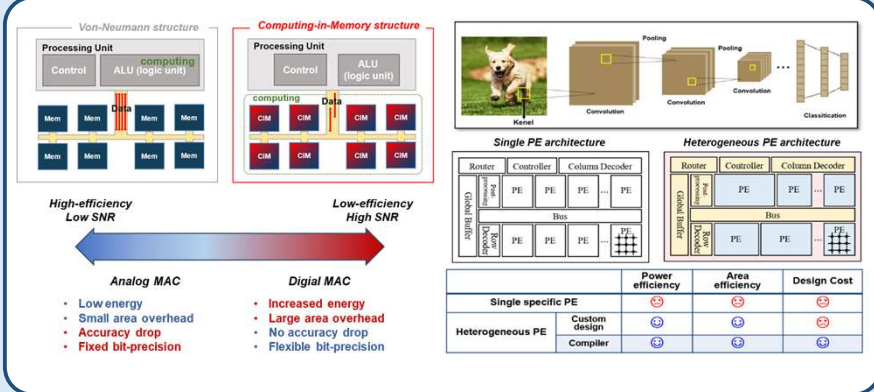


Reconfigurable Hybrid SRAM-CIM for Heterogeneous SRAM-CIM SoC to Achieve High Utilization and Performance

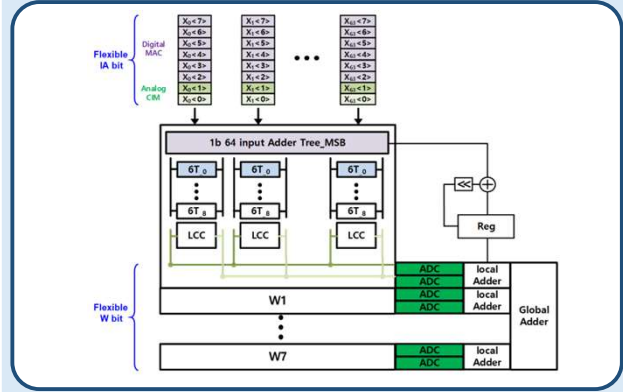
Sekeon Kim and Seong-Ook Jung
Yonsei University, Korea

1. Introduction



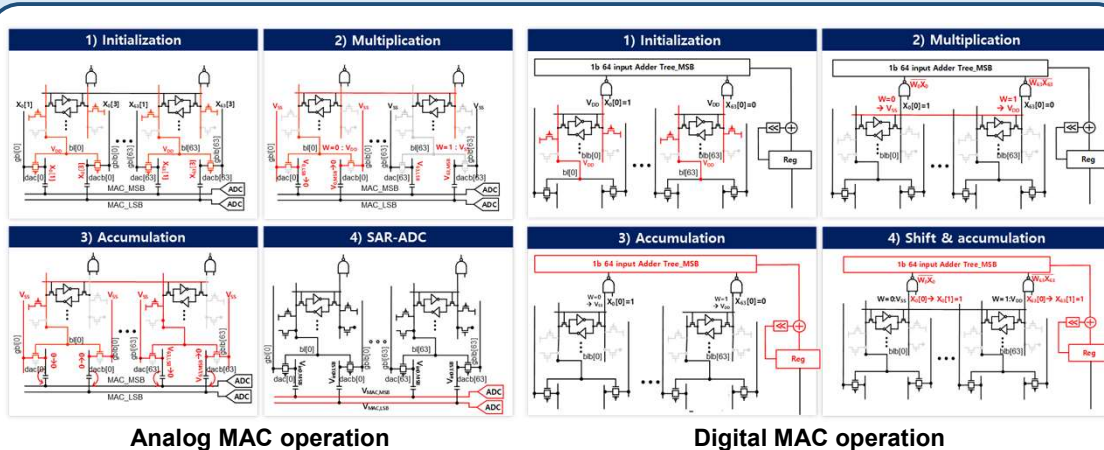
- Computing-in-Memory (CIM)은 메모리 내부에서 병렬 연산을 수행하여 Von-Neumann 구조의 병목현상을 제거하고 에너지 소비를 줄임.
- DNN 모델의 다양한 레이어 크기와 비트 정밀도 요구사항을 맞추기 위해 레이어 별 최적화를 통해 설계 비용을 줄이고 유연성을 높이기 위해 재구성 가능한 SRAM-CIM 매크로를 설계함.
- 또한, 높은 에너지 효율성을 가지는 Analog Multiply-and-Accumulate (MAC) 동작과 낮은 에너지 효율성을 가지나 높은 재구성 가능성을 가지는 Digital MAC 연산을 동시에 수행하는 Hybrid SRAM-CIM을 개발함.

