Korea Supercomputing Conference 2026

KSC2026

2026년 9월 8일(화)~9일(수) | 스위스 그랜드호텔 컨벤션홀(4F)

The Intelligent Frontier: Synergizing HPC, AI and Quantum

ABSTRACT

주최 KISTI 한국과학기술정보연구원 Korea Institute of Science and Technology Information
주관 국가슈퍼컴퓨팅본부
KSF 한국초고성능컴퓨팅포럼
KSCSE 한국계산과학공학회
후원 과학기술정보통신부 Ministry of Science and ICT

Korea Supercomputing Conference 2026

KSC2026

2026년 9월 8일(화)~9월(수) | 스위스 그랜드호텔 컨벤션홀(4F)

The Intelligent Frontier: Synergizing HPC, AI and Quantum

ABSTRACT

주최 KISTI 한국과학기술정보연구원 Korea Institute of Science and Technology Information www.kisti.re.kr
주관 국가슈퍼컴퓨팅본부
KSF 한국초고성능컴퓨팅포럼
KSCSE 한국계산과학공학회
후원 과학기술정보통신부 Ministry of Science and ICT

TABLE OF CONTENTS

Greetings ---------- 04

Summery ---------- 05

Committee ---------- 06

Program ---------- 07

[Tuesday, September 08, 2026]
- Workshop - Day 1
- Keynote Speech
- Plenary 09
  1. HPC+AI 인프라 최신 기술 동향 및 국내 대형 시스템 소개 ---------- 13
  2. AI FOR SCIENCE: 과학적 발견과 혁신 ---------- 24
  3. 양자-HPC 하이브리드 컴퓨팅 기술 및 애플리케이션 ---------- 29
  4. 차세대 메모리·가속 컴퓨팅 기술과 클라우드 운영 혁신 ---------- 39
  5. 초고성능컴퓨팅 SW생태계조성사업 워크숍 (비공개) ---------- 46

- Conference 53
  1. 2026년 한국계산과학공학회 추계학술대회 ---------- 54

[Friday, September 05, 2025]
- Workshop - Day 2 AM 55
  1. HPC 활용을 위한 대규모 데이터 처리 기술 워크숍 ---------- 56
  2. HPC 수치모의실험 천문학 워크숍 ---------- 59
  3. Quantum Intelligence: 양자-AI 융합과 신약·바이오 혁신 ---------- 62
  4. 양자-고전 하이브리드 계산 방법론 및 적용 ---------- 68
  5. 국가슈퍼컴퓨터 6호기 '한강' 및 시범·공식서비스(혁신지원사업) 설명회 ---------- 71
  6. 한국슈퍼컴퓨팅센터 연합 회의 ---------- 73

- Workshop - Day 2 PM 78
  1. HPC 활용 산업 기술 워크숍 ---------- 79
  2. 기상·기후 분야 초고성능컴퓨팅 활용 고도화 워크숍 ---------- 82
  3. Bio-HPC & AI: 차세대 제약·바이오 혁신 ---------- 91
  4. 양자-고전 하이브리드 계산 방법론 및 적용 ---------- 98
  5. 한국초고성능 컴퓨팅포럼기술 교류회 ---------- 100
  6. BIG3 기술교류회 (비공개) ---------- 106

- Tutorial 107
  1. 국가슈퍼컴퓨터 6호기(한강) 튜토리얼: AI 기반 물리 응용 ---------- 108

- Exhibition 109

GREETINGS

안녕하십니까?

2026년 한국슈퍼컴퓨팅컨퍼런스(KSC2026) 조직위원장을 맡은 한국과학기술정보연구원(KISTI) 원장 이식입니다.

올해로 23회를 맞이하는 KSC는 초고성능컴퓨팅(HPC)이 인공지능(AI), 양자컴퓨팅(Quantum)과 긴밀하게 융합하며 새로운 기술 혁신을 이끌어가는 중요한 시기에 개최됩니다. 생성형 AI의 급속한 발전과 초거대 AI 모델의 확산, 양자컴퓨팅 기술의 비약적인 성장, 그리고 과학기술 전 분야에서 데이터 중심 연구가 확대됨에 따라 HPC는 디지털 혁신을 실현하는 핵심 기반 인프라로 자리매김하고 있습니다.

KSC2026은 "The Intelligent Frontier: Synergizing HPC, AI, and Quantum"을 주제로, HPC와 AI, 그리고 양자컴퓨팅이 만들어가는 새로운 지능형 컴퓨팅 시대를 조망하고자 합니다. 서로 다른 기술이 유기적으로 결합하여 과학기술 연구와 산업 혁신을 가속화하고, 미래 사회의 다양한 난제를 해결하는 방향을 함께 모색하는 뜻깊은 자리가 될 것입니다.

이번 KSC2026에서는 AI, Quantum, 디지털바이오, 첨단 제조, 기후·환경, 우주·항공 등 다양한 분야의 전문가들이 최신 연구 성과와 활용 사례를 공유할 예정입니다. 또한 워크숍, 튜토리얼, 특별세션, 기업 기술세션, 전시 프로그램 등을 통해 산·학·연·관이 함께 최신 기술과 미래 비전을 논의하고 협력의 기회를 확대하는 교류의 장을 마련하였습니다.

특히 AI 데이터센터, 차세대 GPU 및 가속기, 고성능 스토리지와 네트워크, AI 네이티브 HPC 플랫폼, 양자-HPC 하이브리드 컴퓨팅 등 차세대 컴퓨팅 기술을 중심으로 국내외 전문가들이 미래 컴퓨팅 생태계의 발전 방향을 제시하고 다양한 협력 방안을 논의할 예정입니다.

KISTI는 앞으로도 국가 초고성능컴퓨팅 인프라를 기반으로 연구자와 산업계, 공공 부문이 함께 혁신적인 연구와 기술 개발을 수행할 수 있도록 적극 지원하며, 대한민국 HPC 생태계 발전을 위해 최선을 다하겠습니다.

KSC2026이 HPC와 AI, 양자컴퓨팅이 함께 만들어갈 지능형 혁신의 미래를 공유하고, 새로운 협력과 도전의 출발점이 되는 뜻깊은 행사가 되기를 기대합니다.

여러분의 많은 관심과 적극적인 참여를 부탁드립니다.
감사합니다.

2026년 한국슈퍼컴퓨팅컨퍼런스 조직위원장

한국과학기술정보연구원 원장 이 식

ABOUT

행사 내용

Korea Supercomputing Conference is...
올해로 23회를 맞이하는 한국슈퍼컴퓨팅컨퍼런스는 HPC 및 계산과학공학 분야의 산·학·연·관 전문가와 학생들이 참여해 연구개발 성과와 최신 기술을 교류하는 국내 최대 규모의 슈퍼컴퓨팅 학술행사 및 전시회입니다.

KSC 2026은 초고성능컴퓨팅, 인공지능, 양자컴퓨팅이 만들어 갈 융합의 미래를 조망하고 새로운 연구 및 산업 협력의 기회를 마련합니다.

행사명	Korea Supercomputing Conference 2026
주 제	THE INTELLIGENT FRONTIER: SYNERGIZING HPC, AI AND QUANTUM
기 간	September 08(Tue)~09(Wed), 2026
주 최	한국과학기술정보연구원 Korea Institute of Science and Technology Information
주 관	국가슈퍼컴퓨팅본부 / KSF Korea Supercomputing Forum 한국초고성능컴퓨팅포럼 / KSCSE 한국계산과학공학회 Korean Society for Computational Sciences and Engineering
후 원	과학기술정보통신부 Ministry of Science and ICT

COMMITTEE

Organizing

위 원 장 : 이 식 원장 (한국과학기술정보연구원)

위　　원 : **곽호상** 총장 (금오공과대학교) **박찬열** 본부장 (한국과학기술정보연구원)

Program

위 원 장 : **김종수** 회장 (한국계산과학공학회)

위　　원 : **최태영** 교수 (이화여자대학교), **허무영** 박사 (기초과학연구원), **국봉재** 센터장 (기상청), **진현욱** 교수 (건국대학교), **권재민** 본부장 (한국핵융합에너지연구원), **김종원** 교수 (광주과학기술원), **염민선** 연구소장 (나무ICT), **홍태영** 센터장 (한국과학기술정보연구원), **정민중** 단장 (한국과학기술정보연구원), **강지훈** 센터장 (한국과학기술정보연구원), **이경하** 센터장 (한국과학기술정보연구원), **이정철** 팀장 (한국과학기술정보연구원)

Workshop

HPC+AI 인프라 최신 기술 동향 및 국내 대형 시스템 소개
분과위원장 **홍태영** 센터장 (한국과학기술정보연구원 슈퍼컴퓨팅 인프라센터)

AI for Science: 과학적 발견과 혁신
분과위원장 **이경하** 센터장 (한국과학기술정보연구원 초거대AI연구센터)

양자-HPC 하이브리드 컴퓨팅 기술 및 애플리케이션
분과위원장 **함재균** 단장 (한국과학기술정보연구원 양자컴퓨팅서비스연구단)

차세대 메모리·가속 컴퓨팅 기술과 클라우드 운영 혁신
분과위원장 **정기문** 센터장 (한국과학기술정보연구원 슈퍼컴퓨팅기술개발센터)

초고성능컴퓨팅 SW생태계조성사업 워크숍 (비공개)
분과위원장 **권오경** 팀장 (한국과학기술정보연구원 슈퍼컴퓨팅지원센터 HPC응용팀)

HPC 활용을 위한 대규모 데이터 처리 기술 워크숍
분과위원장 **이정철** 팀장 (한국과학기술정보연구원 슈퍼컴퓨팅가속화연구단 해석플랫폼팀)

HPC 수치모의실험 천문학 워크숍
분과위원장 **김종수** 회장 (한국계산공학회)

Quantum Intelligence: 양자-AI 융합과 신약·바이오 혁신
분과위원장 **염민선** 연구소장 (나무ICT)

양자-고전 하이브리드 계산 방법론 및 적용
분과위원장 **함재균** 단장 (한국과학기술정보연구원 양자컴퓨팅서비스연구단)

국가슈퍼컴퓨터 6호기 '한강' 및 시범·공식서비스(혁신지원사업) 설명회
분과위원장 **강지훈** 센터장 (한국과학기술정보연구원 슈퍼컴퓨팅응용센터)

한국슈퍼컴퓨팅센터 연합 회의
분과위원장 **심형욱** 센터장 (한국과학기술정보연구원 초고성능컴퓨팅정책센터)

HPC 활용 산업 기술 워크숍
분과위원장 **이정철** 팀장 (한국과학기술정보연구원 슈퍼컴퓨팅가속화연구단 해석플랫폼팀)

기상·기후 분야 초고성능컴퓨팅 활용 고도화 워크숍
분과위원장 **조민수** 박사 (한국과학기술정보연구원 슈퍼컴퓨팅인프라센터)

Bio-HPC & AI: 차세대 제약·바이오 혁신
분과위원장 **이준학** 단장 (한국과학기술정보연구원 디지털바이오컴퓨팅연구단), **이진희**

양자-고전 하이브리드 계산 방법론 및 적용
분과위원장 **함재균** 단장 (한국과학기술정보연구원 양자컴퓨팅서비스연구단)

한국초고성능컴퓨팅 포럼 기술교류회
분과위원장 **심형욱** 센터장 (한국과학기술정보연구원 초고성능컴퓨팅정책센터)

Big3 기술교류회 (비공개)
분과위원장 **홍태영** 센터장 (한국과학기술정보연구원 슈퍼컴퓨팅인프라센터)

Forum,

2026년 한국계산과학공학회 추계학술대회

Conference & Tutorial

국가슈퍼컴퓨터 6호기(한강) 튜토리얼: AI 기반 물리 응용

DAY 1 — 2026년 9월 8일(화) 09:00~19:30

09:00~10:00	등록	
10:00~10:40	**기조 강연 1: 데이터센터 플래시의 재정의 - AI 시대의 낸드, SSD, 그리고 컨트롤러의 미래** 남이현 대표이사 | 파두 (FADU)	
10:40~11:20	**기조 강연 2 : Quantum Computing Hype vs. Reality** 김성혁 상무 | LG전자	
11:20~12:00	**KSC2026 개회식** 개회사 **이 식 원장** | 한국과학기술정보연구원, KSC2026 조직위원장 환영사 **곽호상 교수** | 국립금오공과대학교, 한국초고성능컴퓨팅포럼 운영위원장 **김종수 회장** | 한국계산과학공학회 축 사	
12:00~13:30	Luncheon	
13:30~18:05	Track A	**HPC+AI 인프라 최신 기술 동향 및 국내 대형 시스템 소개**
13:30~18:00	Track B	**2026 한국계산과학공학회 추계학술대회**
13:30~18:00	Track C	**AI for Science: 과학적 발견과 혁신**
13:30~18:00	Track D	**양자-HPC 하이브리드 컴퓨팅 기술 및 애플리케이션**
13:30~18:00	Track E	**차세대 메모리·가속 컴퓨팅 기술과 클라우드 운영 혁신**
13:30~17:30	Track F	**초고성능컴퓨팅 SW생태계조성사업 워크숍** [비공개]
18:00~19:30	Welcome Dinner	

DAY 2 — 2026년 9월 9일(수) 09:00~18:00

09:00~10:00	등록	
10:00~11:40	Track A	**HPC 활용을 위한 대규모 데이터 처리 기술 워크숍**
10:00~12:00	Track B	**HPC 수치모의실험 천문학 워크숍**
10:00~12:05	Track C	**Quantum Intelligence: 양자-AI 융합과 신약·바이오 혁신**
10:00~12:00	Track D	**양자-고전 하이브리드 계산 방법론 및 적용**
10:00~12:00	Track E	**국가슈퍼컴퓨터 6호기 '한강' 및 시범·공식서비스(혁신지원사업) 설명회**
10:00~12:00	Track F	**한국슈퍼컴퓨팅센터 연합 회의**
12:00~13:30	Luncheon	
13:30~18:00	Track A	**HPC 활용 산업 기술 워크숍**
13:30~18:00	Track B	**기상·기후 분야 초고성능컴퓨팅 활용고도화 워크숍**
13:30~18:00	Track C	**Bio-HPC & AI: 차세대 제약·바이오 혁신**
13:30~18:00	Track D	**양자-고전 하이브리드 계산 방법론 및 적용**
13:30~18:00	Track E	**한국초고성능컴퓨팅 포럼 기술교류회**
13:30~18:00	Track F	**Big3 기술교류회** [비공개]

연계 프로그램 | Tutorial

일시 : 2026년 9월 7일(월) 09:30~18:00 | 장소 : 스위스 그랜드호텔, 컨벤션 센터 다이아몬드홀(3F)

| 국가슈퍼컴퓨터 6호기(한강) 튜토리얼 - AI 기반 물리 응용

DAY 1

Keynote & Plenary

Tuesday, September 08, 2026

2026 한국슈퍼컴퓨팅컨퍼런스 기조강연

10:00-10:40	[Keynote Speech 1] A.I 분야 데이터센터 플래시의 재정의 — AI 시대의 낸드, SSD, 그리고 컨트롤러의 미래 남이현 대표이사 (FADU)
10:40-11:20	[Keynote Speech 2] 양자 분야 Quantum Computing Hype vs. Reality 김성혁 상무 (LG전자)
11:20-12:00	KSC 2026 개회식 개회사 이식 원장 (한국과학기술정보연구원, KSC2026 조직위원장) 환영사 곽호상 교수 (국립금오공과대학교, 한국초고성능컴퓨팅포럼 운영위원장) 환영사 김종수 회장 (한국계산과학공학회) 축사
12:00~13:30	Luncheon
13:30~18:00	Workshop, Community

남이현
대표이사 (CEO)
파두 (FADU)

데이터센터 플래시의 재정의 — AI 시대의 낸드, SSD, 그리고 컨트롤러의 미래

AI는 메모리 계층 구조를 근본적으로 바꾸고 있다. 추론(inference)의 시대에 접어들며 KV 캐시와 컨텍스트 데이터가 HBM의 용량 한계를 훨씬 빠르게 넘어서고 있으며, HBM과 NVMe SSD 사이에는 테라바이트급 용량과 메모리급 대역폭, 스토리지급 비용을 동시에 요구하는 새로운 계층의 공백이 생겨났다. 이제 낸드 플래시는 이 공백을 채울 수 있는 유일한 미디어로서, AI 데이터센터에서 가장 중요한 역할을 부여받고 있다.

본 키노트에서는 먼저 AI 데이터센터의 메모리·스토리지 계층 구조 변화와 이에 대한 업계의 대응(GPU 주도 스토리지 아키텍처, High Bandwidth Flash 표준화 등)을 짚어본다. 이어서 SSD가 메모리 티어, 퍼포먼스 티어, 커패시티 티어의 세 가지 형태로 분화되고, 낸드가 대역폭 중심과 밀도 중심으로 갈라지는 기술적 흐름 속에서, 이 모든 복잡성을 흡수해야 하는 SSD 컨트롤러가 스토리지 컨트롤러에서 메모리 컨트롤러로 진화하는 과정을 살펴본다. 마지막으로 PCIe Gen7 컨트롤러, 플래시 기반 추론(decode) 엔진 등 파두가 준비하고 있는 차세대 기술을 통해, 국내 반도체 산업이 AI 인프라 시대에 플래시 기술의 프런티어를 어떻게 주도할 수 있을지 전망을 공유한다.

· Educational History
서울대학교 전기컴퓨터공학부 박사 (2011)

· Employment History
SK텔레콤 융합기술원 Expert PL (2013~2015)
파두(FADU) 대표이사 (2015~현재)

김성혁
상무
LG전자

Quantum Computing Hype vs. Reality

양자기술은 AI 이후의 차세대 혁신 플랫폼으로 주목받으며, 주요 국가들의 대규모 투자와 산업화 경쟁이 가속화되고 있습니다. 경쟁의 초점은 단순한 큐비트 성능 향상을 넘어 오류정정, 양자-AI 융합, 양자센싱, 양자 네트워크, 제조 생태계 및 표준화 확보로 확대되고 있으며, 양자기술은 국가 경쟁력과 산업 혁신을 좌우하는 전략기술로 진화하고 있습니다.

본 세션에서는 글로벌 양자기술의 최신 동향과 산업화 방향을 중심으로 양자컴퓨팅, 양자센싱, 양자통신 분야의 발전 현황과 미래 전망을 살펴봅니다. 특히 산업계의 연구개발 사례를 통해 양자컴퓨팅 기반 유체 해석, 신물질 탐색 및 공학 문제 해결, 양자컴퓨팅 미들웨어 개발, 양자-고전컴퓨팅 하이브리드 프레임워크 구축 등 양자컴퓨팅 분야의 최신 연구 동향을 소개합니다. 아울러 다이아몬드 NV 기반 양자센서의 소형화·제품화와 차세대 보안 네트워크 및 양자인터넷 핵심 기술 개발 사례도 함께 공유합니다.

또한 특허, 표준화, 생태계 구축을 포함한 산업화 전략과 기업이 준비해야 할 방향성을 논의하며, Quantum-AI 융합이 미래 산업 혁신과 새로운 가치 창출을 어떻게 가속화할 것인지에 대한 실질적인 통찰을 제공하고자 합니다.

· Educational History

서울대학교 전기공학부 학사 ('96), 석사 ('98), 박사 ('02)

· Employment History

미국 조지아공대 박사후연구원/연구교수 ('05~'10)
LG 전자 ('10년 11월~) 양자컴퓨팅 연구 리더 ('21년~) 상무 ('24년~)
한국양자산업협회 회장 ('26년 1월~)
K-문샷 양자 분야 추진위원('26년 6월~)
양자전략위원회 실무위원 ('25년 3월~)

DAY 1

Workshop

Tuesday, September 08, 2026

KOREA SUPERCOMPUTING CONFERENCE 2026

The Intelligent Frontier: Synergizing HPC, AI and Quantum

HPC+AI 인프라 최신 기술 동향 및 국내 대형 시스템 소개

계산과학 및 AI 기반 연구를 하기 위해서는 슈퍼컴퓨팅 인프라가 필수적이며, 특히 대형 실험 데이터 분석, 과학·공학 분야 대규모 모델링 및 시뮬레이션, LLM 등 초거대 AI 모델 개발 시 중추적인 역할을 한다. 본 워크숍에서는 최근 주목받고 있는 HPC+AI 인프라 최신 기술과 국내 대표적인 대형 슈퍼컴 운영 사례를 살펴보고자 한다.

