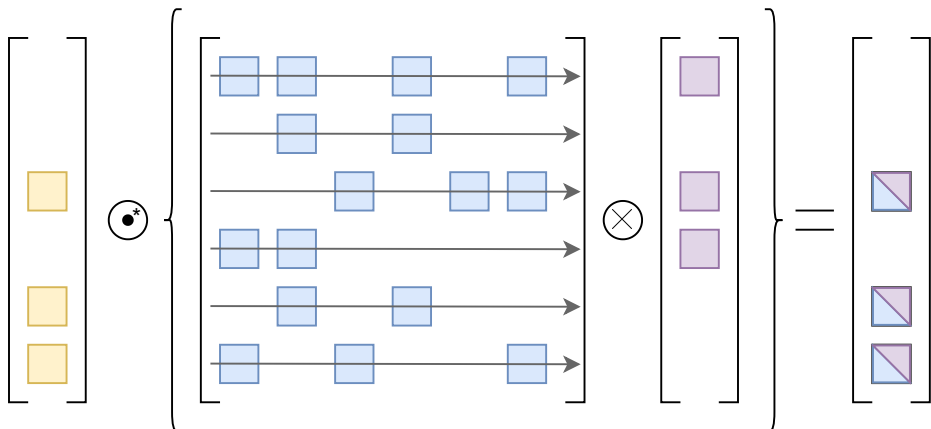


Matrix

Vector

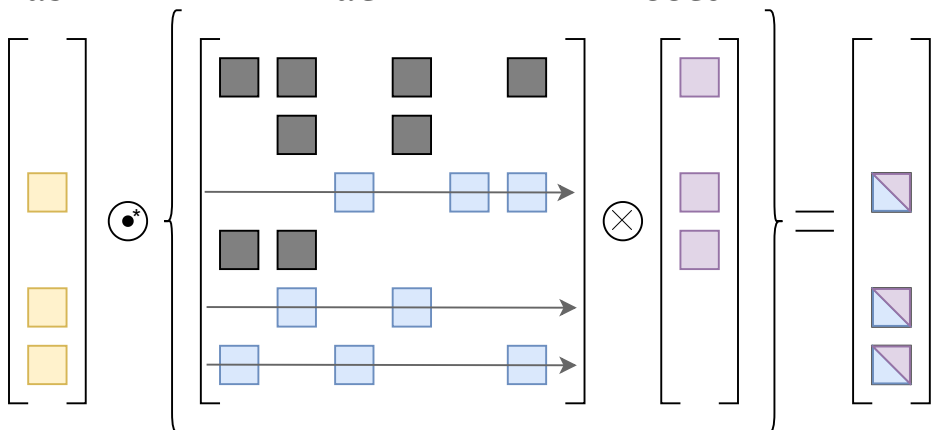


Not-fused masking requires 16 reads and writes to matrix

Mask

Matrix

Vector



Fused masking requires 8 reads and writes to matrix applying mask ahead-of-time