UE8M0-TPU Architecture v0.2 Addendum

# 1) Architecture 수정분 (v0.2)

• 메인 클럭 정책: DDR SKU는 0.50 GHz 고정, HBM SKU는 0.50~0.55 GHz sweet spot.  
• DDR SKU는 Front-side Cache 증설(채널당 16–32 MB + 공유 eDRAM 256–512 MB)로 유효 대역폭 확장.  
• DVFS는 단순화: DDR=0.50 GHz 유지, 필요시 0.60/0.70/0.80 GHz 단계 상승. HBM=0.50 GHz 기본, 연산 부족시 0.55 GHz.  
• Tensor Core Tiles, Codec Cluster, Scratchpad 구조는 v0.1과 동일.

# 2) 성능 스윗스팟 요약

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SKU | 클럭(GHz) | lanes | I=256 Realized PF/s | Perf/W (TF/W, lin) |
| V1-D\_DDR cf=1.0 | 0.5 | 1 | 0.115 | 0.51 |
| V1-D\_DDR cf=1.0 | 0.5 | 2 | 0.115 | 0.51 |
| V1-D\_DDR cf=2.0 | 0.5 | 1 | 0.231 | 1.01 |
| V1-D\_DDR cf=2.0 | 0.5 | 2 | 0.231 | 1.01 |
| V1-H\_HBM | 0.5 | 1 | 0.944 | 2.18 |
| V1-H\_HBM | 0.5 | 2 | 1.088 | 2.51 |
| V1-H\_HBM | 0.55 | 1 | 1.038 | 2.28 |
| V1-H\_HBM | 0.55 | 2 | 1.088 | 2.39 |

# 3) 그래프 삽입

아래 그래프는 I=256 기준에서 클럭 대비 실효 성능/효율의 sweet spot을 보여줍니다.

(그래프 파일을 찾을 수 없음)

# 4) 요약

• DDR SKU: 0.50 GHz 고정, cache\_factor 2.0 이상이면 성능이 배 이상 개선.  
• HBM SKU: 0.50~0.55 GHz가 sweet spot, lanes=1일 때 0.55 GHz에서 roofline 도달.  
• 전반적으로 언더클럭 + 캐시 강화 전략이 Perf/W 최적화를 달성.