13:30~13:55	**Vera Rubin NVL72 시스템 및 레퍼런스 아키텍처 소개** 신민수 이사 (NVIDIA)
13:55~14:20	**AI와 HPC를 위한 AMD Instinct GPU 소개** 김홍필 상무 (AMD)
14:20~14:45	**HyperAccel LPU: 토큰 효율성이 여는 AI 인프라의 미래** 이진원 CTO (하이퍼엑셀)
14:45~15:10	**공장에서 짓는 데이터센터 - 모듈러(PMDC)로 여는 AI 인프라 공기 단축** 전현욱 이사 (엘리스그룹)
15:10~15:35	**NIPA 국가 AI 시스템 1차 구축사례 : 국내 최대 규모 수랭 GPU 클러스터의 도전과 경험** 김종훈 기술본부장 (엠키스코어)
15:35~16:00	Coffee Break
16:00~16:25	**해인 클러스터로 본 GPU 가상화 기술과 차세대 랙스케일 인프라 전망** 이 건 수석연구원 (SK텔레콤)
16:25~16:50	**이상탐지에서 자동복구까지: GPU 504장에서의 학습 워크로드 운영** 황은진 연구소장 (래블업)
16:50~17:15	**GPU의 한계를 돌파하는 AI-Native 데이터 플랫폼: 단순 데이터 서빙을 넘어 토큰(Token) 생산성 극대화로** 노용신 상무 (VAST)
17:15~17:40	**DRAM을 넘어서: AI 추론을 위한 KV 캐시 확장** 신현호 상무 (WEKA)
17:40~18:05	**스토리지가 AI Job을 이해하는 방법 : 스케줄러 연계 데이터 선행 배치와 GPU I/O 최적화** 김태훈 팀장 (글루시스)

신민수
이사
NVIDIA

Vera Rubin NVL72 시스템 및 레퍼런스 아키텍처 소개

본 세션에서는 Hopper, Blackwell, Blackwell Ultra를 거쳐 차세대 Vera Rubin 플랫폼으로 이어지는 NVIDIA 가속 컴퓨팅 아키텍처의 발전 방향을 살펴보고, 이를 랙 스케일 시스템으로 구현한 Vera Rubin NVL72와 레퍼런스 아키텍처를 소개합니다. Vera CPU와 Rubin GPU, 6세대 NVLink 기반의 스케일업 패브릭, ConnectX SuperNIC 및 BlueField DPU를 결합해 72개의 GPU를 하나의 대규모 가속 컴퓨팅 자원처럼 활용하는 시스템 구조와 함께, 여러 랙을 AI 팩토리 규모로 확장하기 위한 네트워크·스토리지·전력·냉각·운영 관리 설계 원칙을 설명합니다.

· Educational History

고려대학교전기전자전파공학부석사(2014)

· Employment History

LG U+ 네트워크 개발 엔지니어 (2014~2019)
Cisco Systems GSP 데이터센터 솔루션스 아키텍트 (2019~2022)
Cisco Systems CX 네트워크 컨설팅 엔지니어 (2023~2024)
Samsung SDS SCP 시니어 엔지니어 (2024~2025)

김홍필
상무
AMD

AI와 HPC를 위한 AMD Instinct GPU 소개

이번 세션에서는 AMD의 차세대 AI 및 HPC GPU 포트폴리오인 Instinct MI455X와 Instinct MI430X를 중심으로, 각 제품이 어떤 아키텍처적 특성을 갖고 있으며 어떠한 워크로드에 최적화되어 있는지를 소개해 드릴 예정입니다.

특히 Helios 플랫폼에 탑재되는 AI 전용 GPU인 MI455X와, 고성능 컴퓨팅(HPC) 워크로드를 위해 설계된 MI430X를 비교하며, 유체역학(CFD), 분자동역학(MD), 기상 예측, 구조 해석 등 전통적인 과학·공학 시뮬레이션부터 대규모 생성형 AI 모델의 학습 및 추론, 그리고 물리 기반 시뮬레이션을 포함하는 디지털 트윈(Digital Twin) 환경까지 다양한 워크로드 요구사항에 따른 최적의 GPU 선택 방안을 살펴볼 예정입니다.

또한 각 워크로드가 요구하는 연산 특성, 메모리 구조, 통신 대역폭 및 시스템 아키텍처 측면에서 MI455X와 MI430X가 제공하는 차별화된 가치를 실제 적용 사례와 함께 소개함으로써, 참석자 여러분이 업무 환경에 가장 적합한 GPU 플랫폼을 선정하고 시스템 자원을 보다 효율적으로 활용할 수 있는 인사이트를 얻으실 수 있도록 돕고자 합니다.

· Educational History

연세대학교 전기전자공학과 학사 (2004)

· Employment History

SK Teletech, SK Telesys 휴대폰 개발 엔지니어 (2004~2012)
NVIDIA Korea 어플리케이션 엔지니어 (2012~2015)
AMD Korea 데이터센터 솔루션 아키텍트 (2016~현재)

이진원
CTO
하이퍼엑셀

HyperAccel LPU: 토큰 효율성이 여는 AI 인프라의 미래

글로벌 인공지능(AI) 산업은 트랜스포머 기반 대규모 언어 모델(LLM)의 등장으로 급격한 변화를 맞이했으며, AI 컴퓨팅 인프라의 패러다임 또한 빠르게 재편되고 있습니다. 이번 강연에서는 LLM 추론(Inference) 서비스가 제기하는 새로운 컴퓨팅 과제와 이에 대응하기 위해 발전해 온 최신 AI 반도체 기술의 동향을 소개합니다. 또한, HyperAccel이 개발한 차세대 AI 추론 반도체 LPU(LLM Processing Unit)의 핵심 기술을 심도 있게 살펴봅니다. LPU는 LLM 추론 워크로드에 특화된 아키텍처를 기반으로, 서비스가 요구하는 지연시간과 처리량을 충족하는 동시에 주어진 비용과 전력으로 더 많은 토큰을 생성하는 '토큰 효율성(Token Efficiency)'을 핵심 목표로 설계되었습니다. 이는 AI 서비스의 총소유비용(TCO)을 획기적으로 낮추고, 폭발적으로 증가하는 AI 추론 수요를 지속가능하게 뒷받침하는 새로운 컴퓨팅 패러다임을 제시합니다.

LPU의 혁신적인 아키텍처와 설계 철학을 통해 AI 시대의 핵심 경쟁력이 단순한 연산 성능이 아니라 '토큰을 얼마나 경제적이고 지속가능하게 생성할 수 있는가'에 있음을 살펴보고, 차세대 AI 컴퓨팅 인프라의 방향과 미래를 함께 전망하고자 합니다.

· Educational History

서울대학교 컴퓨터공학과 석사 (2017)

· Employment History

삼성전자 시스템LSI 책임엔지니어 (2011~2020)
㈜ 뉴블라 CTO (2021.11 ~ 2023.10)
㈜ 하이퍼엑셀 CTO 2023.11 ~ 현재

전현욱
이사
엘리스그룹

공장에서 짓는 데이터센터 - 모듈러(PMDC)로 여는 AI 인프라 공기 단축

이번 세션에서는 초거대 AI 학습과 과학기술 시뮬레이션 등 슈퍼컴퓨팅 및 HPC급 워크로드를 위한 PMDC(Portable Modular Data Center)를 소개합니다.

PMDC는 전력, 냉각, 랙, 네트워크 등 핵심 설비를 공장에서 사전 제작하고 검증한 뒤 현장에 설치하는 방식입니다. 부지 준비와 제작을 동시에 진행해 4개월 이내 구축이 가능하며, 모듈 단위로 확장할 수 있습니다. 기존 건축형 AIDC와 달리 최신 GPU의 전력과 냉각 요구사항을 신속하게 반영할 수 있어, 인프라 기획부터 실제 서비스 개시까지의 기간을 크게 단축할 수 있습니다.

본 발표에서는 NIPA 국가 AI 프로젝트의 NVIDIA B200 GPU 2,560장 규모 클러스터 구축 사례를 중심으로 인프라 설계와 구축 과정을 소개합니다. 이를 통해 PMDC가 대규모 AI와 HPC 인프라를 빠르게 구현하는 현실적인 데이터센터 구축 방식임을 설명합니다.

· **Educational History**

경북대학교 컴퓨터공학과 학사 (2010)

· **Employment History**

IBM Korea Client Technical Sales (2011~2016)
HPE Korea Technical Architect (2016~2019)
NetApp Korea AI Solution Architect(2019~2025)
Cisco Korea AI Solution Architect(2025~2026)

김종훈
기술본부장 (Tech Hub)
엠키스코어

NIPA 국가 AI 시스템 1차 구축사례 : 국내 최대 규모 수랭 GPU 클러스터의 도전과 경험

생성형 AI와 초거대 모델의 확산으로 GPU의 전력 및 발열 밀도가 급격히 증가하면서, 데이터센터는 공랭 중심에서 고집적 수랭 환경으로 빠르게 전환되고 있습니다. 이러한 변화는 냉각 방식뿐만 아니라 전력, 상면, 하중, 배관 및 운영 체계 전반의 통합적인 재설계를 요구합니다.

본 발표에서는 NVIDIA B200 GPU 7,656개와 GPU 서버 957대로 구성된 국내 최대 규모의 수랭 GPU 클러스터 구축 경험을 소개합니다. 랙당 약 74kW의 고밀도 환경에 대응하기 위해 수랭 85%, 공랭 15%의 열부하 분산 구조를 적용하고, 냉각 루프와 인랙 CDU, 매니폴드 및 콜드플레이트를 연계하여 안정적인 냉각 환경을 구현하였습니다.

구축 과정에서 발생한 전력·냉각 용량, 랙 하중, 배관 및 유량 균형, 누수 위험 등의 수랭 관련 과제를 비롯하여 장비 조달과 반입, 시공 간섭, 네트워크·스토리지 연계, 공정 관리, 시스템 통합 및 성능 검증 등 대규모 시스템 구축 전반의 이슈와 해결 경험을 공유합니다.

또한 단계적인 시스템 통합과 성능·안정성 검증 과정에서 얻은 시행착오를 바탕으로, 고집적 수랭 시스템과 대규모 AI 인프라 구축 시 고려해야 할 핵심 요소를 살펴보고 향후 데이터센터의 설계 및 운영 방향을 전망하고자 합니다.

· Educational History

고려대학교 정보통신대학원 석사과정 (2014)

· Employment History

세림INS 삼성전자 SAIT Supercom Admin (2006~2019)
엠키스코어 기술본부 Leader (2019~2026)

이 건
수석 연구원
SK 텔레콤

해인 클러스터로 본 GPU 가상화 기술과 차세대 랙스케일 인프라 전망

대규모 AI 학습·추론 수요가 폭증하면서 GPU 자원을 얼마나 효율적으로 나누어 쓰고, 빠르게 제공하며, 사용자 간에 안전하게 격리하는가가 AI 인프라의 핵심 과제가 되고 있다. 본 발표에서는 SK텔레콤이 국가 AI 파운데이션 모델 사업을 지원하기 위해 구축한 대규모 GPU 인프라 'Haein(海印)'의 사례를 중심으로, 자체 개발한 GPU 가상화 플랫폼 'Petasus AI Cloud'를 통해 1,000장 이상의 NVIDIA Blackwell GPU를 하나의 클러스터로 묶어 여러 사용자에게 클라우드 방식으로 제공하는 기술을 소개한다. 구체적으로는 가상 머신 환경에서도 NVLink 기반 GPU 간 통신과 스토리지 입출력 성능을 베어메탈과 거의 동일한 수준(성능 저하 1% 미만)으로 유지한 방법과 대규모 검증 결과를 공유하고, InfiniBand 네트워크를 사용자별로 완전히 분리하여 보안과 성능 간섭 문제를 해결한 경험을 설명한다. 나아가 8-GPU 서버 단위로 클러스터를 구성하던 기존 방식을 넘어, 랙 전체가 하나의 거대한 가속기처럼 동작하는 차세대 랙스케일 아키텍처(NVL72)의 구조와 비용-성능 특성을 분석하고, 전통적인 HPC 배치 작업 환경과 클라우드 네이티브 AI 플랫폼을 하나의 인프라 위에서 함께 수용하기 위한 Petasus의 발전 방향을 제시한다.

· Educational History

포항공과대학교 컴퓨터공학과 석/박사

· Employment History

포항공과대학교 컴퓨터공학과 연구교수
Open Networking Foundation (ONF) 방문학자
SK텔레콤, 수석 연구원

황은진
연구소장
래블업 주식회사

이상탐지에서 자동복구까지: GPU 504장에서의 학습 워크로드 운영

독자 AI 파운데이션 모델 프로젝트 1·2차 정예팀(업스테이지 컨소시엄) 참여기관으로서, GPU 504장 이상 규모의 클러스터를 4개월 넘게 운영한 경험을 공유합니다. LLM 학습을 위한 대규모 멀티노드 분산학습에서 실제로 마주친 문제들을 유형별로 살펴보고, GPU 장애 시 제한된 자원 내에서 학습 중단 시간을 최소화하기 위해 적용한 MLOps 대응 방안을 소개합니다. 아울러 하드웨어 모니터링 지표를 활용한 장애 예측 가능성을 검토하고, 체크포인트 주기 최적화로 GPU 활용률을 높이는 방안을 함께 탐색합니다.

· **Educational History**
POSTECH 물리학과 박사 (2013)

· **Employment History**
한국과학기술연구원 뇌과학연구소 위촉연구원 (2013-2017)
고려대학교 바이오뇌공학부 겸임교수 (2024-현재)
래블업 주식회사 연구부 (2017-현재)

노용신
상무
VAST Data Korea

GPU의 한계를 돌파하는 AI-Native 데이터 플랫폼: 단순 데이터 서빙을 넘어 토큰(Token) 생산성 극대화로

현대 AI 혁신과 초거대 언어 모델(LLM)의 급격한 발전은 인프라의 전반적인 구조적 변화를 요구하고 있습니다. 기존 HPC 환경을 지배해 온 Lustre와 같은 전통적인 병렬 파일 시스템은 대규모 순차 I/O 처리에는 강력하지만, 수억 개의 소형 파일과 복잡한 임의 무작위(Random) I/O가 폭증하는 현대 AI 파이프라인 및 유전체학(Genomics) 주변 워크로드에서는 메타데이터 서버(MDS)의 잠금 병목(Lock Bottleneck)과 이로 인한 노드 간 통신 오버헤드(East-West traffic)로 한계에 봉착합니다.

본 세션에서는 이러한 기존 병렬 시스템의 한계를 극복하기 위해 고안된 VAST Data의 DASE(Disaggregated and Shared Everything, 분리 및 완전 공유) 아키텍처와 최신 데이터 패브릭 기술 동향을 소개합니다. VAST는 중앙 집중식 잠금 관리자(Central Lock Manager)를 완전히 배제하고, 초저지연 공유 메모리인 SCM(Storage Class Memory) 내에 V-Tree 형태로 파일 시스템 메타데이터를 직접 분산 저장하여 동기화 쓰기 및 읽기 경로의 무장애·무병목 선형 확장을 실현합니다.

특히, AI 스토리지의 역할이 단지 NVIDIA GPUDirect Storage(GDS) 등을 활용하여 GPU에 데이터를 빠르게 공급함으로써 대기 시간을 줄이는(GPU Starvation 방지) 1차원적인 '고속 데이터 서빙'에만 머물러서는 안 된다는 점을 심도 있게 짚어봅니다. VAST가 선도하는 핵심 아키텍처 기술은 GPU의 고대역폭 메모리(HBM) 용량 물리적 한계를 뛰어넘어 고성능 메모리 영역을 확장하는 가상화 가속을 지향합니다.

LLM 분산 추론 환경에서 병목의 주요 원인이 되는 거대한 Key-Value(KV) 캐시 및 대규모 문맥(Context) 데이터를 VAST의 초저지연 패브릭과 무상태(Stateless) 아키텍처로 오케스트레이션하여 지연 시간 없이 오프로딩 함으로써, 하드웨어 증설 없이도 더 넓은 컨텍스트 윈도우와 더 큰 배치 크기를 효율적으로 소화할 수 있습니다. 결과적으로 "동일한 GPU 자원을 가지고도 더 많은 토큰(Token)을 생성하고 추론 연산 효율을 극한으로 끌어올리는" 스토리지 아키텍처의 패러다임 전환과 글로벌 기술 동향을 위한 전략적 제언과 함께 제시합니다.
.

· **Employment History**

전) Ever Pure Systems Engineer

현) VAST Data Korea Sr, Systems Engineer

신현호
상무
WEKA.IO

DRAM을 넘어서: AI 추론을 위한 KV 캐시 확장

이번 세션에서는 대규모 언어 모델(LLM) 추론(Inference) 환경에서 갈수록 심화되는 메모리 장벽(Memory Wall) 문제와, 이를 해결하기 위한 WEKA KV Cache storage solution을 중심으로 효율적인 추론 성능을 위한 KV Cache Storage 전략을 소개드릴 예정입니다. 모델 크기와 사용자 수요가 증가하고 새로운 추론 모델이 등장하면서 KV Cache 용량에 대한 요구가 급격히 확장되고 있으며, 이로 인해 GPU를 과도하게 증설(over-provisioning)하거나 속도·정확성·비용 간의 절충을 감내해야 하는 한계가 발생하고 있습니다.

WEKA KV Cache storage solution은 KV Cache를 위한 페타바이트(PB) 규모의 영속 스토리지를 제공하여, 기존 단일 테라바이트(TB) 규모의 DRAM 대비 메모리 공간을 대폭 확장합니다. 이를 통해 GPU의 HBM에서 KV Cache 데이터를 오프로드(offload)하여 GPU가 가장 중요한 연산에 집중하도록 하고, "불필요한 토큰을 다시 계산하지 않는(Never recompute a token unnecessarily)" 아키텍처로 첫 토큰 생성 시간(TTFT)과 전체 토큰 처리량을 개선합니다.
세션에서는 해당 솔루션의 구현 방식과 추론 인프라 최적화 효과를 살펴봅니다.

· **Educational History**
경희대학교 컴퓨터공학과 학사

· **Employment History**
IBM Korea, Power System(Unix) SE
NetApp Korea, Sr Solution Architect(HPC/EDA)
WEKA.IO, Sr Consulting System Engineer

김태훈
팀장
글루시스

스토리지가 AI Job을 이해하는 방법 : 스케줄러 연계 데이터 선행 배치와 GPU I/O 최적화

AI 학습 환경에서는 GPU 연산 성능뿐 아니라, 학습에 필요한 데이터가 적절한 시점에 준비되는지가 전체 Job 처리 효율을 결정한다. 공유 스토리지의 최대 대역폭이 충분하더라도 데이터셋과 모델이 저속 계층에 위치하거나 여러 Job의 I/O가 특정 경로에 집중되면 GPU는 데이터를 기다리게 된다. 접근 빈도를 관찰한 뒤 데이터를 이동하는 기존 티어링 방식만으로는 첫 번째 Job의 데이터 공급 지연과 불필요한 계층 간 이동을 해결하기 어렵다.

본 세션에서는 Kubernetes 및 Slurm과 같은 배치 스케줄러의 Job 정보를 스토리지 데이터 배치에 활용하는 AI 워크로드 인지형 스토리지 구조를 소개한다. 스케줄러로부터 Job 실행 위치, 필요한 데이터셋, 실행 예정 시점 및 데이터 정책을 전달받아 실행 전에 데이터를 고성능 NVMe 계층으로 선행 배치하고, Job 수행 중에는 GPUDirect Storage 기반의 데이터 경로를 통해 CPU 개입과 GPU 데이터 공급 지연을 줄이는 방안을 살펴본다.

