

2020년 대표성과 후보 추천 요약서

[1] 성과요약서

성과유형	기초·미래선도형 (<input checked="" type="radio"/>) 산업화형 (<input type="radio"/>) 국가·사회문제해결형 (<input type="radio"/>)			
세부과제명 (세부과제번호)	인공지능 프로세서 전문연구실 (1711102885)			
성과목표	인공신경망을 위한 병렬컴퓨팅 연산 능력과 전력효율성을 갖춘 뉴로모픽 프로세서 및 뉴로모픽 컴퓨팅 시작품 개발			
총연구기간	2018년 2월 ~ 2027년 12월			
총연구비	총 20,827백만원 (현재 진행 중)		정부: 19,777백만원 민간: 1,050백만원	
성과책임자 정보	연구자 성명	직할부서	연구본부/연구실	직위/직급
	권영수	인공지능연구소	지능형반도체연구본부	본부장/책임
성과정보				
성과내용	<p>세계 최고수준의 고효율 인공지능프로세서 반도체 칩 AB9 개발 및 AB9 기반의 인공지능 연산 가속 시제품 보드(ABrain-S) 개발</p> <ul style="list-style-type: none"> ○ 세계 최고 수준인 1TFLOPS/Watt 전력효율을 가지는 40TFLOPS, 15Watt 급 고성능·저전력의 인공지능 프로세서 AB9 반도체 칩 제작 ○ 인공지능 서비스 개발을 위한 AB9 컴파일러, 시뮬레이터, API, 다바이스 드라이버를 포함하는 인공지능 최적화 SW 개발 도구(AI-Ware) 개발 ○ AB9 칩 기반의 인공지능 서버 탑재용 인공지능 연산 가속 보드(ABrain-S) 모듈 개발 			
대표성과1	<ul style="list-style-type: none"> • 40TFLOPS, 15Watt 급 인공지능 프로세서 반도체 칩 및 인공지능 연산 가속 프로세서 모듈 <div style="display: flex; justify-content: space-around; align-items: center;">    </div>			
대표성과2	<ul style="list-style-type: none"> • 논문: Automated optimization for memory-efficient high-performance deep neural network accelerator, HyunMi Kim, ETRI Journal, Vol.42, No.4, IF 1.094, 2020년8월 • 논문: 40-TFLOPS Artificial Intelligence Processor with Function-safe Programmable Many-cores for ISO26262 ASIL-D, Jinho Han, ETRI Journal, Vol.42, No.4, IF 1.094, 2020년8월 • 논문: AB9: A Neural Processor for Inference Acceleration, Yong Cheol Peter Cho, Vol.42, No.4, IF 1.094, 2020년8월 • 기술이전: 인공지능프로세서 고속 메모리 인터페이스 및 전력온도 제어 기술, 2.2억원, 2020년12월 			
<p>- 대표성과 선정 시 대상선정을 위해 직원투표(홈페이지 게시)에 동의합니다. 동의 <input checked="" type="checkbox"/></p> <p>※ 공개 항목 : 연구자, 소속, 직위, 성과명, 연구비, 성과내용 등 후보추천서 기재사항</p> <p>- 사실과 다른 내용이 포함되거나 중대한 오류가 발견된 경우에는 선정이 취소될 수 있음을 확인합니다.</p> <p style="text-align: right;">2020년 12 월 16 일</p> <p style="text-align: right;">확인자 권 영 수 (인)</p>				

