

A Scalable, Configurable Neural Network Accelerator Supporting On-device Training

Changsoo Kim, Jaehwa Kwak, and Jongtae Lee

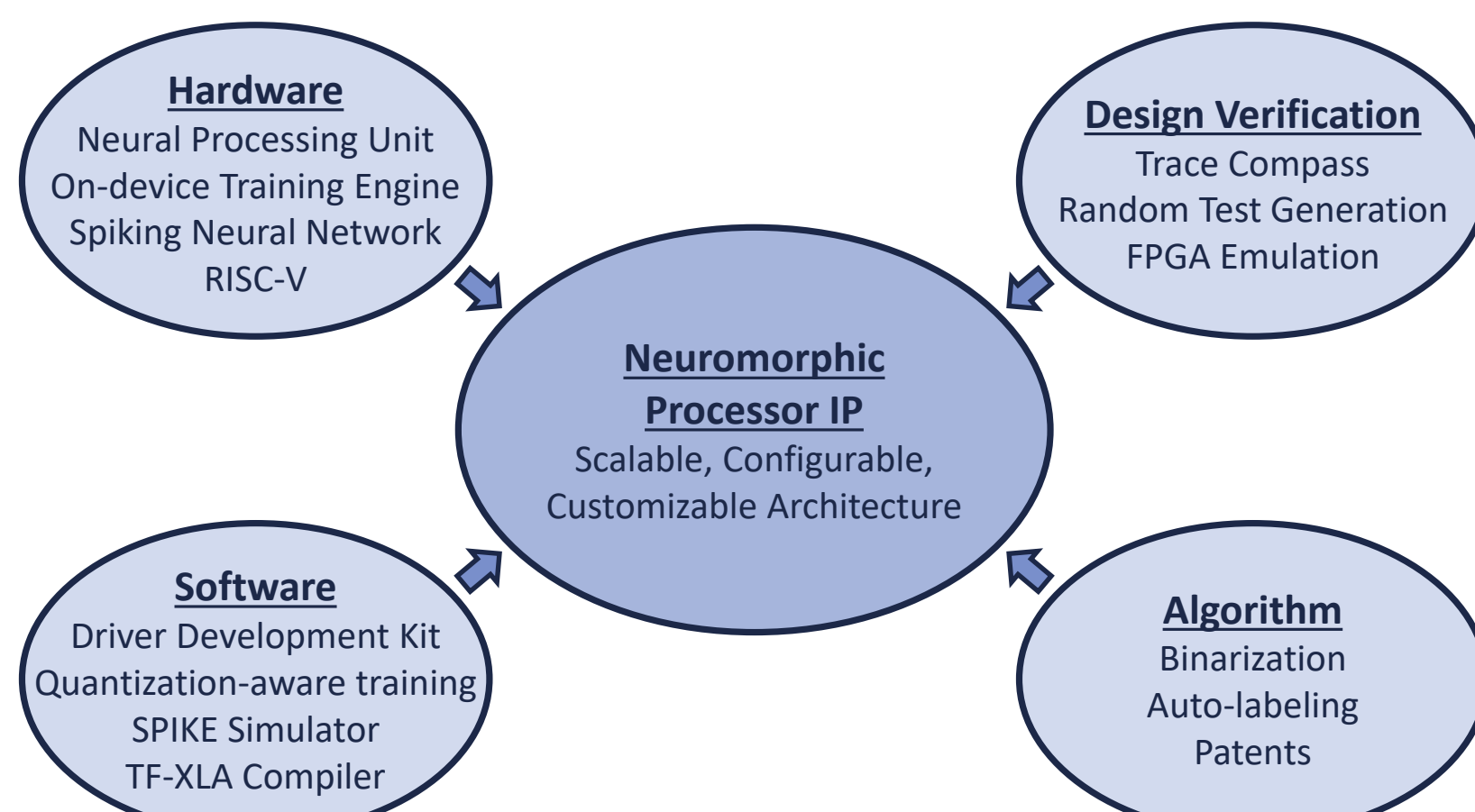
AiM Future

회사 소개

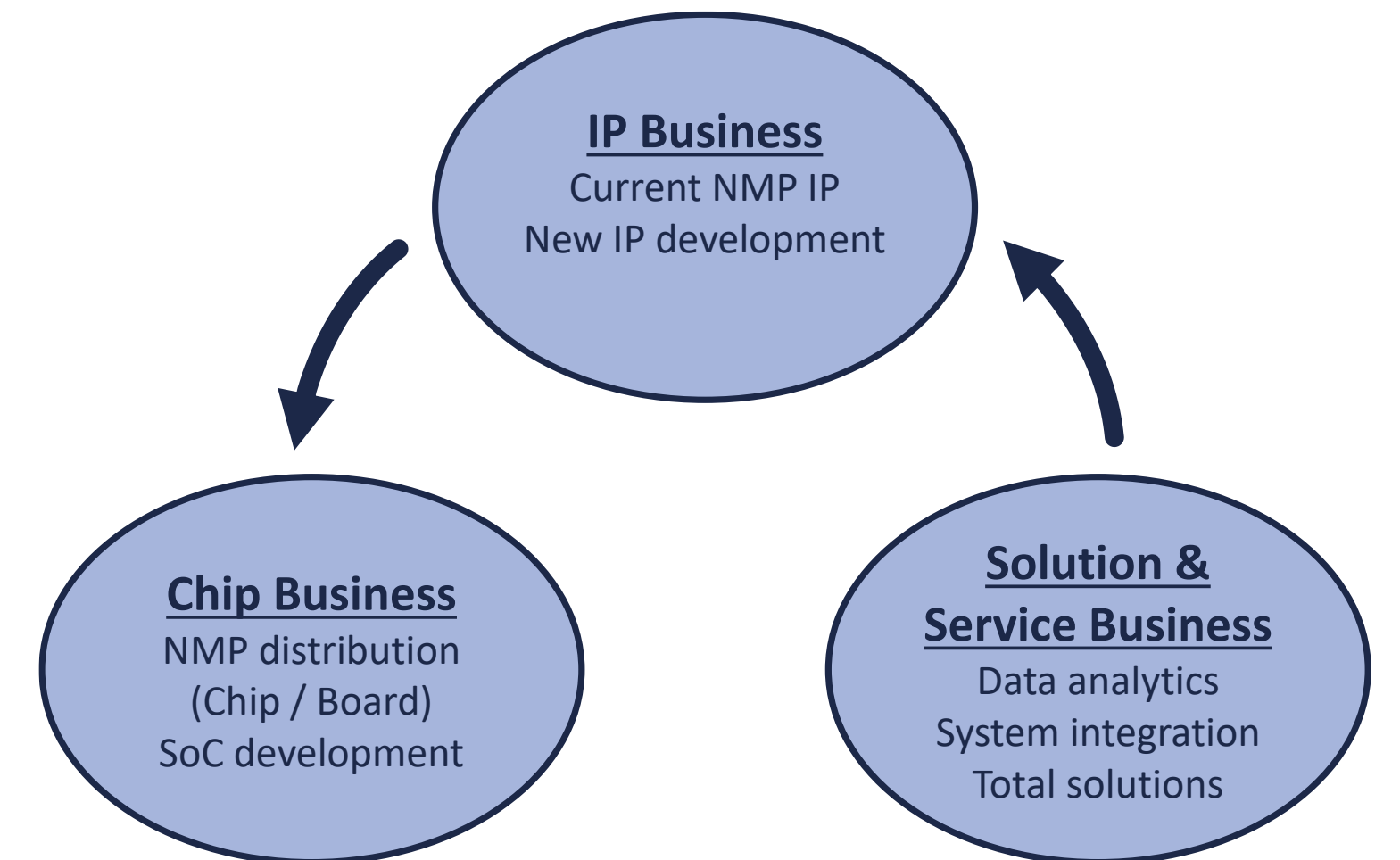
(주)에임퓨처(AiM Future)는 지난 5년여간 실리콘 벨리에서 인공지능용 IP를 개발해 온 팀이 한국에 설립한 회사입니다. 글로벌 비즈니스를 위해 미국 현지에 SilicoNeuro, Inc 라는 법인을 별도로 두고 있습니다. VC로부터 현금 출자를 통해 투자가 완료되면 벤처 기업 인증을 통해 한국과 미국에서 영업 활동을 전개해 나갈 예정입니다.