또한 데이터셋, 모델 및 체크포인트의 접근 특성과 재사용 가능성을 고려해 NVMe, HDD 및 아카이브 계층 간 데이터 위치를 결정하는 정책 기반 배치 구조와, 다수의 GPU Job이 동시에 실행되는 환경에서 스토리지 부하와 데이터 공급 경로를 조정하는 GPU I/O Workload Balancing 기술의 발전 방향을 제시한다. 이를 통해 사후 이동 중심의 티어링을 넘어, AI Job의 실행 계획을 인지하고 필요한 데이터를 먼저 준비하는 AI-aware 데이터 인프라의 구현 가능성을 살펴본다.

· Educational History

세종대학교 컴퓨터공학과 박사 (2025)

· Employment History

세종대학교 - 빅데이터센터 클라우드 아키텍터 (2014 ~2018)
글루시스 - 고성능 스토리지 아키텍터 (2018 ~ 현재)

WORKSHOP 2

Tuesday, September 08, 2026

AI for Science: 과학적 발견과 혁신

13:30~14:30	**AI for Science : Concept & Trends** 안성수 조교수 (KAIST 김재철AI대학원)
14:30~15:30	**Agentic AI for Materials Design** 박찬영 부교수 (KAIST)
15:30~16:00	Coffee Break
16:00~17:00	**Beyond the Boundaries of R&D: Collaborative Strategies for Generative AI Multi-Agents Driving Bio-Innovation** 김태형 대표이사 (BioNexus)
17:00~18:00	**Transforming Scientific Discovery: AI for Science at KISTI** 이경하 초거대AI연구센터장 (한국과학기술정보연구원)

안성수
조교수
KAIST 김재철AI대학원

AI for Science : Concept & Trends

AI for Science는 과학 데이터를 이용한 예측을 넘어, 새로운 후보를 설계하고 연구 과정 자체를 지능화하는 방향으로 발전하고 있다. 본 발표에서는 최근의 발전을 예측 모델, 생성적 설계, AI Scientist라는 세 가지 흐름을 중심으로 소개한다. 신약 개발 분야에서는 단백질 구조와 분자 상호작용 예측, 신약 후보 분자의 생성 및 최적화, 실험 후보 선별에 AI가 어떻게 활용되는지 살펴본다. 소재 개발 분야에서는 결정 구조 생성, 물성 예측, 양자화학 및 분자동역학 시뮬레이션 가속을 통해 방대한 소재 설계 공간을 탐색하는 최근 연구를 소개한다. 이어서 대규모 언어모델과 에이전트 시스템이 문헌 조사, 가설 수립, 계산 실험 설계, 코드 실행 및 결과 해석을 반복적으로 수행하는 AI Scientist로 발전하는 흐름과 현재의 한계를 논의한다.

· Educational History

KAIST 전기및전자공학과 학사 (2015)
KAIST 전기및전자공학과 석·박사통합과정, 공학박사 (2021)

· Employment History

Mohamed bin Zayed University of Artificial Intelligence 박사후연구원 (2021.03–2021.11)
POSTECH AI대학원 및 컴퓨터공학과 조교수 (2021.12–2025.01)
KAIST AI대학원 조교수 (2025.01–현재)

박찬영
부교수
KAIST

Agentic AI for Materials Design

최근 인공지능은 소재의 물성 예측과 구조 생성에 활용되는 수준을 넘어, 새로운 소재를 탐색하고 설계하는 연구 과정 전반을 지원하는 방향으로 발전하고 있다. 특히 대규모 언어모델(LLM)은 방대한 과학 지식을 자연어 형태로 활용하고, 복잡한 연구 목표를 단계적으로 해석하며, 다양한 예측 모델과 데이터베이스, 시뮬레이션 도구를 연결할 수 있다는 점에서 기존 소재 AI와 차별화된 가능성을 보여준다. 이에 따라 LLM은 단순한 질의응답 모델이 아니라, 소재 후보를 제안하고 평가 결과를 해석하며 다음 탐색 방향을 결정하는 연구 에이전트로 활용되기 시작했다.

본 발표에서는 AI for Materials 분야에서의 LLM 활용 방식을 Agentic Orchestration, LLM-guided Evolution, Post-training의 세 가지 패러다임으로 구분하여 살펴본다. 각 패러다임이 소재 탐색과 설계 과정에서 LLM을 어떻게 활용하는지 대표 사례를 중심으로 소개하고, 현재의 가능성과 한계를 함께 논의한다. 이를 통해 LLM 기반 소재 연구의 최근 흐름을 정리하고, 향후 보다 신뢰성 있고 자율적인 소재 발견 시스템으로 발전하기 위한 방향을 조망한다.

· Educational History

서강대학교 컴퓨터공학과 학사 (2014)
포항공과대학교 컴퓨터공학과 박사 (2019)

· Employment History

Microsoft Research Asia 인턴 (2017)
University of Illinois at Urbana-Champaign 박사후 연구원(2019-2020)
University of California San Diego 방문연구원 (2025-2026)
KAIST 산업및시스템공학과 부교수 (2020-현재)

김태형
대표이사
BioNexus

Beyond the Boundaries of R&D: Collaborative Strategies for Generative AI Multi-Agents Driving Bio-Innovation

Scientific research, particularly bio-innovation, is rapidly shifting toward AI-augmented workflows in which autonomous agents support hypothesis generation, literature analysis, experimental design, data interpretation, and decision-making under expert supervision. <hp:lineBreak/><hp:lineBreak/>BioNexusis developing Nexus Co-Scientist, a domain-specialized AI scientist platform integrating LLMs, bio-foundation models, RAG, knowledge graphs, and multi-agent orchestration. It transforms literature, multi-omics, clinical data, biomedical databases, and experimental results into evidence-grounded hypotheses, therapeutic targets, biomarkers, and validation strategies.<hp:lineBreak/><hp:lineBreak/>

Through a human-in-the-loop architecture, specialized agents collaborate across drug discovery, precision medicine, and biological data analysis while ensuring traceability and expert oversight. Nexus Co-Scientist provides a scalable framework for faster, more reliable, and collaborative bio-innovation.

· Educational History

2003년~2005년 부산대학교 생명정보학 박사수료
2001년~2003년 부산대학교 생명정보학 석사졸업

· Employment History

2005 년~2009 년 국가생물연구자원정보센터 (KOBIC) 연구원
2010 년~2024 년 테라젠바이오 본부장(상무)
2020 년~2024 년 과기부 미래인재특별위원회 위원
2021 년~현재 국가생물연구자원정보센터(KOBIC) 운영위원
2023 년~현재 외교부 과기외교자문위원회 자문 위원
2024 년~현재 바이오넥서스 대표이사
2025 년~현재 국가인공지능전략위원회 위원
2026 년~현재 국가과학기술자문회의 위원
2026 년~현재 국가바이오혁신위원회 위원

이경하
초거대AI연구센터장
한국과학기술정보연구원

Transforming Scientific Discovery: AI for Science at KISTI

본 발표에서는 국가 과학기술 경쟁력 대도약을 위한 'K-문샷(K-Moonshot)' 전략과 이에 발맞춘 KISTI의 'AI for Science' 추진 방향및 경과를 소개한다. AI가 단순 보조 도구를 넘어 연구 파트너로 진화함에 따라, KISTI는 국가 초거대 과학기술 데이터와 HPC 인프라를 결합한 자율형 연구 생태계 구축을 위한 지원 체계를 개발하고 있다. 과학기술 특화 파운데이션 모델인 KONI를 중심으로 가설 수립·문헌 분석·실험 설계를 자율 수행하는 'AI 과학자(AI Scientist)' 기술과 연구 행정 업무를 자동화하는 '웹 에이전트(Web Agent)' 등의 기술들을 기반으로 R&D 혁신을 가속화하고, 국가 과학기술 난제 해결 및 글로벌 연구 협력을 이끄는 KISTI의 현재 기술 사항과 비전, 실행 로드맵을 제시한다.

· Educational History

충남대학교 정보통신공학 학사 (1998)
충남대학교 정보통신공학 석사 (2000)
충남대학교 컴퓨터공학 박사 (2006)

· Employment History

KISTI 초거대AI연구센터장/책임연구원 (2014-현재)
UST 응용AI전공 전임교수
한국정보과학회 학회지 편집위원(2020,2026) 및 임원이사(2022)
한국정보과학회 데이터소사이어티 상임이사(2018~현재)
과학기술정보통신부 미래국방기술전략 AI·무인자율분과 위원(2023~2026)
과학기술정보통신부 AI 핵심프로젝트 TF 위원(2025~현재)
과학기술정보통신부 국가과학기술기본계획위원회 위원(2026~현재)
국회도서관 인공지능자문단 위원(2025~현재)
한국과학기술총연합회 AI위원회 위원(2025~2026)

양자 – HPC 하이브리드 컴퓨팅 기술 및 애플리케이션

13:30~13:55	**Quantum circuit optimization using deep reinforcement learning** 이창협 부교수 (한양대학교)
13:55~14:20	**Architecting Quantum Computers with Mobile Trapped Ions** 김준기 부교수 (성균관대학교)
14:20~14:45	**Programmable Nonlinear Photonic Integrated Circuits for Photonic Quantum Computing** 권형한 박사 (KIST)
14:45~15:10	**초전도 양자컴퓨터의 발전 현황** 김요셉 조교수 (고려대학교)
15:10~15:35	**Quantum-AI 기반 신약 타겟 발굴** 한남식 교수 (연세대학교/영국 케임브리지대학교)
15:35~16:00	Coffee Break
16:00~16:25	**Quantum Inside the AI Data Center: IonQ's Hybrid QPU-GPU Accelerated Computing** 예기헌 상무 (IonQ)
16:25~16:50	**Quantum AI: GPU–QPU Integration and Hybrid QML Use Cases** 오성일 디렉터 (콴텔라 코리아)
16:50~17:15	**The Era of the Quantum Supercomputer** Hao-Chen Yeh Technical Sales Engineer (IQM Quantum Computers)
17:15~17:40	**중성원자 양자컴퓨팅 플랫폼의 확장성 및 실용적 HPC-QC 하이브리드 연동** 이우준 책임연구원 (Pasqal)
17:40~18:00	**질의응답 및 마무리**

이창협
부교수
한양대학교

Quantum circuit optimization using deep reinforcement learning

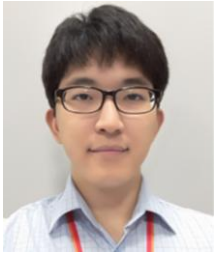
Quantum circuit optimization is essential for reducing noise and decoherence on noisy intermediate-scale quantum devices. While reinforcement learning (RL) has emerged as a promising approach, existing RL agents inefficiently spend substantial training capacity rediscovering elementary circuit reductions that deterministic methods can handle reliably. We propose an RL framework that integrates a deterministic Commutation-and-Reduction (CR) algorithm directly into the training environment. After each agent action, CR automatically performs elementary commutations and cancellations, allowing the agent to focus on non-trivial optimizations. Experiments on Clifford+T and CNOT+Pauli gate sets show that RL+CR consistently produces shorter circuits than standard RL. Moreover, agents trained on small circuits generalize effectively to larger ones: on 20-qubit Clifford+T circuits, five times larger than those used for training, RL+CR removes twice as many gates as standard RL. These results demonstrate that combining deterministic circuit reduction with RL provides an efficient and scalable approach to quantum circuit optimization.

· Educational History

2006 한양대학교 이학사
2011 한양대학교 이학박사

· Employment History

2012~2015 싱가포르국립대학교 양자기술센터 리서치펠로우
2016~2020 독일칼스루헤공과대학교 이론고체물리연구소 연구원
2021 고등과학원 양자우주센터 QUC펠로우
2022~2025 한국표준과학연구원 선임/책임 연구원
2025~현재 한양대학교 물리학과 부교수



김준기
부교수
성균관대학교 양자정보공학과

Architecting Quantum Computers with Mobile Trapped Ions

There has been significant progress in building trapped-ion quantum computers with dozens of ion qubits, yet a clear blueprint for fully fault-tolerant, large-scale systems remains elusive. In this talk, I present a Shuttling-based Distributed Quantum Computing (SDQC) architecture that combines deterministic ion shuttling with distributed entanglement to interconnect multiple processor nodes while keeping data qubits stationary. Architecture-level modeling under realistic error and timing assumptions shows that SDQC enables nearly scale-independent logical clock speeds through aggressive pipelining while maintaining competitive logical error rates. We further evaluate representative workloads, including ECC and Fermi-Hubbard simulations, demonstrating efficient execution with moderate space-time overhead. These results highlight the importance of architectural co-design for scalable trapped-ion quantum computers.

· Educational History

2009.3 ~ 2017.8 서울대학교 물리학 박사
2005.3 ~ 2009.2 서울대학교 물리학 학사

· Employment History

2017.9 ~ 2018.7 서울대학교 자동화연구소 박사후 연구원
2018.9 ~ 2021.2 Duke 대학교 박사후 연구원
2021.2 ~ 현재 성균관대학교 양자정보공학과 부교수

권형한
선임연구원
KIST

Programmable Nonlinear Photonic Integrated Circuits for Photonic Quantum Computing

In this talk, I will present our recent advancements in programmable nonlinear photonic integrated circuits, aimed at translating table-top quantum optics into deployable hardware to address key scalability bottlenecks in quantum information processing. By exploiting the strong nonlinearity, low propagation loss, and fast electro-optic modulation of ferroelectric thin-film platforms such as lithium niobate and lithium tantalate, we establish a unified framework that bridges high-efficiency nonlinear optical interfaces with large-scale programmable architectures. First, I will discuss efficient nonlinear devices that serve as essential building blocks for atom-like quantum systems and bright quantum state generation. Next, I will introduce reconfigurable interferometric meshes co-integrated with high-density electronics, achieving high-speed and precise unitary transformations for quantum state preparation and manipulation. Finally, I will conclude with a future outlook on building deployable and scalable photonic quantum systems.

· Educational History

California Institute of Technology, USA
Ph.D. in Electrical Engineering (2016-2021)
M.S. in Electrical Engineering (2018)
Advisor: Andrei Faraon
Seoul National University, South Korea
B.S. in Electrical Engineering (2016)

· Employment History

Postdoc researcher (2021-2022) Advisor: Jelena Vuckovic
Stanford University, USA
Research Scientist/Senior Research Scientist (2022.10-2023.3/2023.3-Current)
Center for Quantum Technology at Korea Institute of Science and Technology, South Korea

김요셉
조교수
고려대학교 물리학과

초전도 양자컴퓨터의 발전 현황

최근 초전도 양자컴퓨터는 재료 공정, 회로 설계, 극저온 계측 기술의 발전에 힘입어 가장 성숙한 양자 하드웨어 플랫폼 중 하나로 자리잡고 있다. 본 발표에서는 초전도 큐비트 기반 양자 프로세서의 최신 발전 동향을 개괄하고, 양자 우월성(quantum advantage) 실험의 의의와 한계를 비판적으로 조망한다. 특히 현재 구현 가능한 큐비트 규모와 단일/이중 게이트 오류율 수준, 그리고 이들이 실제 알고리즘 수행에 미치는 영향을 정량적으로 살펴본다. 더 나아가 표면 부호(surface code)를 비롯한 양자 오류 정정 연구의 최근 성과를 소개하고, 논리적 오류율을 물리적 오류율 이하로 낮추기 위한 실험적 시도와 그 자원 오버헤드를 논의한다. 마지막으로 대규모 확장을 가로막는 핵심 과제를 정리하고 향후 연구 방향을 전망한다.

· Educational History

울산과학기술원 (학사- 소자물리 / 2010.03 – 2014.02)
포항공과대학교 (박사 – 물리학 / 2014.03 – 2020.08)

· Employment History

박사후 연구원, Lawrence Berkeley National Lab (2020.09 ~ 2021.10)
선임연구원, 한국과학기술연구원 양자정보연구단 (2021.11 ~ 2023.08)
조교수, 고려대학교 물리학과 (2023.09 ~ 현재)

한남식
교수
연세대학교 / 영국 케임브리지대학교

Quantum-AI 기반 신약 타겟 발굴

데이터 기반 방법론은 신약개발의 방식을 근본적으로 바꾸고 있다. 본 강연에서는 Quantum과 AI 멀티오믹스 통합 분석이 치료 타겟의 발굴과 검증을 어떻게 혁신하고 있는지 개관한다. 먼저 AI의 일반 원리가 신약개발에 어떻게 적용되는지를 폭넓은 청중이 이해할 수 있는 수준에서 설명한다.

이어서 양자컴퓨팅 기반 접근법을 소개한다. 양자워크 기반 네트워크 알고리즘은 생체분자 네트워크에서 기존 방법이 포착하지 못하는 숨은 기전적 연결을 드러낸다. 멀티오믹스 데이터와 결합하면 질병 기전의 이해와 신규 타겟 발굴에서 고전적 네트워크 분석과 차별화된 결과를 얻을 수 있다.

특히 현재 연세대학교에서 수행 중인 한국형 ARPA-H 프로젝트를 소개한다. 본 과제에서는 난치성 위암을 대상으로 양자워크 기반 네트워크 분석과 AI 기반 타겟 발굴 플랫폼을 병렬로 운영하고, 두 트랙의 결과를 교차 검증하여 신규 치료 타겟을 도출한다. 양자 알고리즘이 실제 신약개발 파이프라인에서 제공하는 고유한 가치와 실제 양자 하드웨어와의 연계 전략을 함께 논의한다.

마지막으로 CardiaTec Biosciences의 창업과 운영 경험을 바탕으로 학문적 성과가 산업 현장에서 어떻게 사업화되는지를 공유한다. 본 강연은 AI, 양자컴퓨팅, 데이터 기반 방법론이 신약개발의 정밀성과 효율성을 어떻게 높이고 있는지에 대한 전체적인 조망을 제공할 것이다.

· Educational History

박사후 (PostDoc), Computational Epigenomics, University of Cambridge
박사 (Ph.D.), Computational Molecular Biology, University of Manchester 공학석사 (MEng), 인하대학교 공학사 및 법학사 (BEng, LLB), 인하대학교

· Employment History

(현) 연세대학교 양자정보학과 교수
(현) 연세대학교 IBS 교수 (Advanced Science Institute)
(현) 연세대학교 인공지능학과 AI+X 교수, 약학대학 겸임교수
(현) 영국 케임브리지대학교 Milner Therapeutics Institute 그룹리더
(Computational Research & AI 총괄)
(현) 영국 케임브리지대학교 Cambridge Centre for AI in Medicine Faculty
(현) CardiaTec Biosciences Ltd 공동창업자, CTO

예기헌
상무
아이온큐

Quantum Inside the AI Data Center: IonQ's Hybrid QPU-GPU Accelerated Computing

The bottleneck in AI infrastructure has shifted from model architecture to physical constraints — power, cooling, and memory bandwidth. This talk redefines the role of the quantum processor at exactly that point. A QPU does not replace the GPU; it sits beside it and takes on only the calculations where classical scaling breaks down: high-dimensional optimization, quantum chemistry simulation, and QML kernels. With NVQLink and CUDA-Q enabling microsecond-scale GPU-QPU coupling, this architecture is already being deployed across supercomputing centers worldwide rather than merely theorized. We cover why room-temperature trapped-ion QPUs integrate into standard data center racks, IonQ's 99.99% two-qubit gate fidelity, and delivered hybrid results — including a 20x end-to-end acceleration of a drug-discovery workflow and drop-in QML fine-tuning for high-stakes, small-data regimes.