2020년 대표성과 후보 추천서

[2] 2020년 우수성과 내용

1. 성과명

스스로 인공지능 학습/추론을 가능하게 하는 초당 40조개의 연산이 가능한 고성능·고효율 인공지능 프로세서 반도체

2. 성과내용

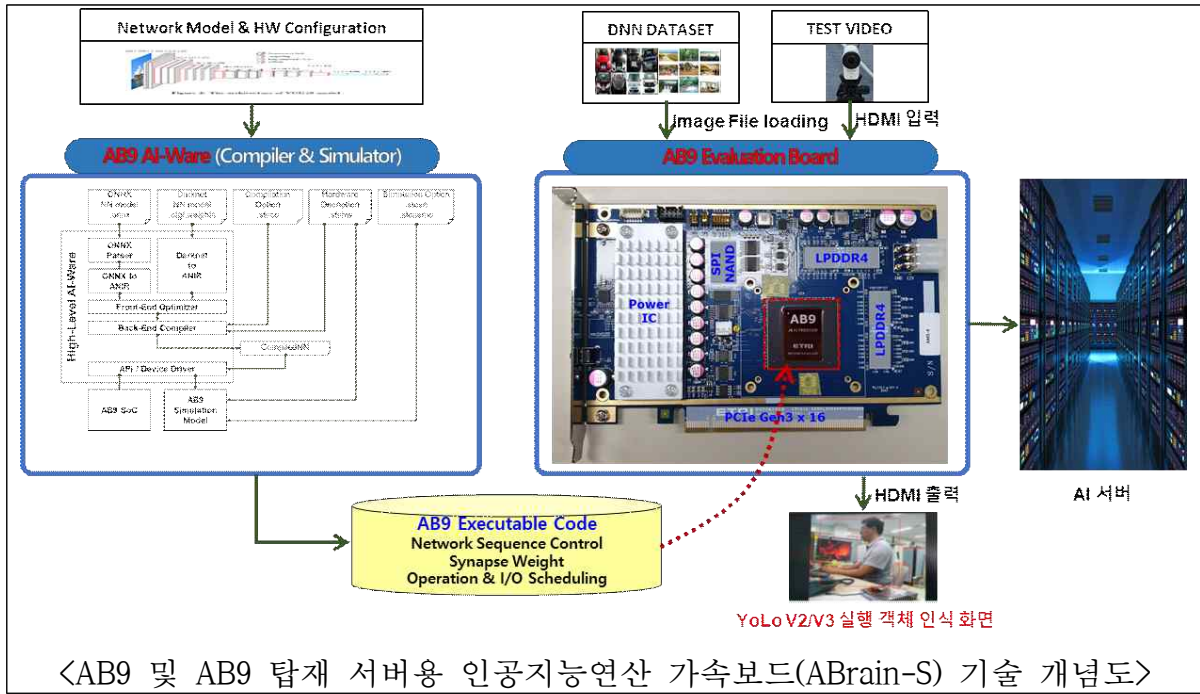
기술개발 목표달성도

기술적 선점이 필요한 분야

- 인공지능 반도체는 4차 산업 인공지능 기반 응용 서비스에 필요한 딥러닝 연산을 최적화하여 고성능, 높은 전력효율로 실행하는 반도체로 기존의 CPU, GPU를 대체하여 사용되고 있는 혁신제품임
- 인공지능 반도체는 인간의 뉴런처럼 연산을 처리하는 PE(Processing Element)를 수천~수만개의 병렬처리 네트워크로 구성한 NPU(Neural Processing Unit) 뿐만 아니라 CPU core, LPDDR4/GDDR5와 같은 고대역폭의 메모리 인터페이스, PCIe Gen3와 같은 고속 외부 장치 인터페이스, 온칩 네트워크, 캐쉬, 전력 제어기 등 최신 SoC 기술이 집약된 고난이도의 최첨단 반도체 기술로서 진입 장벽이 높은 고부가 제품임
- 전세계 인공지능 프로세서의 기술적 중요성 및 시장성이 본격적으로 부각되고 있고, 향후 전세계 반도체 시장을 이끌어 갈 주도적 아이템이 되고 있는 반면, 국내 기술은 매우 부진하므로 ETRI가 선도적으로 기술개발 추진
- 전세계 인공지능 반도체 시장은 2016년 이후부터 본격적으로 (딥러닝은 2013년 이후) 주목을 받고 있으며, 반도체의 다양한 분야 중 가장 성장성(CAGR 5% 이상)이 높은 기술로 평가(출처:Gartner, '20)

기술개발 목표

- 인공지능망을 위한 병렬컴퓨팅 연산 능력과 전력효율성을 갖춘 뉴로모픽 프로세서 및 뉴로모픽 컴퓨팅 시작품 개발
(목표 ①) 40TFLOPS, 15W급 세계 최고수준의 고효율 인공지능프로세서 AB9 개발
(목표 ②) AB9 기반 인공지능 연산 가속 시제품 보드 모듈 ABrain-S 개발



□ 기술개발 목표의 달성성과 및 핵심기술 확보

[개발목표 ①]

- ➔ (달성성과) 국내최초, 세계최고 수준 40TFLOPS, 15Watt 급 고효율 인공지능프로세서 AB9 개발
- ➔ (핵심기술 확보)
 - 인공지능 서비스 가속을 위한 세계 최고 수준의 초저전력 인공지능 반도체 제작을 통해 머신러닝을 위한 기존 AP대비 1,000배 효율의 신개념 컴퓨팅 기술 확보
 - 초소형 16bit 부동소수점 연산기 기반으로 확장가능한 인공지능 반도체 아키텍처 기술 확보
 - 인공지능 연산 성능 극대화 및 전력 소모 최적화 기술 확보
 - 범용 인공지능 서비스 개발을 위한 AB9 NPU 연산 최적화 SW 개발 환경 (AI-Ware) 개발
 - 인공지능 프로세서 명령어의 효율적 스케줄링이 가능한 인공지능 프로세서 시뮬레이터 개발

