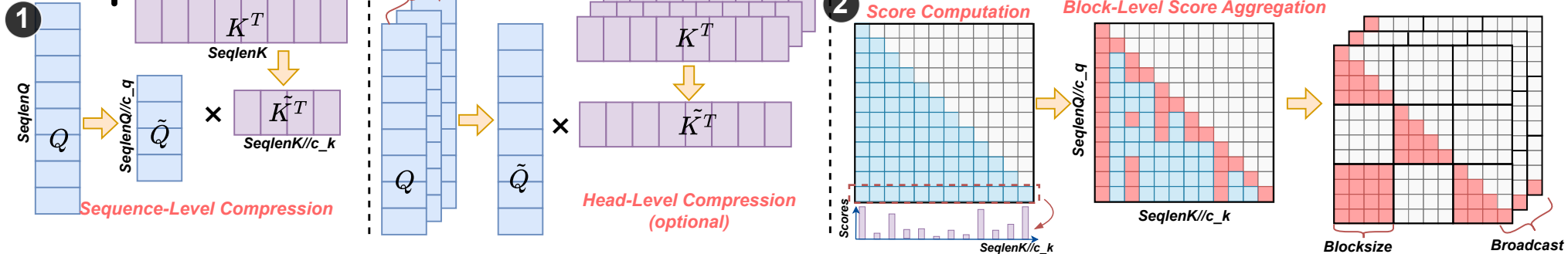


UniSparse



1 Stage1: Multi-Granularity Compression

Sec 4.2

2 Stage2: Dynamic Block Selection

Sec 4.3

3 Block Sparse Attention Computation