· Educational History

University of Waterloo 기계공학 박사 (2008)
York University 컴퓨터사이언스 석사 - AI Specialization (2021)

· Employment History

Lead HPC consultant, SHARCNET-Compute Canada (2008-2021)
Sr. Solutions Architect, Cerebras Systems (2021-2023)
Principal engineer, Technical AI solutions, HPE AMS (2023-2024)
Director, Global Infrastructure Sys. Engineering, Coupang (2024-2025)
present Director, Customer Success, IonQ (2026)

오성일
디렉터
콴델라 코리아

Quantum AI: GPU–QPU Integration and Hybrid QML Use Cases

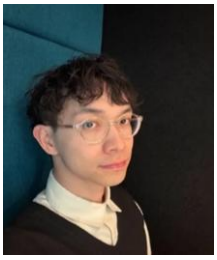
AI와 양자컴퓨팅의 융합이 가속되면서 GPU와 QPU를 하나의 연산 환경에서 활용하는 하이브리드 컴퓨팅이 새로운 Quantum AI 아키텍처로 주목받고 있다. Quandela는 자체 개발한 QML 프레임워크인 MerLin을 기반으로 기존 PyTorch AI 워크플로에 양자 연산을 통합하고, GPU와 광자 QPU가 연계되는 하이브리드 QML 환경을 구현하고 있다. 또한 GPU–QPU 간 연산 지연을 줄이고 QPU를 AI 워크플로의 실시간 가속기로 활용하는 구조를 제시한다. 주요 사례로는 국방 분야 Thales의 고기동 표적 실시간 추적 분석, 통신 분야 Orange의 QML 기반 악성코드 탐지·분류 고도화, 항공·모빌리티 분야 Airbus–BMW Group의 주간 이미지를 야간 이미지로 변환하는 QGAN 기반 합성 데이터 생성 사례를 공유한다. 이를 통해 HPC·AI·Quantum이 결합되는 차세대 이기종 컴퓨팅과 Quantum AI의 산업 활용 가능성을 소개한다.

· Educational History

국민대학교 비즈니스IT대학원 석사

· Employment History

Quandela Korea, Korea-Southeast Asia Sales Director
(현) 한국해군과학기술학회 이사, 대한의용생체공학회 회원
IDBS Korea Country Manager
Alfrica AI division Head, Korea-USA



Hao-Chen Yeh
Technical Sales Engineer
IQM Quantum Computers

The Era of the Quantum Supercomputer

In the post-AI era, quantum computing is expected to become both an important partner and a game changer for High-Performance Computing (HPC). As a leading full-stack quantum computing provider, IQM is committed to shaping the next generation of computational infrastructure. In this presentation, we will share key insights and practical experience in deploying on-premise quantum computing systems and integrating them with HPC environments (HPC+QC). We will explore the technical architecture of superconducting quantum processors and how hybrid algorithms and close integration with classical HPC workflows can help accelerate simulations. Furthermore, we will present real-world simulation results, demonstrate concrete application benchmarks, and introduce IQM's roadmap for scaling quantum computing in supercomputing environments. Attendees will gain a practical understanding of how hybrid classical-quantum systems can be delivered today and how they can help address next generation computational challenges.

· Educational History

PhD Candidate in Microsystems and Microelectronics - EPFL
Master of Science in Physics – National Taiwan University

· Employment History

Technical Sales Engineer / Solutions Consultant – IQM Quantum Computers
Doctoral Assistant – EPFL
Full-time Research Assistant – Academia Sinica

이우준
책임연구원
Pasqal

중성원자 양자컴퓨팅 플랫폼의 확장성 및 실용적 HPC-QC 하이브리드 연동

중성원자 양자 컴퓨팅의 독보적인 물리적 확장성과 임의 형태의 원자 배치 제어 기술을 소개한다. 이러한 하드웨어적 강점을 바탕으로 대규모 조합 최적화, 양자 화학, 그래프 머신러닝 등 실세계 난제를 해결하는 아날로그·디지털 모드의 실용적 활용 사례를 조명한다. 이어 중성원자 양자 프로세서를 heterogeneous 슈퍼컴퓨팅 인프라의 가속기로 통합하기 위한 하이브리드 양자-고전 HPC 시스템 구조를 제시한다. 기존 HPC 자원 관리 스택과의 런타임 미들웨어 통합 아키텍처를 다루며, 이러한 환경에서 실용적 양자 이점을 실현하기 위한 시스템 전략을 논의한다.

· Educational History

KAIST 물리학 박사 (양자컴퓨팅 연구실 / 2019)
KAIST 물리학 학사 (2013)

· Employment History

Postdoc, 서울대학교 양자정보 및 양자컴퓨팅 연구실 (2019 ~ 2024)
책임연구원 (기술사업개발), Pasqal Korea (2024 ~ 현재)

차세대 메모리·가속 컴퓨팅 기술과 클라우드 운영 혁신

13:30~14:10	**퓨리오사 레니게이드: 지속 가능한 프로그래밍형 AI 인프라** 이준원 이사 (퓨리오사에이아이)
14:10~14:50	**CXL 메모리와 Processing Near Memory가 여는 HPC 확장의 새로운 방향** 김주현 CPO/부사장 (XCENA)
14:50~15:30	**차세대 메모리 기반 AI 스토리지 아키텍처** 김경표 연구소장 (㈜글루시스)
15:30~16:00	Coffee Break
16:00~16:40	**AI 시대의 효율적인 GPU 자원 관리와 클러스터 설계** 강대명 SW엔지니어 (래블업)
16:40~17:20	**관측에서 통제로: AI 데이터센터 운영을 바꾸는 차세대 모니터링** 이성림 인프라팀장 (와탭랩스)
17:20~18:00	**UPMEM PIM 기반 응용 파이프라인 재설계: JPEG 디코딩과 k-mer Counting 사례** 남덕윤 교수 (경북대학교)

이준원
이사
퓨리오사에이아이

퓨리오사 레니게이드: 지속 가능한 프로그래밍형 AI 인프라

AI 모델의 성능 향상에는 더 많은 연산이 따르지만, 그 연산의 중심은 학습에서 추론으로 빠르게 이동하고 있다. 사전학습(pre-training)과 사후학습(post-training)의 연산은 모델을 만드는 과정에서 한 번 수행되지만, 모델이 더 오래 추론하도록 하는 Test-Time Scaling이나 여러 모델을 반복 호출하는 AI 에이전트에서는 사용자 요청마다 새로운 연산이 발생한다. 답변 하나에 소비되는 토큰 수가 수십 배에서 수백 배까지 증가하면서, AI 하드웨어의 평가 기준 역시 학습 처리량(training throughput)에서 토큰당 비용(cost per token), 전력 효율, 응답 지연(latency)으로 변화하고 있다.

이러한 추론 중심의 워크로드는 범용 GPU만으로는 비용과 전력 측면에서 비효율적인 경우가 많다. FuriosaAI의 RNGD는 AI 추론 연산의 대부분을 차지하는 Tensor Contraction을 가장 효율적으로 처리하도록 설계된 AI 추론 전용 프로세서이다. 본 발표에서는 이러한 설계 철학을 바탕으로, 실제 고객들과의 PoC 및 서비스 구축 과정에서 얻은 경험을 공유한다. 번역, 챗봇, 검색 증강 생성(RAG), 멀티모달 에이전트 등 다양한 AI 서비스를 RNGD 위에서 구현하며 확인한 성능과 전력 효율을 소개하고, GPU 기반 시스템과의 비교를 통해 추론 인프라의 총소유비용(TCO)과 지속가능성 측면에서 어떤 차이를 만들 수 있는지 살펴본다.

· Educational History

포항공과대학교 전기전자공학과 (2009 ~ 2013)
서울대학교 뇌인지과학과 (박사수료 / 2013)

· Employment History

㈜두다지, 연구소장 (2023 ~ 2025.1)
A2D2, CTO (2025.1 ~ 2025.3)
㈜퓨리오사에이아이, 솔루션 아키텍트 (2025.3 ~ 현재)

김주현
CPO/부사장
XCENA

CXL 메모리와 Processing Near Memory가 여는 HPC 확장의 새로운 방향

HPC 시스템의 성능 확장은 FLOPS를 중심으로 논의되어 왔으나, 실제 병목은 노드 내부의 메모리 용량과 대역폭, 그리고 노드 간 데이터 이동 비용으로 옮겨가고 있다. 현재의 scaling 구조는 대부분의 데이터 이동을 message passing 기반의 transfer semantics 위에서 처리하기 때문에, 논리적으로 동일한 데이터에 대해서도 노드 수만큼의 복제와 그에 수반되는 오버헤드를 반복적으로 지불하게 된다.

본 발표에서는 CXL을 이러한 제약에 대한 memory semantics 기반의 대안으로 소개한다. CXL 프로토콜과 memory sharing 모델의 기본 구조를 정리한 뒤, 메모리 확장이 HPC에 제공하는 기회를 용량 확장 및 tiering, 그리고 공유 메모리 기반의 데이터 접근으로 다시 설계하려는 최근 연구 동향과, 그 과정에서 드러난 가능성과 한계를 함께 다룬다.

이어서 Processing Near Memory를 소개한다. 메모리 device 내부에서 확보 가능한 대역폭은 host link 대역폭보다 크기 때문에, 데이터 이동이 지배적인 연산을 device 내부에서 처리하면 대역폭과 연산 능력을 함께 확장할 수 있다. 실제 CXL device 개발과 시스템 bring-up 경험을 바탕으로, 이러한 확장 방식이 기존 accelerator 중심 접근과 어떻게 다른지, 그리고 실제 시스템에 적용할 때 무엇을 먼저 검토해야 하는지를 논의한다.

· Educational History

서울대학교 전기공학부 학사/석사 (2007)

· Employment History

삼성전자 (2007~2014)
SK Telecom (2014~2017)
SK Hynix (2017~2022) (2022~)

김경표
연구소장
(주)글루시스

차세대 메모리 기반 AI 스토리지 아키텍처

AI 팩토리에서는 대규모 모델 학습과 추론이 동시에 수행되면서 GPU HBM 용량, 데이터 공급 지연, 체크포인트 처리 및 KV-cache 확장성이 전체 시스템 성능을 제한한다. 기존 스토리지 중심 구조는 GPU와 데이터 간 지연시간 차이를 효과적으로 해소하기 어려우며, 학습과 추론에 필요한 서로 다른 데이터 접근 특성을 통합적으로 지원하는 데 한계가 있다.

본 세션에서는 GPU HBM, 호스트 DRAM, CXL 메모리, 저지연 NVMe 및 분산 DAOS 스토리지를 계층적으로 결합한 메모리 중심 AI 스토리지 아키텍처를 제안한다. CXL 메모리는 학습 데이터 버퍼, 모델 상태, 공유 prefix 및 warm KV-cache를 위한 저지연 확장 계층으로 활용하며, DAOS는 모델, 데이터셋, 체크포인트 및 영속 KV-cache를 위한 확장형 공유 스토리지를 제공한다. GPUFlow 제어 계층은 워크로드 특성과 데이터 접근 빈도에 따라 각 계층의 데이터 배치, prefetch, eviction 및 이동을 관리한다.

제안 구조는 학습 환경에서 GPU 데이터 공급과 체크포인트 처리 성능을 높이고, 추론 환경에서는 KV-cache 재사용과 GPU 메모리 부족을 완화하는 것을 목표로 한다. 향후 TTFT, TPOT, 학습 처리량, GPU utilization, cache hit ratio, 계층 간 이동 지연 및 DAOS I/O 성능을 기준으로 효과를 검증할 예정이다. 이를 통해 CXL의 저지연 메모리 확장성과 DAOS의 분산 확장성을 결합한 차세대 AI 팩토리 인프라의 구현 가능성을 제시한다.

· Educational History

성균관대학교 정보공학과 학사 (2002)

· Employment History

고성능 네트워크 스토리지 아키텍처 개발 (2002~2012)
올플래시 스케일 아웃 스토리지 아키텍처 개발 (2013~2020)
AI 스토리지 기술 개발 (2022~ 현재) (2021~ 현재)

강대명
소프트웨어 엔지니어
래블업

AI 시대의 효율적인 GPU 자원 관리와 클러스터 설계

AI 워크로드의 급격한 증가로 GPU는 클러스터 환경에서 가장 중요한 컴퓨팅 자원이 되었다. 그러나 GPU는 CPU나 메모리와 달리 높은 비용과 다양한 하드웨어 제약을 가지며, 단순한 자원 할당만으로는 높은 활용률과 성능을 확보하기 어렵다.

본 세션에서는 AI 시대의 GPU 자원 관리 기술과 클러스터 설계 방법을 중심으로 Device Plugin, GPU Sharing(Time Slicing), MIG, Dynamic Resource Allocation(DRA) 등 최신 GPU 관리 기술을 소개한다. 또한 Container와 Virtual Machine 기반의 GPU 활용 방식, HPC 환경과 Kubernetes 기반 인프라의 운영 모델을 비교하며, AI 워크로드에 적합한 GPU 자원 관리 전략과 앞으로의 발전 방향을 살펴본다.

· Educational History

한국방송통신대학교 대학원 석사 (2018)
부산대학교 정보컴퓨터 공학부 (2004)

· Employment History

파이널데이터 (2002.11 ~ 2008.05)
네이버 소프트웨어 엔지니어 (2008.06 ~ 2012.02)
카카오 소프트웨어 엔지니어 (2013.08 ~ 2017.01)
유데미 데이터 엔지니어 (2017.02 ~ 2020.01)
위버스 데이터 엔지니어 (2020.02 ~ 2021.09)
레몬트리 CTO (2021.10 ~ 2025.02)
래블업 소프트웨어 엔지니어 (2025.07 ~ 현재)

이성림
개발그룹 인프라팀장
와탭랩스

관측에서 통제로: AI 데이터센터 운영을 바꾸는 차세대 모니터링

AI 서비스의 확산으로 GPU는 AI 인프라의 핵심 자원으로 자리잡았으며, 최근에는 Kubernetes, Bare Metal, Virtual Machine 등 다양한 환경에서 GPU 클러스터를 운영하는 사례가 빠르게 증가하고 있습니다. 또한 단일 클러스터를 넘어 여러 GPU 클러스터를 동시에 운영하는 환경이 늘어나면서, 인프라 전반의 상태를 통합적으로 관찰하고 운영하는 기술의 중요성이 더욱 커지고 있습니다.

이번 세션에서는 GPU 서버와 Kubernetes 환경을 중심으로 멀티 클러스터 환경에서의 통합 모니터링 방안을 소개하고, GPU 자원 활용 현황, 클러스터 상태, 애플리케이션 성능(APM), LLM 서비스까지 연결되는 End-to-End 관찰(Observability) 체계를 살펴볼 예정입니다. 또한 AI 기반 운영 기술을 활용하여 대규모 AI 인프라를 보다 효율적으로 운영하고 장애 대응 시간을 단축할 수 있는 방향을 실제 운영 사례와 함께 소개하고자 합니다.

이를 통해 참석자 여러분께서 GPU 기반 AI 인프라 환경에서 다양한 운영 데이터를 통합적으로 분석하고, 멀티 클러스터 환경의 운영 효율성과 서비스 안정성을 향상시키기 위한 실질적인 운영 방향을 얻으실 수 있도록 돕고자 합니다.

· **Educational History**
성결대학교 (2008)

· **Employment History**
GTPlus 미들웨어&APM 엔지니어 (2014~2019)
오픈소스컨설팅 프라이빗 클라우드 엔지니어 (2020~2025)
와탭랩스 컨설턴트&개발팀장 (2025~2026)

남덕윤
부교수
경북대학교 컴퓨터학부

UPMEM PIM 기반 응용 파이프라인 재설계: JPEG 디코딩과 k-mer Counting 사례

Processing-in-Memory(PIM)는 연산 기능을 메모리 내부 또는 인접 영역에 배치함으로써 프로세서와 메모리 사이의 데이터 이동을 줄이는 컴퓨팅 구조이다. 본 발표에서는 PIM 기술을 개괄하고, 실제 UPMEM PIM 시스템을 활용한 응용 파이프라인 최적화 연구를 소개한다. 먼저 메모리 장벽과 PIM의 기본 개념을 살펴보고, UPMEM의 DPU, MRAM, WRAM 및 tasklet 구조와 명시적 데이터 전송 방식을 설명한다. 이를 바탕으로 실제 PIM 응용에서는 연산 성능뿐 아니라 데이터 분할, 호스트-DPU 통신, 메모리 용량 및 실행 중첩을 함께 고려해야 함을 논의한다. 본론에서는 UPMEM PIM의 특성을 고려하여 응용 파이프라인을 재설계한 두 가지 연구를 소개한다. 첫 번째 연구인 NPD는 CPU에서 Huffman 디코딩을 수행하여 MCU 간 의존성을 제거하고, 독립된 MCU의 후속 연산을 여러 DPU에 분산하는 PIM 기반 JPEG 디코더이다. 두 번째 연구인 SCOPE는 k-mer counting 파이프라인의 Sorter 단계에서 CPU와 DPU가 함께 작업하는 shared-stage 공동 실행 구조를 제안한다. 두 연구는 PIM의 효과적인 활용을 위해 개별 커널의 성능뿐 아니라 CPU-DPU 역할 분담, 데이터 통신 및 실행 중첩을 응용 파이프라인 수준에서 함께 설계해야 함을 보여준다.

· Educational History

POSTECH 컴퓨터공학과 학사 (1999)
KAIST 공학 석사 (2001)
KAIST 공학 박사 (2006)

· Employment History

한국과학기술정보연구원 슈퍼컴퓨터SW연구실 실장 (2016~2018)
한국과학기술정보연구원 선임연구원 (2004~2022)
경북대학교 컴퓨터학부 조교수 및 부교수 (2022~현재)

초고성능컴퓨팅 SW생태계조성사업 워크숍 (비공개)

13:30~14:00	**HPC 향 탐색 공간 통합 오토튜닝 프레임워크 개발** 성효진 교수 (서울대학교)
14:00~14:30	**6호기 한강 시스템 소개 및 태스크 기반 MPI 최적화 연구 소개** 권오경 팀장 (한국과학기술정보연구원)
14:30~15:00	**컴파일 타임 전력 모델링 및 최적화 기술** 송진호 부교수 (연세대학교)
15:00~15:30	**Optimizing MPI for the Exascale Computing Era** 진현욱 교수 (건국대학교)
15:30~16:00	Coffee Break
16:00~16:30	**동적 HPC 환경을 위한 불확실성 기반 적응형 스케줄링 프레임워크** 오상윤 교수 (아주대학교)
16:30~17:00	**Auto-Tuning the HPC I/O Stack: Toward Resource-Efficient Parallel I/O at Exascale** 박지성 교수 (포항공과대학교)
17:00~17:30	**토론**

성효진
교수
서울대학교

HPC 향 탐색 공간 통합 오토튜닝 프레임워크 개발

이 세션에서는 HPC 향 오토튜닝 프레임워크의 탐색 공간을 확장하는 최근 연구를 소개하고자 합니다. OpenMP 프로그래밍 모델을 활용하여 반복문 최적화의 탐색 공간을 정의하고 이를 효율적으로 탐색하는 오토튜닝 메커니즘을 제안했던 기존 논문 (HPDC 2025)을 확장하여, 응용과 알고리즘에 특화된 고차원 파라미터 공간과 반복문 최적화와 관련된 저차원 파라미터 공간을 따로 탐색하는 대신 함께 고려하는 통합 오토튜닝 (joint autotuning) 공간에서 오토튜닝을 효율적으로 수행할 수 있는 기법을 제시합니다.

- **Educational History**
  서울대학교 영어영문학과 학사 (2000)
  서울대학교 컴퓨터공학 학사 (2004)
  UC San Diego 컴퓨터공학 석사 (2008)
  UIUC 컴퓨터공학 박사 (2015)
- **Employment History**
  삼성전자 디지털 미디어 연구소 (2004-2006)
  IBM T. J. Watson Research Center (2015-2019)
  포항공대 컴퓨터공학과 교수 (2019-2023)

권오경
팀장
한국과학기술정보연구원

6호기 한강 시스템 소개 및 태스크 기반 MPI 최적화 연구 소개

한국과학기술정보연구원(KISTI)이 구축하고 있는 국가슈퍼컴퓨터 6호기 '한강'은 614.7페타플롭스의 초고성능 슈퍼컴퓨터입니다. 8,336개의 엔비디아 GH200 GPU로 이루어져 있는 GPU 파티션과 804개의 AMD EPYC CPU로 이루어져 있는 CPU 파트션으로 구성되어 있습니다. 대규모 연산에 최적화된 DragonFly 토폴로지 구조의 초고속 네트워크와 205페타바이트 규모의 저장 공간을 제공하며, 직접수냉(DLC) 방식을 적용해 에너지 효율(PUE 1.15)을 극대화한 것이 특징입니다.