보유 기술

- ▶ 칩으로 검증된 Neural Engine 아키텍처 보유
- ▶ NPU 하드웨어 설계 및 최적화 기술 보유
- ▶ Neuromorphic Processor (NMP)와 연동된 DDK 기술 보유
- ▶ 별도의 TF-XLA Compiler 기술 보유
- ▶ 멀티레벨 디자인 검증 기술 보유
- ▶ 인공지능 알고리즘 개발 및 최적화 기술 보유
- ▶ 기계 학습 관련 특허 보유



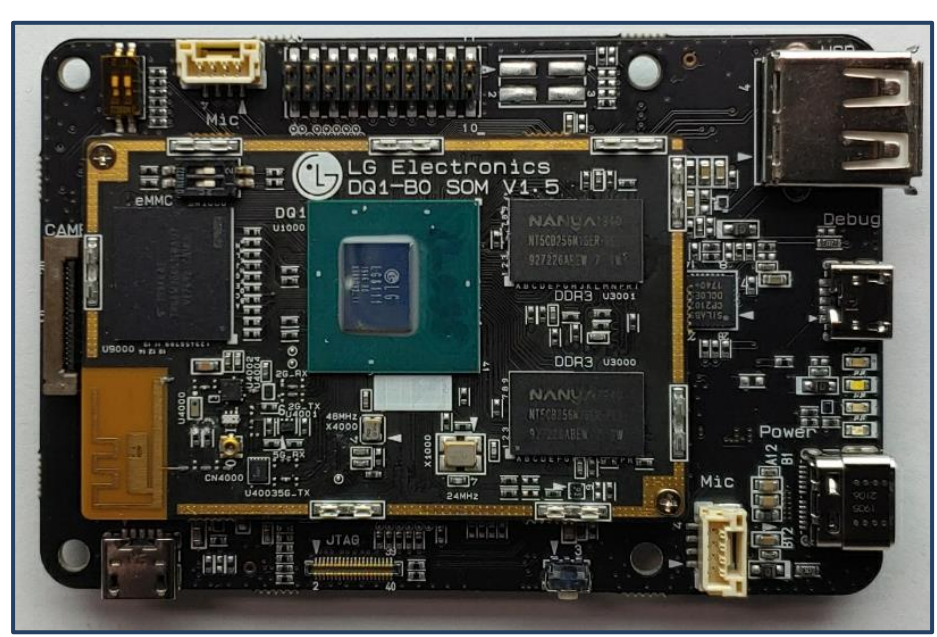
<기술 개요>



<사업 모델>

제품 및 서비스 소개

- ▶ NPU 기반으로 하는 다양한 성능의 NMP HW/SW IP
- ▶ NMP가 내장된 가전용 AI칩을 장착한 다양한 보드
- ▶ NMP IP를 활용한 가전용 AI칩과 TV용 SoC이 개발되었으며 다양한 제품에 적용하여 제품화 될 예정입니다