[개발목표 ②]

- ➔ (달성성과) 서버기반 인공지능 제품/서비스 적용을 위한 시제품 보드 모듈 ABrain-S 개발
- ➔ (핵심기술 확보)

- AB9 Single- Chip, 고속 저전력 LPDDR4(64bit) Dual Channel(332.8Gbps), PCIe Gen3 x 16(128Gbps) 지원
- 서버 탑재를 위한 PCIe 카드 규격의 form-factor 및 Dual PCIe 전원(12V) 입력 지원
- AB9 core 및 LPDDR4 메모리용 1.0V, 1.1V, 1.8V 안정적인 전원 공급
- 최대 2666MHz 고속 동작을 위한 임피던스 매칭을 통한 신호 및 전원부 총 16 레이어 PCB 설계
- AI SW 개발을 위한 다양한 장치 및 인터페이스 지원: 부팅용 플래시 메모리, SD카드, HDMI, JTAG, UART, PMBUS 등

3. 우수성 및 차별성

기술수준 향상 성과

- 인공지능 딥러닝 연산에 최적화된 세계 최고 수준의 국내 최초 인공지능 프로세서 반도체(AB9)
 - 인공지능 알고리즘 최대 연산 성능은 40TFLOPS이고, 전력소모는 15~40 Watt로서 서버용 고성능 제품군 중 세계 최고 수준인 1TFLOPS/Watt 전력효율을 가짐
 - 딥러닝 연산에 최적화된 최대 1.25GHz로 동작하는 128x128개의 16비트 부동소수점(FP16) MAC 연산용 나노코어 시스톨릭 어레이 구조
 - 딥러닝 연산에 최적화된 신경망처리장치(NPU) 아키텍처 및 소모전력 최적화를 위한 누설전력 차단 PG(Power Gating) 기능 지원
 - 데이터 이동 병목현상 최소화 및 효율적인 스케줄링 구조 적용

< AB9 구현 사양 >


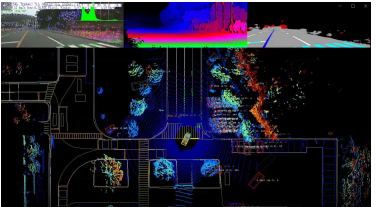
Items	Specification
Max. Performance	40TFLOPS(@FP16)
Power	15~40Watts
Power Efficiency	1TFLOPS/Watt (세계최고 수준)
Max. Core CLK Freq.	~1.25GHz
# of NP Unit	32,768 (16,384 cores)
Size	1-Billion Transistors@28nm
Die Area	17mm x 23mm, 391mm ²
Arithmetic	Half-Precision(16bit) Floating-Point (IEEE-754)
Function	General Object Recognition and Localization
I/F	LPDDR4 2CH, PCIe Gen3 x16, etc.

- NVIDIA의 서버용 고성능 GPU V100과 비교하여 연산성능은 127%, 전력효율은 1000% 우수한 성능 확보

< AB9 과 NVIDIA V100과의 성능 비교 >

성능지표	세계 최고 수준		AB9 기술 수준	
	기관명	기술수준	기술수준	세계최고대비(%)
연산성능	NVIDIA/미국	31.4 TFLOPS (V100,FP16)	40 TFLOPS (AB9,FP16)	127%
전력효율	NVIDIA/미국	0.1 TFLOPS/Watt (V100)	1.0 TFLOPS/Watt (AB9)	1000%

- 인공지능 서버 탑재를 위한 LPDDR4 및 범용 고속 PCIe 인터페이스 기반 고성능 인공지능 가속 모듈(ABrain-S)
 - 현재 개발완료한 인공지능 프로세서(AB9)의 국내 기술 확산 및 상용화를 위해 국내 인공지능 서비스 기업(KT, NBP, SKT 등) 및 연구기관(KARI, ADD 등)과의 시제품 개발을 위한 PoC 진행 중
 - . (KT 협업) 서버기반 영상인식 기술을 구현한 ABrain-S 장착 서버로 PoC 진행 중
 - . (NBP/아토리서치 협업) 안면인식 및 ID추출 기술을 구현한 ABrain-S 장착 서버 기반의 인공지능 서비스 구현 중
 - . (SKT 협업) 음성인식 서비스에 ABrain-S를 적용하기 위한 PoC 진행 중
 - . (ADD 협업) 감시정찰 시스템에 ABrain-S를 적용하여 알고리즘 검증 진행 중
 - . (AI+X, NIPA) 광주집적센터와 연계하여 AI+X 사업에 ABrain-S 기술 적용 진행
 - . (KARI, 항공우주연구소 협업) 고공정찰 군집드론에 ABrain-S 적용 시스템 구성 협의 중