국가슈퍼컴퓨터 6호기 네트워크 토폴로지 환경에 최적화된 통신 성능 예측 및 작업 매핑 모델을 제안한다. 기존의 특정 계층 구조에 종속되었던 통신 평가 지표의 한계를 극복하기 위해, 시스템의 하드웨어 토폴로지 정보(물리적 거리)와 애플리케이션의 실제 통신 패턴(데이터 볼륨)을 통합 분석하는 범용적 평가 지표를 도입하였다. 특히 본 모델은 컴퓨팅 자원의 최적 할당을 위한 복잡한 탐색 및 예측 과정을 고속 연산 기반의 '오프라인 튜닝(Offline Tuning)' 방식으로 분리하여 수행한다. 사전에 도출된 최적의 매핑 정보는 애플리케이션 실행 초기 단계에 즉각적으로 반영된다. 대규모 병렬 컴퓨팅 환경에서 애플리케이션의 통신 효율 및 전체 시스템 성능을 극대화할 수 있는 기반을 마련하였다.

· Educational History

고려대학교 컴퓨터학과 학사(2000)
카이스트 전산학과 석사 (2002)
카이스트 전산학부 박사 (2022)

· Employment History

한국과학기술정보연구원 국가슈퍼컴퓨팅본부(2001~)

송진호
부교수
연세대학교

컴파일 타임 전력 모델링 및 최적화 기술

This talk presents a fast and accurate compile-time power prediction model for machine learning (ML) applications. ML compilers are essential for deploying ML models efficiently across diverse hardware, yet they predominantly optimize for execution latency while overlooking power consumption, which is a growing concern in computing systems. Unlike latency, power dissipation arises from highly nonlinear interactions among computational intensity, memory accesses, parallel execution structures, and hardware-specific dynamics. Capturing these effects from static code-level features is highly challenging and has hindered practical power-aware compilation. To address this, Eyas leverages code features derived from intermediate representations (IRs) that describe ML layer transformations during compilation. These features are aggregated and used to train gradient-boosted trees (GBTs), learning complex relationships between code properties and measured hardware power. Once trained offline, Eyas provides rapid, accurate compile-time power estimation using only code-level features. Experiments on diverse platforms, including CPU, GPU, and embedded devices, show an average prediction error of 6.77% with under 0.4ms inference latency. The practical utility of the developed compile-time power model is demonstrated by integrating it as a low-overhead cost model into a state-of-the-art ML compiler, enabling power-aware optimization, such as power capping and energy minimization.

· Educational History

B.S., Yonsei University, Seoul, South Korea (2007)
Ph.D., Georgia Institute of Technology, Atlanta, GA, USA (2015)

· Employment History

Graduate Intern, Sandia National Labs, Albuquerque, NM, USA (2010-2012)
Graudate Co-op, AMD Research, Bellevue, WA, USA (2013)
Graduate Co-op, IBM TJ Watson Research Center, Yorktown Heights, NY, USA (2014)
Graduate Intern, Qualcomm, San Diego, CA, USA (2015)
Senior Engineer, Intel, Santa Clara, CA, USA (2016)

진현욱
교수
건국대학교

Optimizing MPI for the Exascale Computing Era

엑사스케일 컴퓨팅 시스템은 200 메가와트 이상의 전력을 요구하며, GPU, NPU, DPU 등의 다양한 계산자원을 제공하고 있으며, 다수의 계산 노드와 노드당 CPU 코어를 장착하고 있다. 따라서 엑사스케일 컴퓨팅 시스템은 높은 에너지 효율성, 이종 계산 자원의 효율적 활용, 노드내/노드간 인지 프로그래밍 모델을 위한 기술개발을 요구한다.

본 발표에서는 엑사스케일 컴퓨팅 시스템을 위한 MPI의 에너지 효율성 개선, DPU 기반의 MPI 오프로딩, MPI+X 하이브리드 병렬 프로그래밍 모델을 위한 최근 연구 결과를 소개한다.

· Educational History

고려대학교 전산학 학사 (1997)
고려대학교 컴퓨터학 석사 (1999)
고려대학교 통신시스템공학 박사 (2003)

· Employment History

오하이오 주립 대학교 박사후 연구원 (2003~2006)
건국대학교 조교수/부교수/교수 (2006~현재)

오상윤
교수
아주대학교 소프트웨어학과

동적 HPC 환경을 위한 불확실성 기반 적응형 스케줄링 프레임워크

초고성능컴퓨팅 시스템에서는 워크로드 구성, 실행시간 분포, 대기열 상태와 운영 목표가 지속적으로 변화한다. 이러한 환경에서 point prediction과 고정된 스케줄링 정책만으로는 job termination risk와 자원 활용 효율을 동시에 관리하기 어렵다. 특히 runtime underprediction은 job termination으로 이어지는 반면, overprediction은 backfill 기회를 축소하므로 scheduler 관점의 risk control이 필요하다.

본 발표에서는 runtime uncertainty를 scheduling decision에 반영하는 연구를 소개한다. 개별 job의 tail risk, 시간에 따른 workload drift와 delayed feedback, runtime 구간별 coverage imbalance를 고려하여 safe reservation length를 결정하는 방법을 설명하며, 다양한 워크로드와 운영 목표에 따라 scheduling policy를 적응적으로 선택하는 uncertainty-aware reinforcement learning 기반 최적화 스케줄러를 소개한다. 또한, 초고성능 최적화 프레임워크의 공통 repository를 지원하기 위한 BitLSM 기반 secondary indexing 기술을 간략히 소개한다.

본 발표를 통해 동적인 HPC 환경에서 job reliability, waiting performance, resource utilization을 균형 있게 고려하는 최적화 스케줄러의 설계 방향을 제시한다.

· Educational History

성균관대학교 기계설계학 학사 (1995)
미) 시러큐스대학교 컴퓨터공학 석사 (1999)
미) 인디애나대학교 컴퓨터공학 박사 (2006)

· Employment History

SK텔레콤 전략기술부문 (2006~2007)
아주대학교 소프트웨어학과 (2007~현재)

박지성
부교수
포항공과대학교 컴퓨터공학과

Auto-Tuning the HPC I/O Stack: Toward Resource-Efficient Parallel I/O at Exascale

엑사스케일 컴퓨팅 시스템의 등장에 발맞추어, 양자컴퓨팅 시뮬레이션을 포함한 전례없는 첨단 HPC 응용들이 대거 등장하고 있다. 반면, 개발된 HPC 응용에 대해 최적 I/O 성능을 제공할 수 있는 시스템 설정값을 도출해내는 작업은 여전히 남아있는 초고성능컴퓨팅의 전통적 난제로, 병렬 저장장치 시스템 전반에 대한 전문가적 지식과 많은 계산자원이 소요되는 소모적 작업이다. 이러한 관점에서, 다계층 HPC I/O stack의 파라미터 조합을 자동으로 탐색해 병렬 I/O 성능을 최적화할 수 있는 오토튜닝 프레임워크 개발은 전문가적 지식에 무관하게 HPC 응용에 대한 고성능 I/O를 보장하고 병렬 저장장치 시스템의 자원 효율성을 크게 증대시킬 수 있다는 의의가 있다.

이러한 관점 아래, 본 발표에서는 개발 중인 다계층 병렬 I/O 오토튜닝 프레임워크 관련 연구 결과를 공유한다. 추가적으로, 병렬 저장장치 시스템의 자원 효율성을 함께 증가시킬 수 있는 강화학습, 마르코브 체인 기반 데이터 배치 최적화 관련 최신 연구 성과들을 함께 소개한다.

· Educational History

서울대학교 컴퓨터공학과 학사 (2011)
서울대학교 전기전자및컴퓨터공학과 박사 (2019)

· Employment History

취리히연방공과대학교 박사후연구원 (2019.09~2022.08)
포항공과대학교 컴퓨터공학과 조교수 (2022.09~2026.08)
포항공과대학교 컴퓨터공학과 부교수 (2026.09~현재)

DAY 1

Conference

Tuesday September 08, 2026

KOREA SUPERCOMPUTING CONFERENCE 2026

The Intelligent Frontier: Synergizing HPC, AI and Quantum

2026년 한국계산과학공학회 추계학술대회

13:30~15:30	**2026년 한국계산과학공학회 추계학술대회**
15:30~16:00	Coffee Break
16:00~18:00	**2026년 한국계산과학공학회 추계학술대회**

DAY 2

Workshop - AM

Wednesday, September 09, 2026

KOREA SUPERCOMPUTING CONFERENCE 2026

The Intelligent Frontier: Synergizing HPC, AI and Quantum

HPC 활용을 위한 대규모 데이터 처리 기술 워크숍

10:00~10:40	**대규모 AI 인퍼런스를 위한 Shared KV Cache 아키텍처와 성능 최적화** 김수현 이사 (DDN)
10:40~11:00	Coffee Break
11:00~11:40	**Hybrid Computing on AWS: Scale Beyond Your Data Center** Robert M, Bata Research Lead for APJ (AWS)

김수현
이사
DDN

대규모 AI 인퍼런스를 위한 Shared KV Cache 아키텍처와 성능 최적화

생성형 AI 서비스의 확산과 함께 대규모 LLM 추론 환경에서 성능과 비용 효율성을 동시에 확보하는 것이 중요한 과제로 대두되고 있습니다. 특히 긴 컨텍스트를 사용하는 AI 서비스에서는 KV Cache가 GPU 메모리의 상당 부분을 차지하며, 동일하거나 유사한 프롬프트에 대한 중복 계산으로 인해 자원 활용 효율이 저하될 수 있습니다.

본 발표에서는 대규모 AI 인퍼런스 환경에서 Shared KV Cache 아키텍처를 활용하여 이러한 문제를 해결하는 방안을 소개합니다. 공통 프리픽스와 컨텍스트를 여러 요청 간에 공유함으로써 토큰 재계산을 최소화하고, GPU 메모리 사용량을 절감하며, 처리량을 향상시키는 방법을 살펴봅니다.

또한 분산 추론 환경에서의 KV Cache 관리 방식, 고성능 네트워크를 활용한 캐시 공유 구조, 그리고 캐시 적중률 향상을 위한 최적화 기법을 소개합니다. 실제 적용 사례를 바탕으로 지연 시간 감소, GPU 활용도 향상, 인프라 비용 절감 효과를 분석하며, 대규모 AI 서비스 구축을 위한 실질적인 아키텍처 설계 방향과 운영 전략을 제시하고자 합니다.

· Educational History

숭실대학교 학사 (2007)
숭실대학교 석사 (2009)

· Employment History

2009 ~ 2013 : 한국과학기술정보연구원, 연구원
2013 ~ : DDN Storage, Senior Solution Architect, Manager T3S

Robert M, Bata
Research Lead for APJ
(연구 부문 APJ 총괄)
AWS

Hybrid Computing on AWS: Scale Beyond Your Data Center

Unleash limitless compute power by seamlessly extending your on-premises workloads into AWS, dynamically scaling for peak demands while keeping sensitive data exactly where it belongs.

· **Educational History**

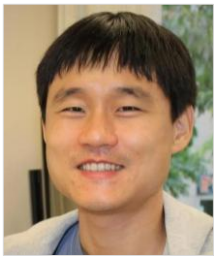
Governors State University and DePaul University

· **Employment History**

Aon Corporation (1994 – 2002)
Chubbs (2002 – 2006)
TWG (2006 – 2017)
AWS(Amazon Web Services) Research Lead for APJ (2017~)

HPC 수치모의실험 천문학 워크숍

10:00~11:10	**전산유체역학 수치모의실험을 통한 별 형성 연구 소개** 문상혁 박사후연구원 (한국천문연구원)
11:10~11:20	Coffee Break
11:20~12:00	**국가 슈퍼컴퓨터 6호기를 위한 비표준 우주론 시뮬레이션: 암흑물질·암흑에너지·수정중력** 김주한 연구교수 (고등과학원 거대수치계산연구센터)



문상혁
박사후연구원
한국천문연구원

전산유체역학 수치모의실험을 통한 별 형성 연구 소개

별과 별 사이의 공간은 기체/플라즈마 상태의 수소, 헬륨 및 소량의 중원소들로 가득 차 있으며, 이러한 "성간물질"이 뭉쳐 새로운 별이 만들어지기도 합니다. 다른 학문 분야에서 찾아볼 수 없는 천문학적 유체만의 독특한 특징이 있다면 바로 유체 자체의 만유인력이 무시할 수 없을 만큼 강하며, 밀도가 낮고 냉각이 빨라 매우 잘 압축된다는 점입니다. 때문에 별의 탄생은 초음속 난류와 중력을 포함한 복잡한 자기유체역학적 과정을 거치며 이를 이해하려면 슈퍼컴퓨터를 동원한 수치모의실험이 필수적입니다. 이 발표에서는 별의 "출생률" 및 "출생 질량 분포"에 관해 저와 공동 연구자들이 수행하고 있는 은하 규모의 거시적 연구 및 개별 별 탄생 영역의 미시적 연구 두 가지를 소개함으로서 천문학에 슈퍼컴퓨팅이 쓰이는 한 가지 예시를 보여드리고자 합니다.

· **Educational History**
서울대학교 천문학 박사 (2022)

· **Employment History**
프린스턴대학교 박사후연구원 (2022~2025)
한국천문연구원 박사후연구원 (2025~)

김주한
연구교수
고등과학원 거대수치계산연구센터

국가 슈퍼컴퓨터 6호기를 위한 비표준 우주론 시뮬레이션: 암흑물질·암흑에너지·수정중력

표준 ΛCDM 우주론은 우주 거대구조를 성공적으로 설명하지만, 암흑물질과 암흑에너지의 물리적 정체는 아직 밝혀지지 않았다. 본 발표에서는 자기상호작용 암흑물질(SIDM), 동역학적 암흑에너지(DE), 비표준 중력이론(nGR), 퍼지 암흑물질(FDM), 원시 블랙홀(PBH)을 하나의 우주론적 중력·유체역학 시뮬레이션 틀에서 연구하는 개발 현황을 소개한다.

기존 RAMSES 계열 코드를 KISTI 슈퍼컴퓨터 6호기의 NVIDIA GH200 환경에서 GPU 가속과 대규모 병렬 실행이 가능하도록 수정·최적화하였다. 이를 이용해 각 모형이 물질의 비선형 군집, 암흑물질 헤일로, 우주 거대구조와 파워스펙트럼에 남기는 차이를 정량화하고, 암흑물질과 암흑에너지의 성질을 연구하고 있다.

· **Educational History**

서울대학교 천문학 박사 (2003년)

· **Employment History**

한국천문연구원 박사후연구원 (2003~2004)
고등과학원 연구원 (2004~2008)
캐나다 이론천체물리연구소(CITA) 방문박사후연구원 (2008~2009)
경희대학교 연구교수 (2009~2010)
고등과학원 거대수치계산연구센터 연구교수 (2010~현재)

Quantum Intelligence: 양자-AI 융합과 신약·바이오 혁신

10:00~10:25	**Quantum state tomography with generative diffusion model** 이창협 부교수 (한양대학교)
10:25~10:50	**조합최적화 문제 해결을 위한 양자 알고리즘과 응용 전망** 최민진 선임연구원 (한국과학기술정보연구원)
10:50~11:15	**난공략성 표적 치료제 발굴을 위한 하이브리드 양자-AI 플랫폼 Q-DrugX** 이수민 연구센터장／전무이사 (삼진제약)
11:15~11:40	**양자최적화알고리즘을 위한 인수분해 기계학습 모델 연구** 이응규 부교수 (경희대학교)
11:40~12:05	**하이브리드 양자컴퓨팅 활용: 알고리즘 동향과 응용 전망** 백경현 조교수 (연세대학교 융합과학기술원)

이창협
부교수
한양대학교 물리학과

Quantum state tomography with generative diffusion model

Quantum state tomography (QST) is essential for characterizing quantum systems but becomes increasingly challenging in the presence of noise and limited measurement resources. In this talk, we present a generative diffusion model approach for QST that reconstructs quantum states from noisy and incomplete measurement data. Numerical results demonstrate improved reconstruction accuracy and robustness, highlighting the potential of diffusion models for scalable quantum state characterization.

· Educational History

한양대학교 물리학과 석사 (2001~2006)
한양대학교 물리학과 공학박사 (2006~2011)
싱가폴 National University of Singapore (2012~2015)

· Employment History

싱가폴 National University of Singapore 연구원 (2012~2015)
독일 Karlsruhe Institute of Technology 선임연구원 (2016~2020)
QUC fellow (2020~2021)
한양대학교 부교수 (2025~ 현재)

최민진
선임연구원
한국과학기술정보연구원

조합최적화 문제 해결을 위한 양자 알고리즘과 응용 전망

이번 세션에서는 물류·배송 경로 최적화, 자원 배분, 신약 후보물질 탐색 등 다양한 산업 및 과학 분야에서 공통적으로 등장하는 조합최적화 문제를 양자컴퓨팅으로 어떻게 접근할 수 있는지 소개합니다. 이론적으로 이차 양자 속도 향상을 제공하는 그로버 알고리즘(Grover's algorithm)부터, 현재의 노이즈가 있는 양자 하드웨어에서도 활용 가능한 양자 근사 최적화 알고리즘(QAOA)까지, 대표적인 양자 알고리즘의 기본 원리와 동작 방식을 살펴봅니다.

또한 실제 조합최적화 문제에는 다양한 제약조건이 존재하는데, 이를 양자 알고리즘에 효과적으로 반영하기 위한 방법을 소개하고, 외판원 문제(Traveling salesman problem) 등의 예시를 통해 설명합니다. 이를 통해 참석자들이 조합최적화 문제 해결을 위한 양자 알고리즘의 핵심 개념을 이해하고, 각자의 연구 및 업무 분야에 어떻게 적용할 수 있을지에 대한 아이디어를 얻을 수 있도록 돕고자 합니다.

· **Educational History**
경희대학교 수학과 학사 (2014)
경희대학교 수학과 박사 (2021)

· **Employment History**
큐노바 컴퓨팅 (2021~2022)
한국과학기술정보연구원 (2023~)

이수민
연구센터장(전무이사)
삼진제약㈜

난공략성 표적 치료제 발굴을 위한 하이브리드 양자-AI 플랫폼 Q-DrugX

신약개발에서는 방대한 화학공간을 탐색하고 단백질과 후보물질 간의 복잡한 상호작용을 정밀하게 예측해야 하지만, 기존의 고전 계산만으로는 정확도와 계산 효율을 동시에 확보하는 데 한계가 있습니다. 이번 발표에서는 이러한 한계를 보완하기 위해 양자컴퓨팅과 인공지능을 결합한 하이브리드 신약개발 플랫폼 Q-DrugX를 소개합니다.

Q-DrugX는 AI 기반 가상 스크리닝, 양자 도킹, 생성형 신약설계 및 약물물성 예측을 통합하여 후보물질의 탐색부터 최적화까지 지원하는 플랫폼입니다. 특히 난공략성 암 대사 표적에 적용하여 저해 유효물질을 발굴한 초기 연구결과를 소개하고, 양자-AI 융합기술의 신약개발 활용 가능성과 향후 발전 방향을 제시하고자 합니다.

