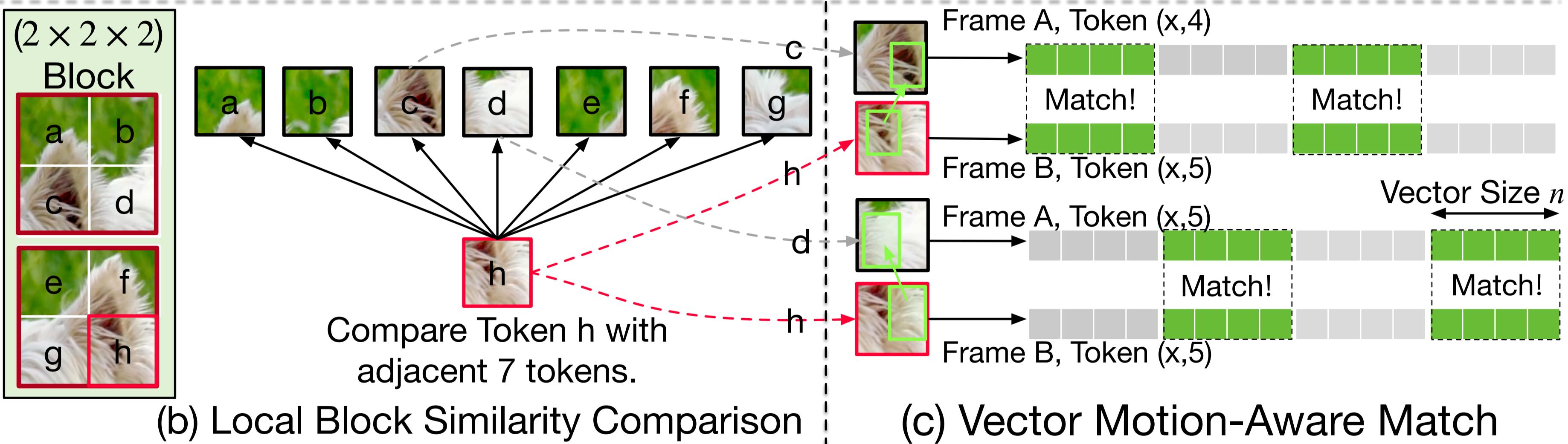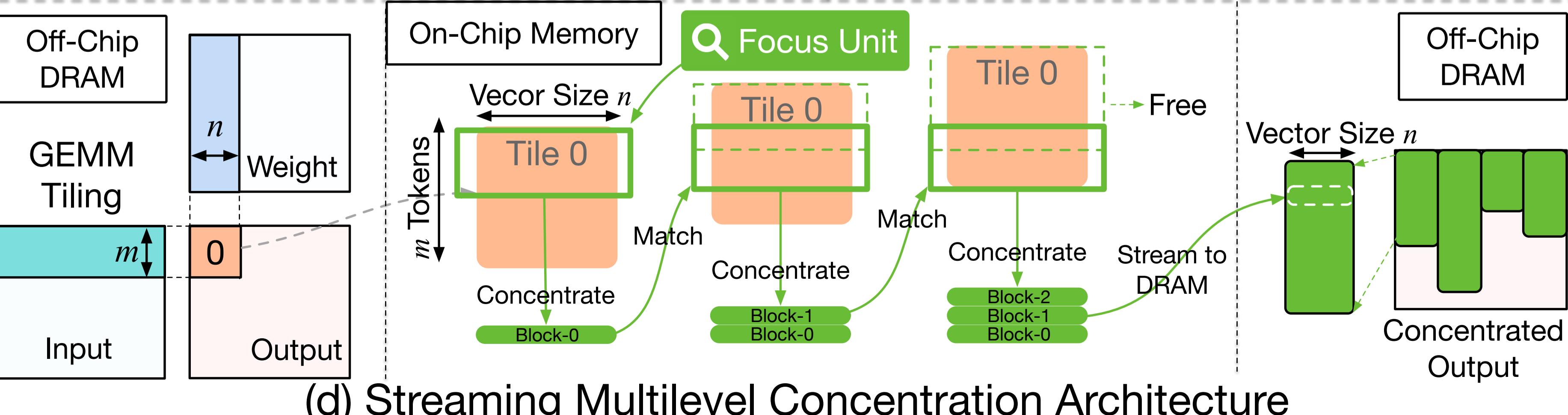(a) Video-Language Understanding: The fluffy white terrier is walking across the grass.

(b) Local Block Similarity Comparison

(c) Vector Motion-Aware Match

(d) Streaming Multilevel Concentration Architecture