AI 칩이 장착된 보드



세탁기



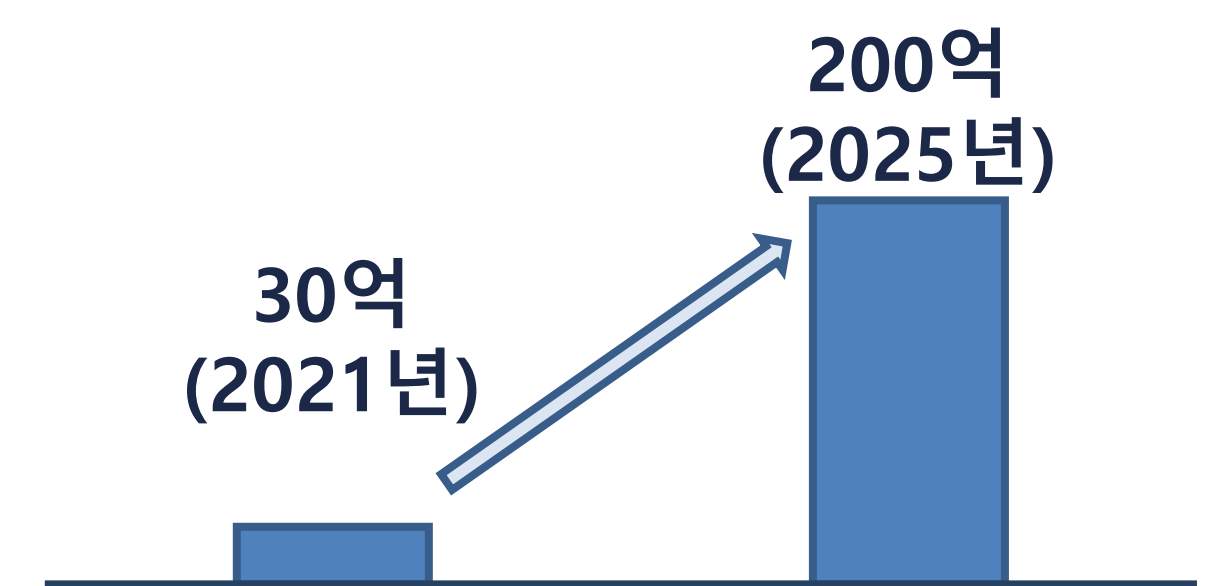
로봇 청소기



스마트TV

목표 시장 및 규모

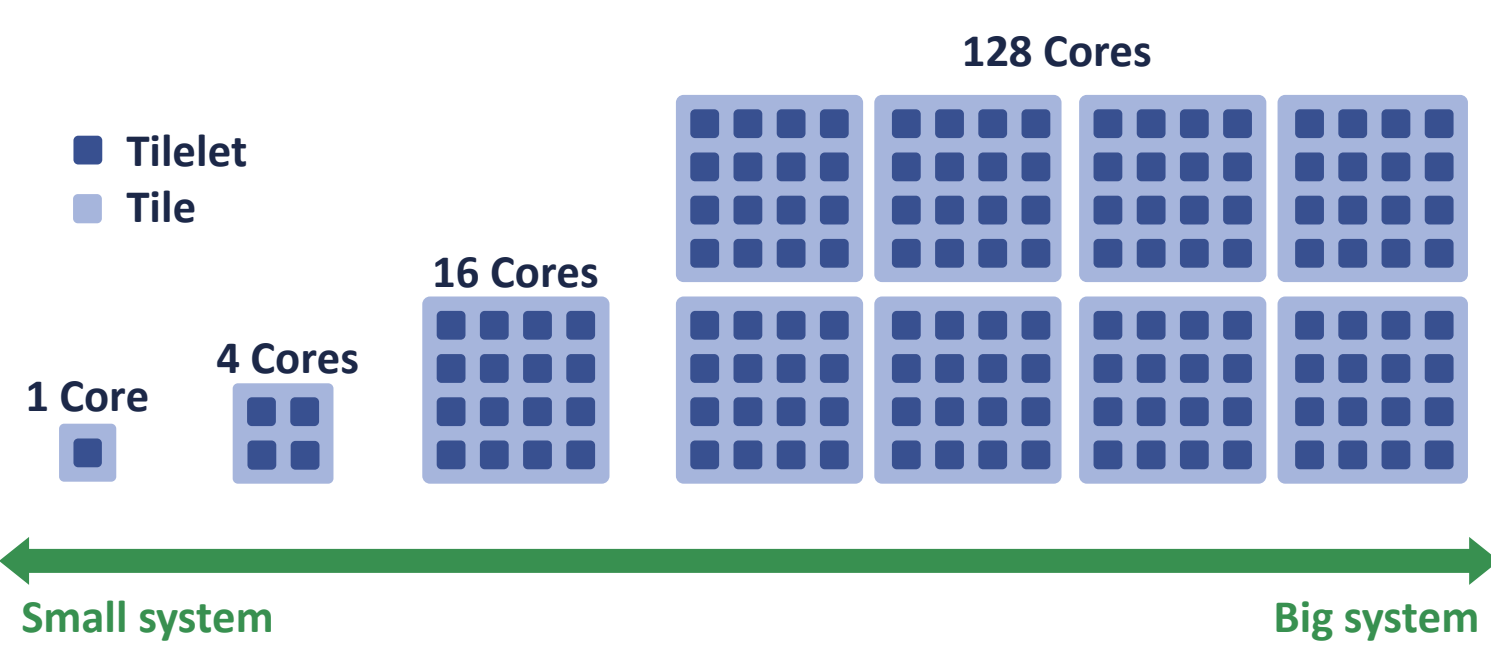
- ▶ 한국/미국/유럽의 자율주행 자동차용 및 가전용 반도체 회사가 주 대상 고객사임
- ▶ IP의 성능에 따라 30~60만불 상당의 License 수입을 통해 3~5년내 매출액 200억원 달성이 목표