· **Educational History**

서울대학교 동물과학과 학사 (2002)
서울대학교 분자생물학 석사 (2004)
University of California, Irvine 약리·독성학 박사 (2016)

· **Employment History**

SK케미칼 차장(책임연구원) (2004~2019)
SK케미칼 평가분석팀장(겸직) (2019~2021)
SK케미칼 Open R&D TF장 (2019~2022)
SK케미칼 오픈이노베이션팀장 (2022)
삼진제약 연구센터장(상무이사) (2022~2024)
삼진제약 연구센터장(전무이사) (2025~현재)

이응규
부교수
경희대학교 전자공학과

양자최적화알고리즘을 위한 인수분해 기계학습 모델 연구

본 세미나에서는 양자 컴퓨팅 기반 이진 최적화(Binary Optimization)를 위한 인수분해 기계학습 모델 연구를 소개하고자 합니다. 양자 컴퓨팅을 통해 이진 최적화 문제를 해결하기 위해서는 해밀토니언(Hamiltonian) 구축이 필수적입니다. 해밀토니언은 이진 변수 간의 상호작용 값을 수치화하여 정의할 수 있으며, 그 방법론은 매우 다양합니다. 본 연구에서는 인수분해 모델을 활용하여 기존 방식보다 유연하고 실용적인 수치화 방법론을 제안하였습니다. 본 세미나에서는 1) 인수분해 기계학습 모델의 수학적 기초 이론, 2) 이를 활용한 양자 최적화의 실용적 적용 사례를 중심으로 발표할 예정입니다.

· Educational History

경희대학교 물리학과 이학사 (2002~2009)
서울대학교 나노과학기술 공학박사 (2009~2015)
미국노터데임대학교 기계항공공학부 박사후 연구원 (2015~2019)

· Employment History

미국노터데임대학교 기계항공공학부 연구교수 (2019~2020)
국립금오공과대학교 기계시스템공학과 조교수 (2020~2021)
경희대학교 전자공학과 조교수 (2021~2025)
경희대학교 전자공학과 부교수 (2025~현재)

백경현
조교수
연세대학교 융합과학기술원

하이브리드 양자컴퓨팅 활용: 알고리즘 동향과 응용 전망

이번 세션에서는 현 세대 양자 하드웨어의 큐비트 수, 오류율 등 물리적 한계를 고전 컴퓨팅 자원과 결합하여 보완하는 하이브리드 양자-고전 컴퓨팅의 개념과 알고리즘 동향을 중심으로 말씀드릴 예정입니다. 특히 VQA(Variational Quantum Algorithm), SQD(Sample-based Quantum Diagonalization)와 같은 대표적인 하이브리드 알고리즘이 어떤 구조로 동작하며, 최적화 문제나 물리·화학 시스템 시뮬레이션 등 다양한 워크로드에 어떻게 적용되고 있는지를 양자컴퓨팅의 역할과 함께 사례 중심으로 소개해드릴 계획입니다. 이를 통해 양자컴퓨팅이 현재 어느 수준까지 발전하였으며, HPC와의 시너지를 통해 앞으로 어떤 방향으로 발전해 나갈지에 대한 전망을 공유하고자 합니다.

· **Educational History**

서강대학교 물리학 학사 (2012)
서강대학교 물리학 박사 (2018)

· **Employment History**

ETRI (2020~2024)
연세대학교 (2024~)

양자-고전 하이브리드 계산 방법론 및 적용

10:00~11:00	**양자컴퓨팅 기초 및 Amazon Braket을 활용한 IonQ 양자컴퓨터 접속** 강유성 박사 (메가존 클라우드)
11:00~12:00	**QPE 및 IQPE** 강유성 박사 (메가존 클라우드)

강유성
Senior Consultant
MegazoneCloud

양자-고전 하이브리드 계산 방법론 및 적용

오류 보정이 완비되지 않은 현 단계의 양자 하드웨어에서 의미 있는 계산을 수행하려면, 양자와 고전 자원을 어떻게 분담시킬 것인가가 핵심 설계 문제가 된다. 본 강의는 이러한 양자-고전 하이브리드 관점에서 위상추정 계열 (QPE)과 변분 (VQE)·샘플링 (SQD) 기반 알고리즘의 원리를 정리하고, 클라우드를 통해 이온트랩 양자컴퓨터에 접속하여 분자 전자구조 문제에 적용하는 과정을 실습한다. 이를 통해 참가자가 현재 하드웨어 수준에서 실현 가능한 계산의 범위와 한계를 스스로 판단할 수 있도록 하는 것을 목표로 한다.

· **Educational History**

2016 연세대학교 물리학과 학사
2025 연세대학교 물리학과 박사

· **Employment History**

2025~현재 메가존클라우드 Senior Consultant

국가슈퍼컴퓨터 6호기 '한강' 및 시범·공식서비스(혁신지원사업) 설명회

10:30~10:40	**인사말** 박찬열 본부장 (한국과학기술정보연구원)
10:40~11:00	**국가슈퍼컴퓨터 6호기 '한강' 소개** 강지훈 센터장 (한국과학기술정보연구원)
11:00~11:30	**6호기 시범·공식서비스(혁신지원사업) 안내** 권오경 팀장 (한국과학기술정보연구원)
11:30~12:00	**질의 응답**

강지훈
슈퍼컴퓨팅응용지원센터장
한국과학기술정보연구원

국가슈퍼컴퓨터 6호기 '한강' 소개

국가슈퍼컴퓨터 6호기 '한강'은 한국과학기술정보연구원(KISTI)이 구축한 국내 최고 수준의 초고성능컴퓨팅 시스템이다. 이론 성능은 614.7페타플롭스(FP64기준)이며, 크게 NVIDIA GH200 (Grach-Hopper) 슈퍼칩 8,336개로 구성된 GPU 파티션과 AMD EPYC 프로세서 1,608개를 탑재한 CPU 파티션으로 구성되어 있다.노드당 최대 ,600Gbps에 이르는 HPE 슬링샷 인터커넥트와 Non-blocking 드래곤플라이 토폴로지를 적용해 대규모 병렬 연산 시 통신 지연을 최소화했으며, 205페타바이트 규모의 러스터 병렬 파일시스템으로 대용량 데이터 처리를 지원한다.

특히 계산노도의 직접수냉(DLC) 방식을 채택해 PUE 1.15의 높은 에너지 효율을 달성했다. 서울을 가로지르는 한강처럼 국가 연구개발 전반에 컴퓨팅 자원을 공급한다는 의미를 담은 이 시스템은 국가 전략기술 확보, 사회문제 대응, 산학연 전 분야 연구, 중소기업 인큐베이팅 등 다양한 목적으로 국가 연구개발을 지원할 예정이다.

반도체·피지컬 AI·AI 데이터센터를 축으로 한 '대한민국 대도약' 전략 속에서, 국가슈퍼컴퓨터 '한강'은 대한민국이 세계 3대 인공지능 강국으로 도약하는 여정에 마르지 않는 연산의 물줄기를 대는 국가 핵심 인프라가 될 것으로 기대된다.

· **Educational History**

카이스트 기계공학과 박사 (2007)
카이스트 기계공학과 석사 (2002)
카이스트 기계공학과 학사 (2000)

· **Employment History**

삼성전자 (2006 ~ 2009)
한국과학기술정보연구원 국가슈퍼컴퓨팅본부 (2009 ~ 현재)

권오경
팀장
한국과학기술정보연구원

6호기 시범·공식서비스(혁신지원사업) 안내

본 세션에서는 국가슈퍼컴퓨팅 자원을 효율적으로 배분하여 과학기술 및 산업 전반의 혁신을 촉진하기 위한 '혁신지원사업'의 세부 구성, 운영 체계 및 일정을 제시한다.

본 사업은 지원 목적과 신청 대상의 특성을 반영하여 다음 세 가지 핵심 프로그램으로 세분화되어 운영된다.

첫째, 'R&D혁신 프로그램'은 산·학·연을 대상으로 계산과학 및 데이터 분석 중심의 과제형 연구를 지원하며, 연구 규모와 목적에 따라 거대연구, 창의연구, 탐색연구의 세 개 분야로 구성된다.
둘째, '산업혁신 프로그램'은 중견 및 중소기업의 산업적 활용과 상용화 연구를 촉진하기 위한 기업 주도형 지원 트랙이다.
셋째, '국가전략 프로그램'은 AI 시대의 국가 과학기술 경쟁력 도약을 위한 K-문샷(K-Moonshot) 추진전략의 일환으로, 국가 과학기술 AI 자원과 역량을 결집하여 국가적 미션을 해결하는 대규모 프로젝트를 지원한다.

혁신지원사업은 이들 3개 프로그램의 고유한 목적, 신청 대상 및 연구수행 특성을 다각적으로 고려하여 공모 방식, 선정 평가, 지원 기간, 모니터링 및 자원 환류 기준을 차등 적용하는 맞춤형 지원 체계를 채택하고 있다. 이를 통해 국가 슈퍼컴퓨팅 자원의 활용 가치를 극대화하고, 전 국가적 차원의 연구개발 역량 및 산업 경쟁력 강화에 기여하고자 한다.

· Educational History
고려대학교 컴퓨터학과 학사(2000)
카이스트 전산학과 석사 (2002)
카이스트 전산학부 박사 (2022)

· Employment History
2001 ~ 현재: 한국과학기술정보연구원 국가슈퍼컴퓨팅본부(2001~)

한국슈퍼컴퓨팅센터 연합 회의

10:00~10:15	**한국핵융합에너지연구원 현황 소개** 권재민 본부장 (한국핵융합에너지연구원)
10:15~10:30	**울산과학기술원 현황 소개** 김용휘 팀장 (울산과학기술원)
10:30~10:45	**자율주행초고성능컴퓨팅전문센터: HPC-AI 공용인프라를 활용한 자율주행 커뮤니티 협력** 김종원 교수 (광주과학기술원원)
10:45~11:00	**IBS HPC 공동활용 인프라 구축 및 국가 공동활용 네트워크로의 확장** 허무영 책임 (기초과학연구원)
11:00~12:00	**KSCA 운영위원회 (closed)**
12:00~13:00	Luncheon

권재민
본부장
한국핵융합에너지연구원

한국핵융합에너지연구원 현황 소개

한국핵융합에너지연구원은 핵융합·가속기 분야 전문센터로 핵융합 시뮬레이션·가상화 기술개발, KSTAR, ITER 연구를 위한 데이터 서비스, 원격 실험을 위한 가상 환경 제공, 가속기 장치 연구자를 위한 자원 제공 등의 역할을 수행하고 있다.

본 발표에서는 한국핵융합에너지연구원의 슈퍼컴퓨팅 시뮬레이션 기술 개발과 이를 활용한 한국형 혁신 핵융합로 설계에 대해 소개한다. 특히 한국형 혁신 핵융합로를 위해 개발 중인 AI 기반의 설계·제어 기술과, 해당 기술의 기반이 되는 슈퍼컴퓨팅 시뮬레이션을 통한 고품질 데이터 생산·활용에 대해 소개한다.

이상의 연구개발은 우리나라 거대 과학 분야 연구 역량 결집을 통해 이루어지고 있으며, 더 나아가 민간-공공 협력 연구를 통한 기술 개발 확대가 필수적이다. 본 발표에서는 핵융합가속기 분야 전문센터를 중심에 두고 우리나라 학연산이 연합하여 유관 분야 협력을 극대화하기 위한 공동 기술 개발 노력과 인프라 공동 활용 계획에 대해서도 소개한다.

· Educational History

한국과학기술원 학사 (전공: 물리학 / 1996.02)
한국과학기술원 석사 (전공: 물리학 / 1998.02)
한국과학기술원 박사 (전공: 물리학/ 2003.02.)

· Employment History

프린스턴 플라즈마 물리연구소 박사후연수원 (2005.10 ~ 2007.9)
국가핵융합연구소 선임연구원 (2008.12 ~ 2012.3)
한국핵융합에너지연구원 책임연구원 (2012.3 ~ 현재)
핵융합가속기초고성능컴퓨팅 전문센터 센터장 (2022.9 ~ 현재)
한국핵융합에너지연구원 인공지능연구본부 본부장 (2026.2 ~ 현재)

김용휘
팀장
울산과학기술원

울산과학기술원 현황 소개

울산과학기술원은 소재·나노 분야 전문센터로 소재·나노 전분야 지원 연구체계 구축, 소재·나노 분야 연구개발 발굴 및 추진, 초고성능컴퓨팅 공동활용 체계 구축 및 안정적인 운영, 소재·나노 산업 생태계 구축 및 확산을 추진하고 있다. 본 발표에서는 울산과학기술원의 현황에 대하여 소개한다.

· Educational History
경희대학교 학사 졸업

· Employment History
2012.05.~ 현재 울산과학기술원

김종원
교수
광주과학기술원

자율주행초고성능컴퓨팅전문센터: HPC-AI 공용인프라를 활용한 자율주행 커뮤니티 협력

광주과학기술원 슈퍼컴퓨팅센터는 자율주행 분야 전문센터로 자율주행/모빌리티/디지털트윈에 특화된 R&D 혁신을 지원하는 Connected Mobility 슈퍼컴퓨팅 서비스를 지원을 목표한다. 국내 대학 유일의 세계 TOP500 등재 Dream-AI 슈퍼컴퓨팅 장비를 운영하고 있으며, 26년 6월부터는 AI기반대학 과학기술 연구혁신을 위한 중앙거점 센터로 활동하면서 차기 슈퍼컴퓨팅 장비인 (가칭) Imagine-AI 구축을 준비하고 있다. 본 발표에서는 상기한 HPC-AI 공용인프라를 활용한 자율주행 커뮤니티 협력에 관하여 다음과 같이 발표하고자 한다. 세부적으로는 커뮤니티 협력을 통하여 가상-실제를 연계하는 디지털트윈 구도에 따라 자율주행 특화 서비스를 제공하기 위한, Dream-AI를 활용하는 데이터셋 공유 기반의 집중훈련 워크플로우를 소개한다.
또한 에지컴퓨팅/디지털트윈/데이터축적을 연계한 오픈소스 ScaleX-POD 멀티-K8S 클러스터를 활용하는 데이터 획득-추론 연계형 자율주행 모델의 검증 과정도 제시한다.

· Educational History

서울대학교 제어계측공학과 공학박사 ('89~'94)

· Employment History

APAN 네트워킹 협의체 기술분야 의장 ('14~현재)
오픈AI에지(舊, MEC)포럼 운영위원장 ('20~현재)
한국슈퍼컴퓨팅포럼 부위원장 ('21~현재)
3기/4기 국가스마트도시위원회 위원 ('22~현재)
GIST 슈퍼컴퓨팅센터 센터장 ('08~현재)
GIST AI대학원 원장 / AI학과 학과장 ('19.10~현재)
GIST AI대학 학장 ('26.07~현재)

허무영
책임
기초과학연구원

IBS HPC 공동활용 인프라 구축 및 국가 공동활용 네트워크로의 확장

최근 초거대 언어모델(LLM) 연구와 우주론적 대규모 수치 모의실험 등 초고성능컴퓨팅(HPC-AI)을 활용한 기초과학 분야의 연산 수요가 기하급수적으로 급증하고 있다. 그러나 기존의 연구실 및 개별 연구단 단위로 분산 구축된 독립 클러스터 환경은 자원의 사일로(Silo)화를 초래하여, 전체적인 가동률 저하, 장비 수급 불균형, 공동활용률 저하라는 구조적 비효율을 낳고 있다.

기초과학연구원(IBS)은 이를 해결하기 위해 개별 연구단 중심의 인프라 조달 패러다임을 개선하여, 본원 리서치솔루션센터(RSC)를 중심으로 HPC 자원을 통합 구축하는 '중앙 허브형 공동활용 서비스 인프라'를 성공적으로 안착시켰다. 이러한 자원 통합 모델을 기반으로 자원의 규모를 매년 체계적으로 증설하는 한편, 내·외부 연구자 대상의 자원 개방 및 공동활용 비율을 극대화하여 연산 자원 활용률이 증가하고 있다. 본 발표에서는 IBS가 수행한 자원 통합과 운영 효율화 성과를 공유하고, 이를 국가적인 초고성능컴퓨팅 공동활용 네트워크로 확장하기 위한 방안을 제시한다.

· Educational History

2004.08. 부산대학교 이학박사 (전공: 통계 및 생물물리학)
2000.02. 부산대학교 이학석사 (전공: 통계물리학)
1998.02. 부산대학교 이학사 (전공: 물리학)

· Employment History

2021.03.~현재 IBS 책임기술원
2019.05.~2021.02. IBS 선임기술원
2017.01.~2019.05. IBS ICCP 프로젝트 리더/슈퍼컴퓨팅 매니저
2011.08.~2016.12. 부산대학교 물리학과 연구교수
2006.05.~2011.08 하버드대학교 CCB, FCSB 박사후연구원
2004.08.~2006.04. 부산대학교 유전체물성연구소 전임연구원

DAY 2

Workshop - PM

Wednesday, September 09, 2026

KOREA SUPERCOMPUTING CONFERENCE 2026

The Intelligent Frontier: Synergizing HPC, AI and Quantum

HPC 활용 산업 기술 워크숍

13:30~14:10	**From HPC Data Pipelines to Production AI: Why AI Requires a Fundamentally Different Data Architecture** Floyd Christofferson VP, Product Marketing (Hammerspace)
14:10~14:50	**AI Infrastructure with Scalable GPU and memory pooling** Martin Cooper SVP (Liqid)

Floyd Christofferson
VP, Product Marketing
Hammerspace

From HPC Data Pipelines to Production AI: Why AI Eequires a Fundamentally Different Data Architecture
(HPC 데이터 파이프라인에서 프러덕션 AI로: AI가 요구하는 새로운 데이터 아키텍쳐)

전통적인 HPC 환경은 단일 도메인의 중앙집중식 병렬파일시스템(Parallel File System)을 중심으로 구축되었으나, AI 및 하이브리드 워크로드의 등장으로 멀티 사이트(Multi-Site), 멀티 클라우드 환경으로 데이터 사일로(Silo)가 심화되어 전체 기업 데이터의 약 73%가 고립되는 현상이 발생하며, GPU 성능을 극대화하는데 데이터 병목과 비효율성이 주요 장애 요소로 작용하고 있습니다.

본 발표는 단순한 성능 개선을 넘어 '데이터 오케스트레이션(Data Orchestration)' 관점에서의 해결책을 제시합니다.

별도의 독자적인(Proprietary) 클라이언트 소프트웨어 설치 없이 표준 Linux 프로토콜을 활용하여 다양한 스토리지 인프라를 단일 글로벌 네임스페이스(Unified Global Namespace)로 통합합니다.

또한, AI 토큰 경제학(Token Economics) 관점에서 AI 환경 구축부터 첫 토큰 생성까지의 시간을 크게 단축하는 'Time to Very First Token' 개념을 강조합니다. 기존 스토리지 방식이 수 개월에 걸친 데이터 마이그레이션과 복사를 요구하는 반면, Hammerspace는 In-Place 데이터 활용 및 자동화된 데이터 배치 오케스트레이션을 통해 수 일 내에 AI 준비 상태를 완료하고 GPU 가동률(Utilization)을 극대화합니다.
결론적으로, 표준 기반 오픈 아키텍처와 지능형 데이터 오케스트레이션을 통해 데이터 거버넌스 및 주권(Sovereignty)을 보호하면서 글로벌 AI 파이프라인의 효율성과 토큰 경제성을 최적화하는 전략을 설명합니다.

· Educational History

University of Northern Iowa

· Employment History

SGI, Director, Product Marketing (2004 ~ 2014)
Arcitecta, CMO (2014 ~ 2016)
StrongBox Data Solutions, CEO (2017 ~ 2021)
Hammerspace, VP, Product Marketing (2022 ~ Present)

Martin Cooper
SVP (Senior Vice President)
Liquid

AI Infrastructure with Scalable GPU and memory pooling

Modern AI workloads are constrained not by compute alone, but by the rigid, server-bound allocation of GPUs and memory that leaves expensive resources stranded. Liqid addresses this with software-defined composable infrastructure that disaggregates GPU and DRAM into shared, dynamically allocated pools. Using its Matrix fabric software over PCIe and CXL fabrics, Liqid enables a single server to scale up to 30 GPUs and 100TB of pooled DRAM, driving device utilization toward 100%.

This talk presents Liqid's pooling architecture and its impact on AI inference. By expanding memory through CXL, KV-cache, RAG, and vector workloads achieve dramatic gains—roughly 7x higher inference throughput and up to 10x faster time-to-first-token—while lowering power and cost per token. Attendees will learn how disaggregated, composable infrastructure delivers the scalability, efficiency, and agility required for next-generation AI data centers.