		
<p align="center">< KT와 공동개발 ></p>	<p align="center">< SKT와 공동개발 ></p>	<p align="center">< 모빌리티 검증 환경 > *ETRI 자율주행지능연구실 협업</p>
<p>고해상도 항공영상을 획득하여 실시간 객체 탐지*, 판단 * 1024x768 VisDrone DataSet, 10 fps</p> 	<p>AI 프로세서 (Low-Power Processing Mode)</p> 	
<p align="center">< ADD, 항공우주연구원과 공동개발중인 이동체 AI 모듈 ></p>	<p align="center">< NBP와 공동개발 ></p>	

기술수준 공인 성과

- AB9 인공지능 프로세서 기술이 과학기술정보통신부 “2020년도 국가연구개발우수 성과 100선” 에 선정(‘20.10)
- ABrain-S 모듈은 KES 2020 전시회에서 “KES Innovation Award 2020” 에 선정 (‘ 20.12)
- 각종 미디어(뉴스, 신문 및 매거진)를 통한 기술 홍보 및 ‘ 인공지능반도체 ‘ 다큐멘터리에서 인공지능 반도체 대표기술로 소개될 예정(TJB, ‘ 20.12)



< 각종미디어를 통한 홍보 >



< 국가연구개발우수성과 100선에 선정 >



< KES Innovation Award 2020 >

4. 성과의 활용도 및 파급효과

경제 활성화 효과

기업 경쟁력 향상

- 인공지능 프로세서 설계 기술을 국내 팹리스 반도체에 업체에 기술이전 및 공개를 통하여 팹리스 업체의 경쟁력 강화를 위한 돌파구를 제공하고, 시작품인 인공지능 반도체를 제공하여 중견, 중소기업과 연계 모듈형태의 인공지능 시스템 개발 추진

산업 경쟁력 향상

- 인공지능 프로세서 기술은 개인화된 지능형 디바이스(모바일 단말기, 지능형 스마트폰), 무인 자율이동체(자율주행 자동차, 자율이동형 드론, 무인 화물수송), 고성능 서버, 인공지능 헬스케어(원격 진료, 인공지능 의료용 영상 분석, 웨어러블 헬스케어 기기), 국가 안보용 기술(무인비행체, 탐지형 로봇), 사회 서비스(금융 서비스, 범죄 감시용 안면/행동 인식) 등의 분야에 적용할 수 있는 국내외 반도체 업계에서 필요로 하는 원천기술
- 정부주도의 인공지능 서비스 기반 기술지원을 통해 인공지능 서비스 산업 에코 시스템 구축

경제적 파급효과

- (파급효과 전망)

- 인공지능 프로세서 반도체는 인공지능 SW와 더불어 인공지능 서비스 창출을 위한 핵심 반도체 부품으로서 해외 글로벌 기업(Nvidia, 구글)에서 선점하고 있는 제품의 국산화를 통해 수십조 이상의 시장 창출 가능
- 지능정보처리기술 서비스의 국산화 및 국내 기술을 통한 시장 창출

국가·사회적 파급효과

○ 해결해야 할 국가·사회문제

- 인공지능 기반 서비스 및 사회 인프라(자율주행 자동차, 도심항공교통(UAM), 헬스케어 등) 구축을 위한 국내 자체 개발 인공지능 프로세서 기술 확보

○ 성과에서 개발된 기술적 솔루션

- 인공지능 반도체 및 인공지능 프로세서 기술에서 국내 최고의 기술을 확보함은 물론, 실시간 학습과 추론이 가능한 ‘인공전자두뇌’를 위한 핵심기술 축적을 통하여 인공지능 반도체 세계 최고 기술 확보
- 자체보유한 AP 및 나노코어 뉴럴넷 가속설계를 바탕으로 인공지능 및 딥러닝 인공지능 프로세서 기반 컴퓨팅 기술을 수요기업(SKT, Naver)과 협력을 통해 상용화 추진

○ 국가·사회적 파급효과

- 해외 글로벌 기업에서 선점하고 있는 인공지능 프로세서 기술의 국내 독자 기술 확보를 통한 기술 의존성 감축 및 정보유출 방지
- 인간수준 또는 인간보다 우수한 성능의 지능을 반도체와 임베디드 SW 융합 기술로 구현하여 모든 사물에 지능을 부여함으로써 제4차 산업혁명을 가속화하는 핵심기술 확보