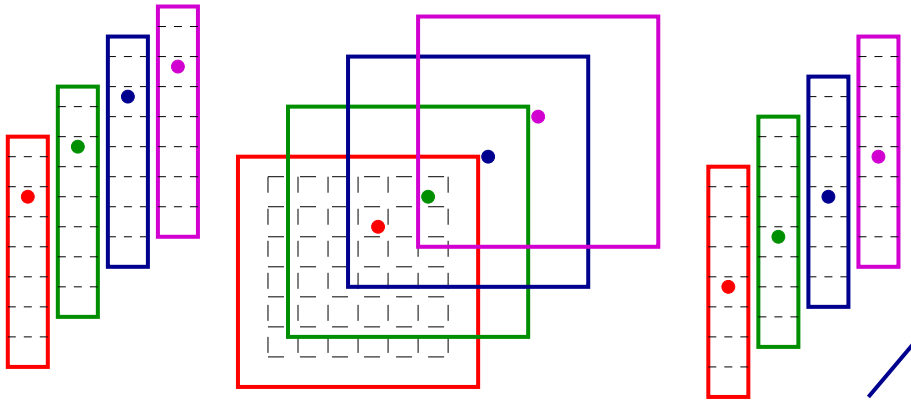
Vector = Matrix x Vector





Reduction of vector sum
is bottleneck for small N

Many vectors = many matrices x many vectors



No reduction or SIMD lane crossing operations.

SIMD interleave