하드웨어 아키텍처

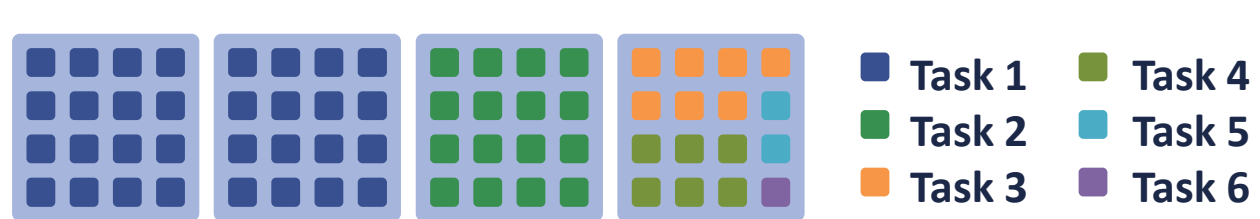
Scalability

- ▶ Tile - Tilelet 으로 명명된 두 단계의 계층 구조.
- ▶ 하나의 시스템에는 1-8개의 Tile이 존재함.
- ▶ 하나의 Tile에는 1-16개의 Tilelet이 존재함.
- ▶ 응용 분야에 따라 손쉽게 시스템을 구성할 수 있음.



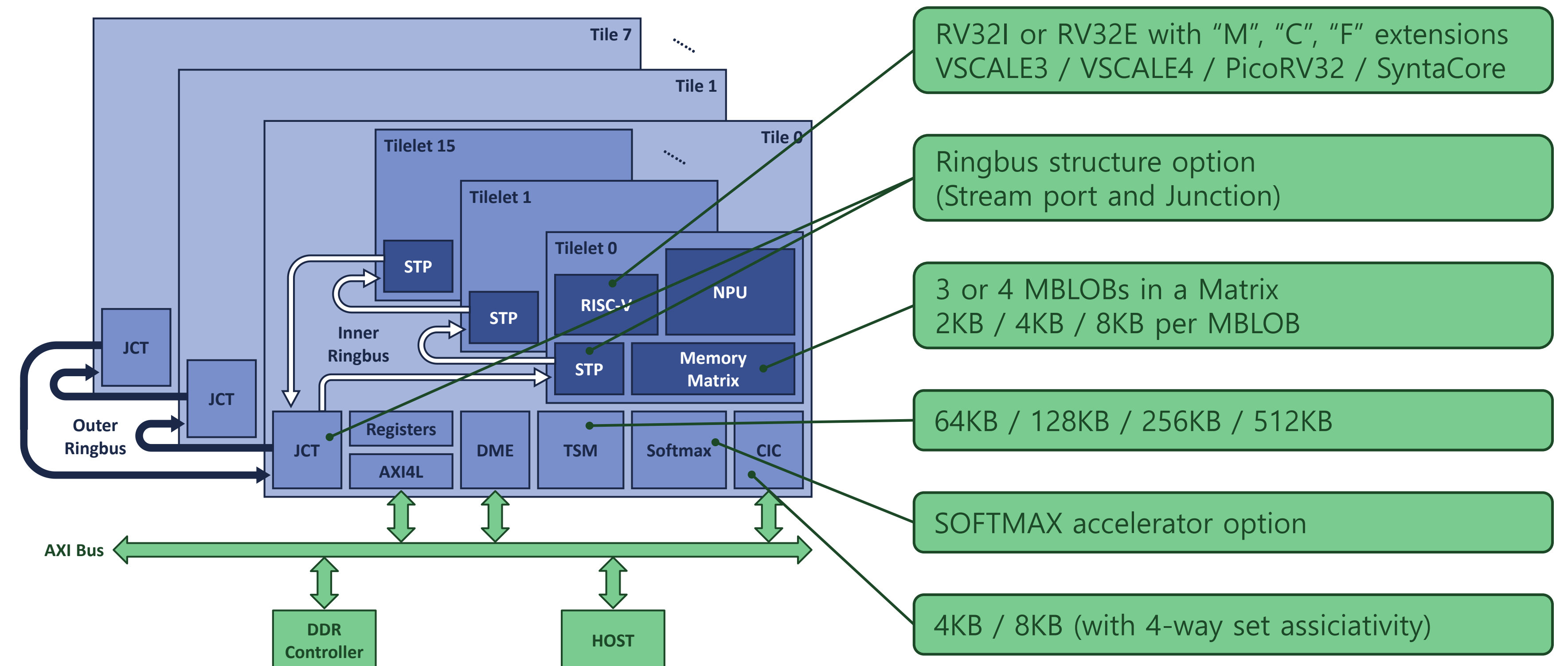
Flexibility

- ▶ 각각의 Tilelet 은 독립적으로 작업을 처리할 수 있음.
- ▶ 실시간으로 작업의 크기 등을 분석하여 연산 그룹을 할당함.