· Employment History

NetApp / SolidFire, Senior Director & Solution Engineering Leadership (2008~2019)
Nebulon, Vice President – Solution Architects & Customer Support (2019~2023)
JetStream Software, Vice President – EMEA (2023~2026)
Liqid, Senior Vice President – WW Solution Architects (2026~)

기상·기후 분야 초고성능컴퓨팅 활용 고도화 워크숍

13:00~13:30	**등록 및 인사** 조민수 박사 (한국과학기술정보연구원)
13:30~14:00	**Assessment of atmosphere-ocean coupling in an AI-based Climate Model** 강대현 선임연구원 (한국과학기술정보연구원)
14:00~14:30	**Advancing High-Performance Computing Applications in Weather and Climate Research: Computing Support for AI-Based Climate Prediction Systems** 박진영 박사후연구원 (한국과학기술정보연구원)
14:30~15:00	**Regional AI Weather Model with Lateral Boundary Conditions for Kilometer-Scale Wind Forecasting over Mountainous Terrain of the Korean Peninsula** 김동훈 교수 (제주대학교)
15:00~15:30	**Developments in HPC and AI for High-Resolution Earth System Modeling** Stan Posey (NVIDIA)
15:30~15:50	Coffee Break
15:50~16:20	**HPC-, AI-, and XR-Based Modeling, Simulation, and Visualization for Wildfire Management: A Research Initiative at KISTI** 조민수 책임연구원 (한국과학기술정보연구원)
16:20~16:50	**Future Change Projections of Wildfires in the Yeongdong Region Using Dynamical and Statistical Models** 차동현 교수 (울산과학기술원)
16:50~17:20	**A High-Resolution AI-Integrated Modeling Framework for Rapid Wildfire Spread and Risk Potential Assessment** 박상훈 교수 (연세대학교)
17:20~17:50	**VAPOR-Based Parallel Visualization of Wildfire Spread Simulations Using the KISTI CAVE Cluster** 김종용 책임연구원 (㈜비즈아이엔에프)

강대현
선임연구원
한국과학기술정보연구원

Assessment of atmosphere-ocean coupling in an AI-based Climate Model

In recent years, deep learning (DL) has shown strong potential for improving weather and climate prediction. Emerging global DL weather prediction models have achieved remarkable accuracy and efficiency for short- to medium-range forecasts of up to 15 days. However, these models remain limited in their long-range forecasting capabilities due to the absence of realistic air-sea coupled processes.

This study aims to resolve both atmospheric and oceanic processes within a DL modeling framework, making it well-suited for subseasonal-to-seasonal (S2S) prediction. Trained on three-dimensional atmospheric and oceanic fields, a convolution-based model successfully reproduces realistic air-sea interactions and wave dynamics. This approach enhances physical consistency by incorporating coupled processes and leveraging large-scale observational and climate model datasets. The results suggest that this DL-based model can extend potential predictability for S2S forecasts while maintaining relatively low computational costs.

· Educational History

울산과학기술원 도시환경공학부 학사 (2012)
울산과학기술원 도시환경공학부 박사 (2019)

· Employment History

University of Washington, 박사후연구원 (2019~2021)
전남대학교 연구교수 (2021-2022)
한국과학기술연구원 선임연구원 (2022~현재)

박진영
박사후연구원
한국과학기술정보연구원

Advancing High-Performance Computing Applications in Weather and Climate Research: Computing Support for AI-Based Climate Prediction Systems

High-resolution numerical weather and climate models require substantial computing resources and power, and their increasing computational cost has become a major constraint on further performance improvement. In contrast, trained AI-based climate prediction models can provide rapid forecasts with relatively low computational and energy costs, offering a promising approach to complement conventional numerical models.

This project aims to support the development of an AI-based high-resolution global climate prediction model and a subseasonal prediction system by enabling efficient training of large-scale climate datasets on massively parallel GPU accelerators. The Korea Institute of Science and Technology Information is conducting the project "Development of an Accelerated Computing Service Environment for AI-Based Climate Prediction Systems" to provide a computing environment in which deep-learning models developed by participating institutions can be ported, parallelized, and optimized.

The main objectives are to establish an energy- and cost-efficient high-performance computing environment, ensure stable operation and convenient access to dedicated deep-learning nodes, and improve model performance through multi-GPU distributed training, mixed-precision training, and hyperparameter tuning.

This presentation introduces the accelerated computing environment and the major parallel optimization activities conducted to support the development of AI-based climate prediction systems.

· Educational History

공주대학교 대기과학과 학사 (2017.02)
울산과학기술원 도시환경공학과 석/박사 (2023.02)

· Employment History

울산과학기술원 도시환경공학과 박사후연구원 (2023.02-2023.09)
한국과학기술정보연구원 슈퍼컴퓨팅 인프라센터 박사후연구원 (2023.10~)

김동훈
학술연구교수
제주대학교 태풍연구센터

Regional AI Weather Model with Lateral Boundary Conditions for Kilometer-Scale Wind Forecasting over Mountainous Terrain of the Korean Peninsula

Accurate and timely prediction of surface wind fields over complex terrain is critical for disaster response, particularly for wildfire management in the mountainous regions of the Korean Peninsula. While global AI weather models have shown remarkable success, they often lack the resolution and regional specificity required for local-scale applications. This study presents a regional AI-based weather model, rFCNx (Regional FourCastNeXt), designed for kilometer-scale (1 km) wind forecasting with 10-minute temporal resolution.

The rFCNx model adapts the global FourCastNeXt architecture to a regional domain (256 × 256 grid) centered on the Gangwon province. To effectively handle lateral boundary conditions (LBC) and mitigate tiling artifacts, we introduce a novel forcing scheme that injects time-evolving WRF-Fire boundary fields into an 18-cell halo, combined with overlapping 4 × 4 Vision Transformer patches. The model is trained on high-resolution WRF-Fire simulation data, incorporating 22 input channels including terrain height, specific humidity, and fire area.

Evaluation results demonstrate that rFCNx successfully reproduces near-surface wind patterns and terrain-locked features consistent with the ground truth WRF-Fire simulations, achieving low error rates for short-term forecasts (0–3 hours). The model significantly reduces computational costs compared to traditional dynamical downscaling, enabling near-real-time prediction. This study confirms the feasibility of using regional AI weather models as fast, high-resolution surrogates for operational wildfire support systems, while also highlighting the need for further refinement in long-term fire spread prediction.

Acknowledgments: 본 연구는 한국과학기술정보연구원(KISTI)의 위탁연구개발과제와

· Educational History

성균관대학교 토목환경공학과 학사 (1994)
성균관대학교 대학원 해안•해양 공학 전공 석사 (1996)
성균관대학교 대학원 해안•해양 공학 전공 박사 (1999)

· Employment History

연세대학교 대기과학과/지구환경연구소 박사후연구원 (1999 ~ 2000)
서울대학교 대기과학과/기후환경시스템연구센터 박사후연구원 (2000 ~ 2002)
(일본) 전력중앙연구소(CRIEPI) 특별계약연구원 (2002 ~ 2005)
기상청/국립기상연구소 기상연구관 (2005 ~ 2012)
한국외국어대학교 WISE사업단 연구개발본부장 (2014 ~ 2016)
인하대학교 인공지능융합연구센터 산학연구교수 (2017 ~ 2022)
제주대학교 태풍연구센터 학술연구교수 (2023 ~ 현재)

Stan Posey
Program Manager, Earth System Science Domain
NVIDIA

Developments in HPC and AI for High-Resolution Earth System Modeling

HPC and AI have become critical components of skillful weather prediction, Earth system modeling and climate change research. NVIDIA has joined withthe weather and climate modeling community towards refinement of horizontal resolution foratmosphere GCMs towards km-scale, in order to explicitly resolve certain small-scale convective cloud processes and provide more realistic local information on forecasting and climate change. At the same time, Exascale HPC systems have arrived and in most cases are designed with GPU accelerators that offer opportunities in reasonable simulation turn-around times balanced with efficiency in energy consumption. Ultimately, output from models at km-scale become the essential databehind deployment of AI models in the NVIDIA Earth-2program. This talk will describe advances in HPC developmentsfor (i) GPU-accelerated numerical models, (ii) AI software and system features for large-scale data handling and ML model training and inference, and (iii) latest NVIDIA development of high resolution AI models that range from storm-resolving nowcasting to long term climate projections.

· Educational History

B.Sc. Mechanical Engineering, University of Tennessee, Knoxville, USA
M.Sc. Mechanical Engineering, University of Tennessee, Knoxville, USA

· Employment History

U.S. DOE Oak Ridge National Laboratory, Research Engineer (1983)
CD-adapco, Engineering Analyst (1985)
Control Data Corporation, HPC Applications Engineer (1987)
Silicon Graphics Inc, Director Worldwide CAE Program (1990)
Panasas Parallel Storage, Director HPC Applications Program (2007)
NVIDIA Corporation, Program Manager, ESS and CFD Domains (2009)

조민수
책임연구원
한국과학기술정보연구원

HPC-, AI-, and XR-Based Modeling, Simulation, and Visualization for Wildfire Management: A Research Initiative at KISTI

In recent years, wildfires have become more frequent, larger, and more difficult to control due to climate change, prolonged droughts, rising temperatures, and stronger winds. Effective wildfire management requires the rapid analysis of complex environmental data, as well as the accurate prediction and advanced visualization of wildfire behavior.

High-performance computing (HPC) and physics-based numerical weather prediction (NWP) models enable high-resolution wildfire simulations with greater accuracy. AI-based weather forecasting and wildfire prediction models can rapidly predict wildfire spread by analyzing large volumes of meteorological and environmental data. Extended reality (XR) can present simulation results in a realistic and immersive environment, allowing researchers, firefighters, and decision-makers to better understand wildfire conditions and make informed decisions.

KISTI's advanced computing and visualization infrastructure provides a platform for integrating HPC, AI, and XR technologies for wildfire management.

This research aims to develop an integrated platform that models and simulates wildfire spread, analyzes and visualizes potential damage in real time, and supports firefighter training, disaster response planning, and scientific decision-making. Ultimately, this research is expected to improve wildfire preparedness, enhance emergency response capabilities, and reduce the loss of life and economic losses.

· Educational History

B.S./M.S./Ph.D in Atmospheric Science, Yonsei University, Seoul, in Korea (1987/1991/1996)

· Employment History

Researcher(Senior/Principal), KISTI (1999~present)
Director of HPC(Service/Research) Center, KISTI (2013~2018)
Vice President, KISTI (2021~2024)
Vice President, Korean Meteorological Society (2018~2019)

차동현
교수
울산과학기술원

Future Change Projections of Wildfires in the Yeongdong Region Using Dynamical and Statistical Models

Future changes in wildfire spread patterns and fire weather risk in the Yeongdong region of Gangwon Province under climate change were projected using the WRF-SFIRE dynamical model and the FWI (Fire Weather Index) statistical model. A fuel map developed by adapting high-resolution terrain data to Korean vegetation characteristics was applied to WRF-SFIRE, and five wildfire cases were reproduced. Using the Pseudo-Global Warming (PGW) experiment with a 23-member CMIP6 model ensemble under the SSP2-4.5 scenario, changes in wildfire spread area were analyzed under future-period (2070-2099) forcing relative to the historical period (1985-2014). The results showed that fire spread area decreased in most cases under PGW forcing due to increased fuel moisture content, whereas in some cases the effect of positive terrain slope offset the increase in fuel moisture, leading to an increase in spread area instead. In addition, the FWI analysis confirmed that rising temperature under PGW conditions was a major contributor to increased FWI, with the pattern of change varying by case and region. However, since precipitation forcing is not currently incorporated into the PGW-based FWI analysis, uncertainty remains in future fire weather risk assessment, indicating a need for further research using approaches such as artificial intelligence techniques.

· Educational History

서울대학교 지구환경과학부 이학사 (2001. 2)
서울대학교 지구환경과학부 이학박사 (2009. 2)

· Employment History

서울대학교 BK21 박사후연구원 (2009. 3-2010.11)
하와이대 국제태평양연구소 박사후연구원 (2011.12~2012.08)
울산과학기술원 지구환경도시건설공학과 교수 (2012.09~현재)

박상훈
교수
연세대학교

A High-Resolution AI-Integrated Modeling Framework for Rapid Wildfire Spread and Risk Potential Assessment

Improving high-resolution wildfire spread forecasting over complex terrain, such as the Korean Peninsula, requires innovative approaches that balance computational efficiency with physical accuracy. While numerical weather prediction (NWP) systems offer detailed atmospheric information, evaluating regional wildfire spread scenarios across various potential ignition conditions often presents challenges in operational flexibility and processing demands.

To address these considerations, this study explores an AI-assisted framework designed to support wildfire behavior and risk potential assessment using high-resolution meteorological fields. Rather than focusing on a single deterministic outcome, this work investigates a post-processing methodology aimed at facilitating multi-scenario evaluations across high-risk areas. By testing the feasibility of AI-guided tools (AI-FIRE) in conjunction with high-resolution atmospheric forcing, we seek to examine how data-driven techniques can complement existing modeling workflows for rapid spatial assessment.

This presentation outlines the conceptual design, data structuring approaches, and preliminary methodology of the framework, providing insights into future directions for high-resolution, hazard-aware decision support systems over East Asia.

· Educational History

B.S. in Atmospheric Sci., Yonsei Univ. (1999)
Ph.D. in Atmospheric Sci., Yonsei Univ. (2009)

· Employment History

Project Scientist in National Center for Atmospheric Research (2011~2017)
Professor, Atmospheric Sci., Yonsei Univ. (2017~)

김종용
책임 연구원
㈜비즈아이엔에프

VAPOR-Based Parallel Visualization of Wildfire Spread Simulations Using the KISTI CAVE Cluster

This study develops a parallel visualization framework that integrates VAPOR with the high-performance cluster nodes of the KISTI CAVE for the immersive analysis of large-scale wildfire spread simulations. Conventional VAPOR outputs are generated from a single viewpoint for desktop displays, limiting the effective use of CAVE's 360-degree multi-surface display and cluster resources. To address this limitation, the framework employs one Master Node and 13 GPU-equipped Display Nodes, applying node-specific camera positions, orientations, and projection matrices to the VAPOR rendering pipeline. The VAPOR Client, Master Server, standalone VAPOR Renderers, shared storage, and a dedicated interface are integrated to support job creation, node-level parallel rendering, result storage, and CAVE display. Each Display Node renders wildfire imagery from its assigned viewpoint, enabling immersive analysis of terrain, wind, smoke, fire extent, and temporal evolution. Future work will focus on automated job distribution, node and failure monitoring, data-loading optimization, and interface enhancement, with the goal of extending the framework into a general-purpose parallel visualization solution for supercomputing-based scientific data in meteorology, oceanography, and atmospheric science.

· Educational History

청운대학교 컴퓨터학 (2014)
동국대학교 멀티미디어학 석사 (2016)
동국대학교 멀티미디어학 박사 (2024)

· Employment History

㈜비즈아이엔에프 17년도 ~ 현재

Bio-HPC & AI: 차세대 제약·바이오 혁신

13:30~14:10	**AI drug discovery: Data, model, and beyond** 신현진 소장 (목암생명과학연구소)
14:10~14:50	**AI agent 기반 합성 가능한 신약 후보물질 설계** 장우대 조교수 (고려대학교)
14:50~15:30	**AI 전환을 향한 여정** 이제현 실장 (KIST)
15:30~16:00	Coffee Break
16:00~16:40	**신약개발에서의 AI의 역할** 김승하 팀장 (LG화학)
16:40~17:20	**DeepZema: AI 기반 합성신약 연구개발 플랫폼을 활용한 Drug Discovery** 임동철 CTO (이노보테라퓨틱스)
17:20~18:00	**in silico Drug Discovery의2026년 트렌드: 'AI 신약개발'의 모호함을 넘어 증명(Proof)으로** 오경석 연구위원 (대웅제약)

신현진
소장
목암생명과학연구소

AI drug discovery: Data, model, and beyond

신약개발은 막대한 시간과 비용이 소요되는 산업으로, 이를 혁신하기 위한 AI 활용이 지난 10여 년간 활발히 시도되어 왔다. 그러나 제한된 데이터와 복잡한 규제 요건으로 인해 가시적 성과는 아직 제한적이다. 최근에는 소량 데이터 기반 학습, 불확실성 정량화를 통한 신뢰도 향상, agentic AI와 human-in-the-loop, 나아가 lab-in-the-loop를 결합한 자율화·자동화 기술이 빠르게 발전하고 있다. 이러한 변화는 좋은 데이터와 좋은 모델이 상호 강화되는 선순환을 가능하게 한다. 본 발표에서는 관련 기술과 산업 동향을 점검하고, 목암생명과학연구소의 주요 활동을 소개하며 AI 신약개발의 근미래 전략을 논의한다.

· Educational History

서울대학교, 전기공학부, 학사 (1998)
The University of Texas at Austin, Electrical and Computer Engineering, M.S. (2001)
The University of Texas at Austin, Electrical and Computer Engineering, Ph.D. (2006)

· Employment History

Dana-Farber Cancer Institute/Harvard University, Postdoctoral Research Fellow (2007~2011)
Takeda Pharmaceuticals, Inc., Principal Scientist (2011~2021)
목암생명과학연구소, 부소장 (2021~2024)
목암생명과학연구소, 소장 (2024~현재)

장우대
조교수
고려대학교 약학과

AI agent 기반 합성 가능한 신약 후보물질 설계

분자 생성모델은 높은 결합력이 예측되는 화학 구조를 쉽게 생성하지만, 그중 상당수는 실제로 합성하기 어렵고, ADMET 특성과 선택성 등 약물이 갖춰야 할 필수 요건을 동시에 만족하지 못한다. 즉 후보물질 설계는 결합력을 높이는 단일 목적 문제가 아니라, 합성 가능성을 제약조건으로 두고 상충하는 여러 목적을 함께 최적화해야 하는 다중 목적 최적화 문제다. 본 발표에서는 reaction template과 building block을 제약조건으로 반영해 탐색 공간을 합성 가능한 영역으로 한정하고, ADMET/PK 예측과 docking 및 MD simulation, 약효 데이터를 AI agent가 통합 해석하여 다음 설계를 제안하는 closed-loop 신약개발 프레임워크를 소개한다.

· Educational History

B.S., 충남대학교 미생물학과 (2014)
Ph.D., KAIST 생명화학공학과 (2021)
M.S., 충남대학교 신약전문대학원 (2016)

· Employment History

KAIST 생물공정연구센터 연수연구원 (2021.02–2021.11)
한국화학연구원 신약정보기술연구센터 선임연구원 (2021.11–2026.08)
UST 과학기술연합대학원대학교 조교수 (2023.08–2026.08)

이제현
AIX전략실장
한국과학기술연구원 AIX전략실

AI 전환을 향한 여정

최근 AI의 파도 속에서 여러 기업이 생존을 위해 AI를 도입하고자 합니다. 그러나 95%가 실패한다는 통계가 말해주듯 결코 만만치 않은 일입니다. AI 도입을 담당한 부서장으로서 저 또한 과거와 현재에 다양한 시행착오를 겪었고, 겪고 있습니다. 아직 성공했다고 말하기 어렵지만, 근 10년간 경험한 일들을 나눔으로써 여러분의 미래로 다가올지도 모르는 실패를 미연에 방지할 수 있는 자리를 만들고자 합니다.

· Educational History

서울대학교 재료공학부 (학사/ 2001)
서울대학교 재료공학부 (박사 / 2008)
Vienna Tech. Univ. (박사 / 2011)

· Employment History

서울대학교 재료공학부 연구교수 (2011-2012)
삼성전자 종합기술원, 반도체연구소 전문/책임/수석 (2013-2017)
한국에너지기술연구원 선임/책임 (2018-2025)
한국과학기술연구원 책임 (2026-현재)

김승하
팀장
LG화학 생명과학 AX팀

신약개발에서의 AI의 역할

AI 기반 단백질 구조 예측 기술이 2024년 노벨화학상 수상으로 기초과학 난제의 해결 수단으로 공인되면서, 신약개발에서 AI 활용은 선택이 아닌 필수 요건이 되었다. 글로벌 선도사들은 설계-제작-시험-학습 사이클을 주 단위로 단축하며 후보물질 발굴 기간을 기존 5~7년에서 2~3년 수준으로 줄이는 것을 공통의 지향점으로 삼고 있으나, 신약 R&D 밸류체인의 단계별로 기술 성숙도와 활용 방식은 여전히 상이한 상황이다.