2. Proposed Hybrid SRAM-CIM



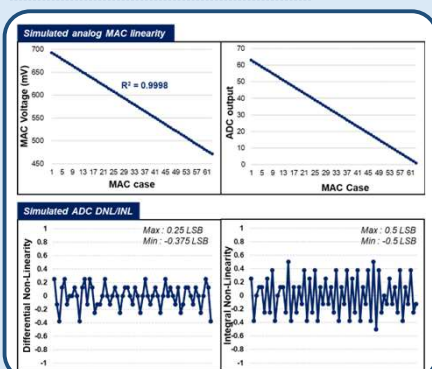
- 6T SRAM 셀을 기반으로하며, 9개의 셀이 local computing cell (LCC)를 통해 그룹화 되어 analog 및 digital MAC 연산 동시 수행.
- MSB 부분은 digital 방식으로, LSB 부분은 charge-coupling 기반 analog 방식으로 MAC 연산을 수행.
- Weight/input bit-precision은 각각 global adder에서 그룹화 되는 LCC 개수와 digital MAC 연산의 accumulation cycle 개수를 통해 조절 가능.

3. Proposed Hybrid SRAM-CIM Operation



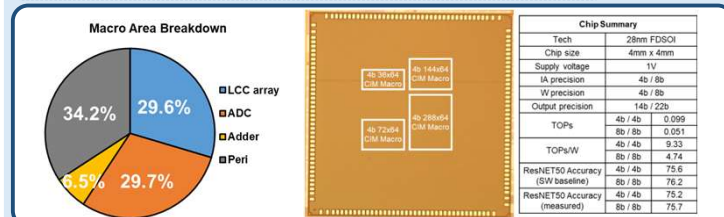
- **Analog MAC 연산:**
 - 1) 내부 비트라인(bl, blb)을 VDD로 충전.
 - 2) 입력 활성화(IA)와 가중치(W)를 곱하여 결과를 DAC 노드에 저장.
 - 3) Capacitor를 활용한 charge-coupling을 통해 연산 수행.
 - 4) 아날로그 MAC 전압을 SAR-ADC를 통해 디지털 값 변환.
- **Digital MAC 연산:**
 - 1) 내부 비트라인을 VDD로 충전.
 - 2) NAND 게이트를 활용하여 Inout과 weight 곱셈 수행.
 - 3) 64개 열의 결과를 "1b 64-input adder"를 통해 누적.
 - 4) 결과를 시프트하여 이전 값과 합산한 후 다음 bit 연산 수행.

4. Simulation Results



- PVT variation에 영향에 취약한 analog MAC 연산에 대한 검증 진행.
- Analog MAC 전압 결과는 $R^2 = 0.9998$ 의 linearity를 보여줌. 6b SAR-ADC에서 analog MAC 전압이 digital code로 변환됨.
- ADC Simulation(1LSB = 3.5mV)
 - DNL: 개별 코드 간 오차는 최대 0.25LSB 최소 -0.375 LSB
 - INL: 전체 구간 누적 오차는 최대 0.5LSB, 최소 -0.5LSB

5. Chip Measurement Results & Conclusion



- 제안한 Hybrid SRAM-CIM 구조는 analog MAC 효율성과 digital MAC 유연성을 결합해 에너지 효율성을 가져가며 다양한 비트 정밀도를 지원하여 Reconfigurability 확보.
- 28nm FDSOI 공정으로 구현된 테스트 칩에서 ResNet-50 기준 0.4~0.5%의 정확도 하락을 보임.