Configurability

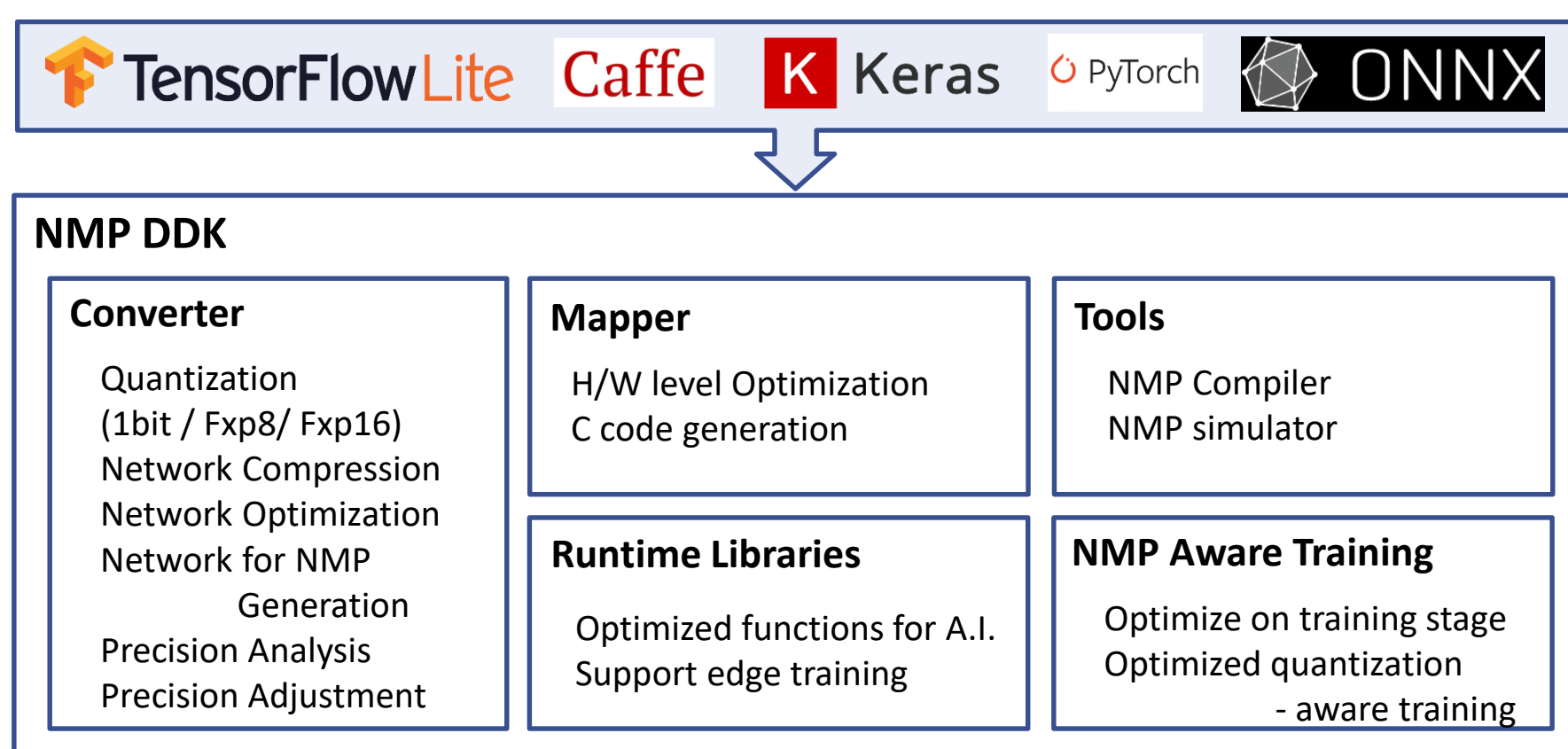
- ▶ 응용 분야에 따라 각각의 구성 요소를 다양하게 설정할 수 있음.