본 발표에서는 먼저 AI 구조 예측·설계 기술의 현재 수준과 남은 한계를 정리하고, 신약 R&D 밸류체인 단계별 AI 활용 영역과 성숙도를 조망한다. 이어 글로벌 AI 신약개발 기업의 접근 방식과 검증된 성과를 비교하여, 모델 성능 자체보다 설계 결과를 실험으로 검증하고 이를 다시 모델에 반영하는 속도가 경쟁력을 좌우함을 논의한다. 마지막으로 LG화학이 자체 구축한 신약 R&D AI 환경인 MediX를 중심으로 그 구성과 적용 사례, 향후 확장 방향을 논의하고자 한다.

· **Educational History**

포항공과대학교 화학공학과, 공학사 (2001)
서울대학교 대학원 응용화학부, 공학석사 (2003)
Iowa State University, Chemical and Biological Engineering, 공학박사 (2011)

· **Employment History**

University of California, Santa Barbara (CA, USA) 화학공학 박사후연구원 (2012-2014)
LG화학 CTO 계산화학 (2014~2023)
LG화학 생명과학 연구개발부문 생명과학 AX팀 팀장 (2024~현재)

임동철
CTO
이노보테라퓨틱스

DeepZema: AI 기반 합성신약 연구개발 플랫폼을 활용한 Drug Discovery

인공지능(AI)은 합성신약 연구에서 후보물질 발굴과 최적화의 효율을 높이는 핵심 도구로 활용되고 있다. 이노보테라퓨틱스는 AI 기반 합성신약 연구개발 플랫폼인 DeepZema를 구축하여 구조 기반 약물설계, 머신러닝 기반 물성 및 ADMET 예측, 단백질-리간드 공동접힘 등 다양한 계산기법을 연구에 활용하고 있다.

본 발표에서는 DeepZema의 개발 배경과 주요 구성 요소를 소개하고, AI와 계산화학 기술이 실제 합성신약 연구 과정에서 어떻게 활용되는지 사례를 중심으로 설명한다. 또한 계산 예측과 실험 결과를 연계한 연구 전략과 AI 기반 신약개발 플랫폼의 발전 방향에 대해 논의하고자 한다.

· **Educational History**

B.S. 서울대학교 (1979–1983)
M.S. 한국과학기술원 (KAIST / 1983–1985)
M.S. University of Illinois, 유기화학 (1988–1990
Ph.D.(with W. L. Jorgensen) Yale University, 계산화학 (1991–1996)

· **Employment History**

Researcher, 한국화학연구원 (KRICT / 1985 ~ 1988)
Post-Doctoral fellow, Yale University, CT, USA (1996 ~ 1999)
Researcher, Schrödinger Inc., NY, USA (1999 ~ 2000)
Researcher, Neurogen Corporation, CT, USA (2000 – 2002)
㈜LG생명과학 연구위원 (2002 ~ 2015)
제마소프트 대표 (2015 – 2019)
LG화학 기술고문 (2017 – 2019)
이노보테라퓨틱스 CTO (2020 – 현재)

오경석
연구위원
대웅제약

in silico Drug Discovery의 2026년 트렌드: 'AI 신약개발'의 모호함을 넘어 증명(Proof)으로

지난 KSC 2025에서 신약개발에 적용되는 다양한 in silico 방법론을 소개하면서, 동적인 생체 환경을 온전히 담지 못하는 물리 기반 계산의 한계, 블랙박스처럼 근거를 설명하기 어렵고 데이터 품질에 좌우되는 AI 예측, 그리고 그 예측이 실제 합성과 실험으로 이어지는 과정의 간극처럼 여전히 남아 있는 과제들을 함께 짚어본지 1년이 지났습니다. 그사이 in silico 신약개발은 물리 기반 계산과 AI를 아우르며 빠르게 넓어졌고, 그 중에서도 'AI 신약개발'은 모든 연구자가 접근할 수 있는 일상이자 업계의 당연한 표준으로 자리 잡았습니다. 그러나 타깃 발굴부터 임상까지 얽힌 다학제적 과정을 'AI 신약개발'이라는 한 단어로 묶기에는 그 개념이 이미 너무 모호해졌습니다. 이제는 AI가 전주기 과정에서 어떻게 세분화되어 적용되는지를 나누어 보고, 그 적용이 신약개발의 생산성, 즉 효과(Effectiveness)와 효율(Efficiency)의 향상으로 실제 이어졌는지를 함께 살펴볼 필요가 있습니다.

올해는 이러한 관점에서 in silico Drug Discovery의 트렌드를 짚어보고자 합니다. 거대 파운데이션 모델(Foundation Model)의 강화, 물리 기반 모델과 AI 기반 모델의 융합, 데이터 주권을 확보하기 위한 소버린 AI (Sovereign AI)의 부상, 그리고 보안 장벽 속에서도 협력을 이끌어내는 연합학습(Federated Learning)의 확산이 그 한 축입니다. 여기에 스스로 가설을 세우고 설계하며 로봇과 협업하는 자율실험실 (Autonomous Lab)과 Agentic AI로의 진화, 나아가 글로벌과 국내의 제약사 및 AI 전문기업이 보여주는 투자와 개발 현황, 그리고 오픈 이노베이션 (Open Innovation)의 방향까지 함께 살펴봅니다.

끝으로 이러한 흐름 속에서 우리가 현장에서 실제로 마주하는 간극을 짚어보고, 앞으로 우리가 갖추어야 할 경쟁력이 무엇인지 함께 고민해 보고자 합니다.

· Educational History

포항공과대학교 화학과 학사 (1996)
포항공과대학교 화학과 이학석사 (1998)
포항공과대학교 화학과 이학박사 (2000)

· Employment History

SK주식회사 수석연구원 (2000~2011)
SK바이오팜 수석연구원 (2011~2021)
일진에스앤티 연구소장 (2021~2022)
대웅제약 연구위원 (2022~현재)

양자-고전 하이브리드 계산 방법론 및 적용

13:30~14:30	**양자 응용 알고리즘 (1): VQE를 통한 분자구조 해석** 강유성 박사 (메가존 클라우드)
14:30~15:30	**양자 응용 알고리즘 (2): SQD를 활용한 CI 문제 해결** 강유성 박사 (메가존 클라우드)

강유성
Senior Consultant
MegazoneCloud

양자-고전 하이브리드 계산 방법론 및 적용

오류 보정이 완비되지 않은 현 단계의 양자 하드웨어에서 의미 있는 계산을 수행하려면, 양자와 고전 자원을 어떻게 분담시킬 것인가가 핵심 설계 문제가 된다. 본 강의는 이러한 양자-고전 하이브리드 관점에서 위상추정 계열 (QPE)과 변분 (VQE)·샘플링 (SQD) 기반 알고리즘의 원리를 정리하고, 클라우드를 통해 이온트랩 양자컴퓨터에 접속하여 분자 전자구조 문제에 적용하는 과정을 실습한다. 이를 통해 참가자가 현재 하드웨어 수준에서 실현 가능한 계산의 범위와 한계를 스스로 판단할 수 있도록 하는 것을 목표로 한다.

· **Educational History**

2016 연세대학교 물리학과 학사
2025 연세대학교 물리학과 박사

· **Employment History**

2025~현재 메가존클라우드 Senior Consultant

한국초고성능컴퓨팅 포럼 기술교류회

13:30~13:45	**개회**
13:45~14:25	**에이전틱 프로그램 기반 계산 재료과학** 한승우 부교수 (서울대학교)
14:25~15:05	**계층형 스케줄러 기반 멀티리젼 GPU 공동활용 방안** 김도하 부문장 (SIEMENS)
15:05~15:45	**날씨 장사 – 기상청 예보에서 국민 맞춤형 기후 서비스로의 전환** 오재호 대표 (나노웨더)
15:45~16:00	Coffee Break
16:00~16:40	**HPC, AI, 양자컴퓨팅 시대의 격자 QCD와 계산핵물리학** 이종완 선임연구원 (기초과학연구원)
16:40~17:10	**HPC 기술분류체계** 진현욱 교수 (건국대학교)
17:10~18:00	**참석자 네트워킹**

한승우
부교수
서울대학교

에이전틱 프로그램 기반 계산 재료과학

계산 재료과학 (computational materials science)는 신소재 발견과 공정의 최적화라는 궁극의 목표를 향해 컴퓨터를 활용해왔다. 전통적으로 사람이 알고리즘과 프로그램을 개발하여 슈퍼컴에 계산을 맡기고 그 결과를 사람이 분석하는 식으로 노동의 분화 (division of labor)가 효율적으로 이루어져왔다. 최근 agentic AI가 발전함에 따라 기존에 사람이 맡았던 판단의 일부를 AI가 맡는 agentic program 이 가능하게 되었고 기존의 deterministic program으로 탐구하기 어려웠던 영역을 개척하면서 신소재 발견과 공정 최적화라는 목표에 한걸음 다가갈 수 있게 되었다. 본 발표에서는 특히 graph neural network 기반의 범용 기계학습 포텐셜과 agent를 결합한 새로운 형태의 agentic program의 구체적 사례를 소개하고 향후 국가 슈퍼컴 활용에 미칠 영향에 대해 논의해보고자 한다.

· Educational History

서울대학교 물리학 학사 (1993)
서울대학교 물리학 석사 (1995)
서울대학교 물리학 박사 (2000)

· Employment History

프린스턴 대학교 연구원 (2001-2003)
이화여자대학교 전임 교원 (2003-2009)
서울대학교 재료공학부 부교수, 교수 (2009-현재)
고등과학원 AI 및 자연과학센터 겸직 교수

김도하
부문장
SIEMENS

계층형 스케줄러 기반 멀티리젼 GPU 공동활용 방안

본 발표에서는 계층형(Hierarchical) AI Scheduler 기반의 멀티리전 GPU 공동활용 아키텍처를 소개합니다. 기본구조는 중앙의 Global Policy 을 담당하는 최상위스케줄러가 여러지역으로 분산된 GPU 클러스터의 자원상태, 데이터 지역성(Data Locality), 네트워크 상태 및 운영 정책을 종합적으로 분석하여 최적의 실행 위치를 결정하고, 각 지역의 Local Scheduler는 PBS Pro, Slurm, Kubernetes 등 기존 워크로드 관리 시스템과 연계하여 작업 실행 및 자원 관리를 수행하는 구조를 설명합니다. 이를 통해 지역별로 분산된 GPU 자원을 하나의 논리적 컴퓨팅 환경으로 통합하여 사용자에게 일관된 서비스와 최적의 자원 활용 환경을 제공할 수 있습니다.

또한 다음단계로 AI Agent를 기반으로 과거 작업 이력과 사용자 패턴을 학습하여 GPU 수요를 예측하고, 작업 우선순위와 SLA(Service Level Agreement)를 고려한 정책 기반 스케줄링을 수행함으로써 GPU 활용률을 극대화할 수 있습니다. 이를 통해 단순한 장애 복구 기능을 넘어 예측(Predictive) → 판단(Decision) → 실행(Action) → 학습(Learning)의 순환구조를 기반으로 동작하는 AI 운영 플랫폼(AIOps)으로 발전할 수 있으며, 국가 AI 컴퓨팅센터 및 대규모 HPC 인프라의 운영 비용 절감과 안정성 향상에 기여할 것으로 기대합니다.

· Educational History

단국대학교 컴퓨터과학부 학사 (1998)

· Employment History

T scientific 모바일컨텐츠 관리 플랫폼 개발(1999~2004)
알테어엔지니어링 한국지사장 (2005~2026)
지멘스 인더스트리 소프트웨어 SAE 사업총괄 (2026~)

오재호
대표
나노웨더

날씨 장사 – 기상청 예보에서 국민 맞춤형 기후 서비스로의 전환

기후변화로 폭염, 홍수, 가뭄, 산불과 같은 극한기상이 더욱 빈번하고 강하게 발생하고 있지만, 그 피해는 모든 사람에게 동일하게 나타나지 않는다. 고령자, 장애인, 저소득층과 기상관측이 부족한 지역은 신속하고 실질적인 재난정보를 제공받기 어렵다. 우리나라의 기상정보는 주로 정부가 생산하여 일반 국민에게 일괄적으로 제공하고 있다. 이러한 공공서비스는 국가적으로 필수적이지만, 개별 지역과 산업, 다양한 이용자의 세부적인 요구까지 충족하는 데에는 한계가 있다.

기존의 공급자 중심 체계를 국민과 수요자 중심의 기상·기후서비스 생태계로 전환되어야 할 것이다. 정부는 신뢰할 수 있는 국가 관측자료와 예보를 제공하고, 민간 부문은 이를 바탕으로 초고해상도 분석, 신속한 정보 갱신, 수요자별 맞춤정보와 다양한 전달 수단을 개발함으로써 공공서비스를 보완할 수 있다.

이러한 접근의 사례로 알파멧(AlphaMet)을 제시한다. 알파멧은 실제 관측센서와 가상센서 자료를 계산 효율성이 높은 상세화 기술과 결합하여, 관측자료가 부족한 지역에서도 초국지적 기상상태를 재현하는 플랫폼이다. 이를 인공지능, 디지털 트윈, 모바일 앱, 스마트 안내판 및 다국어 정보전달 기술과 연계하면 복잡한 기상자료를 특정 지역·계층·산업에 필요한 실용적인 재난정보로 전환할 수 있다.

기상·재난·보건·사회정보를 통합한 플랫폼은 예·경보의 정확성, 신속성과 접근성을 높이고, 재난관리·스마트시티·재생에너지·농업 등 다양한 분야에 활용될 수 있다. 궁극적으로 공공과 민간의 협력은 국가 기상서비스를 강화하고, 과학기술의 혜택을 누구나 누릴 수 있는 포용적이고 공정한 기후회복력 사회를 구축하는 데 기여할 것이다.

· Educational History

Doctor of Philosophy (Ph.D.) in Atmospheric Sciences, University of Oregon, USA

· Employment History

(유)나노웨더 대표

이종완
선임연구원
기초과학연구원

HPC, AI, 양자컴퓨팅 시대의 격자 QCD와 계산핵물리학

자연계의 강한 상호작용과 원자핵의 구조 및 반응을 제일원리로 이해하고 예측하기 위한 대표적 계산과학 방법론인 격자 QCD 시뮬레이션과 계산핵물리의 과거, 현재, 미래를 HPC의 관점에서 조망하고자 합니다. 먼저, 이들 분야가 대규모 병렬컴퓨팅 아키텍처와 알고리즘의 발전과 함께 어떻게 정밀 이론도구로 자리잡아 왔는지를 살펴보고, 현대 핵·입자물리학의 발전에 어떤 방식으로 중요한 기여를 해왔는지를 소개합니다. 이어서 다체 핵계산의 계산 복잡성, 실시간 동역학, 유한 밀도, 부호 문제(sign problem), 임계 감속(critical slowing down)과 같은 난제들이 왜 차세대 계산 패러다임을 요구하는지를 논의합니다. 특히, 물리 정보 기반 기계학습과 양자컴퓨팅을 활용한 새로운 접근법이 전통적인 HPC 기반 계산을 대체하는 것이 아니라 상호보완적 도구로서 격자장론과 계산핵물리의 지평을 어떻게 넓힐 수 있는지를 전망합니다. 마지막으로, 국내의 HPC, AI, 양자컴퓨팅 역량을 바탕으로 한국이 이 분야에서 어떤 전략적 역할을 수행할 수 있을지 비전을 제시하고자 합니다.

· **Educational History**
University of Washington, Ph.D. in Physics

· **Employment History**
기초과학연구원(2023 ~)

진현욱
교수
건국대학교

HPC 기술분류체계

HPC 기술분류체계는 HPC 활용실태조사, 기술수준평가, 연구개발과제기획 등에서 다양하게 활용되고 있다. 따라서 HPC 기술분류체계를 현대의 기술에 맞게 최신화하는 것은 올바른 정책수립을 위한 중요한 시작점이며, 새로운 기술의 등장이 점점 가속화되는 현시대에서는 더욱 중요하다.

본 발표에서는 HPC 기술분류체계의 최신화를 위해서 다음의 내용을 발표하고 토의하는 장을 마련한다.

1) HPC 기술분류체계 현황 및 점검

2) 2026년 HPC 기술분류체계 최신화 방향

3) HPC 기술분류체계의 최신화 체계

· Educational History
고려대학교 전산학 학사 (1997)
고려대학교 컴퓨터학 석사 (1999)
고려대학교 통신시스템공학 박사 (2003)

· Employment History
오하이오 주립 대학교 박사후 연구원 (2003~2006)
건국대학교 조교수/부교수/교수 (2006~현재)
건국대학교 정보통신처장 (2022 ~ 2024)
한국정보과학회 부회장 (2015, 2020, 2024)
한국정보과학회 컴퓨터시스템소사이어티 회장 (2020~2022)

BIG3 기술교류회 (비공개)

13:30~15:30	**BIG3 기술교류회**
15:30~16:00	Coffee Break
16:00~18:00	**BIG3 기술교류회**

Tutorial

Monday, September 07, 2026

KOREA SUPERCOMPUTING CONFERENCE 2026

The Intelligent Frontier: Synergizing HPC, AI and Quantum

국가슈퍼컴퓨터 6호기(한강) 튜토리얼: AI 기반 물리 응용

슈퍼컴퓨터 6호기 GH200 환경에서의 AI 활용 방법을 소개합니다.
NVIDIA PhysicsNeMo를 활용한 AI 기반의 물리 분야 활용 실습 교육 과정입니다.

09:30~10:00	**등록**
10:00~10:50	**슈퍼컴퓨터 6호기(한강) 소개 및 접속 안내** 강지훈, 윤태호 (한국과학기술정보연구원)
11:00~12:00	**GH200 시스템 이해 및 활용법 소개** 유현곤 (NVIDIA)
12:00~13:30	점심
13:30~14:30	**GH200 시스템 실습 - 컴파일/프로파일링/튜닝 기본** 유현곤 (NVIDIA)
14:40~16:00	**AI 기반 물리 활용 실습 - NVIDIA PhysicsNeMo 기본** 양민규 (NVIDIA)
16:10~17:30	**PhysicsNeMo 기반 Neural Operator/FNO 실습** 양민규 (NVIDIA)
17:30~18:00	**마무리**

한국과학기술정보연구원

국가슈퍼컴퓨팅센터

HPC이노베이션허브

(주)엠키스코어

(주)엠키스코어

(주)이코어

(주)리더스시스템즈

아이크래프트

(주)테라텍

(주)유니트론텍

HAMMERSPACE

(주)글루시스

AWS

(주)삼보에드텍

(주)씨플랫폼

콴델라코리아

Korea Supercomputing Conference 2026

KSC2026

2026년 9월 8일(화)~9일(수) | 스위스 그랜드호텔 컨벤션홀(4F)

The Intelligent Frontier: Synergizing HPC, AI and Quantum

주최 KISTI 한국과학기술정보연구원 Korea Institute of Science and Technology Information www.kisti.re.kr

주관 국가슈퍼컴퓨팅본부

KSF 한국초고성능컴퓨팅포럼

KSCSE 한국계산과학공학회

후원 과학기술정보통신부 Ministry of Science and ICT

Korea Supercomputing Conference 2026

KSC2026

2026년 9월 8일(화)~9일(수) | 스위스 그랜드호텔 컨벤션홀(4F)

The Intelligent Frontier: Synergizing HPC, AI and Quantum

주최 KISTI 한국과학기술정보연구원 Korea Institute of Science and Technology Information www.kisti.re.kr

주관 국가슈퍼컴퓨팅본부

KSF 한국초고성능컴퓨팅포럼

KSCSE 한국계산과학공학회

후원 과학기술정보통신부 Ministry of Science and ICT