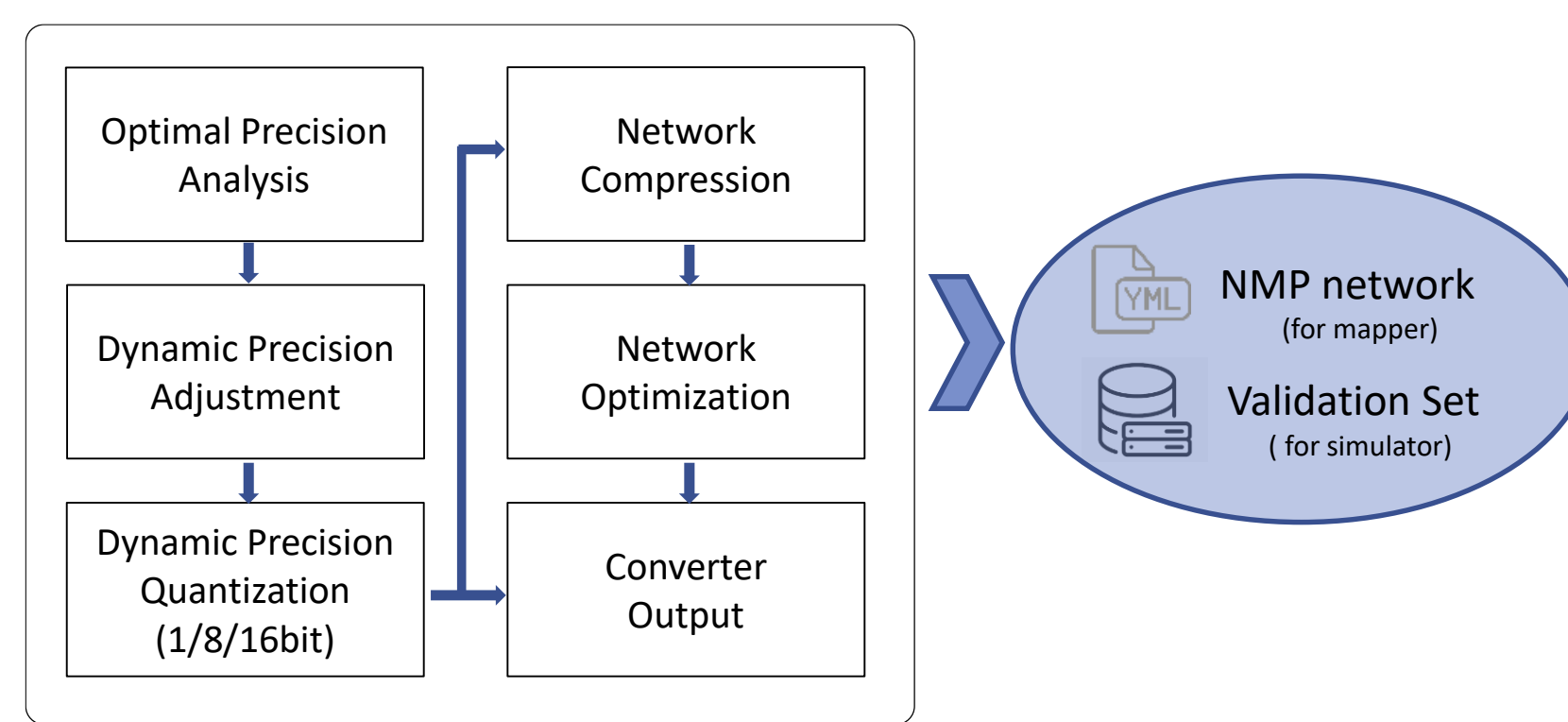
소프트웨어 솔루션

NMP DDK

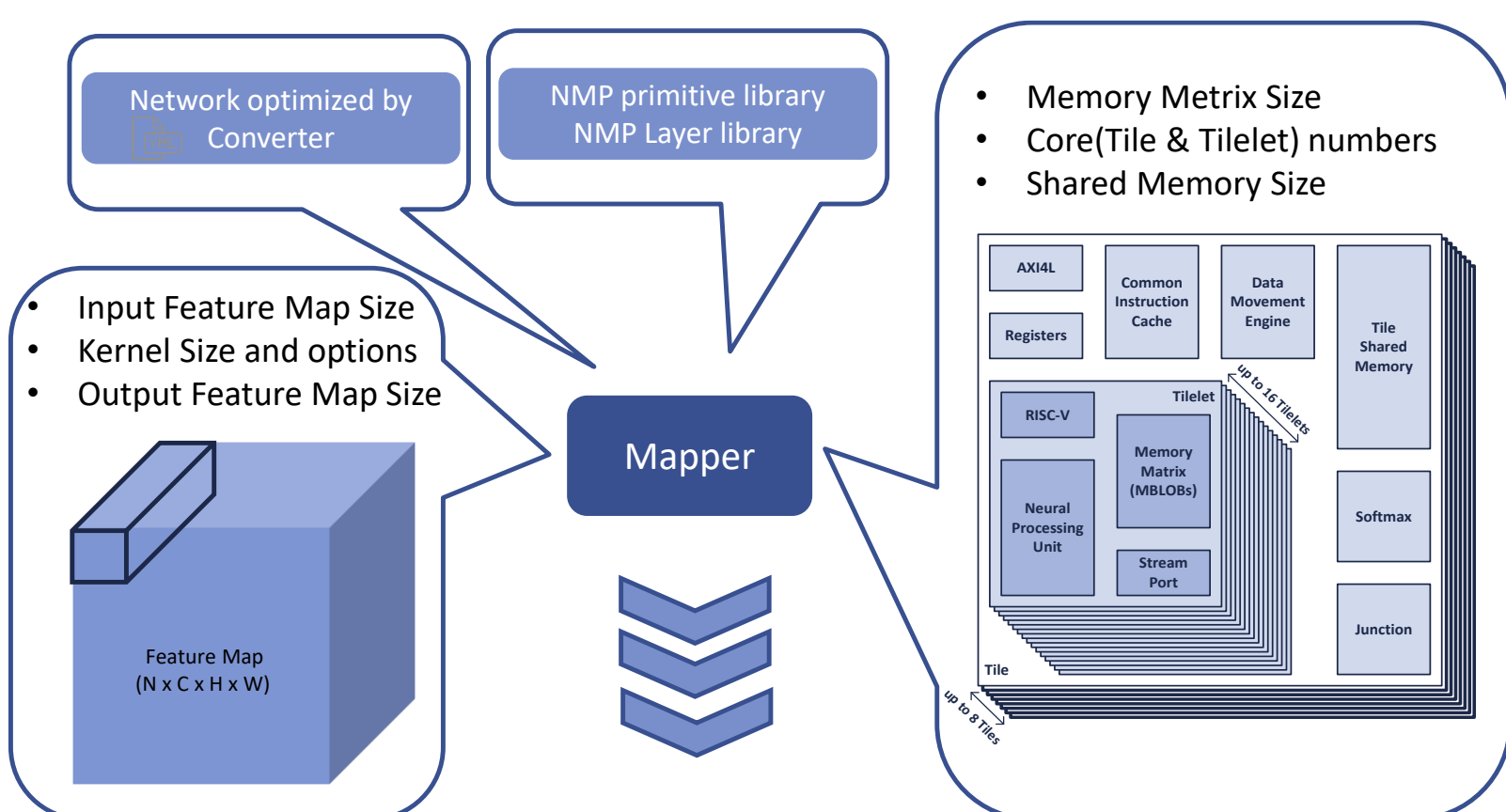
- ▶ Caffe, TensorFlow 등 프레임워크로부터 생성된 인공지능 모델을 NMP 기반의 바이너리 파일로 변환하여 커스텀 보드 및 FPGA 보드에 적용하는데 필요한 라이브러리와 컴파일러 등 일체의 개발 환경을 제공.
- ▶ 이미 학습된 모델을 추론용으로 사용할 수도 있으며, 기초적인 학습만 끝난 모델을 NMP에서 On-Device 학습을 하여 추론의 성능을 높일 수 있도록 사용할 수도 있음.



<NMP DDK 개요>



<NMP DDK Converter>



<NMP DDK Mapper>

NMP TF-XLA

- ▶ TF-XLA를 NMP에 적용하여 인공지능 모델을 보다 쉽고 빠르게 NMP에 적용할 수 있는 컴파일 기법.

